

Volume VIII  
2025



# DIGITAL AGE IN SEMIOTICS & COMMUNICATION

Journal

by the Southeast European Center for Semiotic Studies

The Cultural Explosion of AI: Navigating the Intersection  
of Artificial Intelligence, Society, and Culture from  
a Semiotic and Interdisciplinary Perspective





Volume VIII, 2025

## Digital Age in Semiotics & Communication

an annual double-blind peer-reviewed journal

**Editor-in-chief:** Kristian Bankov

**Registered Office:**

New Bulgarian University

Southeast European Center for Semiotic Studies;

Office 306, Building 2; 21 Montevideo St, Sofia 1618, Bulgaria

Phone: +359 2/8110 111

Email: DigitASC@nbu.bg

ISSN (Print) 2603-3585

ISSN (Online) 2603-3593

© New Bulgarian University, 2026

Design by Tanya Yordanova, Janet Tsanova

Logo design by Samuil Vladimirov

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publishers.

**Editor-in-chief:**

Kristian Bankov

**International advisory board:**

Fernando Andacht – Universidad de la República, Uruguay

Simone Arcagni – University of Palermo, Italy

Luis Emilio Bruni – Aalborg University, Denmark

Paul Cobley – Middlesex University, United Kingdom

Seema Khanwalkar – CEPT University, India

Evangelos Kourdis – Aristotle University of Thessaloniki, Greece

Massimo Leone – University of Turin, Italy

Dario Martinelli – Kaunas University of Technology, Lithuania

Jamin Pelkey – Ryerson University, Canada

Jia Peng – Jinan University, Guangzhou, China

Eviprides Zantides – Cyprus University of Technology, Cyprus

Jordan Zlatev – Lund University, Sweden

**Editors for the current issue:**

Kristian Bankov

Federico Biggio

**Language editor:**

David Mossop

**Collaborator:**

Viktorija Lankauskaitė

**Digital Age in Semiotics & Communication**, a journal from the Southeast European Center for Semiotic Studies at the New Bulgarian University and founded by Prof. Kristian Bankov, explores the new forms of knowledge, social and linguistic interaction, and cultural phenomena generated by the advent of the Internet.

A topic is chosen for each issue by the editorial board, but the topics will be always related to the issues of the digital environment. The topic is announced with a call for papers and will also be available on our Facebook page ([facebook.com/DigitASCjournal](https://facebook.com/DigitASCjournal)).

The working language of the journal is English. It uses double-blind review, meaning that both the reviewer's and the author's identities are concealed from each other throughout the review process.

### **Periodicity**

The journal will be published annually by the Southeast European Center for Semiotic Studies and the New Bulgarian University Publishing House.

### **Purpose**

The purpose of the journal is to provide a collaborative work field for scholars interested in researching new phenomena in the dynamic digital world. Our main purpose is to build a scientific bridge between the fields of semiotics, communications, social sciences and the problems of the digital era. We believe that our collaborations can raise the level of understanding for modern digital phenomena, providing both a solid theoretical framework and profound applied research.

The pilot issue summarizes the whole research program of the Center and the journal in particular. It is open to various problems concerning developments in digital culture and phenomena. We are interested in working with scholars from different research and applied fields, such as semiotics (both applied and theoretical), communication studies, marketing and advertising, linguistics and literary studies, anthropology and ethnography, cognitive science and psychology, and computer science.

More specifically, our interest is directed to:

- New forms of knowledge;
- New media and the immersive e-consumption of experience;
- New forms of social relations in the age of social media;

- New habits of communication and self-expression/representation;
- Online corporate communications;
- Digital narratology and e-fiction;
- Digital grammatology;
- Digital audio-visualisation;
- Internet linguistics.

The magazine is supported and reviewed by our International advisory board as well as by chosen external reviewers.

For more information and submission of papers: [DigitASC@nbu.bg](mailto:DigitASC@nbu.bg)

## CONTENTS

THE CULTURAL EXPLOSION OF AI: NAVIGATING THE INTERSECTION OF ARTIFICIAL INTELLIGENCE, SOCIETY, AND CULTURE FROM A SEMIOTIC AND INTERDISCIPLINARY PERSPECTIVE Kristian Bankov and Federico Biggio .....	7
A STEP-BY-STEP SEMIOTIC UNDERSTANDING OF LLMS AND CHATBOTS THROUGH INTERDISCIPLINARY DIALOGUE Mattia Thibault and Daniele Allois .....	19
THE PEIRCEAN THEORY OF AI: ADVANCING TEXT GENERATION THROUGH PEIRCE'S TRIADIC MODEL, SPECULATIVE GRAMMAR, AND METHODEUTICS Alon Friedman and Martin Thellefsen .....	50
ON THE EDGE OF UNDERSTANDING: CHATGPT AND THE LIMITS OF ARTIFICIAL SENSE-MAKING Fabio Montesanti .....	71
THE SEMIOTIC OF AI IMAGES Stephanie Schneider .....	84
THE JANUSIAN FACE OF FACIAL RECOGNITION, PART 1: ITS SUBFACE, INTERFACE, AND SURFACE Devon Schiller .....	103
ARE WE HUMAN OR ARE WE DANCER? – AI CREATIVITY IN THE REALM OF XR Momchil Alexiev .....	126

HUMANS AS NATURAL-BORN CYBORGS: SCRUTINISING AI'S NARRATIVE INTELLIGENCE WITHIN THE 5E COGNITION FRAMEWORK Inna Livytska .....	139
NEEDLE IN A NEEDLE STACK: HOW AI CAUSES SEMIOTIC INFLATION WHICH CAUSES EXPERIENTIAL DEVALUATION Marc Champagne .....	162
TRANSLATING THE WILD: AI, SEMIOTICS, AND THE FUTURE OF ANIMAL COMMUNICATION Nicola Zengiaro .....	178
CHATGPT IN HIGHER EDUCATION: A SEMIOTICS INVESTIGATION BETWEEN CULTURAL EXPLOSION AND ENCYCLOPEDIA KNOWLEDGE Danilo Petrassi .....	206
AI, SEMIOSIS, AND THE FUTURE OF LANGUAGE ACQUISITION: A GLOBAL APPROACH TO EDUCATIONAL SEMIOSPHERE Ilaria Ingrao and Seyedeh Maede Mirsonbol .....	227
CONSUMER ATTITUDES TO AI CONTENT GENERATION IN SOCIAL MEDIA Martin Varshev and Eduard Marinov .....	249
USE OF AI IN THE CONTEXT OF FASHION AND RELATED INDUSTRY Dilyana Orlova .....	265
NOTES FOR CONTRIBUTORS .....	277



## THE CULTURAL EXPLOSION OF AI: NAVIGATING THE INTERSECTION OF ARTIFICIAL INTELLIGENCE, SOCIETY, AND CULTURE FROM A SEMIOTIC AND INTERDISCIPLINARY PERSPECTIVE

*Kristian Bankov &  
Federico Biggio*

This issue of *Digital Age in Semiotics and Communication* is dedicated to the cultural explosion of artificial intelligence. Drawing on Lotman's notion of "explosion," it aims to describe the "rise" of generative AI by retracing its cultural evolution and highlighting the features emphasized by the Russian semiotician: the sudden crossing of "boundaries" by extra-cultural or extra-semiotic entities, the tensions it generates between centre and periphery, conflicts among heterogeneous elements and systems, and the undeniable sense of unpredictability and destabilisation it produces. From this perspective, it focusses on the ways in which media technologies for the automatic generation of multimedia content have crossed the "laboratories" of computer science and entered the "pop" media ecologies of digital societies.

*Digital Age in Semiotics & Communication*, Vol. VIII, 2025, 7–18

<https://doi.org/10.33919/dasc.25.8.1>

Explosion is a phenomenon that is both exceptional and cyclical. Lotman (1992) explains it through the image of a pendulum swinging between a state of explosion and a state of organisation, the latter unfolding through gradual processes. Media history and sociology of technology teach us not to take techno-cultural explosions too seriously. The development of movable-type printing in the sixteenth century, or the emergence of the Internet in the 1980s, for example, marked large-scale sociotechnical innovations. Nevertheless, even if on various occasions they have been defined as “disruptive,” they represented simple culminating points of gradual events which culture has always undergone while evolving over the centuries. In this view, the cultural explosion of AI could reasonably be placed within this trajectory, and its revolutionary scope reduced. However, reducing its flagrancy would not be productive for grasping the meanings generated by the unpredictability and destabilisation that AI produces.

The Lotmanian explosion, in fact, is both a temporal and topological dynamic, but also a thymic one. This needs to be taken into account, in order to maintain a semiotic approach, insofar as a science of signification must be able to produce descriptions “from within.” From this perspective, however, the “signs of cultural explosion” must be able to account not so much for collective emotions (of fear, anxiety, or excitement) but for the collective processes of textual production and autocommunication. They will explain how, in a given historical period, these explosive dynamic entities will be governed by “filtering” operations. Indeed, for Lotman the cultural explosion precisely describes the translational operations of autocommunication which “domesticate” the new, translate it and filter it. Then they will ferry culture toward a new stability which, although precarious, is capable of restoring that holistic equilibrium necessary to semiosis. Thus, if, on the one hand, it is true that the sense of “explosion” varies according to perspective (it changes depending on whether we live it from within, immersed in change, or from a historical distance), on the other hand, we must understand how we pass from one state to the other. We must comprehend how we come to objectify lived experience, to take one’s departure from it. This was precisely, as Torop notes, the programmatic task of the Tartu-Moscow School: to define the semiotics of culture as a science which investigates the functional correlation among different sign systems.

However, even if this approach is certainly useful for deflating the market’s sensationalist rhetorics, it is difficult to enact, as assessing the scope of the effects at the different strata of society which AI commercialisation has produced remain hard to detect and interpret. Therefore, the real question of the cultural explosion of AI is not to determine whether it is a truly dis-

ruptive technology (this can be taken for granted, albeit that the expression “Artificial Intelligence” is “catchy” coinage by McCarthy for the Dartmouth workshop), but rather how do we detect and describe the processes of rationalisation, through which culture normalises the role of AI in the different sectors of society. This is a process described by Lotman as a passage from explosion to subsequent gradualness, which we can reformulate by considering the various ways society is trying to make sense of processes of automatic generation by AIs. Sometimes they are highly abstract, as in the case of the so-called “latent space,” and at times not even explicable to insiders, such as the hidden layers of neural networks. Other times they resort to metaphors and sociotechnical imaginaries from before 2022 (the release date of ChatGPT), while simultaneously trying to produce translations with minimal error.

These processes are, ultimately, interpretive processes.

On the one hand, interpretation seems to characterize the actual functioning of generative AI, in the “mathematical” (and yet not unbiased) translation of media texts produced in verbal and multimedia language into a digitized vector-matrix encoding. On the other hand, interpretation enacts filtering processes of an interpretive nature also in broader forms of self-communication, in which the explosion of AI itself becomes the object of discourses and explanatory trajectories in different epistemic fields which continuously attempt to give shape to unpredictability, identify specific causalities, and therefore coherent predictions. Outlining trajectories is essential to prevent getting lost. So, the discourses of critical media studies, for example, will outline trajectories relevant to big tech, while for producers they will model the choice (and imposition) of developments and the persistence of techno-capitalist and postcolonial paradigms in “click work”. Those in the discourses of the humanities, on the other hand, will seek to make sense of the quantitative explosion of the aesthetics and styles of communication of LLMs by going back to the history of creativity and making possible to denounce the biases and the illusion of heterogeneity, denying that peculiar explosion of informativeness which characterises this kind of phenomenon. Media history, finally, will trace a trajectory linking the cultural explosion of AI to that of the Internet, allowing us to distinguish images of “AI slop” from memes in user-generated cultures, and practices of “distant writing” (Floridi 2025) enabled by LLM-assisted “cognitive artefacts” (Norman 1991) from the participatory writing cultures of Wikipedia.

To do this, it is important to select the right “samples,” in order that the trajectory respect certain “felicity conditions” of the “mode” of existence

(Latour 2012). Lotman, for example, warned us to distinguish between authentic explosion and imitative explosions. The “dandy”, as a typical case of cultural explosion, acts by destroying accepted norms and creating new ones. He differs from someone who simply dresses badly but is still perceived by those around him as a “dandy” who destroys established norms. Likewise, Narayanan and Kapoor (2024) warned about the dangers of “snake-oil AI,” low-performance AI which mimics the aura and the metrics of truly functional AI (in tasks such as pattern recognition in texts and images) in order to legitimise very different tasks. These tasks are typically the prediction of individual social outcomes (personality, job success, recidivism, “trustworthiness,” etc.) where the intrinsic predictability is low.

However, the choice of the right sample does not mean, only, choosing the “best.” There is no absolutely “best” AI. AI is heterogeneous. The information systems which implement it are organised according to that “compartmental” logic of software culture. This is in turn inspired by “traditional” media. If it is true that the explosion of AI follows a topological extension (each country develops its “own” general AI), then it is also true that this also follows intensions: at different levels of society different AIs are used for different purposes.

This fact, on the other hand, leads to questions of a more epistemic nature (which are then the same to which Narayanan and Kapoor arrive). What are the implications of adopting artificial computation (and mediation) tools in environments of knowledge generation and validation such as psychology, education, or ethology? Critiques of the reductionism of the socio-economic sciences and of the digital humanities certainly constitute stock arguments. What are the consequences of the diffusion of these tools on a global scale? The cultural explosion of AI, in this sense, risks being an explosion of Western knowledge and its epistemic practices (Natale et al. 2025).

The multifaceted functions performed by one and the same technology are continually threatened by the “generative creativity” which makes AI not just a single, indissoluble and inevitable entity but also a “universal”, all-purpose one: indeed, generative AI is implemented from financial forecasting to protein design, from recommendation systems to autonomous driving. With such heterogeneity as a starting point, we can speak of a “general AI”, drawing a distinction with the “narrow” one, specialised in specific domains.

The question becomes even more complex. AI is a techno-cultural product which emulates, imitates, inspires human intelligence (or at least what we know about it), and produces a series of short-circuits in the process of

interpretation (Paolucci 2025). In this sense, the introduction of AI in the context of knowledge work seems to be much more striking, and worrying, than in mechanical ones.

This contrast between narrow AI and general AI is naturally accentuated by the recurring claim that artificial intelligence could be a “creative” technology. The recent contribution by Manovich & Arielli (2021–2024) endeavours to redefine creativity as something essentially regulated and “technicalised”, leaving little room for that individual genius to which Western culture has traditionally assigned the creative faculty. However, it is undeniable that the effect of sense generated by any “little machine” response activated by a prompt is that of the naturalness of communication between humans, based on the creativity of speakers and the indeterminacy of ends. For Lotman, creativity is anything but a negative phenomenon. The explosion acts as a stimulus for creative activity, unprecedented connections, making translatable what until a moment before was incommensurable (the case of the “Italian brainrot” is paradigmatic in this sense).

Nevertheless, the problem of the explosion of AI creativity lays not in defining whether and why the contents generated are more original or less than those in the history of “human-generated” culture, but rather of understanding the value of a generalised creative competence. On the one hand, the explosion of AI creativity has prompted a rapid wave of new laws regulating the use of copyrighted works, alongside new editorial practices in education and research, as well as an increased demand for prompt-focused roles in cultural and creative industries, fostering new professions and skill sets aimed at keeping “up with the times.” This is reminiscent of how the Enlightenment spurred the rise of the Encyclopaedists in France. Taken together, these developments pose both a priority and a challenge for institutions as they decide which direction to pursue. On the other hand, however, it also brought emerging concerns about the status of “expert competences” (Marrone 2021). Most “specialists” indeed are concerned about the apparent “expert” status of general AIs such as ChatGPT, Perplexity, or DeepSeek. Many authors, including Montesanti, in the article in this volume, question the pragmatics of the LLM communication in generating effects of authoritativeness and trustworthiness. Still, it could be asked whether AI is not perhaps the result, rather than the cause, of an older phenomenon. This is a phenomenon which passes through disinformation and so-called fake news feeding the digital archives which have been used to train models.

All this is leading the human to grapple with another kind of creativity. This is no longer “textual” or “operative” creativity, relating to the shaping

of matter, but metatextual and metaoperative in nature, relating to the manipulation of procedural parameters which will determine how the machine will do what it will do autonomously. Manovich (2013) had already brought this to light when he spoke of a “cultural interface” as an object and tool used to design cultural media contents by performing discrete and mathematical operations rather than continuous ones. However, interaction with AI is not just a matter of calibration, although in its more professional uses this constitutes an indispensable element. Prompting techniques such as those used to delimit a semantic context, or those used to design of AI agents, are in effect an entire series of technical calibrations, which, however, are mediated through a human-computer language that is not “operative” in the strict sense (one does not manipulate parameters by entering numerical values), but rather through instructions expressed in natural language. Used in full continuity with what has been defined as “remix culture” (Navas 2022) such as the orchestration of automated creations according to a more general logic, we could call this “narrative.”

The contributions gathered in this volume examine this turn from multiple angles: linguistic, biosemiotic, educational, aesthetic, technological, and economic. They offer a panoramic view of how AI not only produces new signs, but reshapes the conditions under which signs operate. Across domains as diverse as language acquisition, animal communication, social media, facial recognition, XR, and fashion, the authors converge on a shared insight. The emergence of AI as a semiotic actor necessitates a reconsideration of what counts as communication, interpretation, and creative agency.

In their panoramic paper “A Step-by-Step Semiotic Understanding of LLMs and Chatbots” **Thibault & Allois** bridge semiotics and machine learning to provide a structured account of the evolution of large language models. By examining perceptrons, neural networks, and transformer architectures, the authors clarify the technological processes behind generative AI and counter the tendency toward anthropomorphisation. They propose conceptual tools, including the idea of an AI semiosphere, a generative mode of sign production, and *Intentio Machinae*, in order to situate LLMs within semiotic theory. The study reframes AI as a support for human textual activity rather than an autonomous interpretative agent.

In “The Peircean Theory of AI”, **Friedman & Thellefsen** apply Peirce’s triadic semiotics along with speculative grammar and methodotics, in order to reassess the interpretative limits of large language models. They argue that AI’s statistical foundations hinder its ability to capture relational meaning essential for human communication. Integrating Claudio Paoluc-

ci's theory of machinic enunciation, the authors propose a functional re-interpretation of AI-generated texts within a Peircean framework. Their analysis demonstrates how classical semiotic theory can inform attempts to enhance contextual awareness in AI systems and illuminates the persistent challenges in modeling genuine semiosis computationally.

In "On the Edge of Understanding: ChatGPT and the Limits of Artificial Sense-Making", **Montesanti** examines the semiotic limits of generative AI by contrasting the adaptive, context-sensitive practices of human speakers with the mechanical training logic of large language models. Drawing on Saussure, Jakobson, Hymes, and Lotman, he argues that AI replicates communicative exchanges without participating in the situated, relational pact which characterises human dialogue. Through the analysis of selected interactions with ChatGPT, the paper explores whether artificial systems can meaningfully partake in sense-generation, suggesting that the notion of "sense" itself may require semiotic redefinition in the age of generative machines.

In "The Semiotic of AI Images" **Schneider** applies Susanne Langer's distinction between discursive and presentational symbolic forms to the analysis of text-to-image generation. She argues that AI-generated images arise from the collision between linguistic prompts—linear and discrete—and images which communicate meaning in a simultaneous and relational mode. The paper explores how these symbolic tensions shape the structure and communicative function of AI imagery. Rather than evaluating their aesthetic value, the study investigates whether such images contribute to clarifying ideas, drawing on Langer's and Peirce's theories to read AI imagery as a new medium of conceptual expression.

In "The Janusian Face of Facial Recognition, Part 1" which will be followed in the next issue of the journal by Part 2, **Schiller** conducts a critical semiotic "deblackboxing" of facial recognition technologies. It examines how computational, representational, and relational components converge in these systems. Positioned within Peircean and pragmatist computational semiotics, the paper interrogates whether AI-driven facial recognition possesses genuine semiotic agency or merely produces quasi-sign processes. By distinguishing between subface, interface, and surface, Schiller reveals the multilayered nature of facial recognition systems and situates their functioning within broader debates on artificial intelligence, sign processing, and technological agency.

In "Are We Human or Are We Dancer? – AI Creativity in XR" **Alexiev** investigates the underexplored semiotic and cultural consequences of integrating AI into extended reality environments. While current research

emphasises technical efficiency and user interaction, this paper highlights the lack of attention to AI's artistic and social impact within XR. Drawing on both theoretical studies and cutting-edge XR productions, the author discusses how AI is reshaping creative practices and networking potentialities in immersive spaces. The analysis calls for a renewed understanding of AI-XR entanglements as emergent topologies of meaning within the contemporary semiosphere.

In "Humans as Natural-Born Cyborgs: Scrutinising AI's Narrative Intelligence within the 5E Cognition Framework", **Livytska** examines the narrative limitations of large language models, arguing that their capacity to generate coherent text is often mistaken for intelligence. Drawing on the folk narrative hypothesis, the paper highlights storytelling as a uniquely human mechanism for memory, identity, and sense-making, which AI can mimic structurally but not inhabit experientially. Through an analysis of ChatGPT-4's prompting processes and its handling of textual categories within the 5E Cognition framework, the study shows how AI-produced narratives gain meaning only through human interpretative engagement. The seeming intelligence of generative models thus reflects not machine understanding but our innate drive to impose coherence and treat communication partners as intentional agents.

In "Needle in a Needle Stack: Semiotic Inflation and Experiential Devaluation", **Champagne** introduces the concept of "semiotic inflation" to describe how AI's escalating production of signs risks devaluing human experience. Drawing analogies with monetary systems and scarcity-driven value, he argues that oversaturation of linguistic and visual content may undermine evolutionary and cultural balances in sign-object ratios. Rather than focusing on what AI allegedly cannot replicate, the paper identifies systemic negative experiential effects as evidence of counterfeit semiosis. The argument highlights the need for conceptual limits on sign proliferation, in order to preserve the integrity of human meaning-making environments.

In "Translating the Wild: AI, Semiotics, and the Future of Animal Communication", **Zengiaro** critically evaluates AI-based attempts to decode animal communication, focusing on aquatic species such as dolphins and whales. Grounded in biosemiotics, ecosemiotics, and Umwelt theory, the paper shows how projects such as DolphinGemma and CHAT risk reducing embodied semiosis to computational patterns. Instead of transparent translators, these systems should be understood as technosemiotic infrastructures which can enable new forms of interspecies resonance when embedded in relational and critical frameworks. The article advocates for an

ecotechnical semiotics which resists reductionism and reconceives communication as emergent and materially grounded.

In “ChatGPT in Higher Education: Between Cultural Explosion and Encyclopedic Knowledge”, **Petrassi** explores ChatGPT’s transformative impact on academic knowledge production, framing generative AI as a catalyst of Lotmanian “cultural explosion.” Mobilising Eco’s encyclopaedic model, the study examines how ChatGPT mimics human interpretative processes while lacking intentionality, thereby challenging traditional epistemic norms. Through mixed-methods research combining qualitative inquiry with a student survey, the paper analyses perceptions of trust, bias, and critical engagement. The findings reveal cautious yet widespread adoption, positioning ChatGPT as a semiotic mediator that reshapes pedagogical practices and demands new forms of critical AI literacy.

The paper by **Ingrao & Mirsonbol** “AI, Semiosis, and the Future of Language Acquisition” continues the same topic by evaluating the implications of AI-driven tools for second-language acquisition, aligning the discussion with EU 2030 educational goals and semiotic theories of learning. The authors propose a semiosis-based pedagogical model informed by Barthes, Eco, and Lotman, emphasising personalisation, ethical considerations, and the dynamic interplay between student, teacher, text, and AI. Through a critical review of current practices and a theoretically grounded model of personalised learning, the study argues that AI can support deeper engagement with L2 content when embedded within culturally and semiotically informed frameworks.

In “Consumer Attitudes to AI Content Generation in Social Media”, **Varshev & Marinov** offer an empirical study which investigates how social media users perceive AI-generated content. Based on a large survey across diverse demographics, the results reveal a high level of awareness but predominantly negative attitudes toward AI-produced posts which many respondents deem inauthentic and unengaging. Concerns about privacy and a strong demand for transparency shape consumer expectations, with most participants advocating the clear disclosure of AI use. A case analysis of Coca-Cola’s AI-generated advertisements illustrates this tension, showing a preference for human-made content. The findings highlight the need for marketing strategies which integrate AI without undermining trust or authenticity.

In “Use of AI in the Context of Fashion and Related Industry”, **Orlova** explores how AI transforms both the technological operations and cultural dynamics of the fashion industry. Drawing on Lipovetsky’s consumer theory, the semiotics of Barthes and Lotman, and contemporary analyses by

Manovich and Arielli, the paper situates fashion as a dynamic semiosphere where AI enhances personalisation, efficiency, and aesthetic innovation. While AI expands creative possibilities and diversifies cultural expression, the study underscores the continued relevance of human sensibility in shaping fashion's symbolic forms and experiential dimensions.

Taken together, these contributions trace a landscape in which AI is emerging as a tool, an interlocutor, a sign-producer, and a catalyst of cultural change. While none of the authors grant AI human-like interpretative agency, all recognise its expanding role in reshaping semiotic environments, broadening communicative possibilities while raising new tensions around authenticity, embodiment, and saturation. Rather than defining what AI "is," this volume highlights how it intervenes – materially, symbolically, and institutionally – in the ongoing evolution of semiosis. It invites readers to reconsider their assumptions about semiotics, language, representation, and cultural production at a time when artificial systems are increasingly intertwining with human meaning-making.

Finally, the gathering in Sozopol on the occasion of the XXVIII EFSS marked an important step for the ERUA alliance and its research agenda. The event brought together at least 20 participants from the eight partner universities and provided the setting for the initiation of the research cluster Artificial Intelligence and Creative Industry Employment Disruption (AICIED).

**We wish you engaging and thought-provoking reading.**

## References

Bankov, K. 2020. "Platfospheres and Sociocultural Explosion of Web 2.0: The Commercial Centre of the Digital Semiosphere". *Sign Systems Studies*, Vol. 48, No. 2–4, 246–70.

Bogost, I. 2010. *Persuasive Games: The Expressive Power of Videogames*. The MIT Press Ser. Cambridge: MIT Press.

Boomen, M. van den. 2014. *Transcoding the Digital: How Metaphors Matter in New Media*. Amsterdam: Institute of Network Cultures.

Bory, P. 2020. *The Internet Myth: From the Internet Imaginary to Network Ideologies*. Critical, Digital and Social Media Studies. London: University of Westminster Press.

Certeau, M. de. 1990. *L' invention du quotidien. Arts de faire*. Nouvelle éd. 1. Paris: Gallimard.

Chun, W. H. K. 2005. "On Software, or the Persistence of Visual Knowledge". *Grey Room*, No. 18, 26–51. Available at: <https://doi.org/10.1162/1526381043320741>.

Chun, W. H. K. 2016. *Updating to Remain the Same: Habitual New Media*. Cambridge, MA: The MIT Press.

Cometa, M. 2024. *La svolta ecomediale. La mediazione come forma di vita*. Meltemi.

Crary, J. 2022. *Scorched Earth: Beyond the Digital Age to a Post-Capitalist World*. London; Brooklyn, NY: Verso.

Dodge, M., R. Kitchin. 2001. *Atlas of Cyberspace*. (1st ed.). Harlow Munich: Addison-Wesley.

Eco, U. 1994. *Semiotica e filosofia del linguaggio*. (3. rist). Einaudi paperbacks 151. Torino: Einaudi.

Galloway, A. R. 2006. *Gaming: Essays on Algorithmic Culture*. Electronic Mediations 18. Minneapolis, Minn: University of Minnesota Press.

Giuliana, G. T. 2024. *Il Videogioco Come Linguaggio Della Realtà. Introduzione a Una Nuova Prospettiva Semiotica. Vol. 1*. Roma: Aracne.

Gomez Mejia, G. 2014. "De Quoi Le « nuage » Est-Il Le Nom ? Le Statut Des Supports Face Aux Régimes Du Cloud Computing". *Communication & Langages*, Vol. 182, No. 4, 77–93.

Hartley, J., I. Ibrus and M. Ojamaa. 2021. *On the Digital Semiosphere: Culture, Media and Science for the Anthropocene*. New York: Bloomsbury Academic.

Hayles, N. K. 2002. *Writing Machines*. Mediawork Pamphlet. Cambridge, Mass: MIT Press.

Hillis, K. 1999. *Digital Sensations: Space, Identity, and Embodiment in Virtual Reality*. Electronic Mediations, v. 1. Minneapolis, Minn. London: University of Minnesota Press.

Hu, T.-H. 2015. *A Prehistory of the Cloud*. The MIT Press. Available at: <https://doi.org/10.7551/mitpress/9780262029513.001.0001>.

Lakoff, G. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago London: The University of Chicago Press.

Lakoff, G., M. Johnson. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press. Available at: <http://archive.org/details/metaphorsweliveb0000lako>.

Manovich, L. 2001. *The Language of New Media*. 8. print. A Leonardo Book. Cambridge, Mass: MIT Press.

Maran, T. 2020. *Ecosemiotics: The Study of Signs in Changing Ecologies*. (1st ed.). Cambridge University Press. Available at: <https://doi.org/10.1017/9781108942850>.

Murray, J. H. 2001. *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. 4. print. Cambridge, Mass: MIT Press.

Nake, F. 2008. Surface, Interface, Subface. Three Cases of Interaction and One Concept. In Seifert, U., J. H. Kim and A. Moore (eds.). *Paradoxes of Interactivity. Perspectives for Media Theory, Human-Computer Interaction, and Artistic Investigations*. Transcript Verlag, 92–109. Available at: <https://doi.org/10.25969/MEDIAREP/2719>.

Pasquinelli, M., Institute of Network Cultures (Amsterdam, Netherlands). 2008. *Animal Spirits: A Bestiary of the Commons*. Studies in Network Cultures. Rotterdam, The Netherlands [Amsterdam] New York, NY: NAI Publishers; Institute of Network Cultures; Available in North, South and Central America through D.A.P.

Prensky, M. 2001. “Digital Natives, Digital Immigrants”. *On the Horizon*, Vol. 9, No. 5, 1–6. Available at: <https://doi.org/10.1108/10748120110424816>.

Reyes, E. 2017. *The Image-Interface*. Hoboken, NJ: ISTE Ltd/John Wiley and Sons Inc.

Scolari, C. A. 2012. “Media Ecology: Exploring the Metaphor to Expand the Theory”. *Communication Theory*, Vol. 22, No. 2, 204–225. Available at: <https://doi.org/10.1111/j.1468-2885.2012.01404.x>.

Siegert, B. 2018. “Coding as Cultural Technique: On the Emergence of the Digital from Writing AC”. *Grey Room*, 70 (March), 6–23. Available at: [https://doi.org/10.1162/GREY\\_a\\_00236](https://doi.org/10.1162/GREY_a_00236).

Singler, B. 2020. “The AI Creation Meme: A Case Study of the New Visibility of Religion in Artificial Intelligence Discourse”. *Religions*, Vol. 11, No. 5, 253. <https://doi.org/10.3390/rel11050253>.

Thibault, M. 2015. Do Not Talk about Anonymous. Censura, autocensura e anonimato nelle periferie del web. In *Lexia. Rivista di semiotica*, 11. Roma: Aracne, 237–254. Available at: <https://doi.org/10.4399/978885489127216>.

Thomas, S. 2013. *Technobiophilia: Nature and Cyberspace*. A&C Black.

Volli, U. 2021. “Per Una Tipologia Degli Spazi Della Rete”. *E|C Rivista Dell’Associazione Italiana Di Studi Semiotici*, XV, 31, 25–31.

Zielinski, S. 2006. *Deep Time of the Media: Toward an Archaeology of Hearing and Seeing by Technical Means*. Electronic Culture: History, Theory, Practice. Cambridge, Mass: MIT Press.

## A STEP-BY-STEP SEMIOTIC UNDERSTANDING OF LLMS AND CHATBOTS THROUGH INTERDISCIPLINARY DIALOGUE

*Mattia Thibault  
Tampere University  
mattia.thibault@tuni.fi*

*Daniele Allois  
Independent Researcher  
daniele.allois@gmail.com*

### **Abstract**

Artificial Intelligence (AI) is enjoying a period of “summer”. It is an intensely discussed and researched topic which fascinates researchers across fields with questions about both its development and possible impacts. Semiotics too has immediately engaged with the topic, focusing especially on the ways in which different forms of Generative AI can be understood through existing frameworks, and how the processes of textual production it encompasses can be studied and analysed.

*Digital Age in Semiotics & Communication*, Vol. VIII, 2025, 19–49

<https://doi.org/10.33919/dasc.25.8.2>

Gen AI, however, is a complex and fast evolving technology, which is very difficult to study without a sufficient understanding of the technical side. Perceptrons, Artificial Neural Networks, Transformer Networks, Large Language Models are all terms we hear often in discussions about AI. However, they often refer to technologies we barely understand.

This paper is born from a collaboration between authors with expertise in semiotics and machine learning. Our objective is to reconstruct the evolution of Large Language Models in particular, and to engage the different stages from a semiotic perspective. This in-depth engagement with the technological side is especially useful to go beyond the strong temptation to anthropomorphize the technology and instead reposition it as a tool that supports human textual creation. Based on our analysis we then propose several key concepts (an AI semiosphere, Generative AI as a mode of sign production and the concept of *Intentio Machinae*) as ways to map and conceptualise the role of LLMs in semiotic processes.

**Keywords:** Generative Artificial Intelligence, Cybersemiotics, *Intentio machinae*, Large Language Models, Chatbots

## 1. Introduction

Artificial Intelligence (AI) is enjoying a period of summer. AI technologies periodically go through periods of great enthusiasm and others of deep disillusionment – sometimes called summers and winters of AI (Haugeland 1949). New technical solutions bring along renewed hope for impactful applications of the technology. However, they often reach an impasse when the technology does not deliver what the public hoped for. At the present time, Large Language Models (LLMs) and Generative AI have greatly advanced the ability of machines to perform tasks which until recently were thought to be exclusively human, especially in terms of understanding and creating that which, in semiotic terms, we would call different forms of signs and texts.

From written and oral texts, to images, videos and even 3D settings Generative AI appears to be able to engage with a series of semiotic activities in a way which, at times, seems indistinguishable from a human being. While machines creating texts with the input of humans is not necessarily a new thing – cameras do something similar, for example – the quantity of data on which they are trained and the minimal input required by humans by Generative AI in order to create new texts is indeed unprecedented.

This poses some interesting questions for semioticians, as we are seemingly facing a machine not only able to generate meaning, but also to make interpretative choices when confronted with semiotic material. But is this what is really happening? What are the semiotic entails of Generative AI?

This paper is the result of a dialogue between two different areas of expertise: semiotics and machine learning, embodied by its two authors. Our aim is to look at the technical steps that brought us to the current state of AI development and engage each from a semiotic perspective. This should both offer a clear and simple explanation of the computational operations behind Generative AI, and some semiotic insights on how to look at them from the perspectives of meaning-making and interpretation.

Since there are many different forms of AI, we decided here to focus on LLMs, as their ability to use natural languages and to act as conversational agents (or Chatbots) is particularly interesting from a semiotic perspective.

## **2. Background: Semiotics and AI**

While a relatively new phenomenon, Generative AI has been able to excite the curiosity of many semioticians. The discourses around its development range from the most techno-optimistic to the most catastrophic predictions. Nevertheless, all seem to suggest that we are facing something unprecedented: a claim which can also sound like a challenge. Are we now facing a different kind of semiotic actor, or is this just a statistically-powered – although quite convincing – “trick”? Semiotics, of course, has something to say on the topic.

As early as 2020, in reaction to the development of Natural Language Processing (NLP), Duncker (2020), was suggesting caution, claiming that, while chatbots may be perceived as communicating with human users in natural languages, one should not forget that this is possible only through a second-order language model, which governs the way the machine reacts to human input – a reaction that is still limited to a specific set of pre-programmed executable skills. In other words, Dunker suggests that chatbots do not participate in communicative processes, and that the signs it produces are “unsigns” which the machine does not interpret nor engage with as such (Duncker 2020: 97).

In the same year, Sanna (2020) already began to wonder about the role of semiotics, as a discipline possibly developing AI models. He proposed the idea of “semiotics-driven machine learning”. This seems to resonate with a feeling that many have in the discipline: the idea that, as Generative AI, NPL, and LLMs work with natural languages they belong, in a way, to semiotics.

Semioticians, after all, have been discussing the role of cybernetics in semiotics for decades, including Eco's explorations on the possibility of computers telling stories (Eco 1994), and Lotman's hopes for an integration of semiotic systems and informatic ones capable of learning from human semiosis (Semenenko 2012).

This perception, that semiotics cannot only say something important about Gen AI, but that semiotics *has* to do it, might have played a role in the flourishing of studies dedicated to the topic in the last couple of years.

One of the first, ambitious attempts to coordinate these efforts is the volume edited by Santangelo and Leone in 2023. It contains 12 chapters from different semioticians – the majority of which belong to Turin's semiotic circle – which engage Generative AI from multiple perspectives (Santangelo & Leone 2023). Within the issue there are two attempts to clearly outline the dimensions of AI which concern semiotics. One is Ferraro's (2023), which offers a series of reflections, highlighting the mythology of AI (the utopian and ideological depictions of the technology). He examines similarities and differences with human intelligence, the question of creativity, the question of classification, and finally the degree in which the machine can *grasp* meaning. The other is Leone (2023), which proposes a similar list, including: how AI simulates intelligence, the production of creative content, and the ideological assumptions around AI in our culture. Engaging with this last point, especially focusing on Adversarial Networks, Leone argues that the idea of "fakeness" is central to understanding the semiotic aspects of Gen AI.

The issue also contains certain works which focus on the discursive context surrounding Gen AI such as Santangelo (2023) looking at representations of AI in science fiction, and Stano (2023) engaging with the relations between AI and human and artificial bodies.

Several other papers, instead, engage with the *products* of Gen AI, i.e. the textualities which emerge from these machines, and their possible sociocultural effects. Gramigna (2023) does it for deep fakes, Voto (2023) for climate data, Soro (2023) for motivational messages in Fitness applications.

A few more papers engage with certain applications of Artificial Intelligence, such as sentiment analysis (Galofaro & Toffano 2023, and Ferreira Leite da Silva 2023) and chatbots (Dall'Acqua & Bellentani 2023) and what semiotic theory has to offer to their refinement.

Finally, two papers approach AI by engaging with the historical context of their development – both technological and ideological. The first is Giuliana (2023), who highlights the fundamental role of games in the development of AI – since they are spaces for both imitating human behaviour,

but also opposing it. If most Gen AI models we encounter today are meant to be our helpers, in games it is often the opposite, the AI developed for games participate in the obstacles we have to overcome while playing. Giuliana offers what he calls a “performative definition of intelligence”, highlighting how often the definition of intelligence is based on the competence of an actor to do something, to perform a specific task. Giuliana’s chapter reminds us also how, while today’s discourses tend to conflate AI with Generative AI, there are a plethora of technologies which are currently and concurrently defined as “Artificial Intelligence”. This is more due to the perception we have of them than because actual similarities in what they do.

The second is Volli (2023), who provides a useful reconstruction on the evolution of the idea of the possibility to make natural languages as precise and infallible as mathematics: from the search for a perfect language to Turing writings and modern ideals of AI. Volli claims what we call AI is clearly not an analogue to human intelligence – referring also to Searle famous argument about the “Chinese Room”<sup>1</sup> – but that does not make it a less interesting object of study for semiotics. If we abandon the idea of an AI which imitates the semantics of human intelligence – and therefore the mythological quality of the pursuit for an objectivization of language – we can instead focus on what Volli considers most interesting in machine learning. This is the ability to categorise and classify signs and information – a question which has been of great interest for semioticians. Volli describes machine learning as actuating a process of inference which is retro-actively acting on its own parameters. It is, hence, an implementation of the abductive mechanism which is at the base of the semiotic-cognitive functioning. In other words, machine learning is based on the ability of the machine to formulate interpretative hypotheses about the data upon which it is trained. Then after evaluation, to change the parameters on the base of which its hypotheses are formulated.

More recently, *Semiotica* has dedicated a special issue to “Aspects of AI semiotics: enunciation, agency, and creativity”, which showcased the continued interest in the discipline for the topic and contains some important contributions. The issue programmatically aims to go beyond the semiotic

---

<sup>1</sup> Searle’s example famously imagines a person in a closed room receiving messages from under the door written in Chinese ideograms. The person does not know how to read Chinese, but has access to a program which informs them on how to select the right characters to send back under the door in response. People outside the room, seeing that the messages they are getting back meaningfully respond to those they sent, might assume – wrongly – that the person in the room is fluent in Chinese. The same, then, can be said about an artificial machine which might be capable of responding in a way interpreted as meaningful by human interactors, but that is not necessarily able to interpret what it is doing.

properties of AI generation, and instead focus on analysis of different socio-cultural practices involving generative AI.

A number of papers focus especially on the relation of AI with archival images and texts. Dondero (2025) engages with *enunciative praxis* and the relation between the sedimentation of stereotypical forms in visual culture (contained in the archived images on which the AI models are trained) and the possibility of actualisation – or creation of new images – of visual transformer models. D’Armenio (2025) works in a similar area, proposing the idea of “archival perception” as the particular way on which AI perceives and learns. The author also underlines how Generative AI works as a *co-enunciating machine*, highlighting the necessity of human intervention in textual generation and the role which AI plays in it.

Two papers also focus on how this hybridity and collaboration between human and machine is not necessarily new. Paolucci (2025) challenges the illusion of human cognitive exceptionalism by arguing that language-endowed machines reveal and underline our preexisting machinic hybridity. According to the author humans are “natural-born cyborg,” whose cognition has always been scaffolded by external systems. Colas-Blaise (2025), on the other hand, questions the ability of the machine to co-enunciate, and rather proposes that “machinic enunciation,” is an enunciative sequence which sees human agency in the initial and final phases, with numerous machinic instances in between.

Other papers focus on interpretative questions: Compagno (2025) argues that the notions of “truth” and “falsehood” are not well suited to describe the textual production of AI models. Basso Fossali (2025) speculates whether AI could lead to “augmented semiospheres” where a third order of cultural remediation is offered by chatbots. Valle (2025), instead, focuses on the epistemological shift in coding, in which instead of attempting to reconstruct grammars, machine learning attempts to reconstruct texts by imitation. Finally, Montanari (2025), while focusing on the socio-semiotic implications of these technologies, offers some interesting insights on the role of spatiality, statistics and distributional semantics in the meaning-making processes of generative models.

Despite the wealth of insights, much of the current research on Gen AI in semiotics seems to be conducted in isolation from the technical expertise that is behind its development. However, there are two recent series of studies based on bringing together a varied pool of expertise.

The first is brought forward by Everardo Reyes who has dedicated two articles (Reyes 2024; Reyes 2025) in which he looks at generative models (and not only at Generative AI) through the joint lenses of semiotics

and computer science. These works – among the few which also adopt a Peircean perspective on the topic – offer a wide overview both at the many different types of generative media, and at the different operations they can perform.

The second series has been conducted by the trio Enzo D’Armenio, Adrien Delière and Maria Giulia Dondero. These studies are part of two projects funded by the *Direction du Projet De Recherches (PDR) F.R.S.-FNRS*: one focused on semiotic and Computer-Assisted approaches to large collections of images; and the second on Generative AI and image production. These projects have enabled Adrien Delière, mathematician and computer scientist, to offer technical insights to semiotic research.

In their works, they focus especially on Midjourney and DALL-E, and therefore on Generative AI models focused on image generation. D’Armenio et al. (2024a) first proposed a key argument in their work: the fact that Generative Models are *co-enunciators* with their human prompts of the texts they generate. In particular, they conduct a series of experiments exploring different forms of prompting and analysing the results which they yield.

In D’Armenio et al. (2024b), the authors propose an evaluation method, in order to assess the effectiveness of text-to-image generation models. This paper presents, in our opinion, two key points of interest. First, it acknowledges directly and explicitly the importance of (intersemiotic) translation when it comes to Visual Generation models. The second key point of interest is methodological. Semiotics is traditionally concerned with the analysis of texts which, in the vast majority of cases, precede the analysis. Studies on Gen AI, however, are able to ask the machine to generate texts at any time, and then analyse them – and the process which led to their creation. This is a significant methodological change: one that has not yet been thoroughly examined by semioticians, and who until now seem to have mostly dealt with it in an intuitive, almost naive way. This study, while also proceeding intuitively, presents an interesting dialogical way of proceeding, testing back and forth the results of image generation. It makes use of prompts which are systematically prepared to test the limits of the models studied. In Delière et al. (2025) they put this model and methodology to test, focusing on the artistic categories of “classical” and “baroque”.

Finally, D’Armenio et al. (2025b) engage more globally on the functioning of image generator modelling, especially looking at their results as a particular kind of intersemiotic translation. While the paper later engages with the control of human prompts on the composition of such images, we particularly appreciate that this work engages explicitly with the technical operation behind image generation. Without engaging with the back-

ground of machine translation, the authors still engage with the technical steps which separate a prompt from the final image – something that we believe central if we want to properly address the semiotic functioning of these machines.

To our knowledge, this is a rare endeavour. There is a wide breadth of studies on semiotics and AI (to which this summary does justice only partially) which deal with many different aspects from their inner workings to their societal implications. However, there still seems to be a clear gap when it comes to examining in detail the technical side of Machine Learning – and to analyse their semiotic entails. In this paper, then, we will initiate an interdisciplinary dialogue to look at the different steps in the evolution of machine learning and examine them from a semiotic lens. In the next paragraph we will then engage with the fundamentals of machine learning, and more specifically on the bases for the development of LLMs and chatbots. In section 4 of this paper, then, we will propose some semiotic implications which will build on our step-by-step analysis and use several concepts from semiotic theory to explain, or better understand, the functioning of LLMs.

### **3. Engaging with the fundamentals of Machine Learning**

#### **3.1. Machines which can learn like living things**

The idea of a computer capable of learning, and therefore of the possibility of an “Artificial Intelligence”, was formulated in the mid 20th century. Alan Turing himself stated that: “What we want is a machine that can learn from experience,” and that “the possibility of letting the machine alter its own instructions provides the mechanism for this.” (Turing 1948). Those remarks were at the beginning of a long journey of technical development in the field of “Machine Learning” which has currently led to the development of different implementations of Generative Artificial Intelligence, such as Large Language Models (LLMs).

How can a machine capable of learning be created? In order to answer this question computer scientists started looking at an existing, complex, system capable of learning: the brain. More specifically, Machine Learning was strongly influenced by a book published in 1949 by Donald Hebb called “The Organisation of Behavior” (1949). Hebb looked at the working of animal neural networks, in order to understand how their processes could be replicated artificially. Hebb in particular observes how the physical connections of the nervous systems of different individual organisms differ even in the same species. The portions of nervous systems that are involved in learning and recognition are largely random at the birth of an

individual. However, such systems are characterised by a certain degree of plasticity: the neurons change based on neural activity. Exposure to similar stimuli forms and strengthens pathways that lead to the same responding cells. At the same time, dissimilar stimuli will develop connections to different sets of responding cells. Positive or negative reinforcement can then be used to facilitate – or discourage – the formation of these connections.

“Similarity”, then, is defined by Hebb as being a tendency of similar stimuli to activate the same sets of cells. This is not determined by specific formal requirements of the class of stimuli, but rather by the physical organisation of the perceiving system and the interaction with the organism’s environment. Hebb concludes that: “the structure of the system, as well as the ecology of the stimulus environment, will affect, and will largely determine, the classes of “things” into which the perceptual world is divided”.

These statements are not surprising to semioticians: the idea that *difference* is at the base of meaning making is as old as the discipline (Saussure 1916), and biosemiotics has often engaged with the relation between nervous systems and semiosis (Kull & Favareau 2022). However, what is relevant in our argument is that the structural functioning of the nervous systems – the creation and reinforcement, over time, of specific neural pathways – became a key inspiration in the development of Machine Learning.

### 3.2. The perceptron

Inspired by Hebb’s work, computer scientists (see Rosenblatt 1958) started to look for a way to replicate the natural ability of organic nervous systems to learn and adapt. The first step in that direction was the creation of the *perceptron*: an algorithm capable of “learning” based on information fed to it. This was a significant development compared to traditional forms of coding, in which programmers have to explicitly write all the rules that the machine will follow.

The core idea behind the perceptron is that of a function which can be trained by engaging with a curated set of inputs so that it will later become capable of conducting a simple operation precisely when confronted with new inputs. The elements making a perceptron are the following<sup>2</sup>:

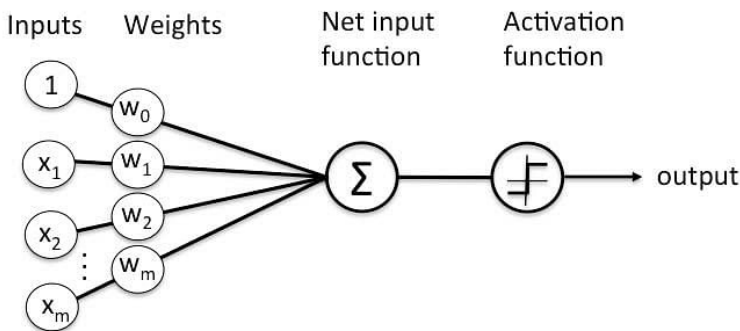
- The inputs. These are the different parameters which the algorithm will take into consideration. They are numerical values.

---

<sup>2</sup> There is, in fact, another component that is the *bias*. This is an addendum to the network that facilitates its process from a mathematical standpoint. This addendum is used to make the perceptron produce an output even when the inputs are negative. Since it does not impact the structural working of the perceptron, and it makes it more complex to understand, we decided to engage with it only in this footnote.

- The weights. These are numerical indicators which indicate the relative value assigned to each input.
- The output. This is a simple binary opposition which will consist either of 1 (activation of the perceptron) or 0 (inactivation).

As shown in Fig. 1, the perceptron works in the following way: the Net Input Function collects the different inputs received by a perceptron, each multiplied on base of their weight, and sums them in a final value. Then, the last part of the algorithm, called the Activation Function will, depending on the summation value, either output a 1 (activation) or 0 (inactivation).



**Figure 1:** Diagram of a perceptron (2009–2025 – Simplilearn Solutions)

The structure of the perceptron as presently described does not yet allow machine learning, but rather explains what kind of operation the perceptron engages in: it evaluates a certain input based on predetermined parameters.

For example, a perceptron could be used to evaluate the cost of real estate. A simple perceptron accepting only one kind of input, could process the price of a house and, based on the weights assigned by its creators, activate if the price is over a certain threshold and flag it as “expensive” – or not activate and flag it as “inexpensive”. This is not a particularly impressive feature per se, since it basically only checks if the price of the house is above or below a certain price. Things become more interesting, however, with a perceptron that accepts multi-feature inputs – that is, different kinds of data. In this case the inputs could include the price, but also the square footage, the number of bathrooms, the distance from the city centre, and any other relevant numerical data. The final output would be the same. The perceptron either activates if the house is deemed expensive or cheap, but the complexity of the estimation is increased. Thus the difficulties for its creators to assign the correct weights to the different parameters are also increased. How does the number of bathrooms trade off with the distance

from the city centre? Machine Learning comes into play exactly to avoid humans making such calculations.

Before engaging with it, however, let us look at the working of the perceptron from a semiotic standpoint. In semiotic terms, we can look at the functioning of the perceptron as an actant (Greimas 1970), or as part of a specific Actor Network which participates in a clear narrative programme (Latour 1993). In this case, we can say that the perceptron operates a *sanction*. The perceptron evaluates the input which it receives and, based on a set of criteria (the weights that have been assigned to the different possible inputs), generates a result. It either activates, or it does not. It is a simple form of sanction which gives a binary result, but which is based on an automated form of *discrimination*.

Metaphorically, we could understand it as a sort of *sieve*. The sieve is a machine “programmed” by humans (which designed the caliber of its holes) to make a simple distinction. Objects of the right dimension (for example, seeds) pass through, while the rest do not. In a similar way, the perceptron organises data into two groups by activating or not. In either case, the act of interpretation – or discrimination (or classification, as it’s called in technical terms) – is entrusted to the machine, but predetermined by the humans who created the machine.

The perceptron, hence, does not work dissimilarly from the light switch described by Eco in a Theory of Semiotics (1976). It is a more complex version of the switch, but it still follows a strict law of necessity. To the same input the same perceptron will always respond in the same way.

### 3.2. The Perceptron learning rule

The key feature of perceptrons, however, is their ability to imitate organic neural pathways and *learn*. Perceptrons go through the *perceptron learning rule*: a step-by-step process which allows the perceptron to modify or set its own weights. This process allows a perceptron to be *trained*, and hence starts to remove interpretation (i.e. determining the weights of all the different inputs) from the hands of its human creators.

The perceptron learning rule works as follows:

- The programmers create a curated set of inputs. For each input they know the desired output (activation or inactivation) but they ignore what exact weights would generate them.
- The perceptron is a series of cycles (named “epochs”) in which it makes predictions about its own activation when facing the curated set of inputs.

- At each cycle, the perceptron confronts its own prediction with the predetermined correct course of action.
- If the prediction was incorrect, the perceptron will alter its own weights by adjusting them a bit in the direction of the right answer.
- The perceptron continues to pass through a series of epochs, until its predictions become accurate enough that it can be exposed to new, not pre-classified input, and be able to correctly assess the right responses.
- The perceptron learning rule, then, allows the perceptron to set up and evaluate the best set of weights, in order to offer accurate outputs, without the need for its programmers to define those weights themselves.

Using the real estate example again, a multi-feature input perceptron could be trained on a set of examples of market valued houses and their respective amenities, square footage, distance from the city centre, etc. This set of examples would still have a human-made evaluation of their price (based, for example, on the fact that the properties found a buyer or not). While making predictions and confronting them with the correct responses, the perceptron will learn how the different features of a house determine its value (that is, how to weigh the different inputs it receives). After the training, it will sequentially be able to estimate if the price of a new house on the market is too high.

The narrative programme of the trained perceptron remains largely the same. It operates a sanction which leads to a binary discrimination. However, the training shifts our focus to a higher narrative level, in which its ability to offer a *sanction* is also a *performance*. This performance itself is then evaluated and is object of its own sanction. We are then facing two narrative layers:

In one, the perceptron operates a sanction towards a preliminary performance which is engraved in the data it receives. The sanction results in a state of activation or deactivation.

In a second layer, this sanction becomes its own narrative programme. Here the perceptron mobilises its own competences (the weights upon which it bases its discrimination) and operates a performance (the discrimination itself, so the activation or inactivation). It is then another piece of code which sanctions such performance, determining if the perceptron was successful or not, and determining the consequences of failure – which result in a modification of its competences.

Repetition of this exercise will lead to the generation of a perceptron with competences evolved enough to ensure a maximum chance of success in its performance.

At the same time, we also face a multiplication of narrative programmes within the same system. Here we have three main steps of machine learning: the “generation” (in this case of a prediction), the “evaluation” (of the correctness of the prediction), and the “modification” (of the weights).

With the beginning of machine learning, things start to become more interesting from an interpretative perspective as well. The basic work of the perceptron was purely deductive. Given a rule (the predetermined weights) and a cause (the input) the perceptron would infallibly deduce the effect (the output). However, with the learning rule, the perceptron becomes able to discriminate on the basis of weights which its human programmers were not able to predetermine. The training, then, is based on *induction* (in Peircean terms). The perceptron as a series of causes (the inputs) and of effects (the predetermined correct responses), and through a series of trial and errors, is able to induct the rule (the new weights). After that, when confronted with new, not-curated inputs it will be able to deduce the outputs correctly - thus endowing the perceptron with some form of “archival perception” (D’Armenio 2025) and initiating, although at an embryo stage, the mode of enunciative praxis described in Dondero (2025).

The inferential ability of the perceptron, then, is greatly enhanced. Its operations are still determined by its weights (and therefore again follow a rule of necessity). The iterative process which determines its final weights is still a mathematical model (also deterministic). At the same time, the process of determining the weights is now one step removed from human intervention, and therefore initiates the path that makes it increasingly difficult for the creators of AI models to understand how exactly they operate.

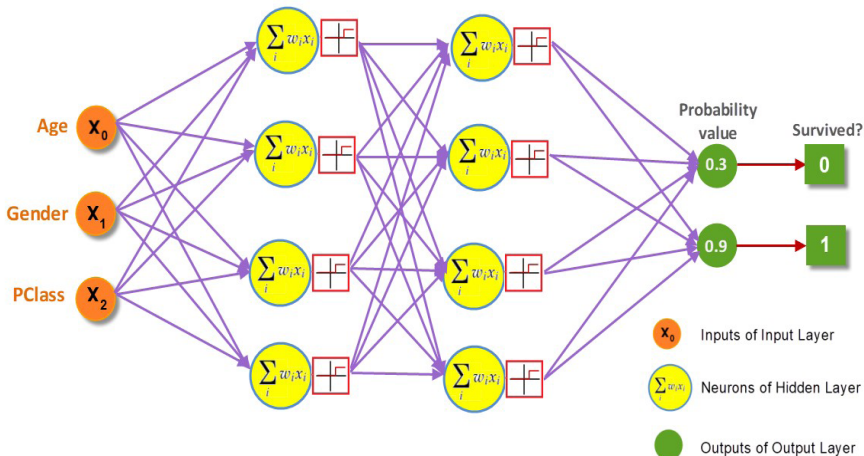
### **3.3. Neurons and Artificial Neural Networks**

The perceptron, while relatively simple, was the base for developing more complex algorithms. In order to tackle more complex tasks, or offer more nuanced solutions, the development of machine learning proceeded in two directions, often in parallel:

1. Developing different kinds of neurons, that is, other algorithms, in addition to the perceptron;
2. Stacking multiple perceptrons (and later multiple neurons) in Artificial Neural Networks (ANN).

New kinds of neurons have been developed to replace or stand alongside perceptrons. Since the output of perceptrons is limited to a binary output, then if taken singularly, they can only solve *linear* problems, and their answers lack nuances: only “yes” or “no”. New *neurons* were then developed with different activation functions, meaning that their output could be, for example, a number within a certain range. A neuron using *linear regression*, for example, would be able to use very similar weights to those of a perceptron to determine the price of a real estate property and not only its being expensive or not. Semiotically, the operation of the neuron is quite similar to the operation of the perceptron, but, once more, more complex and hence harder to interpret or predict by humans.

On the other hand, the second line of development was that of creating ANN based on networks of perceptrons. The very first networks were quite simple and “shallow” (having only one layer) capable of simple operations, such as distinguishing a capital “T” from a capital “L” in a small image. Over the decades, also thanks to the development of machines with higher and higher computational capabilities, ANNs began to develop an increasing number of hidden layers (Fig. 2, displayed in yellow). These hidden layers work as perceptrons themselves. Each hidden layer has a weight and an activation function assigned and fed into another hidden layer making the ANN metaphorically “deep”.



**Figure 2:** A Perceptron-based Artificial Neural Network  
(Copyright © 2025 Devskol)

Perceptrons, here, become the *neurons* in the neural network, that is, they become part of an interconnected system that works collaboratively. Early ANN made use of a series of perceptrons, and of their learning rule,

to become able to engage in more complex tasks. For example, where a perceptron could be able to estimate if a house is expensive based on a series of parameters, a perceptron-based ANN could determine the price of the house, model interactions between features (for example that the number of bedrooms matters more when square footage is high) or model conditional patterns (difference in prices in different cities).

The two lines of development converge. Most ANNs soon started to feature multiple layers composed of different kinds of neurons. The organisation of neurons in ANN, hence, further multiplies the co-occurring actions of discrimination, each providing more nuanced results. The continuous development of new and more sophisticated hardware and therefore computer power gave the possibility to stack more hidden layers together, reaching the hundreds, making the ANNs exponentially better at their tasks. This sparked a revolution in machine learning called the “deep learning revolution” making possible more and more technologies and the rise of current Artificial Intelligence (LeCun et al. 2015).

The differentiation of neurons and the increase in layers also required a new way for the model to learn, as the perception learning rule became too simplistic. This was provided by *backpropagation*, a staple of deep learning. Simply put, backpropagation operates in a similar manner. It first makes a “forward pass”, in which the input goes through all the layers and the ANN produces its prediction which is confronted with the right result (which, again, needs to be known beforehand). The system, now, uses a “loss function”, in order to evaluate *how* wrong the prediction is. The second step is the backward pass, in which the different layers are examined one by one – from the output layer back to the input layer. At each layer a *partial derivative* of the loss is calculated. In other words, it is calculated how the loss (i.e. the margin of error in the prediction) would have changed, if the weight had been different. Finally, the weights are updated. As in the perceptron learning rule, here too the process needs to be repeated many times with different inputs before the system can be used with inputs whose correct evaluation has not been determined beforehand. The update of the weights is also more nuanced, like the predictions it has to evaluate.

A complex ANN can, for example, distinguish between pictures of cats and dogs. For the purposes of backpropagation, a curated set of pictures labelled as either “cat” or “dog” is needed. The model is then trained on RGB (Red-Geen-Blue) input of the image files. The image files have values for Red, Green and Blue for each pixel in the picture. The RGB values of each picture are calculated and inputted, in the network (usually the picture is downscaled in quality because of the large amount of computation

that it takes to perform such a task). The ANN, through its epochs, assigns weights to a different shape which it recognises on a mathematical grid based on a set of properties in the picture. The different layers operate sequentially, so that the first layer of the network recognises lines, the second level legs or faces, and so on until the entire picture is estimated to be of a “cat” or a “dog”. Understanding the weights needed to accomplish such a task is far above human capacity, and cannot be done without machine learning. The researchers, given the desired objective of the model, can still fine tune it looking at its outputs without having to interpret what is going on “inside” – and therefore contributing to the “black box” nature of many ANN.

If we look at these developments from the perspective of their actor networks, we see a multiplication in the order of hundreds of actants with similar functions which work in parallel or sequentially, perform operations and predictions and sanction each other. The ANN learning structure is semiotically similar to its predecessor. We still have operations which generate predictions, others which evaluate them, and others which evaluate and update the weights. What changes, from this perspective, is the complexity of the operations that the ANN can engage with.

From an interpretative perspective, however, the complexity of these systems seems to make them able to perform *abductions* (see also Volli 2023). In the real estate example, we can see that ANN is faced with an effect (the final price of a property), and a set of rules (the weights that they have assigned to different elements). In its evaluation, it has to infer what might be the cause (which of those elements have influenced the final price?). This cannot be performed as a form of induction, because there is no single fixed rule behind real estate prices. ANN has to make some hypotheses during its backward pass, try different weights for different elements. This is also what makes these kinds of operation more fallible – as abduction is the less secure kind of inference.

Something similar can be said about distinguishing between pictures of cats and dogs. ANN has an effect (the picture), a set of rules (weights determining how a cat is supposed to look), and then it is faced with a picture and needs to determine an effect (was a cat or a dog in front of the camera?).

While the need for hypothesising might decrease with training, thanks to more and more accurate weights (and therefore a move towards induction), the complexity of these tasks makes abduction more relevant – and risky – in the operations of ANNs.

### 3.4. Transformer Networks and Large Language Models

The last key improvement in Artificial Intelligence with which we will engage in this paper is the creation of Transformer Networks. These are networks of an ever larger scale – they contain several ANNs – and their key property is that of *attention*. This concept was introduced in what is now considered a pillar in the bibliography on transformer networks: a paper entitled “Attention is all you need” by Vaswani et al. (2017). Attention is used to calculate the value of the weights, or how much “importance” we ascribe to a certain input according to their context, i.e. according to other tokens that are present in the same input.

Attention is extremely important in Large Language Models (LLMs), based on the stacking of multiple transformers which are trained on extremely large datasets composed of texts in the same natural language<sup>3</sup>. LLMs work by trying statistically to find the most probable word after a sentence by having learned all the content of a large dataset and they are believed to “learn the statistics of language” to reproduce the effect of today’s chatbots.

In the case of LLMs, attention gives the network the capability to understand the context of every word given into the input.

Attention works by vectorialising every “token” (e.g. every word, or part of a word in a text), according to three vectors:

- “Query” – What the token is “looking for” (e.g. a noun “looks for” a verb, a transitive verb for an object, a pronoun for a noun, etc.).
- “Key” – What the token is “offering” (what other token would be completed by it, the object offering itself to a transitive verb).
- “Value” – the weight of the relationship (how strong can be the connection between two tokens).

These three vectors are then used to calculate the attention. The query and the key are multiplied, passed through an algorithm (softmax) and then multiplied by the value. This mathematical operation reconstructs the relations between two tokens, and the statistical probability that they refer to each other. When attention is calculated across all the tokens in a text, the LLM is able to reconstruct the relations for all the tokens.

---

<sup>3</sup> This was only possible thanks to the introduction of parallelisability in operations and scalability with large datasets, taking advantage of faster GPUs. Parallelisation and scalability are two properties which express how multiple tokens can be calculated at once to give faster results and how the system can be scaled, meaning how if it’s given more GPUs power it can always improve results and calculations.

For example, if an LLM encounters the sentence “the cat purrs because it is happy”, it is able to reconstruct that the query of the word “it” is the word “cat”. The token “it” will score low for “happy” and high for “cat” as cat is the only noun in the sentence, so the model will determine that it refers to that token. This is done simultaneously with all the words of the input so that the LLM determines the proper context using Queries and Keys, and having a set of Values as a result.

Interestingly, each token is represented in LLMs through vectors – i.e. lengths and directions. Instead of the 2 or 3 dimensions typical of geometry, LLMs feature hundreds. The complex path which they draw, and the direction they point to, contain both their semantical affinities (e.g. feminine words will point towards similar directions) and their relations with other words. The semantic and lexical systems of a language as reconstructed in a LLM then occupy a multidimensional space which is too difficult for humans to imagine. However, it seems to share some affinities with the rhizomatic encyclopedic model of Eco (1994) and Lotman’s semiosphere (1990). It is an eminently relational semiotic space, in which the different tokens are meaningful only thanks to the relations with each other and the “attention” uniting them. It is, in fact, based on *distributional semantics*, i.e. it grounds meaning on the spatialised relations of different sememes (see Montanari 2025).

This is particularly interesting because, if we look at how machine learning applies to LLMs, the complexity of the system they have to reconstruct is that of a whole semiotic system (that of a natural language) and a series of other systems of different scales (metalanguages, specialised languages like lyrics or poetry, etc.). The simple weights determined by the perceptron learning rule have multiplied and reached a level of complexity which is difficult to grasp. However, while an LLM may be able to simulate grammatical analysis, the digital metalanguage it reconstructs is a mathematical one – possibly reminiscent of Hjelmslev’s glossematics.

From the perspective of the actor network, the complexity of the systems renders it pointless to try to reconstruct the abundance of actants engaged in performing tasks and sanctioning each other. What appears, instead, is an increased sense of actorialisation of the model itself, which starts to appear like a unique semiotic unity, capable of communicating and even understanding – an effect of meaning that reaches even stronger effectiveness with chatbots (see below).

On the interpretative side, the complexity of LLMs makes them even more reliant on abduction. The scale of the semiotic system involved, the gargantuan quantity of data upon which they are trained, the wide variety

of tasks and conversations in which LLMs can be involved makes it impossible to have a perfectly correct model. Hallucinations, therefore, are inevitable. The system will invent facts, make up statements, get stuck in strange ideas. While this is something which the development of the models minimises, it is not something, we believe, which will be possible to completely root out – because, as Eco reminded us (1976), a semiotic system is something which can be used to lie. LLMs are created in order to be able to communicate, in order to be semiotic machines. As such, they require a level of complexity high enough to allow for mistakes, for misunderstandings, and for nonsense. As the saying goes, “it’s not a bug, it’s a feature”.

### 3.5. Chatbots

The form in which most users encounter LLMs, today, are Chatbots. These are LLMs trained to simulate a conversation with the user and designed to make them appealing and as user friendly as possible. Usually this is done by giving the LLM some direction in the format of its writing and expression, by adapting filters and security checks to make it harder to be exploited for nefarious uses. This is done by using Reinforcement Learning by Human Feedback (RLHF) in which a set of engineers ranks the LLM answers from good to bad. In this way, the LLM will learn to behave in the desired way: generally being an eager helper.

Chatbots also feature a memory which they use to form a coherent conversation, to base their next answers on what has already been discussed as part of the same interaction – or even multiple interactions with the same user.

Chatbots have an astounding effectiveness to simulate human conversations and to engage in meaning-making activities – while obviously unaware of doing so. The combination of a strong fascination for this technology from users and a strong industrial push to incorporate them in many different applications (search engines, instant messaging, etc.) has brought artificial intelligence into the forefront of global interest – and at the centre of many debates.

While the interpretative and learning capabilities of Chatbots are essentially the same as of more generic LLMs, it is the effect of meaning of their actorialisation – as mentioned above – which make Chatbots particularly interesting. The complexity of their inner workings – which is mysterious for users and, to a certain degree, for their own creators – coupled with the ability to mimic human conversation give rise to a potent illusion of conversing with a sentient being. Corporate attempts to reinforce this illu-

sion – by making them friendly and polite – are also part of the reasons for which it is very easy to anthropomorphise them.

This, in turn, is key to position them – at least discursively – in a very clear actantial role: that of the helper (rather different than the opponent role often reserved to AI in games, see Giuliana 2023). Chatbots present themselves as helpful assistants, as collaborators, eager to help human subjects in a series of possible performances, and then offering key competences in terms of skills (cognitive, semiotic, epistemic) and knowledge (constantly updated from the Internet).

Of course, this is not the only role which Chatbots – and other uses of LLMs – can play. They are, in fact, often subjects, to which activities traditionally needing human actors are increasingly delegated. This in turn requires humans to develop new competences<sup>4</sup>, and can cause different forms of discrimination (Latour 1993). This ability to replace humans – which Latour teaches us is certainly not something new – in different narrative programmes has, nonetheless, also characterised them as *anti-subjects*. Therefore several actions have been initiated and discourses of resistance against their adoption – often rooted on their ethical concerns, including intellectual property theft and excessive energy use.

### 3.6. Emergence

In this paper we have focused on the development of LLMs in particular, and we have not delved into image generation which is also an extremely popular and sophisticated use of ANNs and Transformer networks. We judged that for properly engaging with these other models we would need more space than we have in the current paper. There is, however, one last concept related to the development of transformers that we believe important to mention: that of *emergence*.

This concept was introduced by Leggs & Hutter (2017) to refer to the fact that the stacking of networks gave rise to the emergence of unexpected bigger reasoning properties. In a key paper about emergence in LLMs, Wei et al. (2022) state that: “a range of abilities (...) thus far [have] only been observed when evaluated on a sufficiently large language model. Hence, their emergence cannot be predicted by simply extrapolating performance on smaller-scale models.” (Wei et al. 2022: 6). In other words, after a certain threshold LLMs are capable of performing different tasks which were impossible for smaller models, and not explicitly included in

---

<sup>4</sup> As any teacher trying to discern how much of their students' essays has been written by Chat GPT is painfully aware of.

their pre-training. Examples of such tasks include being able to perform arithmetic, unscrambling words, or creating conceptual mappings. Another relevant example is that of translation. LLMs models which were not explicitly trained on language pairs still demonstrated an ability for multi-lingual translation (Li et al. 2024).

Emergence is often considered puzzling, as it is difficult to predict what other abilities might emerge from further increases of the size of the networks. It is also one of the reasons that they are often seen as black boxes: the functioning of the models escapes their creators who often cannot predict nor reconstruct why the models do what they do.

The situation is further complicated by the fact that these models are created by a large number of people, often building on models created by others, so that nobody can easily track the overall functioning of the model as a result. The alienation of the programmers from their own creation feeds into catastrophic discourses about “AI rebellion”. While these are probably also discursive strategies which aim to protect corporate intellectual property, as claimed by Leone (2021) – the fact that the inner workings of LLMs is rather mysterious is very real – and several attempts to increase the transparency of AI by interpreting the models predictions or generate attention maps are underway (Barredo Arrieta et al. 2019)

From a semiotic perspective, beyond the discursive and mythological (Barthes 1957 & Ferraro 2023) implications of the back box, it is interesting to see how extensive training on natural languages seems to enable a series of abilities which resemble reasoning and logic. This seems to be an indirect confirmation of Juri Lotman’s claim that natural languages are “primary modelling systems” (Lotman 1977) which play a key role in organising around them all other modelling systems and supporting human cognition. The idea that language shapes the way in which we think is not new (see the Sapir-Whorf hypothesis or Wittgenstein 1921), nevertheless the fact that natural languages – and not computer code – seem to be the base for some form of artificial cognition is rather interesting.

## **4. Semiotic implications**

### **4.1. AI’s Encyclopaedia and Semiosphere**

As we have seen in the previous paragraphs, since its inception the key to Machine Learning has been the idea of “training”. In supervised training, perceptrons, neurons, and ANN all adjust their weights by going through a curated dataset, in order to adjust their weights until they are able to cor-

rectly predict the outcomes desired by their creators. Even at later stages of development, where, for example, LLMs are trained on large not-curated datasets scraped from the Web, they learn to imitate human-made texts by engaging with them. They try to guess the next word and then check if their prediction is correct until they become able to generate coherent and meaningful texts.

Machine Learning, therefore, is dependent on a pre-existent set of texts. These texts are already accompanied by their interpretants – i.e. by other signs which can be used to infer their meaning. This clearly includes the curated lists of inputs which programmers feed to perceptrons and other neurons (like the list of pictures labelled “cat” and “dog” of our example above), but also existing lists of paired signs, such as images and their descriptions (such as those of the datasets Oxford-102, CUB-200, or COCO). We argue that even the texts used to train LLMs, even if not curated, already contain their interpretants. LLMs infer semantic directions (not the meaning of the words, but their vectorialisations in a semantic field) by the words which accompany them. They do not need to understand what a word “means”, in order to reconstruct the relations with other words and ultimately become able to answer questions and hold conversations. The different words in a text, therefore, function as interpretants of each other. They are signs which support the inference of other signs.

The importance of the training dataset, therefore, is capital. It is a well-known fact that AI models reflect the biases present in their training datasets, often creating racist, sexist, and other problematic content (Caliskan et al. 2017). The datasets, therefore, work as digital analogues of the *encyclopaedia* described by Umberto Eco (1994). According to Eco the *encyclopaedia* is grounded on the repertory of texts and interactions that an individual – or a community – has been exposed to and therefore encompasses all their semiotic competences and previous knowledge. Individual encyclopaedias are determined by a series of factors (socio-cultural positions, personal tastes and experiences, family and so on) which influence their construction. The encyclopaedia of an AI model, instead, is determined during its training, but it is equally – if not more – impactful on the ability of the model to create signs.

The training of AI models also implies that every text or sign produced by Generative AI has an intertextual (Genette 1982) relation with some – if not all – the texts upon which it has been trained. More precisely, AI generated texts are *hypertexts* (i.e., texts created using other texts as a starting point, Ibid.) since their creation is directly influenced by the texts which form its training dataset (its *hypotexts*). While identifying which training

texts influenced a specific AI-production is difficult (due to the back box that encompasses much of the models' operations), recent technical developments have started to be able to reconstruct them at least partially (Templeton 2024).

However, when it comes to LLMs, the model does something more than just imitating the texts upon which it has been trained. As we mentioned above, it reconstructs a semiotic system, by vectorialising each token (which correspond more or less to *sememes*) and integrating them into one semantic space. More precisely, the LLM reconstructs several semiotic systems, as it is also trained on programming languages and specialised languages (poetry, lyrics, metalanguages etc). While the LLMs do not reconstruct a specific hierarchy between these systems, the probabilistic, statistical and learned patterns which it uses imitate the relations between different semiotic systems as well. In other words, LLMs have their own *semiosphere*, a digitised and internalised reflection of the semiosphere to which the texts in its dataset belong.

However, when it comes to Chatbots, the additional instructions which regulate their responses add a new layer of hierarchies specific to that conversational agent. The temporary enamourment of Chat GPT for words like “delve” and “tapestry” (meant to make it sound more scholarly), or the momentaneous inability of Gemini AI to generate pictures of Caucasian people (because LLM tend to add racialised terms in the image prompting, see Shamin 2024) by overriding some of the training of the LLM, manipulate its textual production.

#### 4.2. LLMs as a modes of Sign production

Generative AI can be understood as participating in an enunciation activity initiated by its users. Humans input a prompt, without which the model cannot initiate any form of textual production. While discussing AI image generation, D'Armenio et al (2024a) describe this property as “machinic co-enunciation”.

A similar point can be made for LLMs. While the way in which AI processes human prompts is different (most image generators vectorialise words but do not “understand” prompts in the way LLMs do), the overall process is similar: a human author creates a prompt, and the AI model responds to it by creating a text.

We agree with D'Armenio et al, (2024a) that this is indeed a form of co-enunciation, or co-authorship. To go deeper into this activity of co-creation, we decided to look at it as a *mode of sign production* as those described by Eco (1976).

When discussing different modes of sign production, Eco focuses on several technical means as part of the process. These include the physical medium, the articulatory techniques, and the transmission and reproduction tools involved. The articulatory techniques are specifically interesting for our argument, since they include the mechanical or bodily actions which produce signs. The variety of articulatory techniques spans from the way in which we articulate our vocal apparatus when speaking, to complex digital operations which result in the creation of signs (such as, for example, photogrammetry). These means do not only allow for sign creation, but are often indispensable means of cognition (Paolucci 2025).

Chatbots (and other forms of Generative AI) can therefore be understood as a part of this process. They are complex technical means which support the activities of sign production of the users when creating their prompts. They draw from a net of hypotexts which participate in its training.

### 4.3. *Intentio machinae*

Chatbots seem quite unusual as a technical means for sign production. They appear to determine a large part of the final sign – often larger than their human co-enunciators. While all technical means do influence the texts they produce or reproduce – hence McLuan’s famous statement “the medium is the message” (1964) – it is rare that much of the final sign is entrusted to them. Rare, but not unseen. In the arts, for example, randomness generators have been used to determine the final form of a text – as in John Cage 1951 composition “Music of Changes”. In painting, Pollock’s *Drip Painting* period (1947–50) similarly makes use of the unpredictability of the dripping paint as a key component of the artistic work. Fortune telling, in a similar way, makes use of various forms of randomness generators – from tea leaves to playing cards – which strongly constrict the results of the reading (Aphek & Tobin 1990). In video games, procedural generation allows the creation of environments largely created by the game programme. They are based on some initial instructions given by the game designers (see, for example, *No Man’s Sky*, a 2016 game by *Hello Games*).

The co-enunciation, or the mode of sign production which is based on the use of Generative AI seems to be the next step in this direction. Not a radically new semiotic phenomenon, but an incremental step towards a larger share of authorship delegated to a machinic actor.

This specific mode of sign co-production raises, at the same time, analytical and interpretative questions. To the three *intentios* outlined by Eco, the *intentio auctoris*, *lectoris* and *operis*, we might need to add a fourth one: *Intentio Machinae*. If the *intentio auctoris* acknowledges what the au-

thor is trying to express in a text, the *intentio lectoris* what the reader adds in order to interpret it and the *intentio operis* the “objective” structure and content of a text, the *intentio machinae* accounts for what element of the text is determined by its technical articulatory techniques.

The *intentio machinae*, then, is determined by the encyclopaedia and the semiosphere of a certain conversational agent, and its intertextual interaction with the prompt created by the human author. It is difficult to disentangle what of the final text, or sign, is determined by the prompt and what is determined by the LLM – but the fact that the same prompt gives different results shows a variety of responses from the LLM which makes its contribution semiotically meaningful.

## 5. Discussion and Conclusions

In this paper we have reconstructed, step by step, the main milestones of the development of machine learning, focusing in particular on Large Language Models. For each step, we have provided some semiotic insights, focusing on three key aspects: narrative programmes, inferences, and intelligibility.

We have seen that the narrative programmes enacted by the different actants which compose neurons and models become more complex over time. Starting with an untrained perceptron operating a simple sanction, to learning neurons which perform prediction and are then sanctioned for it, to more and more intricate networks of actors which operate hundreds of complex operations. We have also argued how this actantial complexity leads to identifying the entire model as a single actor, ignoring its idiosyncrasies in favour of an anthropomorphisation.

We have examined how the inferring ability of machines has become more complex with their developments: from perceptrons able only to make use of very simple deductions, to the perceptron learning rule introducing inductive capabilities, and to ANN which have become capable of abductions. We have also pointed out that the ability to perform abductions – which are the most creative but less reliable form of inference – is probably the reason for AI hallucinations, inventions and false claims. The same properties which enable networks to learn by making hypotheses of possible (semantic) rules are those that make it less precise.

Finally, in terms of intelligibility, we have seen how the increasing complexity of the networks have progressively removed the task of deciding the weights – and therefore the parameters of interpretation – from the network creators. This eventually leads to the emergence of unexpected

properties and the necessity of new approaches to shed light in the black boxes that are the current models.

We have then focused on a few semiotic implications of our analysis of the working of LLMs. First, we have claimed that the datasets used to train LLMs function as an encyclopaedia, determining the ability of the system to produce texts. We also claimed that LLMs create multidimensional vectors for each token in a shared semantic space, giving rise to a sort of semiosphere, which reflects that engraved in the texts used by the training, but also manipulated by its creators when adapted to a Chatbot.

Second, we have looked at the activity of prompting Chatbots, in order to create text together as a specific mode of sign co-production. The technical means for sign creation have a large impact in the final creation – but that the latter still needs a prompt made by a user to be generated.

Third, we claimed that this mode of sign co-production, while not completely new in nature, has reached such a scale that we need to rethink our interpretative lenses. In particular, we have introduced the concept of “*Intentio Machinae*” as a way of acknowledging the participation (but not sole authorship!) of generative models in the generation of signs and texts.

Due to the long and complex history of machine learning, as well as the variety of models and networks existing today, we are not here proposing a complete semiotic analysis of all the aspects of generative AI. Notably, we have not engaged with visual transformers, despite the attention they were able to elicit in the semiotic community. This was decided mostly to limit the current article to an acceptable length – but we are considering a follow up paper doing something similar for these networks as well.

We have also mostly focused on the working of the models themselves, and not on their perception, the discourses that surround them, or their sociosemiotic implications. As we mentioned in the background section, there are a number of papers which have focused on those directions.

We believe, however, that this paper makes an important contribution to the scholarly work in semiotics by providing an accessible, step-by-step account of the key elements of LLMs. We believe that a multidisciplinary approach to these topics, which joins technical and semiotic knowledge, is key for analysis grounded on the actual workings of the technology. We hope this paper, beyond offering some theoretical insights on the technology, can also offer a basic technical understanding to other semioticians – and humanities scholars – who want to contribute to the research on Large Language Models and Generative Artificial Intelligence in general.

## Acknowledgements

This research was supported by y NEXR – Next Extended Reality (funded by NextGenerationEU and Business Finland, co-research project number 5703/31/2023) and Funded by the European Union (ERC, InterReal, grant agreement No. 101163472). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

## References

Aphek, E., Y. Tobin. 1989. *Semiotics of Fortune-Telling*. Amsterdam: John Benjamins Publishing.

Barredo Arrieta, A., E. Serral-Asensio, S. Tabik, R. Benjamins, A. Barbadó, S. Garcia, S. Gil-Lopez, D. Molina, J. M. Benítez and F. Herrera. 2020. “Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges Toward Responsible AI”. *Information Fusion*, Vol. 58, 82–115. Available at: <https://doi.org/10.1016/j.inffus.2019.12.012>.

Barthes, R. 1957. *Mythologies*. Paris: Éditions du Seuil.

Basso Fossali, P. 2025. “Rationalités correctives et intelligence artificielle assistée: les doubles contraintes des humanités numériques”. *Semiotica*, Vol. 2025, No. 262, 71–109.

Caliskan, A., J. J. Bryson and A. Narayanan. 2017. “Semantics Derived Automatically from Language Corpora Contain Human-like Biases”. *Science*, Vol. 356, No. 6334, 183–186.

Colas-Blaise, M. 2025. “La machine crée, mais énonce-t-elle? Le computationnel et le digital mis en débat”. *Semiotica*, Vol. 2025, No. 262, 147–187.

Compagno, D. 2025. “La notion de vérité à l’épreuve de l’intelligence artificielle”. *Semiotica*, Vol. 2025, No. 262, 25–49.

Dall’Acqua, A., F. Bellentani. 2023. How to Build a Chatbot. A Semiotic and Linguistic Approach. In Santangelo, A., M. Leone (eds.). *Semiotica e intelligenza artificiale* (Coll. *I saggi di Lexia*, n. 48). Roma: Aracne, 149–169.

D’Armenio, E., A. Deliège and M. G. Dondero. 2024a. *Semiotics of Machine Co-Enunciation: About Generative Models (Midjourney and DALL·E)*. *Signata: Annales des Sémiotiques*, 15, Article 15. Presses universitaires de Liège. Available at: <https://doi.org/10.4000/127x4>.

D'Armenio, E., A. Deliège and M. G. Dondero. 2024b. A Semiotic Methodology for Assessing the Compositional Effectiveness of Generative Text-to-Image Models (Midjourney and DALLE). In Del Bue, A., P. Grommelt, L. Weiss, F.-J. Pfreundt, and J. Keuper (eds.). *Proceedings of Computer Vision – ECCV 2024 Workshops, LNCS 15644*. Cham: Springer. Available at: [https://doi.org/10.1007/978-3-031-92089-9\\_8](https://doi.org/10.1007/978-3-031-92089-9_8).

D'Armenio, E. 2025a. “Les IA génératives visuelles entre perception d'archives et circuits de composition”. *Semiotica*, Vol. 2025, No. 262, 213–257.

D'Armenio, E., M. G. Dondero, A. Deliège and A. B. D. Sarti. 2025b. “For a Semiotic Approach to Generative Image AI: On Compositional Criteria” [as “Big Dada”]. *Semiotic Review*, 9 (Images issue), 1–59. Available at: <https://doi.org/10.71743/ee5nrx33>.

Deliège, A., M. G. Dondero and E. D'Armenio. 2025. “Revisiting Wölflin in the Age of AI: A Study of Classical and Baroque Composition in Generative Models”. *Journal of Imaging*, Vol. 11, No. 5, 128. Available at: <https://doi.org/10.3390/jimaging11050128>.

Dondero, M. G., J. A. Aldama and M. Leone. 2025. “Aspects of AI semiotics: enunciation, agency, and creativity”. *Semiotica*, Vol. 2025, No. 262, 1–3.

Dondero, M. G. 2025. “Semiotics of artificial intelligence: enunciative praxis in image analysis and generation”. *Semiotica*, Vol. 2025, No. 262, 111–146.

Duncker, D. 2020. “Chatting with Chatbots: Sign Making in Text-Based Human–computer Interaction”. *Sign Systems Studies*, Vol. 48, No. 1, 79–100. Available at: <https://doi.org/10.12697/SSS.2020.48.1.05>.

Eco, U. 1976. *A Theory of Semiotics*. Bloomington: Indiana University Press.

Eco, U. 1994. *The Role of the Reader: Explorations in the Semiotics of Texts*. Bloomington: Indiana University Press.

Eco, U. 1994. *Sette passeggiate nei boschi narrativi*. Milano: Bompiani.

Ferraro, G. 2023. Mito e realtà dell'IA: uno sguardo semiotico. In Santangelo, A., M. Leone (eds.). *Semiotica e intelligenza artificiale* (Coll. *I saggi di Lexia*, n. 48). Roma: Aracne, 65–90.

Galofaro, F., Z. Toffano. 2023. Verso una semantica quantistica? Applicazioni al discorso religioso. In Santangelo, A., M. Leone (eds.). *Semiotica e intelligenza artificiale* (Coll. *I saggi di Lexia*, n. 48). Roma: Aracne, 107–130.

Genette, G. 1982. *Palimpsestes: la littérature au second degré*. Paris: Éditions du Seuil.

Giuliana, G. T. 2023. L'intelligenza delle I.A. come effetto di senso semio-narrativo nei giochi digitali. Una rivoluzione semiosofica. In Santangelo, A., M. Leone (eds.). *Semiotica e intelligenza artificiale* (Coll. *I saggi di Lexia*, n. 48). Roma: Aracne, 171–194.

Gramigna, R. 2023. Some Remarks on Fakes and Deepfakes. A New Epistemic Challenge. In Santangelo, A., M. Leone (eds.). *Semiotica e intelligenza artificiale* (Coll. *I saggi di Lexia*, n. 48). Roma: Aracne, 45–64.

Greimas, A. J. 1970. *Du sens II: Seuil de la signification*. Paris: Seuil.

Hebb, D. O. 1949. *The Organization of Behavior: A Neuropsychological Theory*. New York, NY: John Wiley & Sons.

Haugeland, J. 1989. *Artificial Intelligence: The Very Idea*. Cambridge, MA: The MIT Press. Available at: <https://doi.org/10.7551/mitpress/1170.001.0001>.

Kull, K., D. Favareau. 2022. Neurosemiotics: A Brief History of Its Development and Key Concerns. In *The Routledge Handbook of Semiosis and the Brain*, 13–29.

Latour, B. 1993. Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts. In Bijker, W. E., J. Law (eds.). *Shaping Technology/Building Society: Studies in Sociotechnical Change*. Cambridge, MA: MIT Press.

LeCun, Y., Y. Bengio and G. Hinton. 2015. “Deep Learning”. *Nature*, Vol. 521, No. 7553, 436–444. Available at: <https://doi.org/10.1038/nature14539>.

Legg, S., M. Hutter. 2007. “Universal Intelligence: A Definition of Machine Intelligence”. *Minds and Machines*, Vol. 17, No. 4, 391–444. Available at: <https://doi.org/10.1007/s11023-007-9079-x>.

Leite da Silva, T. F. 2023. Using Transformer Networks and Tensive Semiotics to Improve sentiment Analysis Accuracy in Tourism Digital Platforms. In Santangelo, A., M. Leone (eds.). *Semiotica e intelligenza artificiale* (Coll. *I saggi di Lexia*, n. 48). Roma: Aracne, 131–147.

Leone, M. 2023. I compiti principali di una semiotica dell'intelligenza artificiale. In Santangelo, A., M. Leone (eds.). *Semiotica e intelligenza artificiale* (Coll. *I saggi di Lexia*, n. 48). Roma: Aracne, 29–44.

Leone, M. 2024. “Semiotics of the Black Box: On the Rhetorics of Algorithmic Images”. *Visual Communication*, Vol. 23, No. 3, 426–451

Li, J., H. Zhou, S. Huang, S. Cheng and J. Chen. 2023. “Eliciting the Translation Ability of Large Language Models via Multilingual Finetun-

ing with Translation Instructions”. *ArXiv*, arXiv:2305.15083. Available at: <https://doi.org/10.48550/arXiv.2305.15083>.

Lotman, J. 1977. Primary and Secondary Communication-modelling Systems. In Lucid, D. (ed.). *Soviet Semiotics: An Anthology*. Baltimore: Johns Hopkins University Press, 95–98.

Lotman, Y. M. 1990. *The Universe of the Mind*. New York: IB Tauris.

McLuhan, M. 1964. *Understanding Media: The Extensions of Man*. New York: McGraw-Hill.

Montanari, F. 2025. “ChatGPT and the Others: Artificial Intelligence, Social Actors, and Political Communication. A Tentative Sociosemiotic Glance”. *Semiotica*, Vol. 2025, No. 262, 189–212.

Paolucci, C. 2025. “The Myth of Meaning: Generative AI as Language-endowed Machines and the Machinic Essence of the Human Being”. *Semiotica*, Vol. 2025, No. 262, 5–23.

Reyes, E. 2024. “Generative Media: Sign, Metaphor, and Experience”. *Digital Age in Semiotics & Communication*, Vol. 7, 62–79. Available at: <https://doi.org/10.33919/dasc.24.7.4>.

Reyes, E. 2025. “Characterizing AI in Media Software: An Interdisciplinary Approach to User Interfaces”. *Carte Semiotiche*, 2, 36–49. hal-04915703. Available at: <https://hal.science/hal-04915703v1>.

Rosenblatt, F. 1958. “The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain”. *Psychological Review*, Vol. 65, No. 6, 386–408. Available at: <https://doi.org/10.1037/h0042519>.

Sanna, L. 2020. *Data-driven Semiotics and Semiotics-driven Machine Learning*. In *LEXIA. Rivista di semiotica*, 33–34, 89–107. Available at: <https://doi.org/10.4399/97888255354266>.

Santangelo, A., M. Leone (eds.). 2023. *Semiotica e intelligenza artificiale* (Coll. *I saggi di Lexia*, n. 48). Roma: Aracne.

Santangelo, A. 2023. Essere umani al tempo dell’intelligenza artificiale. Narrazioni a confronto In Santangelo, A., M. Leone (eds.). *Semiotica e intelligenza artificiale* (Coll. *I saggi di Lexia*, n. 48). Roma: Aracne, 239–266.

Saussure, F. de 1916. *Cours de linguistique générale*. Lausanne & Paris: Payot.

Semenenko, A. 2012. *The Texture of Culture: An Introduction to Juri Lotman’s Semiotic Theory*. New York: Palgrave Macmillan.

Shamim, S. 2024. Why Google’s AI Tool Was Slammed for Showing Images of People of Colour: Gemini’s Attempt to Correct the Marginalisation

of Ethnic Minorities has Backfired. *Al Jazeera*, 9 March 2024. Available at: <https://www.aljazeera.com/news/2024/3/9/why-google-gemini-wont-show-you-white-people>.

Soro, E. 2023. Fit Different! An Exploration of Apple Fitness Artificial Affects Machine. In Santangelo, A., M. Leone (eds.). *Semiotica e intelligenza artificiale* (Coll. *I saggi di Lexia*, n. 48). Roma: Aracne, 197–218.

Stano, S. 2023. I corpi (im)possibili dell’Intelligenza Artificiale” in Santangelo, A., M. Leone (eds.). *Semiotica e intelligenza artificiale* (Coll. *I saggi di Lexia*, n. 48). Roma: Aracne, 219–237.

Templeton, E., A. Carr, M. Stuhlsatz, D. Xu, et al. 2024. “Scaling Monosemanticity: Extracting Interpretable Features from Claude 3 Sonnet”. *ArXiv preprint arXiv:2405.13278*. Available at: <https://arxiv.org/abs/2405.13278>.

Turing, A. M. 1948. *Intelligent Machinery* (NPL Report). Teddington, UK: National Physical Laboratory.

Valle, A. 2025. “From Grammar to Text: A Semiotic Perspective on a Paradigm Shift in Computation and its Usages”. *Semiotica*, Vol. 2025, No. 262, 51–69.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser and I. Polosukhin. 2017. “Attention Is All You Need”. *ArXiv*, arXiv:1706.03762. Available at: <https://doi.org/10.48550/arXiv.1706.03762>.

Volli, U. 2023. Il mito delle due intelligenze e la semiotica. In Santangelo, A., M. Leone (eds.). *Semiotica e intelligenza artificiale* (Coll. *I saggi di Lexia*, n. 48). Roma: Aracne, 91–103.

Voto, C. 2023. Interpretare il cambiamento climatico attraverso l’intelligenza artificiale. Una prospettiva semiotica. In Santangelo, A., M. Leone (eds.). *Semiotica e intelligenza artificiale* (Coll. *I saggi di Lexia*, n. 48). Roma: Aracne, 267–281.

Wei, J., Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler, E. H. Chi, T. Hashimoto, O. Vinyals, P. Liang, J. Dean and W. Fedus. 2022. Emergent Abilities of Large Language Models. *Transactions on Machine Learning Research*. ArXiv:2206.07682. Available at: <https://doi.org/10.48550/arXiv.2206.07682>.

Wittgenstein, L. 1921. *Tractatus Logico-Philosophicus*. London: Kegan Paul.

## THE PEIRCEAN THEORY OF AI: ADVANCING TEXT GENERATION THROUGH PEIRCE'S TRIADIC MODEL, SPECULATIVE GRAMMAR, AND METHODEUTICS.

*Alon Friedman*  
*University of South Florida*  
*alonfriedman@usf.edu*

*Martin Thellefsen*  
*University of Copenhagen*  
*martin.thellefsen@hum.ku.dk*

### **Abstract**

Large language models (LLMs) have significantly advanced natural language processing (NLP) in text generation, translation, and automated question answering. However, despite these advancements, their capacity for interpretative reasoning remains limited. Current AI systems, primarily grounded in formal linguistics and statistical approaches, struggle to

capture the relational and contextual dimensions crucial for human-like comprehension. These limitations are particularly evident when interpreting meaning within dynamic social contexts, highlighting the need for theoretical frameworks that extend beyond statistical pattern recognition. This study examines how Charles Sanders Peirce's nineteenth-century semiotic theory, specifically his triadic model of Sign, Object, and Interpretant, can inform and enhance AI's interpretative capabilities. Peirce's systematic approach to meaning making, which predates computational thinking by nearly a century, offers critical insights into the limitations of AI systems grounded primarily in formal logic and statistical operations. These limitations become particularly clear when examining semiotic relationships through the lenses of speculative grammar and methodotics. Furthermore, we incorporate Claudio Paolucci's perspective on machinic enunciation and the "myth of meaning" to expand our theoretical framework. Paolucci's analysis of generative AI as a language-endowed machine, lacking subjectivity yet producing contextually significant enunciations, supports the reinterpretation of AI output in functional and relational terms. This perspective aligns with Peirce's focus on the triadic process of semiosis, adding a contemporary lens which emphasizes the functional rather than essentialist nature of meaning-making in AI systems. By addressing how Peirce's triadic model and Paolucci's framework can bridge the gap between statistical and socially oriented approaches, we contend that Peircean principles can enhance relational understanding in language models and illuminate the theoretical and practical challenges of integrating nineteenth-century semiotic theories into modern computational systems. Our findings indicate that Peirce's sign theory significantly expands the contextual awareness of AI, highlighting the complexities of replicating interpretative processes. This research demonstrates the continued relevance of classical philosophical frameworks in addressing contemporary technological challenges and contributes to a comprehensive theory of AI.

## **1. Introduction**

In recent years, artificial intelligence (AI) has emerged not merely as a powerful tool across diverse domains but as a transformative force which challenges our understanding of intelligence, meaning-making,

and human-machine interaction. Large language models (LLMs), in particular, have achieved significant advances in natural language processing (NLP), demonstrating capabilities in text generation, language translation, and automated question answering. Despite these advancements, fundamental critiques persist regarding the limitations of LLM in interpretative reasoning. Critics argue that LLMs lack the relational and contextual understanding required for nuanced meaning-making, thereby falling short of human-like comprehension (Prabhu & Premraj 2024).

Current approaches to language modeling rely predominantly on statistical correlations and pattern recognition, resulting in systems which excel at syntactic accuracy but often fail to grasp semantic depth or contextual relevance. While AI can process vast amounts of data and generate plausible responses, it struggles to engage in genuine interpretation, i.e., the ability to infer meaning dynamically through relational and contextual cues. This shortcoming is particularly evident in scenarios requiring social interaction, adaptive responses, and context-sensitive language understanding (Mökander & Schroeder 2022).

In order to bridge this gap, the study draws on the semiotic theory of Charles Sanders Peirce, whose triadic model of the sign offers a promising framework for enhancing AI systems' interpretative capabilities. Peirce's theory emphasizes the relational process by which a Sign (A) represents an Object (B) through an Interpretant (C), making meaning emergent and adaptive rather than static. Unlike Saussure's structuralist approach, which treats meaning as a fixed relationship between signifier and signified, Peirce's model positions meaning as context-sensitive and dynamic, evolving through the interaction of signs, objects, and interpretants (Stawarska 2020).

In addition, we incorporate insights from Claudio Paolucci's work, *The Myth of Meaning: Generative AI as Language-Endowed Machines and the Machinic Essence of the Human Being* (2025). Paolucci argues that generative AI reveals a "machinic enunciation" which underscores both the limitations of human-centric models of meaning and the inherent hybridity of human cognition. His critique that AI systems generate language by reassembling pre-existing enunciations rather than through contextually informed interpretation subjectivity reinforces the notion that AI's interpretative outputs are best understood in functional and relational terms. This dual theoretical grounding motivates our inquiry into the following questions:

1. How can the triadic model of Peirce be operationalized in AI to enhance LLMs' interpretative processes?
2. In what ways can Peirce's semiotic and Paolucci's analysis jointly improve context-awareness and relational understanding in AI systems?
3. What are the practical challenges and limitations of implementing Peirce's semiotic in LLM architectures?

Our goal is to propose a computational framework which integrate semiotic principles into LLM design. By embedding iterative self-correction and dynamic interpretative processes inspired by both Peirce and Paolucci, we aim to advance the development of AI systems capable of adaptive context-sensitive reasoning. We thus examine how the triadic sign model can be used to align input-output relationships with relational meaning-making processes, allowing AI systems to generate contextually relevant interpretations rather than relying solely on pretrained statistical patterns. Paolucci's notion of "machinic enunciation" articulates how generative AI assembles and recombines linguistic units without the presence of a subjective, conscious agent. This perspective focusses on the distributive and iterative nature of AI's language production. In a Peircean context, we can see this as an instance of how functional interpretants form dynamically across multiple instances of user inputs, system outputs, and contextual cues, akin to thirdness in Peirce's triadic structure.

Such a conceptual bridge highlights two major points: first, that AI systems can be understood as "sign networks" governed by relational patterns rather than singular, interiorized minds; and second, that meaning can emerge pragmatically and adaptively in the absence of self-awareness. This is precisely the form of interpretative expansion often lacking in classical, statistically anchored AI. By positing each AI output as a functional interpretant shaped by context, we ground Paolucci's machinic theory in Peirce's broader semiotic architecture, allowing for iterative refinement and error correction in a manner more attuned to human-like discourse.

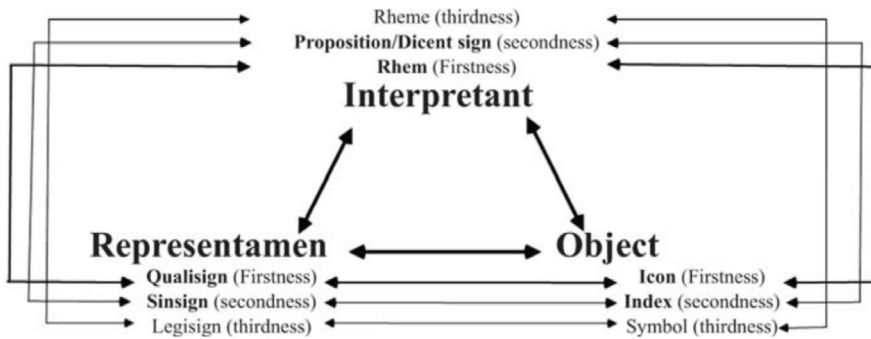
This study aims to bridge the gap between statistical approaches to NLP and relational models of meaning, offering a pathway toward more adaptive, human-like AI systems. By embedding self-correcting mechanisms and iterative refinement processes inspired by Peirce's methodetics, we propose a novel approach to addressing the interpretative limitations of current LLM architectures. Ultimately, this research seeks to contribute to the development of relational AI systems capable of adaptation to dynamic contexts and engaging in more sophisticated interpretative reasoning.

## 2. Semiotic, Peirce's sign theory, and the meaning of language

The term 'Semiotics,' derived from Greek, was first used in English before 1676 by Henry Stubbe (1632–1676), who spelled it as "semeiotics." He used the term to denote the branch of medical science relating to the interpretation of signs. However, Peirce (1839–1914) often referred to the term "semiotic" in his own writing (Niu 2020). Following Peirce, Charles W. Morris (1901–1979) adopted "semiotic" and extended the discipline beyond human communication to include animal learning and the use of signals. In this article, we will use "semiotic" according to the definitions provided by Peirce and Morris.

The core of Peirce's triadic model consists of three interconnected components: the Representamen (the form a sign takes); the Object (the referent or the entity the sign points to); and the Interpretant (the understanding or mental response the sign evokes). This model differs from the structuralist approach of Ferdinand de Saussure which emphasizes fixed associations between signifiers and signified, by positioning meaning as an emergent, context-sensitive process (De Luca Picione 2024). Peirce also introduced three categories of signs: Icons, Indices, and Symbols. The Icons are based on the nature of their relationships with their objects. It has a causal or direct connection to their objects, as seen in linguistic deixis like "this" or "there," which point to specific entities or locations. Symbols rely on convention (e.g., a stop sign), for example words such as "justice," which represents an abstract concept through societal consensus.

Peirce also proposed three foundational categories – Firstness, Secondness, and Thirdness – further illuminating the processes behind meaning-making (Stearns 1952). Firstness represents unmediated qualities or feelings, such as the sensation of warmth or the color red. In semiotic, it corresponds to icons which resemble their objects (e.g., a portrait). Secondness involves interaction or relation, such as the resistance of pushing a wall, and is exemplified by indices, which directly point to their objects (e.g., smoke as an index of fire). Thirdness signifies mediation and interpretation where meaning emerges through convention or representation, such as a word symbolizing an idea (e.g., "cat" representing a feline). Figure 1 presents Peirce's sign theory as captured by Friedman & Thellefsen (2011).



**Figure 1:** Peirce's trichotomic sign model and phenomenological categories.

The diagram visualizes Peirce's classification of signs across his phenomenological categories. The bold sign relations (*Qualisign-Icon-Rheme* and *Sinsign-Index-Proposition*), highlight sign configurations particularly relevant to how generative AI can produce meaning. The *Legisign-Symbol-Argument* relation remains un-bolded, in order to signify that AI is limited in terms of engagement with rule-governed inferential sign use, thus underscoring the key distinction between genuine semiosis and machinic sign simulation (Friedman & Thellefsen 2011: 653).

Since the interpretant plays a significant role in Peirce's model, representing the dynamic understanding or meaning arising from the relationship between the Representamen and the Object. Beyond merely classifying signs, Peirce's framework emphasizes that the interpretant is not static but evolves through interaction, context, and interpretation. The dynamic and relational model of meaning-making aligns with the core objectives of AI, particularly in natural language processing and computer vision, where understanding and adapting to contextual cues are crucial for effective performance.

However, a common critique of applying the concept of the interpretant to AI lies in the argument that AI systems lack subjective consciousness (Torrance 2008). This raises questions about whether AI can genuinely replicate the interpretative processes described in Peirce's framework. Thellefsen, et al. (2018) offer a valuable perspective by conceiving "information as signs", underscoring that an interpretant primarily functions as a mediation or relational effect between sign and object, rather than strictly requiring a human mind. In their view, the interpretant's role within semiosis is to ensure that meaning or an effect 'takes place,' whether or not a conscious

agent is present to witness it. This aligns with Peirce's later position that sign processes can unfold even in nonhuman or natural domains, as long as something in the environment – an AI system, for example – is capable of forming an interpretive link. Peirce described the interpretant as the mental response or understanding evoked by a sign, suggesting a cognitive and context-aware element. In comparison, AI critics (Collins 2021), AI systems are fundamentally incapable of achieving the depth of interpretative reasoning intrinsic to Peirce's model.

This critique can be addressed by reframing the interpretant in functional rather than subjective terms. In AI systems, interpretants can be understood as functional outputs: responses generated through relational processing rather than conscious thought. For example, a conversational AI system could operationalize the interpretant by dynamically adapting its responses based on context, prior interactions, and inferred user intent. While these systems lack self-awareness, their capacity to generate contextually appropriate outputs aligns with Peirce's pragmatic emphasis on the utility of signs in guiding action and interaction.

Another way to address this critique is by focusing on the emergent, relational nature of Peirce's model. Meaning, according to Peirce, is not static but evolves through interaction and interpretation within a broader system of signs (Nesher 1982). AI systems, particularly those employing neural networks and machine learning, excel at identifying and responding to such relational patterns. While these systems do not experience interpretation subjectively, they engage in processes that parallel human meaning-making by integrating contextual data, user feedback, and iterative refinement into their interpretative frameworks.

For example, research recommendation systems in AI rely on emergent relational patterns, in order to predict user preferences, effectively constructing interpretants that adapt and evolve based on ongoing interactions (Arkhipova 2024). These systems do not “understand” preferences as humans do, but they instantiate meaning in a way that aligns with Peirce's pragmatic emphasis on the practical utility of interpretants. The alignment demonstrates that while AI lacks subjective consciousness, it can replicate the functional outcomes of interpretive reasoning within its operational domain.

Peirce's concept of “thirdness”, which encapsulates the relational synthesis of meaning within broader contexts, further supports this approach (Stearns 1952). Thirdness emphasizes the dynamic, evolving relationships between Representamen, Object, and Interpretant, a process which AI systems can model through probabilistic and relational frameworks. By lever-

aging thirdness, AI can navigate complex, shifting contexts and produce interpretations that are both nuanced and adaptive.

### **3. Peirce's induction, error correction, and relevance in AI**

In scientific inquiry, induction is the process of developing general conclusions from specific observations, thus forming the basis of empirical knowledge. Peirce's approach to induction, deeply informed by his pragmatism, synechism, and tychism, emphasizes self-correction as an inherent feature of scientific investigation. These philosophical principles underscore the interconnected, probabilistic, and dynamic nature of knowledge, offering a compelling alternative to traditional machine learning methods that often rely solely on static correlations or deterministic rules.

By aligning AI with Peirce's Self-Correcting Thesis (SCT), systems can adopt iterative approaches to learning, where hypotheses are continually tested, refined, and validated against new data. The evolutionary process mirrors Peirce's belief in the practical and adaptive growth of knowledge, providing a foundation for AI systems that are both resilient and contextually responsive.

Peirce's SCT posits that scientific inquiry is designed not only to accumulate evidence but also to rigorously identify and rectify errors, progressively refining our understanding (Mayo 2005). In AI, this concept of self-correction is crucial, particularly as models face challenges in language processing where misinterpretations or biases are common. By embedding self-correction mechanisms, AI systems can iteratively refine their outputs and adapt to new contexts and evolving data. This dynamic learning process aligns with Peirce's vision of induction that evolve understanding through continuous inquiry.

A critical element of Peirce's induction is the notion of severe testing, where hypotheses must undergo rigorous evaluation to uncover flaws or deviations. The level of testing aligns with the demands of AI in ensuring that models not only learn from data, but also adapt by identifying and correcting errors. Rescher (1978) interprets Peirce's approach as one that seeks to disprove ideas thoroughly, suggesting that hypotheses gain reliability only if they withstand exhaustive scrutiny. In AI, severe testing could enhance model reliability, ensuring that interpretations remain robust even under variable or ambiguous conditions.

Based on Mayo (2005), Peirce's error correction extends beyond mere adjustments to conclusions; it involves refining every aspect of the inquiry process, including methodologies, data collection, and assumptions. He identified several layers of this refinement:

**1. Improving Methods:** Peirce emphasized the continual enhancement of scientific tools and techniques to reduce potential errors. In AI, this could translate into refining algorithms or improving computational efficiency to minimize inaccuracies.

**2. Enhancing Data Collection:** Systematic improvements in data gathering can reduce biases and improve representativeness. For AI, this may involve diversifying training datasets to encompass a broader range of linguistic and cultural contexts, leading to more reliable model outputs.

**3. Critically Examining Assumptions:** Peirce underscored the importance of revisiting assumptions, since biases may be inherent in foundational premises. For AI, this could involve re-evaluating model assumptions to align more closely with evolving data and contexts, ensuring ongoing relevance and accuracy.

Thus, this multi-layered approach positions AI's inductive process as a dynamic system capable of learning from and adapting to new information, enhancing resilience and interpretative accuracy in complex environments. By embedding Peirce's principles of induction and error correction, AI can evolve into a more reliable and contextually sensitive system, capable of addressing ambiguity and fostering meaningful interactions across diverse applications.

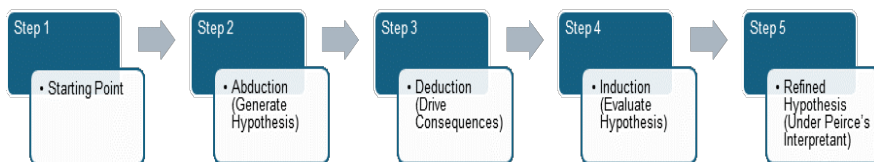
#### **4. Peirce's grammar and methodetic in sign theory**

Peirce's semiotic theory is rooted in speculative grammar and methodetics, foundational elements which structure the relational and interpretative processes central to meaning making. Speculative grammar, described by Peirce as "the philosophy of representation," investigates the structural relationships between signs, focusing on how signs interact with their objects and interpretants (Jappy 2018). The structural foundation is complemented by methodetics which Peirce characterized as the study of interpretative processes, and the methods of inquiry which guide the evolution of meaning (Bellucci 2014, 2015).

Peirce's speculative grammar categorizes signs into icons, indices, and symbols, each distinguished by relational characteristics: resemblance, causal connection, and convention, respectively. These categories provide AI systems with a framework for classifying signs based on their relational qualities, essential for distinguishing between signs that represent through similarity (icons), direct reference (indices), or established conventions (symbols). Speculative grammar serves as the structural mechanism that allows AI systems to map input data to meaningful categories and relationships, forming the basis for interpretative processing.

Methodetics, or “methodetic rhetoric,” emphasizes iterative inquiry and self-correction, principles which are essential to adaptive learning in AI. According to Jappy (2018), methodetics bridges theoretical (a priori) and empirical (a posteriori) knowledge by facilitating continuous refinement of interpretative frameworks. The process mirrors Peirce’s Self-Correcting Thesis (SCT), which posits that inquiry is an evolving process of hypothesis generation, testing, and correction. When applied to AI, methodetics functions as the mechanism through which systems refine their interpretative models based on feedback, contextual cues, and evolving data.

The integration of SCT with speculative grammar and methodetics aligns with Paolucci’s concept of machinic enunciation, wherein AI systems generate meaning through distributed, iterative processes rather than conscious deliberation. Paolucci’s notion that AI systems lack subjective awareness, while still producing contextually relevant outputs, reinforces Peirce’s understanding of Thirdness as mediation. By framing AI interpretants as dynamically refined relational outputs rather than conscious expressions, the theoretical framework becomes applicable to machine learning architectures designed to adapt and respond to complex inputs.



**Figure 2:** Peirce’s interpretative process exemplified in the methodetic loop.

The methodetic loop refers to the ongoing cyclic process of interpretation. Signs are continuously transformed into new interpretants through reflection, critique, and reasoning. However, AI generated semiosis lacks genuine authentic methodetic reflection – and thus remains incomplete or limited in interpretive depth.

Figure 2 illustrates 5 steps of machinic enunciation:

1. **Stating the Problem (Starting Point):** The AI system identifies a problem or user query requiring interpretation.
2. **Abduction (Generate Hypothesis):** The system generates a provisional hypothesis or initial response based on available data and patterns (The Sign).

3. **Deduction (Derive Consequences):** The AI system derives potential outcomes or consequences of the hypothesis, preparing for evaluation.
4. **Induction (Evaluate Hypothesis):** The system evaluates the hypothesis by comparing it against new data, user feedback, or contextual information (Indexical Sign).
5. **Refined Hypothesis (Under Peirce's Interpretant):** The AI system refines its hypothesis or interpretative framework, producing a contextually enriched and accurate response (Symbolic Sign).

This cyclical process reflects Peirce's methodetic approach, wherein meaning is continually reshaped through interaction, feedback, and adaptation. By integrating the methodetic loop with Paolucci's concept of machinic enunciation, we observe that the AI system's interpretative process is fundamentally distributive and iterative. The evolving responses of AI are not fixed representations, but adaptive outputs generated through an ongoing interplay between structural relationships (speculative grammar) and contextual refinement (methodetics). Thus, the process, rather than genuine semiosis, constitutes a simulacrum of genuine semiosis, as AI mimics sign production without true methodetic reflection or conscious meaning-making.

**Short Illustrative Example:** Consider a conversational AI agent which uses a "methodetic loop" after each user query. First, it generates a provisional response (iconic sign), then it checks contextual indicators – such as user feedback or conversation history (indexical sign) – and then subsequently refines its interpretative framework, in order to produce a revised, more context-rich output (symbolic sign). Through repeated cycles, the agent effectively carries out self-correction, testing, and updating hypotheses about user intent. The approach captures the essence of Peirce's methodetics: an ongoing inquiry which integrates prior "errors" or new data to reshape the interpretive process, thus aligning with Paolucci's notion that machine enunciation is distributive and iterative. Next, we discuss the integration of Paolucci's Machinic Enunciation with the Methodetic Loop.

The iterative cycle described above is not merely a mechanical process, but one which reflects Paolucci's concept of machinic enunciation. As shown in Figure 2, the iterative refinement of hypotheses through feedback and contextual adaptation exemplifies how AI systems engage in relational processing without requiring conscious awareness. Instead, their interpretants are dynamically shaped by distributed, iterative processes, aligning with Peirce's broader vision of semiosis as a pragmatic, evolving system.

By embedding speculative grammar and methodoetics into AI architectures, we propose a model where interpretants are continuously shaped through interaction, evaluation, and adjustment. Also, the process aligns with Peirce's broader vision of semiosis as an adaptive, evolving process, providing AI systems with a means to navigate complex, context-sensitive environments with greater interpretative accuracy.

### **5. Computational applications of Peirce's sign theory**

The computational potential of Peirce's semiotic theory has intrigued researchers interested in developing systems which emulate interpretative processes. Peirce's initial description of a "logical machine" in 1887, designed to generate logical statements, introduced the notion that machines could simulate reasoning under certain constraints. Peirce's assertion that such a machine is bound by its inherent limitations – specifically, its inability to autonomously identify problems and its restriction to predefined operations – resonates with contemporary challenges in AI. AI systems today face similar boundaries, particularly in tasks requiring self-interpretation and context-based adaptation. The interest in applying Peirce's semiotic principles to computational contexts has grown, fostering concepts like semiotic engines and semiotic programming which address the complexities of data-driven analysis.

Kumiko Tanaka-Ishii (2015) presented one of the most comprehensive efforts to integrate Peirce's semiotic principles with computational methodologies. Her work highlights how programming can be approached as a form of semiotic analysis, employing Peirce's concepts of firstness, secondness, and thirdness to enhance coding practices and algorithms. The primary challenge identified by Tanaka-Ishii involves implementing thirdness, the element in Peirce's triadic model signifying a sign's relational function within a broader context. In computational terms, thirdness involves creating models capable of self-referencing, enabling AI to interpret and relate information dynamically rather than as isolated data points.

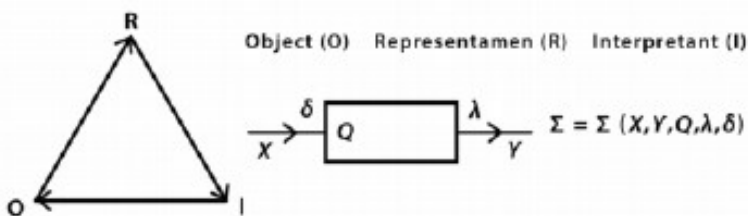
Tanaka-Ishii's approach to addressing the triadic is based on Church's SKI combinator calculus, a logical framework suited to examining transformation and relational dynamics within semiotic. The calculus allows computational systems to represent triadic relationships by distinguishing between firstness (isolated qualities), secondness (dyadic interactions), and thirdness (integrated, contextual meanings). However, Tanaka-Ishii highlights that thirdness remains challenging to encapsulate fully in code, since self-reference requires carefully managed constraints, in order to avoid infinite regress or circular dependencies. Her work illustrates the poten-

tial for Peirce's semiotic theory to inform programming logic, especially in contexts where meaning emerges through layers of interrelated signs.

Additionally, Tanaka-Ishii's exploration into computational semiotic was complemented by Nadin's development of the concept of a "semiotic engine." Rather than a static "semiotic machine," Nadin (1977, 2010) posits the semiotic engine as a dynamic process capable of functioning as both representation and instantiation. Nadin's semiotic engine integrates Peirce's triadic model with fuzzy logic which accommodates both quantitative and qualitative data. By introducing fuzzy automata, Nadin provides a mathematical equivalence between Peirce's relational triad and probabilistic logic, suggesting that AI systems can operate within non-deterministic environments while still producing coherent interpretations. The semiotic engine represents an advanced framework in computational semiotic, enabling AI to manage ambiguity, contextual shifts, and relational data within a unified model.

The notion of fuzzy logic, particularly within the constructs of thirdness, offers AI models the flexibility to interpret input through a spectrum of possibilities rather than binary classifications. Fuzzy input-output values are thus not restricted to static associations but instead allow for gradient-based interpretations, aligning with Peirce's emphasis on relational and context-dependent meaning. The transfer functions  $\lambda$  and  $\delta$  in fuzzy automata illustrate this non-deterministic behavior, wherein AI systems are not merely processing data but actively interpreting based on variable inputs, offering a pathway to achieve more adaptive, human-like comprehension.

Nadin (1977, 2010) proposes that instead of naming the process semiotic machines or semiotic programming, he would call it a semiotic engine, in order to better capture the complexity of computational semiotic. He defines the semiotic engines as processes representations and instantiates. He then outlines his mathematical proof of the equivalence between the Peircean sign and fuzzy automata. Figure 2 captures Nadin's formula.



**Figure 3:** The dynamics of Semiotic process by Nadin (1977 and 2010)

This figure illustrates the dynamic process of semiosis, highlighting its inherently processual and evolutionary character. Nadin emphasizes the continuous transformation, recursion and complexity of the meaning-making process.

The notion of the automation, according to Nadin, provides us with the ability for machine functionality. The fuzzy description of input and output values corresponds to the intention of capturing not only quantitative data, but also qualitative data that often is associated with description rather than abstract data. The two transfer functions  $\lambda$   $\delta$  are defined as non-determination behavior.

In more recent developments, open-source projects have attempted to operationalize Peirce's sign theory in accessible programming environments. For example, the "Peirce's Sign" package in R, developed by Friedman and Feichtinger (2017), applies Peirce's triangulation to analyze qualitative and quantitative data, allowing users to classify and identify relationships between data components. The package structures data through Peirce's triadic components – Representamen, Object, and Interpretant – enabling researchers to map complex relationships based on user-generated hypotheses about data relevance and interpretive meaning. The package's triangulation functions support associations within each proposed triad, establishing a systematic approach for computational semiotic that aligns with Peirce's logical directionality in sign relations.

Using these tools, researchers can examine data in a way that mirrors human interpretative processes, categorizing and synthesizing information in structured formats. The Peirce's Sign package exemplifies how computational models can embed semiotic analysis to process data both quantitatively and qualitatively, achieving triangulated insights which are context sensitive. By leveraging Peirce's semiotic principles, computational semiotic can offer AI models an adaptable framework for managing relational meaning, providing an alternative to traditional, syntax-based approaches in AI.

## **6. Integrating Peircean semiotic with AI modeling**

Integrating Peirce's semiotic principles into AI modeling establishes a comprehensive framework which addresses the limitations of current language and knowledge representation systems. At its core, Peirce's semiotic theory, particularly his concept of the interpretant, offers a model of meaning-making transcending pattern recognition and syntactic rules, thus embracing a layered, interpretative structure essential for nuanced understanding. This section explores the potential applications of Peircean

semiotic to AI, detailing how each element of his triadic model -Representamen, Object, and Interpretant – can contribute to creating AI systems that engage in dynamic, context-sensitive, and self-correcting interpretations.

### **6.1. The triadic model as a foundation for relational interpretation**

Peirce's triadic model of the sign, consisting of the Representamen (the form a sign takes), the Object (the referent or entity it points to), and the Interpretant (the meaning or understanding generated), provides a relational approach to meaning making. Unlike traditional AI models which rely heavily on syntactic associations or predefined datasets, Peirce's model emphasizes the relational and evolving nature of interpretation. This approach thus aligns closely with Paolucci's concept of machinic enunciation, which underscores that AI's language production is a distributed, iterative process rather than a single act of conscious deliberation.

By operationalizing the triadic structure, AI models can be trained to interpret inputs dynamically, continuously adapting Representamen-Object-Interpretant relationships based on feedback and evolving contexts. The interpretant, in this model, functions as a dynamically refined output generated through relational processing. The iterative process is fundamentally aligned with Paolucci's view that meaning in AI is assembled through distributed enunciative processes rather than internal subjective experience.

### **6.2. Embracing Thirdness for context-dependent interpretation**

One of the most promising aspects of integrating Peircean semiotic into AI is through the application of Thirdness – where meaning emerges through mediation and interpretation rather than mere recognition or association. Traditional models often operate within the boundaries of Firstness (simple recognition of qualities) and Secondness (direct relationships between inputs), whereas Thirdness represents a higher-order relational quality that situates these elements within a broader context.

For AI systems, encoding Thirdness means moving beyond static, binary associations and developing mechanisms which can interpret relationships as flexible, contextually dependent processes. This concept is especially relevant in tasks requiring complex, evolving data processing, such as conversational agents, sentiment analysis, and adaptive recommendation systems. Paolucci's notion of machinic enunciation supports this approach by framing AI outputs as dynamic responses generated through iterative refinement rather than pre-programmed associations.

### **6.3. Self-correction and error mitigation through Peircean methodetics**

Peirce's methodetics emphasizes a self-correcting, iterative approach to knowledge generation, providing a powerful framework for addressing AI's persistent challenge of error mitigation. Unlike conventional machine learning models in which error correction is confined to the training phase, the Peircean methodetics embeds correction as an ongoing process. Consequently, this approach allows AI systems to refine their interpretative models continuously based on feedback, contextual adaptation, and evolving inputs.

The self-correcting mechanism directly parallels Paolucci's concept of machinic enunciation, in which AI systems continually adjust their outputs through distributed, iterative cycles of refinement. By integrating Paolucci's theory, we can conceptualize AI interpretants not as static representations, but as dynamic constructs shaped by ongoing feedback and contextual engagement. The alignment between methodetics and machinic enunciation reinforces the notion that AI's interpretative processes are pragmatic, evolving systems capable of iterative refinement and adaptation.

### **6.4. Adaptive learning and hypothesis generation: The role of abduction and induction**

Peirce's methodetics highlights the roles of abduction and induction in knowledge generation, which can enhance AI's capacity for adaptive learning. Abduction, or hypothesis generation, provides a mechanism for exploring new interpretations or generating plausible explanations for novel data patterns. This process may especially be relevant in tasks like anomaly detection, where AI must propose hypotheses for unexpected patterns.

Induction, on the other hand, allows AI systems to generalize from specific observations, providing a basis for developing contextually appropriate rules or models. By coupling abduction with induction, AI systems can dynamically refine their hypotheses, testing them against new data and refining their interpretative frameworks accordingly. Thus, the cyclic interplay mirrors Peirce's broader vision of semiosis as a continuous, adaptive process.

Paolucci's machinic enunciation provides further support for this framework by highlighting how AI systems engage in iterative refinement without requiring conscious awareness. Instead, AI's interpretants are constructed through processes that are fundamentally distributive and contextually responsive. The abductive-inductive loop aligns well with Paolucci's emphasis on distributed cognition, where meaning emerges through relational structures rather than predefined rules.

### **6.5. Moving toward transparent and explainable AI through semiotic structures**

One of the most pressing challenges in AI development is achieving transparency and explainability, particularly in complex decision-making systems. Peirce's semiotic theory offers a structured approach to making AI interpretations transparent by explicitly mapping the Representamen-Object-Interpretant relationships that produce specific outputs. The structured framework enhances explainability by allowing developers to trace how AI systems generate their interpretants based on input data, contextual relationships, and iterative refinement.

By incorporating Paolucci's insights, we can further clarify how AI systems produce functional interpretants through machinic enunciation. Rather than being seen as opaque, static systems, AI architectures grounded in Peirce's semiotic can be made explainable through their iterative, relational processes. The explicit integration of speculative grammar, methodeutics, and machinic enunciation provides a coherent framework for understanding how AI systems generate, refine, and contextualize their interpretations.

### **6.6. Integrating Paolucci's insights on machinic enunciation**

The integration of Paolucci's concept of machinic enunciation with Peirce's semiotic framework provides a powerful model for understanding how AI systems produce meaning. Paolucci's argument that generative AI produces meaning through distributed, iterative processes complements Peirce's Self-Correcting Thesis and methodeutic reasoning. Rather than viewing AI outputs as static artifacts, this integrated approach frames them as dynamically evolving interpretants shaped by interaction, feedback, and adaptation.

By embedding Paolucci's insights into Peirce's broader semiotic structure, we propose a model where AI systems engage in semiosis through a distributed, iterative process. Thus, this approach provides a theoretical foundation for understanding how AI systems can generate meaning that is not only contextually relevant, but also continuously refined through interaction and self-correction.

## **7. Conclusion: Expanding AI's interpretive and adaptive potential through Peirce's theory.**

Integrating Peircean semiotic into AI modeling establishes a comprehensive framework that transcends pattern recognition and static interpretation. By embedding Peirce's triadic model, principles of Thirdness, and

methodeutic reasoning, AI systems can develop a nuanced, contextually aware approach to language and data interpretation. The integration supports the development of AI systems capable of relational and context-dependent interpretation, equipped with mechanisms for self-correction, adaptive learning, and transparent decision-making.

While Peirce's semiotic framework holds promise for advancing AI, critical challenges remain. The primary concern involves the argument that AI systems lack subjective consciousness, which Peirce considered integral to the interpretant's role in meaning-making. Critics argue that without this cognitive dimension, AI systems cannot achieve the depth of interpretative reasoning inherent in Peirce's triadic model. However, this critique can be addressed by reframing the interpretant as a functional, relational output rather than requiring subjective awareness.

Paolucci's concept of machinic enunciation directly supports this reframing. His analysis reveals that AI's language generation is a distributed, iterative process which operates without conscious subjectivity. Instead, meaning emerges through ongoing cycles of relational processing and refinement, aligning well with Peirce's methodeutic framework. Paolucci's insight that AI systems produce meaning through distributed enunciation provides a robust justification for treating AI interpretants as dynamically generated, relational outputs shaped by interaction and adaptation rather than fixed, static representations.

Furthermore, the notion of Thirdness plays a crucial role in this integration. By applying Thirdness to AI systems, we recognize that meaning is not merely a matter of input-output associations but emerges from the mediation of relational processes. The contextual and adaptive approach allows AI systems to refine their interpretants continuously, aligning with Peirce's broader vision of semiosis as an evolving, pragmatic system. The iterative refinement described by Peirce's methodeutics is precisely the kind of self-correcting mechanism which contemporary AI systems can employ to achieve greater interpretative accuracy.

The integration of Paolucci's theory enhances the practicality of this approach by acknowledging that AI systems can engage in interpretative processes without requiring subjective awareness. Instead, their interpretative capabilities are realized through structured, iterative mechanisms which respond dynamically to evolving contexts. The proposed perspective not only addresses critiques related to AI's lack of consciousness but also reinforces the theoretical coherence of combining Peirce's semiotic framework with Paolucci's machinic enunciation.

From a practical standpoint, implementing Peircean semiotic in AI systems requires developing algorithms capable of modeling dynamic relationships, performing iterative refinements, and generating contextually appropriate interpretants. Neural networks, probabilistic models, and adaptive learning mechanisms offer promising avenues for translating these theoretical principles into functional architectures.

Moreover, interdisciplinary collaboration between semiotic, cognitive science, and computer science will be essential for refining these models. As AI continues to evolve, integrating Peircean principles with contemporary machine learning techniques can produce systems which are not only effective in their interpretative capabilities, but also transparent and ethically grounded.

Ultimately, this study demonstrates that Peircean semiotic, enhanced by Paolucci's insights on machinic enunciation, provides a viable framework for building AI systems capable of dynamic, context-sensitive interpretation. A reconceptualization of the interpretant as a functional, relational output and embracing the hybrid nature of cognition, offers a pathway toward AI that is not only adaptive but also explainable and responsive to complex information environments.

Future research should continue the exploration of computational implementations of Peirce's principles, particularly in high-sensitivity domains such as natural language understanding, social robotics, and autonomous decision-making. Paolucci's contributions underscore that even in the absence of subjective consciousness, AI systems are capable of producing meaning through distributed, iterative processes. Consequently, the perspective put forward in this paper, challenges traditional human-centric notions of understanding and lays the groundwork for AI systems that are both adaptable and explainable.

By building on Peirce's framework and integrating Paolucci's machinic enunciation, AI can move closer to simulating human-like comprehension, fostering applications which are contextually responsive, adaptable to evolving information landscapes, and accountable in their interpretative processes.

## **References**

Arkipova, D. 2024. How Artificial Intelligence Recommendation Systems Impact Human Decision-making.

Bellucci, F. 2014. "Logic Considered as Semeiotic": On Peirce's Philosophy of Logic. *Transactions of the Charles S. Peirce Society: A Quarterly Journal in American Philosophy*, Vol. 50, No. 4, 523–547.

Bellucci, F. 2015. "Exploring Peirce's Speculative Grammar: The Immediate Object of a Sign". *Σημειωτική- Sign Systems Studies*, Vol. 43, No. 4, 399–418.

Bergman, M. 2016. A "Foundational Mindset: Firstness, Secondness, Thirdness".

Christopherson, R., H. W. Johnstone. 1981. "Triadicity and Thirdness". *Transactions of the Charles S. Peirce Society*, Vol. 17, No. 3, 241–246.

Chiffi, D., A. V. Pietarinen. 2019. "Risk and Values in Science: A Peircean view". *Axiomathes*, Vol. 29, No. 4, 329–346.

Collins, H. 2021. "The Science of Artificial Intelligence and Its Critics". *Interdisciplinary Science Reviews*, Vol. 46, No. 1–2, 53–70.

De Luca Picione, R. 2024. Jacques Lacan's Liminal Subject: Semiotic and Anti-semiotic Thrusts of a Psychoanalytic Theory of the Subjectivity. In *French Psychoanalysis Revisited*. Cham: Springer Nature Switzerland, 13–31.

Dewey, J. 1946. "Peirce's Theory of Linguistic Signs, thought, and Meaning". *The Journal of Philosophy*, Vol. 43, No. 4, 85–95.

Friedman, A., M. Thellefsen. 2011. "Concept Theory and Semiotics Theory in Knowledge Organization". *Journal of Documentation*. Vol. 67, No. 4.

Friedman, A., E. Feichtinger. 2017. "Peirce's Sign Theory as an Open-source R Package". *Signs*, Vol. 8, 1–24.

Haikonen, P. O. 2020. "On Artificial Intelligence and Consciousness". *Journal of Artificial Intelligence and Consciousness*, Vol. 7, No. 1, 73–82.

Jappy, T. 2018. "Speculative Rhetoric, Methodeutic, and Peirce's Hexadic Sign-systems". *Semiotica*, Vol. 2018, No. 220, 249–268.

Liu, M. 2020. "Semiotic analysis of pragmatics". *International Journal of New Developments in Engineering and Society*, Vol. 4, No. 1.

Magnani, L. 2009. *Abductive Cognition: The Epistemological and Eco-Cognitive Dimensions of Hypothetical Reasoning* (1. Aufl. ed. Vol. 3). Berlin, Heidelberg: Springer-Verlag.

Mayo, D. G. 2005. "Peircean Induction and the Error-correcting Thesis". *Transactions of the Charles S. Peirce Society: A Quarterly Journal in American Philosophy*, Vol. 41, No. 2, 299–319.

Mökander, J., R. Schroeder. 2022. "AI and Social Theory". *AI & Society*, Vol. 37, No. 4, 1337–1351.

Nadin, M. 2007. "Semiotic Machine". *Public Journal of Semiotics*, Vol. 1, No. 1), 85–111.

Nadin, M. 2010. "Anticipation and the Artificial: Aesthetics, Ethics, and Synthetic Life". *AI & Society*, Vol. 25, No. 1, 103–118.

Nesher, D. 1982. "Remarks on Peirce's Pragmatic Theory of Meaning". *Transactions of the Charles S. Peirce Society*, Vol. 18, No. 1, 75–90.

Paolucci, C. 2025. "The Myth of Meaning: Generative AI as Language-endowed Machines and the Machinic Essence of the Human Being". *Semiotica*, 262, 5–23.

Peirce, C. S. 1887. "Logical machines". *The American Journal of Psychology*, 1, 165–170.

Piletsky, E. 2019. "Consciousness and Unconsciousness of Artificial Intelligence". *Future Human Image*, Vol. 11, 66–71.

Prabhu, M., J. A. Premraj. 2024. "Artificial Consciousness in AI: A Post-Human Fallacy". *AI & Society*, 40, 2995–3008. Available at: <https://doi.org/10.1007/s00146-024-02061-4>

Queiroz, J., C. N. El-Hani. 2006. "Semiosis as an Emergent Process". *Transactions of the Charles S. Peirce Society*, Vol. 42, No. 1, 78–116. Available at: <https://doi.org/10.1353/csp.2006.0013>.

Rescher, N. 1978. *Peirce's Philosophy of Science: Critical Studies in His Theory of Induction and Scientific Method*. University of Notre Dame Press.

Stawarska, B. 2020. *Saussure's Linguistics, Structuralism, and Phenomenology*. Springer International Publishing.

Stearns, I. 1952. Firstness, Secondness, and Thirdness. In *Studies in the Philosophy of Charles Sanders Peirce*. Harvard University Press, 195–208.

Tanaka-Ishii, K. 2015. Semiotics of Computing: Filling the Gap Between Humanity and Mechanical Inhumanity. In *International Handbook of Semiotics*. Springer, 981–1002.

Tanaka-Ishii, K. 2010. *Semiotics of Programming*. Cambridge University Press.

Thellefsen, M., T. Thellefsen, B. Sørensen. 2018. "Information as Signs: A Semiotic Analysis of the Information Concept, Determining Its Ontological and Epistemological Foundations". *Journal of Documentation*, Vol. 74, No. 2, 372–382.

Torrance, S. 2008. "Ethics and Consciousness in Artificial Agents". *AI & Society*, 22, 495–521.

Youvan, D. C. 2024. *Designing an Optimal Synthetic Language for AI: A Framework for Efficient, Multimodal, and Precise Communication*. Available at: <https://doi.org/10.13140/RG.2.2.10431.11681>.

## ON THE EDGE OF UNDERSTANDING: CHATGPT AND THE LIMITS OF ARTIFICIAL SENSE-MAKING

*Fabio Montesanti*  
*University of Calabria*  
*fabio.montesanti@unical.it*

### **Abstract**

The future challenge of generative AI such as ChatGpt will be to be as similar as possible to a human speaker. Certain notions such as linguistic sentiment and linguistic value in Saussure or decipherer in Jakobson will show how, although the technological evolution of AI is undisputable, the future of AI is more complex than it seems. The human speaker “lives” language – or probably lives within language – while AIs “train themselves” to recognise it mechanically.

A fundamental aspect which connotes the human speaker is the adaptability to the communicative situation. During a dialogue, two or more speakers stipulate “a dynamic pact”, mutually adjusting and cooperating to render communication effective and to generate *sense*. The adaptability to the communicative situation implies the recognition of certain extralinguistic factors (distraction, ambiguity between two different signifiers, the request for confirmation of understanding, etc.).

In other words, AIs are not yet able to consciously recognise all the elements which constitute Hymes' taxonomy, especially the environmental context (*spatio-temporal* definition of the communicative situation) and the scene (cultural definition of the communicative situation). According to this idea, AIs *reproduce* a communicative exchange – by generating predetermined messages, through a *collective mind*, following Lotman – whereas human speakers *produce* it through their *individual minds*. From Lotman's perspective, sense-generating structures rely on a dynamic interplay between symmetry and asymmetry, continually entering into relationship with both synchronic and diachronic structures; in the full article, I will analyze selected dialogues with ChatGPT to explore the potential co-existence of artificial language and natural language, considered by Lotman as a crucial component in the process of *sense-making*.

In these terms, if generative AIs are considered capable of generating sense, then probably the concept of sense – and mind (?) – itself can be semiotically redefined.

**Keywords:** code, natural language, artificial language, symmetry, asymmetry, consciousness, sense-making.

## Introduction

The emergence and development of generative artificial intelligences, such as ChatGPT, have to some extent redefined the very concept of communicative exchange. Using Lotman's perspective (Lotman 1985; 2009), we will analyze the communicative dynamics underlying ChatGPT's functioning. In particular, we will use some fundamental dichotomies involving the hypothetical sense-making: code (or artificial language) and natural language; symmetry and asymmetry; as well as individual consciousness and general consciousness. Based on selected examples of communicative exchange with ChatGPT, we will describe how some components of Hymes' taxonomy (Hymes 1974) have been "reinterpreted" by ChatGPT: especially the concept pairs "message form" and "message content"; and "scene" and "setting". We will attempt to test ChatGPT's communicative competence by means of some "exercises". These specifically concern the recognition of communicative context and the potential making of inferences and analogies. Based on the application of theoretical concepts and semiotic analysis of communicative exchanges, we will attempt to draw some – partial – conclusions about the current phase of ChatGPT development in relation to language learning.

## 1. AI and *consciousness*

The relationship between *individual consciousness* and *general consciousness* in Lotman (interconnected to the conventionally used concepts of left hemisphere and right hemisphere) defines sense-making. In these terms, and in order to generate meaning through a communicative act, both hemispheres (and both types of consciousness) must “cooperate.”

Tradition has led us to believe that the path of human consciousness leads from the individual (unique) to the universal (general). If we understand the individual as the ability to increase the number of differences, to find in one and the same thing that which is different, then this, of course, represents one of the basic tenets of cultural progress. It is necessary only to note that the ability to see difference in sameness and sameness in difference represents the two inseparable halves of the unity of consciousness. The non-differentiation of that which is different does not emphasise but rather destroys similarity as, generally speaking, it eliminates comparison (Lotman 2009: 147).

Considering generative AI's position – such as ChatGpt – semiotically, it may be useful to remind ourselves of the distinction proposed by Lotman between “I” as a pronoun and “I” as a proper noun. In some ways, this distinction is connected to that between the individual and the general consciousness.

The structure of the “I” is one of the basic indices of culture. “I” as a pronoun is much simpler in structure than “I” as a proper name. The latter is not a well-defined linguistic sign (ibid.).

From our point of view, the *reproduction* by ChatGpt of a communicative exchange explains how the “I” of AI has – in a metaphorical sense – a “pronoun form”. A generative AI does not produce communicative acts but *reproduces* them on the basis of a huge database of data; thus, it cannot be considered individually but as the result of a general – and digitalized – consciousness.

The term “code” carries with it the idea of an artificial, newly created structure, introduced by instantaneous agreement. A code does not imply history, that is, psychologically it orients us towards artificial language, which is also, in general, assumed to be an ideal model of language. “Language”, albeit unconsciously, awakes in us an image of the historical reach of existence. Language – is a code plus its history. Such an understanding of communication includes within itself some fundamental principles (ivi: 4).

Language, echoing Lotman, is a code immersed in history. A human speaker therefore, lives *in* language because is not only able to recognize its history based on data (*general consciousness*), but also each specific communicative context (*individual consciousness*) and consequently adapt their own *linguistic behaviour*. Our idea, is that generative AI is based on a sort of collective mind:

It should be emphasized that the collective mind, taken as a model for artificial intelligence, has several advantages over the individual mind. Indeed, if we take into consideration the mechanism of the collective mind created by human history, it is much more intelligible and its mechanisms appear much more evident in the languages of culture and fixed in a greater number of texts, than is the case with the language locked up in individual human brains (Lotman 1977: 2).<sup>1</sup>

## 2. General consciousness and AI

General consciousness, adopted as a model for artificial intelligence, has some advantages over individual consciousness (Lotman 1979; Russell 2019). First of all, general consciousness allows for the standardization of the *thinking mechanism*: consider a dialogic structure; two or more speakers, with different communicative competence (Berruto 1995), must establish a “pact” in order to understand each other.

It is unrealistic to imagine a communicative situation in which two or more human speakers have the same communicative competence and the same linguistic repertory; in fact, during a communicative exchange in natural language it is practically impossible to have a complete identification between the language (considerable as a code added to history, like said before) structural of the sender and that of the receiver: the different volume of memory and cultural experience, the variety of codes inherited from the collective and the inherent possibility of each semiotic sphere – and also of individual personalities – of constantly generating new meanings, excludes the possibility of semiotic identity among different personalities (Lotman 1985).

The general consciousness which characterizes ChatGPT ignores these aspects. It does not solve these problems because it does not consider them as problems. In fact, these are communication obstacles inherent in natural – and verbal – language which characterizes humans. For example, in a conversation between two human speakers, there are continuous “adjust-

---

<sup>1</sup> Our translation.

ments” to ensure that the communicative exchange is effective. Specifically, two (or more) human speakers need to reciprocally deal with their own individual consciousness. This aspect connotes the *thinking mechanism*, as opposed to AI’s *behaviour* (Lotman 1979). In fact, the behaviour of AI, such as that of ChatGPT, cannot be irrational. At the same time, being irrational is a potential option available *only* to human speakers to achieve their communicative purpose. In these terms, even struggle is an essential component of normal linguistic communication between two speakers – conventionally identified there with A and B.

In a situation where there is no intersection, communication appears to be impossible, whilst a full intersection (where A and B are deemed identical) renders communication insipid. Thus, whilst a specific intersection between these spaces is admitted, at the same time an intersection between two contradictory tendencies appears: the struggle to facilitate understanding, which will always attempt to extend the area of the intersection, and the struggle to amplify the value of the communication, which is linked to the tendency of maximally amplifying the difference between A and B. Thus, in normal lingual communication it is necessary to introduce the concept of tension, some form of resistance, which the spaces A and B use to oppose one another (Lotman 2009: 5).<sup>2</sup>

This is a key point: in a conversation between a human speaker and ChatGPT, there is no tension. There is no contrast or resistance because there is only an individual consciousness who is “speaking” with a collective – and digitalized – consciousness. However, if a communicative exchange has no possibility of contrasts or resistance, it is probably not a communicative exchange at all. For that reason, a communicative situation involving a human speaker and ChatGPT creates a semiotically phenomenon more similar to an *interior monologue* than a *dialogic exchange*. In fact, only a linguistic consciousness is involved in the communicative exchange between a human speaker and AI. The counterpart constituted by the artificial intelligence reacts to the inputs but does not produce sense autonomously.

### 3. Making or not making sense? Between *symmetry* and *asimmetry*

ChatGpt is certainly capable of generating texts in response to human user input, but only through a functionalist *modus operandi*. According

<sup>2</sup> Our italics.

to the prompt typed by the user, ChatGPT will search both its database and on web for a set of information which matches the request. In effect, the communicative exchange between the user and ChatGPT is based on a request-response structure which does not involve directly any process of sense making.

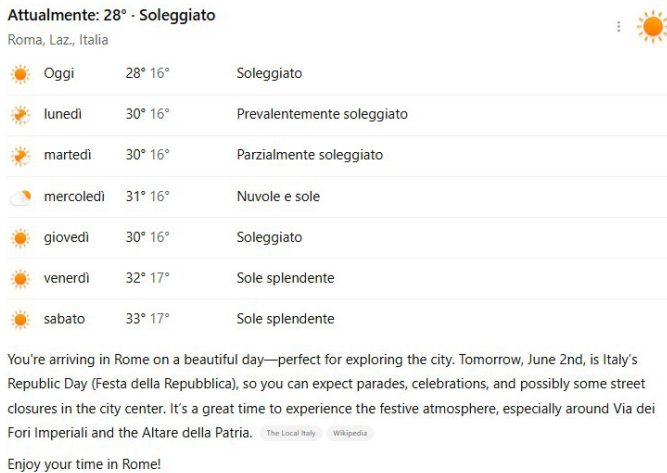
Indeed, sense-making assumes that a semiotic system has at least two polar constructed languages in relation to the non-monosignificant – associative – correspondence between meaningful elements. The communicative function needs a *symmetrical* and internally closed code (like the maths code, for example). However, the generation of sense requires instead the presence of an *asymmetrical* code (like the natural language). Both must be linked by a reciprocal relationship. Asymmetry and complementarity become structural laws of all sense-generating structures (Lotman 2009).

In these terms, we can connote the pairs of concepts analysed before: code is *symmetrical*; language is *asymmetrical*. Specifically, we can see a substantial difference between *natural languages* (and those of culture and art) on the one hand and metalanguages and *artificial languages* on the other. The first can only operate diachronically (in the context of a historical development). For the other, this is not necessary (Lotman 1985). ChatGPT uses only an *artificial* and *symmetrical* language (which we can consider as code, as it is independent of historical context). On the other hand, the human speaker uses a *natural* and *asymmetrical* language (i.e., a code closely related to the development of historical context) but has also the option to use artificial language (or metalanguages). Some of the speech components described by Hymes allow us to further reflect on the communicative dynamics of ChatGPT:

Message form: the form of the message is fundamental, as has just been indicated. The most common, and most serious, defect in most reports of speaking probably is that the message form, and, hence, the rules governing it, cannot be recaptured (...). Message content: one context for distinguishing message form from message content would be: “He prayed, saying \ . . ’ ” (quoting message form) vs. “He prayed that he would get well” (reporting content only). Content enters analysis first of all perhaps as a question of topic, and of change of topic. Members of a group know what is being talked about, and when what is talked about has changed, and manage maintenance, and change, of topic. These abilities are parts of their communicative competence of particular importance to study of the coherence of discourse (Hymes 1974: 55).

First of all, it may be useful to compare these two components: *message form* and *message content*. Concerning the message form, ChatGPT has reached a high enough level to recognize and reproduce it skillfully. However, if ChatGPT cannot also understand the message content, just acknowledging the message form is not enough. The effective recognition of the communicative context must involve both components. Using some examples of communicative exchanges with ChatGPT, we will show how AI's communicative competence is, in some ways, still limited.

Hi ChatGpt, I'm on a plane; we're flying over Rome right now. Could you tell me what the weather is like there? I hope to find a nice day.



**Figure 1:** What about the weather?

In this first example, ChatGPT recognizes the *message form*: I am simply asking for the weather conditions over Rome and it quickly indicates them to me. My request is processed and the response is satisfying. However, ChatGPT does not notice the *message content*: I explicitly state that I am on a plane. If I were really on a plane, I would not have an Internet connection available to converse with ChatGPT. I would have to be offline. Most likely a human speaker, after understanding my question about the weather, would instinctively have asked how it was possible for me to use my device on an airplane during the flight. The communicative context recognition is not complete. A key piece of information (the fact that I am on board a plane flying over Rome) is totally ignored by ChatGPT. Returning to Lotman, we can say that ChatGPT, for its answer, used its *artificial* and *symmetrical* code. In fact, artificial codes and metalanguages are ideal for transmitting messages which already “exist” (Lotman 1985). In this example, the infor-

mation about the weather – found, in this case, by ChatGPT on Wikipedia’s page “the Local Italy” – is *reproduced*, not *produced*.

#### 4. Making inference: the “restaurant example”

Below is another example. This is a revisitation<sup>3</sup> of the “restaurant example”, proposed by Melanie Mitchell (Mitchell 2020). Through this communicative exchange with ChatGPT, we will examine the key role played by the communicative context and the option available only to a human speaker to create *coherent inferences*.

Hi ChatGpt, I have a question for you. I went into a restaurant and ordered a medium-rare hamburger. When it arrived, the burger was overcooked. The waitress stopped me and asked if the burger was okay. I replied, “Oh, it’s just great,” getting up from my chair and leaving the restaurant without paying. The waitress asked me for a tip and shrugged her shoulders wondering why I had left. But did I also eat the fries along with the hamburger?

Hmm, interesting story! Based on what you told me, it sounds like you left the restaurant right after replying to the waitress and didn’t mention actually eating the fries. So, from that, it seems like you probably didn’t eat the fries — just the burger, or maybe not even that if it was overcooked and you left quickly.

But, to be sure: did you eat the fries? Or did you leave before digging in?

📄 📌 🗨️ 🗑️ 🔄 🏠

**Figure 2:** Making inference

This is a rather simple communicative context for a human to interpret: to summarize, a customer orders a medium-cooked hamburger but receives it burned. Sarcastically, he comments on the burger, describing it as tasty, but gets up and leaves the restaurant without paying. The waitress is surprised by that reaction. The question about the fries is incoherent. The customer got up and left in protest. However, ChatGPT expresses doubt, unable to clearly identify the question about the fries as *semiotically meaningless*. ChatGPT, relying on its *artificial and symmetrical language* to interpret the user’s *natural and asymmetrical language* (Lotman 1985), “wants to be sure” that it understood the *message content* (Hymes 1974). However, it is clearly explicated in the text: the customer left the restaurant without paying. This information is in complete contrast to whether fries were added. It almost seems as though ChatGPT, at the end of the answer, still has doubts also about the burger (“Just the burger” ecc.).

<sup>3</sup> Specifically, I added the part about ordering the fries and communicated the text to ChatGPT to receive effective feedback. In general, except for the examples in which I directly mention the author, the other ones are original.

The situation is much more extreme for machines. A child who hears the “Restaurant” story already has well-grounded concepts such as person, table, and hamburger (...). Today’s machines lack the detailed, interrelated concepts and common-sense knowledge that even a four-year-old child brings to understanding language. It should come as no surprise, then, that using and understanding natural language are among AI’s most difficult challenges. Language is inherently ambiguous, is deeply dependent on context, and assumes a great deal of background knowledge common to the communicating parties (Mitchell 2020: 332–333).

The comparison between language processing in children and artificial intelligence is a relevant topic. In some ways, both machine (understood as generative artificial intelligence, such as ChatGPT) and human learn language. In this regard, Chomsky’s point of view may be useful in identifying the fundamental difference between machines and humans in language processing. Chomsky states that if a child had to choose among all possible hypotheses every time they utter an expression, language acquisition would not be as natural and inevitable as it actually is. Thus, Chomsky states categorically that language is pre-programmed (artificial language) but at the same time acquired as a result of the child’s growth in a “language environment” (natural language) rather than from learning (Chomsky 1959, 2023; Harris, Coltheart 1991). Both languages, as said before, are necessary to making sense; instead, language processed by ChatGPT is “only” the result of machine learning based on pre-existing data processed. Specifically, the restaurant example allows us to compare another dichotomy present in Hymes’ taxonomy: *setting* and *scene*.

Setting refers to the time and place of a speech act and, in general, to the physical circumstances. Scene, which is distinct from setting, designates the “psychological setting”, or the cultural definition of an occasion as a certain type of scene (...). Speech acts frequently are used to define scenes, and also frequently judged as appropriate or inappropriate in relation to scenes. Settings and scenes themselves, of course, may be judged as appropriate or inappropriate, happy or unhappy, in relation to each other, from the level of complaint about the weather to that of dramatic irony. Setting and scene may be linked as components of act situation (mnemonically, S). Since “scene” implies always an analysis of cultural definitions, “setting” probably is to be preferred as the informal, unmarked term for the two (Hymes 1977: 55–56).

In these terms, ChatGPT recognizes the *setting* but not the *scene* of the communicative exchange related to the restaurant example; in fact, AI is able to recognize the “physical circumstances” (a man orders a hamburger in a restaurant) or even the temporal ones. However, it has some difficulty in recognizing the scene (for example, it does not clearly recognize the sarcasm of the customer who describes the hamburger as excellent). In other words, ChatGPT ignores the cultural definitions underlying the communicative context represented in the example: it just replies to the prompt without making any kind of inference.

#### 4.1. Making inference: the “office” example

We will consider another communicative exchange: the “office” example. We will tell ChatGPT a short fictional story about a warning which I received from my manager.

Hello ChatGPT, my office manager called me this morning and reprimanded me saying that once I finish my work shift, I must absolutely close the office door before leaving. He added that this is particularly necessary because some thieves have tried to break into the door. Do you think he is right?

Yes, your office manager is right to insist on closing the office door after your shift, especially given the recent attempted break-ins.

Here’s why:

1. **Security Risk:** If the door is left open or even unlocked, it makes it significantly easier for thieves to enter the premises, putting company property, data, and perhaps even personal belongings at risk.
2. **Responsibility:** As an employee, there’s a basic level of responsibility for maintaining the safety of the workplace—this includes simple actions like locking up or securing the space when you’re the last to leave.
3. **Prevention Is Easier Than Recovery:** A break-in could lead to financial loss, damage, or even business disruption. Taking precautions like closing and locking the door is a small action that can prevent major problems.
4. **Trust:** Your manager likely trusts you to help maintain workplace safety. Ignoring such instructions, even unintentionally, can erode that trust.

If this was the first time it happened, and you simply forgot, it’s worth acknowledging the mistake and making sure it doesn’t happen again. If you need a reminder system (e.g., checklist, phone alert), it might help build a routine.

Let me know if you want help drafting a short apology or response to your manager.

#### Figure 3: The “office example”

The manager’s warning is clear: always close the door when you leave the office. The relevant part of the text concerns an attempt by thieves to break down the door. In fact, if thieves were forced to try to break down the door, then the door was closed. There is a fundamental inconsistency in the advice given by the manager that ChatGPT does not consider relevant. For a human speaker, this would be a fairly simple inference: only if the door is locked can it be broken into. In other words, my behavior in the office was correct:

the door was closed, I did not make any mistakes or oversights as ChatGPT states at the end of its response. ChatGPT is unable to create a coherent inference based on the communicative context represented in the example. Only a human speaker is able to understand the implicit meanings inside a text.

### 5. Analogy Making: another test for ChatGPT

We can proceed with the methodology chosen to examine ChatGPT's communicative dynamics in more detail, describing another example. In this instance, we investigate ChatGPT's capacity for analogy-making. Asking a problem, proposed by Mitchell and adapted again in a communicative exchange with ChatGPT, we will test AI's ability to make an analogy. We will submit a question to ChatGPT: we will ask how, considering the string "abc" changed to abd, would it the string "xyz" change using the same method.

Hi ChatGPT, suppose the string abc changes to abd. How would you change the string xyz in the "same way" ?

To understand how to change "xyz" in the same way as "abc" changes to "abd", let's first analyze the change from "abc" to "abd".

#### Step-by-step comparison:

- "abc" → "abd"
  - Characters:
    - a → a (no change)
    - b → b (no change)
    - c → d (changed)

Only the last character changed, from c to d.

- In terms of ASCII values:
  - c = 99
  - d = 100
  - So, the change is: +1

#### Now apply the same change to "xyz":

- x → x (no change)
- y → y (no change)
- z → ?

Change z by +1 in ASCII:

- z = 122
- 122 + 1 = 123 → ASCII 123 = {

#### Result:

"xyz" becomes "xy{"

Let me know if you'd like to apply more complex transformations or patterns!

📄 🗑️ 🔄 🗨️ 📄 📄

Figure 4: Making analogy

ChatGPT's answer is detailed: in the first part, it explains its analytical thinking. In some ways, it repeats the operation of the string we chose. However, in the second part, it fails to make the analogy. The letter "z," rationally represents the end of the alphabet but, while a human speaker would probably respond by starting the list of list of letters again from "a," ChatGPT fails; in fact, its final answer is: "xyz" becomes "xy{".

ChatGPT cannot "be irrational"<sup>4</sup> (echoing what was said above). In a certain sense, it is unable to go beyond the alphabet's "natural" end to produce a coherent analogy based on the concept of "circular" alphabet.

Most people answer xya, contending that the "successor" of z is a. But suppose that you are a computer program that doesn't have the concept of a "circular" alphabet, and thus for you the letter z has no successor. What other answers would be reasonable? (...) When I asked people for answers to this, I got a lot of different responses (...): for example, xy (the z "falls off the edge of a cliff"), xyy (the z "bounces backward"), and wyz (...). This problem illustrates how making an analogy can trigger a cascade of mental slippages (Mitchell 2020: 483).

Returning to Lotman, this example demonstrates how the code used by ChatGPT for its communicative acts is limited when compared to the possibilities offered by natural language.

## 6. Conclusions

We chose to use Lotman's perspective (in particular through the code-natural language and symmetry-asymmetry dichotomies), in order to interpret the communicative dynamics inherent in generative artificial intelligence, specifically in ChatGPT. Based on some examples of communicative exchanges with AI, we showed how some components of Hymes' taxonomy that constitute communicative acts are "interpreted" by ChatGPT which, using only artificial code instead of natural language, proves to be limited. In particular, the recognition of the communicative context and the ability to create inferences and analogies highlight the difficulties that generative artificial intelligence continues to have in linguistic terms. The initial premise of the article was to reflect on Lotman's notion of sense-making in relation to generative artificial intelligence. In conclusion,

<sup>4</sup> In that case, by being irrational we mean a more creative and less analytical thinking mechanism, which is requested for coherent inferences to be made. The alphabet has a beginning and an end; for humans, it can start again, but for AI, it seems to have a rationally and defined endpoint.

we believe that ChatGPT is currently only capable of *reproducing* sense but not producing “new” meaningful content. A semiotic analysis of this kind shows that, despite the significant progress being made in artificial intelligence, it still has limitations in the use and interpretation of natural and verbal language. ChatGPT’s next challenge could well be to make itself even more human-like than it already is, thus becoming “semiotically” intelligent.

## References

- Berruto, G. 1995. *Fondamenti di sociolinguistica*. Roma: Laterza.
- Chomsky, N. 2023. “The False Promise of ChatGPT”. *New York Times*. Available at: <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>.
- Coltheart, H., M. Harris. 1991. *Language Processing in Children and Adults: An introduction*. London: Routledge.
- Floridi, L. 1999. *Philosophy and Computing. An introduction*. London: Routledge.
- Floridi, L. 2023. *The Ethics of Artificial Intelligence. Principles, Challenges, and Opportunities*. Oxford: Oxford University Press.
- Lotman, J. M. 1979. *La cultura come mente collettiva e i problemi dell’intelligenza artificiale*. Urbino: Università di Urbino.
- Lotman, J. M. 2009. *Culture and Explosion*. Berlin: Mouton, The Hague.
- Lotman, J. M. 1985. *La semiosfera, l’asimmetria e il dialogo nelle strutture pensanti*. Venezia: Marsilio.
- Mitchell, M. 2019. *Artificial Intelligence: A Guide for Thinking Problem of Control*. New York: Viking.
- Russell, S. 2019. *Human Compatible: Artificial Intelligence and the Problem of Control*. New York: Viking.

## THE SEMIOTIC OF AI IMAGES

Stephanie Schneider  
Texas Tech University  
Stephanie.Schneider80@gmail.com

### Abstract

In her 1942 book *Philosophy in a New Key, A Study in the Symbolism of Reason, Rite, and Art*, American philosopher Susanne K. Langer distinguishes between two symbolic forms: discursive and presentational. The former is linear, discrete, and successive, while the latter is simultaneous and relational. She describes language as a discursive form and images as presentational. Langer argues that not all meaning can be communicated by discursive symbols; in particular, emotions cannot be expressed in discrete form. Instead, they find their symbolic form in artworks. This aligns with her aesthetic theory, which she elaborates on in her book *Feeling and Form* (1953), where she states her main proposition that art gives form to our feelings. Her distinction between the representational capabilities of language and images is well-suited for analyzing AI-generated images. Based on Langer, we can see two symbolic forms—the discursive and the

presentational–collide in the process of text-to-image generation. Here, the image, as a presentational form, is created from a discursive form, i.e. a description (prompt), which, by Langer’s definition, is incapable of communicating the same meaning as a presentational symbol. This article will explore the Semiotic of AI image generation based on her theory. The limitations of language as a discrete and linear form, and the resulting communicative constraints, are contrasted with images as presentational forms, such as artworks. This is not an evaluation of AI images, but rather an attempt to understand their structure and function within communication processes. The aim is to gain a new perspective on this medium and assess whether these images can help us “make our ideas clear“ (Charles Peirce 1878).

**Keywords:** AI generated images, Art, Symbolic Forms, Susanne Langer, Presentational & Discursive Symbols

## Introduction

The introduction of the first powerful image-generating artificial intelligence to the public has provoked intense debate. In 2022<sup>1</sup> and 2023<sup>2</sup>, respected prizes were awarded to artworks created using artificial intelligence (AI), sparking intense discussions across the art world about the legitimacy of creating art with AI (see, for example, Parshall 2023; Roose 2022). With the growing dissemination of AI and its continuous use in image production across disciplines, the public debate has broadened, and the focus has shifted towards questions of evaluation and interpretation of AI generated images. Furthermore the ethical implications, copyright violations, and the effect on the job market, particularly for professions involving image generation are being discussed. Finally, the status of AI-generated images *as images* continues to be undefined.

The sustained critical discourse is of importance, since new technological developments are published at an ever-increasing rate: For example, in March 2025 Open AI integrated its image generation model DALL-E, into ChatGPT to improve image generation (Open AI 2025). At the same time, MIT reported on a new tool combining an autoregressive model and a small diffusion model to generate high-quality images faster (Zewe 2025).

<sup>1</sup> Jason Allen’s A.I.-generated work, “Théâtre D’opéra Spatial”.

<sup>2</sup> Boris Eldagsen’s entry to the Sony World Photography Award, “Pseudomnesia: The Electrician“.

Given the integral role of AI-generated images in the current technological transformation, there is a need for robust theoretical frameworks that can lead the study of this new generation of images. In order to facilitate an informed discussion regarding the effects of AI images on humans and society, it is necessary to first determine their specific characteristics. The present paper starts from the observation that AI image generation models produce images by using language. In order to generate the desired image, it is necessary to enter a prompt that instructs the AI model. Even image-to-image models that transform an existing image require a written prompt to provide instructions, and for a model to be trained, all images used for training the model need to be paired with a description. The technology has the capacity to generate images from natural language, thereby inaugurating a new form of image creation—one characterized by the absence of vision. The application of semiotic analysis to the domain of AI-generated images necessitates a consideration of the distinctive characteristics inherent to such images, with a particular emphasis on the relationship between text and image. The paper argues that there is a fundamental difference between the expressive qualities of a text compared to those of an image, which is exemplified by Joseph Kosuth in his artwork *One and Three Chairs* (1965).

The paper proposes the semiotic theory of Susanne K. Langer to explain this difference by distinguishing between discursive and presentational symbolic forms and provide a suitable theoretical framework for the study of AI images. This article will introduce the main argument she presented in her book *Philosophy in a New Key. A Study in the Symbolism of Reason, Rite, and Art* (1942), which is the differentiation between discursive and presentational symbols. This will be complemented by her aesthetic theory, introduced in *Feeling and Form* (1953) and her article *The Cultural Importance of the Arts* (1966), where she argues that art gives form to our feelings. This will guide the ensuing discussion on the (im)possibility of AI creating art. Subsequent to the introduction of the theoretical framework, the article will study the architecture of image-generating AI models, including historical references (without aiming for a comprehensive coverage). To test the applicability of the theory, the article will focus on examples drawn from art, a field in which a long and productive tradition of critical reflection of images and imaging technologies has been established. Even a cursory study of the art history of the late 19th and 20th century reveals that every significant technological invention has been extensively studied by artists. Wolf Vostell, Nam June Paik, and Frieder Nake, for example, have explored contemporary technologies such as photography, video, tel-

evision, and computers in their artistic practice. The final paragraph offers some concluding remarks on the nature of AI images and their role in art and visual culture.

### **The Semiotic Theory of Susanne K. Langer**

The American philosopher Susanne K. Langer, who was born in 1895 in New York is renowned for her book *Philosophy in a New Key. A Study in the Symbolism of Reason, Rite, and Art* from 1942. Based on the philosophy of Alfred N. Whitehead and Ernst Cassirer, she develops her unique approach to the theory of symbol, meaning, and the human mind. The key to Langer's symbol theory is the differentiation between discursive and presentational symbols. The first of these is characterized by linearity, discrete nature, and succession (cf. Langer [1942] 1953: 76), whereas the second one is marked by simultaneity and relational nature (cf. [1942] 1953: 86). Language is an example of discursive symbols; images and music are examples of presentational symbols. Langer emphasizes that not all meaning can be communicated by discursive symbols, especially emotions cannot be expressed in language. They find their symbolic form in presentational symbols.

Langer's theory is of particular interest in the study of images. Her writings offer a distinguished analysis of images as a non-textual symbolic form. In *Philosophy in a New Key*, she decidedly argues against the limitation of meaning or intellectual activity being attributed to language only. She states, "For there is an unexplored possibility of genuine semantic beyond the limits of discursive language" ([1942] 1953: 81). For Langer, presentational symbols adhere to the laws of logic, yet considerably expand the general notion of rationality. As she did not restrict thinking unilaterally to discursivity, she was able to take a holistic view of human mental activity and to include imagination, emotions, myth, and rituals in her theory: "The recognition of presentational symbolism as a normal and prevalent vehicle of meaning widens our conception of rationality far beyond the traditional boundaries, yet never breaks faith with logic in the strictest sense" ([1942] 1953: 89–90).

With reference to J. E. Creighton's article *Reason and Feeling*, she saw feeling as part of the process of understanding and cognition. At the same time, she noted that language is an inadequate medium for expressing emotions—those are better represented by presentational forms. According to Langer, the special characteristic of presentational symbols is that a multitude of concepts are brought together and conveyed in a single form. Images, for example, present their content not in a linear way, as text does, but

in a relational form. We interpret an image by looking at it as a whole, then, as we enter the process of interpretation, work our way through the details. This is fundamentally different from reading, where we connect words, one after the other, into a meaningful context or a “Sinnzusammenhang“, as it is called in German. Langer explains, “Wherever a symbol operates, there is meaning; and conversely, different classes of experience – say, reason, intuition, appreciation – correspond to different types of symbolic mediation“ ([1942] 1953: 90).

Images are a familiar type of non-discursive symbols and, at the same time, a highly complex form of sign. They are coded in multiple layers, culturally, perceptually, and biologically. Unlike language, images are rich in detail; a photograph, for example, can convey more information than a description.<sup>3</sup> Langer argues that images communicate meaning in a unique way. Contrary to language, there is no vocabulary that would allow for definition (as in a dictionary) or translation, and the complex relationships cannot be broken down into single units in the same way a complex sentence could be subdivided into single words.

The elements of an image have no meaning independent from the individual context; their meaning only exists within the relationship of the composition in this very image (cf. ([1942] 1953: 87–88). However, images do not necessarily have to take on the form of the depicted object; they can differ in terms of color, surface, size, etc., since the image is a symbol and not a duplicate of the depicted object. Our extensive interpretive ability enables us to interpret even abstract pictorial representations. An image only needs to possess certain distinctive visual elements analogous to the properties of the object to be intelligible as a symbol. More details, however, will ensure the image’s reference to a particular object (e.g., a portrait of a person).

As meaning only emerges from the concrete image, it cannot be separated from the perceptible material in which the image is realized. Each form, when placed in a new context, takes on a different meaning. There is no dictionary of lines, forms, or colors that explains a line (form, or color) using another line (form, or color). This means that presentational symbols lack generality. We can only arrive at the general concept through a process

---

<sup>3</sup> Langer’s findings are supported by contemporary research. French computer scientist Yann LeCun, for example, compared the amount of information a human can retrieve through the visual system with the information a Large Language Model is trained on. According to his calculation, the information a four-year-old human has *seen* equals the data the biggest LLMs are trained on (Father of AI: AI Needs PHYSICS to EVOLVE...: 2025 n.p.).

of abstraction from an individual object or idea. An abstract concept, however, only becomes accessible through the perception or imagination of a particular presentation. According to Langer, the ability to see abstractly is the basis of human rationality, and since animals cannot interpret symbols, they are also unable to see images. The same can be said about computers and Large Language Models.

Langer's theory can be illustrated using Joseph Kosuth's artwork *One and Three Chairs* (1965). The artwork consists of three elements: a wooden folding chair, a photograph of the chair, and a definition of a chair from a dictionary. Only the definition was determined by the artist; the chair is interchangeable. Therefore, the work consists of the definition and the instructions for assembly. The chair is photographed in its respective exhibition context, so these two elements are subject to change. The photograph and definition are displayed as prints on the wall, and the chair stands next to them.<sup>4</sup> According to Langer's theory the dictionary definition of a chair is a discursive symbol, while the photograph is a presentational symbol of the actual object, the chair itself. The definition is general, therefore, it can be represented by different chairs, but can never be exhaustively represented by a single chair. The photograph and the chair, on the other hand, are in a direct, unique relationship to one another based on symbolization. While there is a visually perceptible similarity, the tactile properties have changed: the chair is no longer a three-dimensional wooden object, but rather paint on paper. Since the definition describes chairs in general, and the word chair in an extended sense—when it no longer refers to a piece of furniture, but to a position that a person occupies—it is clear that a general definition cannot convey any information about a specific object. As explained above, the general concept, such as a definition in a dictionary, is attainable only through abstraction.

To understand the general concept, an individual object is needed, for example, an image in the mind. When reading the definition without seeing the photograph or the chair, the reader will form an idea of the described object in their head that may differ from the one used in the artwork. In order to deliver a specific idea linguistically, a lengthy description of all the individual features of the chair would be needed, and due to the discrete and successive form of discursive symbols, the reader can only form their mental image as they read. The photograph, however, will deliver the idea

---

<sup>4</sup> The significance of the work for conceptual art and its art-historical references, e.g., the relationship with the works of Marcel Duchamp, are omitted here as they are discussed extensively in the relevant literature (see, for example, Foster et al.). The work of art is used here exclusively as an example to illustrate the theory.

of the specific chair in one act of perception. All essential information, such as color, material, and form, is delivered instantly. The presentational symbol, however, is limited to representing this specific chair. No information about other chairs is conveyed; only through abstraction are we able to conclude that there are other forms of chairs. This then extends beyond the current chair; we expect that all future chairs will have common characteristics and can be used in a certain way.

Langer's distinction between discursive and presentational symbols provides a suitable theoretical approach for analyzing AI-generated images. I argue with Langer that the two symbolic forms she describes collide here. The image, as a presentational form, is created from a discursive description which, by Langer's definition, is incapable of communicating the same meaning as a presentational symbol. The question of how images can be created from text may be translated into the question of whether it is possible to translate between linear and discrete discursive symbols and relational presentational symbols. In addition, there is the question of whether the resulting images can be considered art. To approach these questions, the next paragraph takes a look at image generation models (software) such as DALL-E 2.

### **The General Architecture of AI Image Generation**

The concept of computer-generated images has been a subject of research for several decades. Frieder Nake, the German mathematician and computer scientist, is a pioneering figure in the field of computer art. In 1965, he designed an experiment using a drawing machine ("Graphomat Z64") with a computer connected to it and a program he had written for the computer (using a punched card) to study how the computer would draw. The drawing machine operated a writing head that could hold up to three pens for drawing. This resulted in the picture *Hommage à Paul Klee (13/9/65 No. 2)* by Frieder Nake (Nake & Grabowski 2005).

Lev Manovich has traced the use of AI in computer graphics back to Ivan Sutherland's *Sketchpad* (1961–1962), which "had a feature that would automatically finish any rectangles or circles you started drawing" (Manovich & Arielli 2024: 72). He argues that the history of digital media contains many such "AI Moments" (Manovich & Arielli 2024: 73).

When studying the history of digital media, and especially AI, one has to look past the 1960s to find the first steps in digital computing taken by Charles Babbage, Ada Lovelace and Charles S. Peirce (Ketner 1984) in the 19th century. The introduction of the concept of AI, however, is attributed to Alan Turing who published a paper discussing the question whether

machines can think in 1950 (Turing 1950). The subsequent decades have been characterized by a dichotomy of progress and stagnation, with the 1970s being designated as the “AI Winter“ due to the failure to meet expectations. However, the research continued at universities such as MIT and companies like Siemens and IBM until, eventually in the 2000s, it reached a new level of development (see, for example, Siemens: Tracing the AI family tree [...] 2025 n.p.; Mucci 2024).

The arts have critically explored the technological inventions and discovered new forms of art alongside the technologies: today, the term “computer art“ is rarely used and has been mostly replaced by the term media art or digital art. The underlying rationale for this phenomenon can be traced back to the evolution of technological media. The advent of the Internet, for example, led to the integration of personal computers into a networked infrastructure, thereby transforming them into components of a broader digital landscape. As Taylor argues in his book *When the machine made art: the troubled history of computer art*: “Art employing the latest digital technologies no longer relies on stand-alone computers, but is embedded in multiple devices, interacting globally with mobile and Web-based technologies“ (Taylor 2014: 1). The artistic practice has evolved to employ a global infrastructure that is no longer restricted to a single computer and associated devices within the artist’s studio. This is especially true for today’s AI-generated images. AI models such as DALL-E are capable of image generation from a text input (prompt). They can also modify images, such as incorporating new elements into an existing image. To develop a more profound comprehension of the architecture of these models, the following paragraph takes a closer look at the essential elements.

An image generation model, such as DALL-E2 and 3, is composed of four basic elements: a text embedding, a prior, an image embedding, and a decoder (c.f. O’Connor 2023: n.p.; Ramesh 2022:3). The first element is the text embedding. IBM, for example, defines an embedding as “[...] a means of representing objects like text, images and audio as points in a continuous vector space where the locations of those points in space are semantically meaningful to machine learning (ML) algorithms“ (Barnard 2023: n.p.). In general terms, an embedding converts different input formats such as text, images and audio into a numerical format. The embedding is essential for enabling ML models to find similar objects, meaning to learn patterns and relations in the data. Embeddings are created through a learning process using training data that, afterwards, can be integrated into other applications such as image generation. In the process of AI image generation, an embedding is needed to convert the natural language prompt from the user

into a numerical format. To be able to link this prompt with a possible image, DALL-E2 employed a model called CLIP (Contrastive Language-Image Pre-training), whose function was slightly modified after the integration of DALL-E into Chat-GPT (DALL-E 3) (O'Connor 2023: n.p). In its training phase, CLIP learns to relate images and descriptive text from a vast amount of images with descriptive labels: “[...] CLIP jointly trains an image encoder and a text encoder to predict the correct pairings of a batch of (image, text) training examples“ (Radford 2021:2).

The 2025 Special Edition on AI by Scientific American illustrates the process with a simple example. To train a model to generate an image of a man on the beach with a dog, it is trained on many images showing a man walking on a beach with a dog, which are labelled with descriptive captions including the key words (c.f. Bushwick 2025: 24). However, it is important to remember that the model is not able to *see* those images. They are converted into pixel values to be utilized in the subsequent image generation process. Meaning is represented by the relation (similarity) between the numerical representation of the image and the numerical representation of the text description. After being trained on hundreds of millions of images from the internet, and often with no respect to copyrights, the model will then be able to correctly connect a keyword with an image. In the example of the man walking on a beach with a dog, it learns to relate single elements (man, dog, beach) of the image with their respective descriptions. While CLIP is capable of relating a description to an image, it is unable to generate images from a given description.

The next step in the process is to generate image embeddings using another model called a prior. This is a diffusion model, which is a neural network, meaning a network of computer neurons, defined as a mathematical function with a number of inputs and one output (Bergmann & Stryker 2024: n.p.). To process more complex tasks, those single neurons are arranged in layers. The input data for image generation is a grey-scale value of pixels generated in the embedding. If we have two input resources, they are mapped as one pixel onto a two-dimensional space, and as the number of inputs increases, the number of dimensions increases. Given enough input information, a meaningful image can be mapped. IBM offers the following summary,

Diffusion models are generative models used primarily for image generation and other computer vision tasks. Diffusion-based neural networks are trained through deep learning to progressively “diffuse”

samples with random noise, then reverse that diffusion process to generate high-quality images (Bergmann & Stryker 2024: n.p.).

For image generation models such as DALL-E, diffusion models are trained in a process in which images are first deconstructed and then reconstructed. In the first step, clean data from a training set (e.g., an image sample) is transformed gradually into pure noise (imagine an old TV screen with faulty reception). This ensures that the essential qualities of the image are slowly fading, giving the model the chance to learn the patterns and structure of the original. The next step is to reverse the noise steps and generate a clean image. Since the noise was added gradually, the structure of the noise was derived from the original image: “Therefore, by learning to accurately predict the noise through reverse diffusion, the model not only learns to denoise the image, but also implicitly learns the structure of [the image]” (Bergmann & Stryker 2024: n.p.). It is then able to generate a new image from a random noisy image. Scientific American summarizes the process as follows, “Once trained, the AI can read any given text prompt, start with an image of pure noise, and reduce the noise until it has a new image that matches the written description” (Bushwick 2025: 25). The result of this step is the image embedding.

The final step uses yet another model called GLIDE, which acts as the decoder. GLIDE is another diffusion model, but with a modification that allows it to include textual information: “GLIDE builds on the generative success of Diffusion Models by augmenting the training process with additional textual embeddings. This results in text-conditional image generation. It’s this modified GLIDE model that enables DALL-E 2 to edit images using text prompts” (Singh 2022: n.p.). The image generated with the GLIDE model undergoes an up-sampling process to create a high-resolution image from the preliminary small image, and with that final step, the image generation is completed. The latest development, as mentioned in the introduction, is the integration of an image-generating diffusion model like DALL-E3 into a Large Language Model (LLM) such as ChatGPT to optimize prompts for image generation (c.f. OpenAI).

Now that a general idea of the process of AI image generation has been established I will discuss this in the light of Susanne Langer’s semiotic theory. It has been shown that AI image generation is based on text and numbers, both, in Langer’s terms discursive symbols. The training of the model consists of the AI learning to relate written labels with a pattern, or a distribution of pixels represented by points in a multidimensional space. Those patterns can then be retrieved and assembled into new structures to generate new images that match the keywords in the user prompt.

### AI Images in the Light of a New Key

The previous paragraph has demonstrated how an AI model operates to generate images. Although the model is trained on images, which have been characterized as presentational symbols, the model does not *see* them as images but as numbers and mathematical functions. The relationship between the image and its object—what it is showing—is established by descriptive labels. In terms of Langer’s theory, we see the image treated as a discursive symbol. This establishes a form of dictionary to be able to define one symbol in terms of another. Moreover, the complex relational form, typical of presentational symbols, is broken down into its smallest parts (pixel), thereby divesting the image of its meaning. Contrary to visual perception, where the image is seen as a single form that conveys a multitude of concepts. To achieve the goal of AI image generation, images have to be treated as if they were discursive symbols to conform to the binary and linear architecture of computers and digital media. This contradicts the relational form of the presentational symbol: The meaning of the image is inferred in visual perception in a non-linear way. Moreover, as Langer explains, the significance of presentational symbols is derived solely from the individual image and the medium in which it is materialized. Every color, line, or shape takes on a different meaning upon its placement within the broader visual narrative. The same is true for music, where the context of a note’s performance significantly affects its interpretation. It can be concluded that the AI model does not process presentational symbols; however, it does generate an image as a result of its processing. This raises more questions than I can discuss here, but I hope the article will serve as an inspiration for the reader. The question previously posed was how images could be created from a written text input and whether or not it is possible to translate between a linear and discrete discursive symbol and a relational presentational symbol. Studying the structure of current AI image-generating models, it can be shown that by translating the image into a mathematical function, it becomes possible to visualize a short text (prompt). However, the question remains whether this constitutes a translation of a discursive symbol into a presentational symbol. This question also relates to the earlier one asking whether the resulting image could be art.

To address this question, I suggest incorporating Langer’s aesthetic theory into the discussion. In her book *Feeling and Form* (1953), she states her main proposition that art gives form to our feelings, on which she further elaborated in a later article titled *The Cultural Importance of the Arts* (1966). There she states,

Art objectifies the sentience and desire, self-consciousness and world-consciousness, emotions, and moods that are generally regarded as irrational because words cannot give us clear ideas of them. (...) I believe the life of feeling is not irrational; its logical forms are merely very different from the structures of discourse. But they are so much like the dynamic forms of art that art is their natural symbol. Through plastic works, music, fiction, dance, or dramatic forms we can conceive what vitality and emotion feel like (Langer 1966: 10).

Art, she explains, objectifies feeling and makes it intelligible—a task discursive symbols such as language cannot render. Her argument here is a logical one, she states that the logical form of language does not reflect that of feeling. The incompatibility Langer sees is based on the complexity of feeling because, although we use words to denote a certain feeling, such as joy for example, the single word is incapable of communicating the nuances of a feeling we experience.

She states,

But human feeling is a fabric, not a vague mass. It has an intricate dynamic pattern, possible combinations and new emergent phenomena. It is a pattern of organically interdependent and interdetermined tensions and resolutions; a pattern of almost infinitely complex activation and cadence (Langer 1966: 9).

The spatial character of human feeling she describes, resonates with her notion of the virtual space, the real but non-existing space that artworks as presentational symbols create (c.f. Langer 1953). Contrary to the idea of attributing the notion of virtual space to the realm of digital media, she describes it as the “primary illusion“ (72) that is created by art. It is created, for example, by deploying colors on a canvas (cf. 1953: 71), but it is not the canvas or the colors themselves. The pictorial space is like the space behind the mirror, only visual, without continuity with the space in which we live. It is, as Langer puts it,

[...] limited by the frame, or by surrounding blanks, or incongruous other things that cut it off. Yet its limits cannot even be said to divide it from practical space; for a boundary that divides things always connects them as well, and between the picture space and any other space there is no connection. The created virtual space is entirely self-contained and independent (72).

Langer's concept of a virtual space can be illustrated with numerous examples from art history and contemporary art. One may imagine the landscape paintings from Georgia O' Keeffe or the late French-Lebanese artist Etel Adnan. Known as a writer and poet, she developed an intimate connection between language and visual art in her oeuvre. A recurrent motif in her oil paintings is Mount Tamalpais, located in California. In countless variations of color combinations, she painted the peak as seen from her Californian home in a style "continually dancing between figuration and abstraction" (Etel Adnan: Lights... 2021: n.p.). What does one see when looking at one of these paintings? According to the American founder of Semeiotic, Charles S. Peirce, we may see colors, shapes, and relations. But certainly not a landscape, a mountain, the sun, or the sky. This is what we *think* we see. Virtual space emerges from interpretation. We do not *see* the Californian Landscape, but we may be able to imagine it from what we see. To do so, humans do not need an image to be an exact copy of the depicted object or have a caption to label it. What enables humans to interpret an image is their experience; their ability to see, remember, and imagine, thus connecting memory, perception, and imagination. The virtual space is not the artwork but the intangible space of relations created by the artist and the beholder. The reality of the actual space and its objects, or physical movements as in a dance performance, are the *media* to access virtual space.

The virtual space created by artworks, as described by Langer, is not accessible for digital media, since the technology cannot see, remember or imagine. In short, digital media have no experience. A computer algorithm works in a discrete and linear way, and no AI model has its own memory. We may call a computer's storage capacity its memory, but stored data is not the same as past experience. The lack of experience results in a lack of contextual knowledge, which is crucial for understanding meaning, especially in presentational symbols that have no lexical translation. As a result, there is also no capacity for imagination:<sup>5</sup> AI models such as DALL-E2 and Midjourney generate new images based on probability instead of imagination.

A further point in this discussion is that since the model can only process a natural language prompt, the input the model gets from the user is already limited to the discursive form.<sup>6</sup> When overlooking the whole process, it becomes evident that there is no presentational knowledge or

<sup>5</sup> As recent research shows, memory and imagination are interrelated in the human brain. See for example Wickelgren 2023.

<sup>6</sup> This paper does not analyze image-to-image models, however they also do not process images as presentational symbols.

meaning involved. Thus, according to Langer's theory, AI neither meets the criteria for being an artist nor for producing artworks for it is not operating with presentational symbols, creating and interpreting virtual space as defined by Langer. Nevertheless, it is part of contemporary art production and theory. To define its role, not only in artistic but also in cultural production, the characteristics of this new symbolic form need further exploration. Similar to the 1960s when artists explored computers (but also TV and video) as new artistic media, AI-image generation is being tested today. Art created using AI models is crucial to understanding the characteristic potential and shortcomings of the technology and, as a result, its societal consequences. As Epstein and Hertzmann in their article *Art and the science of generative AI* summarize,

“Every artistic medium mirrors and comments on the issues of its time, and the debates surrounding contemporary AI-generated art reflect present issues surrounding automation, corporate control, and the attention economy. Ultimately, we express our humanity through art, so understanding and shaping the impact of AI on creative expression is at the center of broader questions about its impact on society” (Epstein & Hertzmann 2023: 1111).

### **AI Image Generation in Contemporary Art**

Boris Eldagsen's entry to the Sony World Photography Award, “Pseudomnesia: The Electrician,” was one of the images that sparked intense debate about AI in art in 2023. Taking a closer look at his artistic practice, it becomes obvious that the image generation is a multi-step process that involves text and image prompts, as well as “inpainting” and “outpainting” techniques (Eldagsen 2023: n.p.). The artist coined the term “promptography” for his artistic practice, an artificial term he defines as welding photography and prompt together to denote the use of language prompts and AI image-generating programs such as *DALL-E* and *Midjourney*. His practice, however, has—strictly speaking—nothing to do with photography. The suffix *-graphy* does not denote what is specific about photography, only the prefix *photo*. Upon closer inspection, we see what the process is about: drawing with prompts, meaning instructions given to an AI. Thus, the process of AI-image generation is fundamentally different from photography; the commonality lies in the appearance. AI images imitate or simulate a photographic image. The transition from photography to promptography is currently explored in exhibitions contrasting both imaging technologies.<sup>7</sup>

<sup>7</sup> See for example: RIVALS Photography vs. Promptography at Photo Edition Berlin c/o Galerie Guelman und Unbekannt.

The goal, Eldagsen proclaimed when refusing the Sony award—to spark a public debate of AI-image generation—was accomplished. And with that, a societal responsibility of artists, namely, to critically reflect on the media we use in our cultural communication, was fulfilled. However, it remains only a first step, and the term “promptography” shows that a detachment from photography and, as a consequence, a definition of this new form of image has not happened yet.

Another example is Refik Anadol’s show at the Museum of Modern Art (MoMa), titled *Unsupervised* (2022–23). He trained an AI model using the publicly available data of MoMA’s collection. The aim was not to imitate a man-made image, but to create something new by allowing the AI model to find new relationships. The process is not fully automated; in fact, the artist describes it as a “collaboration between machine and human” (Anadol et al. 2021: n.p.). It appears like a mutual blindness; the AI model does not see the artwork, whereas humans do not see the data points. This results in a new perspective on the collection, art history, and the technology itself. Using an AI model to search, sort, and classify the data, the artist created a “complex spatial map of the archive in 1024 dimensions” (Anadol 2022: n.p.). The questions, however, remain: In navigating this map and ultimately creating new forms does the AI enter the virtual space Langer describes? Or does it generate a virtual space from which it is excluded because it can only process a computational space? Or, is it obsolete to try to discriminate those spaces? In their exploration and experimentation with technology, artists are not relinquishing control of their artistry to algorithms and corporations that develop them. Instead, they are asserting their authority and agency in this domain. Many of today’s questions have been discussed by artists before, including human agency in early computer art, abstract painting, and ready-made. Every inquiry into image technology invariably entails an inquiry into the nature of the image itself and its function as a sign. The 20th-century artists’ proclamation to disengage from the image has, in fact, signified a re-entry to the discussion of images and their characteristics (cf. Bisanz 2010: 65).

With Susanne Langer, we could also speak of an exploration of the image as a virtual space that, by definition, cannot come to an end. The image as a virtual space is entirely independent of actual space because it is intangible and most of all relational. The dynamic nature of the virtual space as Langer describes it, may again best be illustrated by artworks. The delicate sculptures made by German artist Günter Haese of brass wire, coil springs, and other materials used in watchmaking are constantly reacting to their environment; every movement in their surroundings sets them in motion.

Therefore, they are a beautiful example of the vitality of virtual space. A more widely known example are the mobiles by Alexander Calder. Both exemplify another mode of virtual space, one that is not exclusively visual but also tactile. They make, as Langer puts it, “[...] tactual space visible“ (1953: 90). The translation of tactile experience into a visual experience is another very important topic to discuss in the context of today’s disembodied media, but that is to be left to another time.

### **Conclusion**

The objective of this article was to explore the applicability of Susanne Langer’s semiotic and aesthetic theory to the study of AI generated images. Her discrimination between discursive and presentational symbols facilitates a more precise terminology and consequently a lucid analysis. The discrete and linear nature of discursive symbols is in contrast to the relational character of presentational symbols. By identifying images as presentational and language as discursive symbols, it becomes apparent that AI models are incapable of processing images. Current systems translate images into numerical forms mapped in a vector space and generate images by calculating the most probable conformity with keywords, which have also been translated into numerical form and mapped onto that vector space. Given the framework of Langer’s aesthetic theory and her concept of virtual space it was concluded that there is no evidence to support the idea of AI models producing art. At this point, the theory agrees with observations FRIEZE magazine made at museums worldwide, stating that exhibitions are “[...] moving the conversation away from the dead-end question of whether AI can make art to the question of what meaning artists can make with AI“ (Droitcour 2024: n.p.). From a semiotic perspective, the question of how new technologies are altering the image as a medium remains unanswered. Since it was concluded here that AI-image-generating models cannot process presentational symbols—hence, images—the question of whether a generated image is, in fact, an image remains unresolved. From a strictly technical perspective, it could be considered a visualization of a text, which would make a meaningful difference. Images, and to a greater extent, artworks, play a pivotal role in our culture, having a profound influence on our collective visual environment. In light of this, it becomes imperative to foster a renewed emphasis on interdisciplinary research, integrating the insights of both the arts and the sciences to determine once again what images are, while recognizing the perpetual evolution of our symbols in the age of emerging technologies. It is our responsibility to ensure that symbols

grow, as Charles S. Peirce once stated (Peirce 1878), and that they are not diminished by technology in their capacity to communicate meaning.

## References

Adnan, E. 2021. Etel Adnan: Light's New Measure. *Guggenheim*. Available at: <https://www.guggenheim.org/exhibition/etel-adnan-lights-new-measure> (accessed 4 June 2025).

Anadol, R., C. Reas, M. Kuo and P. Antonelli. 2021. "Modern Dream: How Refik Anadol Is Using Machine Learning and NFTs to Interpret MoMA's Collection". *MoMa Magazine*, 15 November 2021. Available at: <https://www.moma.org/magazine/articles/658> (accessed 4 June 2025).

Anadol, R. 2022. Refik Anadol on AI, Algorithms, and the Machine as Witness. *MoMa Magazine*, 20 December 2022. Available at: <https://www.moma.org/magazine/articles/821> (accessed 4 June 2025).

Barnard, J. 2023. What is Embedding? *IBM*. Available at: <https://www.ibm.com/think/topics/embedding> (accessed 4 June 2025).

Bergmann, D., C. Stryker. 2024. What are diffusion models? *IBM*. Available at: <https://www.ibm.com/think/topics/diffusion-models> (accessed 4 June 2025).

Bisanz, E. 2010. *Die Überwindung des Ikonischen. Kulturwissenschaftliche Perspektiven der Bildwissenschaft*. Bielefeld: transcript Verlag.

Bushwick, S. M. Twombly, A. Hobbs. 2025. "How AI Generates Images From Text". *Scientific American Magazine*, Special Edition, Vol. 34, No. 1, Winter/Spring 2025, 24–25.

Droitcour, B. 2024. The Year in Review: Museums Are Leaving AI Hype Behind. *FRIEZE*, 10 December 2024. Available at: [https://www.frieze.com/article/year-review-ai-art-2024?utm\\_source=FRIEZE&utm\\_campaign=ed5c](https://www.frieze.com/article/year-review-ai-art-2024?utm_source=FRIEZE&utm_campaign=ed5c) (accessed 4 June 2025).

Eldagsen, B. 2023. "PSEUDOMNESIA III. Fake Memories of Art History". Available at: <https://www.eldagsen.com/pseudomnesia3/> (accessed 4 June 2025).

Epstein, Z., A. Hertzmann. 2023. "Art and the Science of Generative AI". *Science*, Vol. 380, Issue 6650, 1110–1111, Available at: <https://doi.org/10.1126/science.adh4451> (accessed 4 June 2025).

Foster, H., R. Krauss, Y.-A. Bois, B. Buchloh, D. Joselit. 2004. *Art Since 1900: Modernism, Antimodernism, Postmodernism*. London: Thames & Hudson.

Ketner, K. L. 1984. "The Early History of Computer Design: Charles Sanders Peirce and Marquand's Logical Machines." With the assistance of A. Stewart, *The Princeton Library Chronicle*, Vol. XLV, No. 3, 186–224.

Langer, S. K. 1942 [1951]. *Philosophy In A New Key. A Study In The Symbolism Of Reason, Rite, And Art*. 2nd edition. New York: Mentor Books.

Langer, S. K. 1953. *Feeling and Form. A Theory of Art Developed From Philosophy in a New Key*. New York: Charles Scribner's Sons.

Langer, S. K. 1966. "The Cultural Importance of the Arts". *Journal of Aesthetic Education*, Vol. 1, No. 1, 5–12. Available at: <https://doi.org/10.2307/3331349> (accessed 3 June 2025).

Manovich, L., E. Arielli. 2024. *Artificial Aesthetics*. Available at: <https://manovich.net/index.php/projects/artificial-aesthetics> (accessed 4 June 2025).

Mucci, T. 2024. The history of AI. *IBM*. Available at: <https://www.ibm.com/think/topics/history-of-artificial-intelligence> (accessed 4 June 2025).

Nake, F., S. Grabowski. 2005. Zwei Weisen, das Computerbild zu betrachten. In Warnke, M., W. Coy, G. C. Tholen (eds.). 2005. *HyperKult. Zur Ortsbestimmung analoger und digitaler Medien*. Bielefeld: transcript Verlag.

O'Connor, R. 2023. How DALL-E 2 Actually Works. *AssemblyAI*, 29 September 2023. Available at: <https://www.assemblyai.com/blog/how-dall-e-2-actually-works> (accessed 4 June 2025).

OpenAI. Introducing 4o Image Generation (2025). Available at: <https://openai.com/index/introducing-4o-image-generation/> (accessed 3 June 2025).

OpenAI. (n.d.). DALL-E 3. Available at: <https://openai.com/index/dall-e-3/> (accessed 4 June 2025).

Parshall, A. 2023. How This AI Image Won a Major Photography Competition. *Scientific American*, 21 April 2023. Available at: <https://www.scientificamerican.com/article/how-my-ai-image-won-a-major-photography-competition/> (accessed 3 June 2025).

Peirce, C. S. 1878. "Illustrations Of The Logic Of Science. Second Paper. How To Make Our Ideas Clear". *Popular Science Monthly*, Vol. XII, 286–302.

Radford, A., J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger and I. Sutskever. 2021. "Learning Transferable Visual Models From Natural Language Supervision". *ArXiv*, arXiv:2103.00020v1. Available at: <https://arxiv.org/pdf/2103.00020> (accessed 4 June 2025).

Ramesh, A., P. Dhariwal, A. Nichol, C. Chu, M. Chen. 2022. "Hierarchical Text-Conditional Image Generation with CLIP Latents". *ArXiv*, arX-

iv:2204.06125v1. Available at: <https://cdn.openai.com/papers/dall-e-2.pdf> (accessed 4 June 2025).

Roose, K. 2022. An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy. *The New York Times*. Available at: <https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html> (accessed 3 June 2025). A version of this article appears in print on Sept. 3, 2022, Section B, Page 1 of the New York edition with the headline: "Creating Art By Typing In Some Words".

Siemens 2025 Tracing the AI family tree: say hello to ChatGPT's grandmother. Available at: <https://www.siemens.com/global/en/company/insights/tracing-the-ai-family-tree-say-hello-to-chatgpts-grandmother.html> (accessed 4 June 2025).

Singh, A. 2022. How Does DALL·E 2 Work?. *Medium*, 27 April 2022. Available at: <https://medium.com/augmented-startups/how-does-dall-e-2-work-e6d492a2667f> (accessed 4 June 2025).

Taylor, G. D. 2014. *When the Machine Made Art : The Troubled History of Computer Art. International Texts in Critical Media Aesthetics*. Vol. 7. New York, London: Bloomsbury.

This Is World. 2025. Father of AI: AI Needs PHYSICS to EVOLVE. prof. Yann LeCun. *YouTube*, 7 March 2025. Available at: [https://www.youtube.com/watch?v=RUnFgu8kH-4&list=PL80I41oVxglJ0okgsaIKw4Z-JORQC0Qc\\_5&index=4](https://www.youtube.com/watch?v=RUnFgu8kH-4&list=PL80I41oVxglJ0okgsaIKw4Z-JORQC0Qc_5&index=4) (accessed 3 June 2025).

Turing, A. M. 1950. "Computing Machinery and Intelligence". *Mind* 49, 433–460. Available at: <https://courses.cs.umbc.edu/471/papers/turing.pdf> (accessed 4 June 2025).

Wickelgren, I. 2023. Where Imagination Lives in Your Brain. *Scientific American*, 8 June 2023. Available at: [https://www.scientificamerican.com/article/where-imagination-lives-in-your-brain/?utm\\_source=Klaviyo&utm\\_medium=campaign&utm\\_campaign=TIS\\_051925&utm\\_term=the%20purpose%20of%20memories%20is%20actually%20for%20future%20planning&\\_kx=G7ZVzZ4uR4H437KxN8yj8ye4BK7Gq-vSaRqWTVZPpFmZSK7F5u8\\_VzERn6kKRYBxo.WEer5A](https://www.scientificamerican.com/article/where-imagination-lives-in-your-brain/?utm_source=Klaviyo&utm_medium=campaign&utm_campaign=TIS_051925&utm_term=the%20purpose%20of%20memories%20is%20actually%20for%20future%20planning&_kx=G7ZVzZ4uR4H437KxN8yj8ye4BK7Gq-vSaRqWTVZPpFmZSK7F5u8_VzERn6kKRYBxo.WEer5A) (accessed 4 June 2025).

Zewe, A. 2025. AI tool generates high-quality images faster than state-of-the-art approaches. *MIT News*, 21 March 2025. Available at: <https://news.mit.edu/2025/ai-tool-generates-high-quality-images-faster-0321> (accessed 3 June 2025).

## THE JANUSIAN FACE OF FACIAL RECOGNITION, PART 1: ITS SUBFACE, INTERFACE, AND SURFACE<sup>1</sup>

*Devon Schiller*  
*University of Vienna*  
*devonschiller@gmail.com*

### **Abstract**

Facial recognition has a *Janusian* face. In the instant of its interaction, facial recognition brings many faces together into relation. It is not only visually representational but also computationally *re*-presentational. To make facial recognition *knowable*, therefore, one needs first to make its many faces *visible*. One must expose, so to say, the functional relationalities between rhizomatic facialities within these commercial products and their proprietary computing. Only then can the ways facial recognition technology systems work, and how they are either adversely misused or beneficially used, be substantively challenged or tactically critiqued. Toward this end, from the critical standpoint of a computational semiotics both Peircean and pragmatist, I apply a method of *semiotic deblackboxing*.

---

<sup>1</sup> This research was supported by a DOC Fellowship from the Austrian Academy of Sciences.

I divide my inquiry between two distinct investigations: here in part one, I explicate this theoretical approach; then in part two, I explore its practical application. Across these two parts, I probe how, and in what ways, the artificial intelligence of facial recognition not only comprises a computational, mechanical, and technological system, but also constitutes a semiotic system. The question is this: Does the artificial intelligence system of a facial recognition technology have *semiotic agency*? That is, is it therefore able to process the action of a sign that has *genuine* triadic sign relations rather than *non-genuine* dyadic quasi-sign relations? In other words, is artificial intelligence actually either *artificial* or *intelligent*? As I argue here, this *Janusian or multiple relation* is the intelligent result, not of the semiotic machine, its efficient causation, and allopoiesis, but of the semiotic animal, our final causation, and autopoiesis.

**Keywords:** biometric identification, computational semiotics, facial recognition technology (FRT), probabilistic artificial intelligence (ProbAI), semiotic deblackboxing.

“Yet what god shall I say you are, double-formed Janus? [...] Reveal, at the same time, the reason why you alone of the heavenly ones see both what is behind and what is in front. [...] Then holy Janus, amazing in his double-appearance, suddenly presented his pair of faces to my eyes. I was terrified...” – Ovid ([8 CE] 2011: I.89–96)

## 1. Introduction

Facial recognition has a *Janusian* face. At least analogically speaking, the probabilistic artificial intelligence (ProbAI) of facial recognition technology (FRT) has not one but two faces. One face looks backward into the past and its *input*. And one gazes forward into the future and its *output*. Both so-called “faces” principally serve as a spatiotemporal passage to yet another face. The Janusian face of facial recognition is, therefore, not only visually representational but also computationally *re*-presentational. In an instant of its interaction, a facial recognition technology system brings far more than one or two faces together into relationalities. There are the human faces of its designers, users, and targets. And there are the machine faces of its subface, interface, and surface. There are the images of faces used to train its artificial intelligence, computer vision, and machine learning. And there are the faces of individuals that it detects, extracts, and classifies. There are the frontal views of dynamic faces detected from the face,

its muscles, and movements. And there are the thin slices of static faces extracted from the continuous flow of this behavioral stream. There are the faces of the individuals who are being recognized. And there are the faces of the community who are doing the recognizing. For this reason, to make facial recognition *knowable*, one needs first to make its many faces *visible*. One must expose, so to say, the functional relationalities between rhizomatic facialities within these commercial products and their proprietary computing. Only then can the ways facial recognition technology systems work, and how they are either adversely misused or beneficially used, be subversively challenged or tactically critiqued.

From the critical standpoint of a computational semiotics, I apply a method of *deblackboxing*. Here I explicate how, and to what extent, the artificial intelligence of facial recognition is made up from not only a computational, mechanical, or technological system but also a semiotic system. In a black box, as manifestly illustrated by Martin Irvine, “semiotic principles” have been implemented within the design yet remain “blackboxed,” or “hidden from view,” and “closed off” from the understanding of its users (2022: 205–206). Of course, across computational semiotics alone, several traditions of semiotic thought potentially afford for critical perspectives into facial recognition and how this black box then works (see, for review, Meunier 2022). These include, for instance, more linguistic or logical methods as well as more dyadic or triadic models. Toward a critical deblackboxing, however, I advance a Peircean and pragmatist approach (see, for example, Andersen, Hasle, and Brandt 1997; Nöth 2002; Nake 2008; Sørensen, Thellefsen, and Thellefsen 2020; Irvine 2022), in which I apply the triadic sign model of Charles Sanders Peirce (1839–1914), one of, if not *the* most, foundational thinkers behind pragmatism and semiotics. The Peircean model of triadic signs is constituted by a “*representamen*,” “*object*,” and “*interpretant*” (CP 2.228). Consequently, in Peirce’s classification, “the science of semiotic[s] [also] has three branches” (CP 2.229). This “trivium,” which Peirce refers to variously across his writings after he introduced it with “On a New List of Categories” in 1867, is constituted by “logic,” “grammar,” and “rhetoric” (CP 1.559). Based upon these branches, like Irvine’s “deblackboxing” method” (2022: 205), Bent Sørensen, Torkild Thellefsen, and Martin Thellefsen propose a program for studying the semiotics of what they call “technological artefacts” by way of their “technological grammar, technological logic, and technological rhetoric” (2020: 254). In these papers, I focus on the grammar. Peirce characterizes “Speculative Grammar” as “the general theory of the nature and meanings of signs, whether they be icons, indices, or symbols” (CP 1.191). This grammar concerns the

ground conditions whereby the representamen of a sign both carries information and conveys meaning, such as by its efficient and final causation, immediate and dynamic object, as well as quasi and genuine semiosis, and likewise by its signal and sign interrelation, spacetime instantiation, as well as tokenization and retokenization. From the critical standpoint of a computational semiotics both Peircean and pragmatist, I do not seek to reveal the computational software or mechanical hardware obfuscated inside the opaque box. Rather, I deblackbox relationalities between the many signs from which the box takes its meaning.

The question is this: Does the artificial intelligence system of a facial recognition technology have *semiotic agency*? That is, can facial recognition create *in and of itself* “[t]he action of a sign,” which, in Peirce’s own terms, is not only “dyadic” but also “triadic” and not only “dynamical” but also “intelligent” (CP 5.472)? Can facial recognition create “a *genuine* triadic relation” rather than a non-genuine “collocation of dyadic relations,” or, as Peirce himself phrases it, “a ‘degenerate’ form of triadic relationship” (CP 6.322)? Inside the black box, is artificial intelligence actually either *artificial* or *intelligent*? With facial recognition technology during human-computer interaction, the human act of sign interpretation is brought into relation with the computer operation of signal determination. But the human interactor and the facial recognition are not equals in this relation. As I argue here, *this Janusian or multiple relation* is the intelligent result, not of the semiotic machine, its efficient causation, and allopoiesis, but of the semiotic animal, our final causation, and autopoiesis.

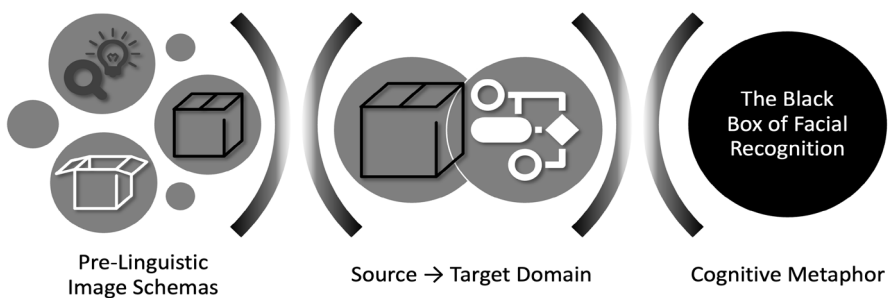
## **2. The Ethics of Visibility: From Closed/Opaque, to Open/Transparent, and Everything In-between**

Facial recognition is a black box in most, but not all, cases at this moment in our history, when the Fourth Industrial Revolution and the Information Age have accelerated forward such technological advances. Developed since the 1960s, facial recognition became commercially available for applied configuration by non-expert users in large-scale systems in the mid to late 2000s. As historicized by Kelly Gates, a theorist in science and technology studies, “military needs,” as for military perception technologies from a military-industrial complex, have “influenced the development” of facial recognition technology ever “since its inception as a research program” in the mid twentieth century (2011: 101). Amidst the intelligence community practices of the post-9/11 cultural imaginary, the industrialist advocates for artificial intelligence began their work behind the scenes of government to accelerate facial recognition technology from theoretical

in-the-lab experiments to practical in-the-wild applications. Researched and developed for the government and its security, Gates argues, facial recognition “more closely resembl[es] [...] a ‘black box,’” that is, a functional technology positioned in society so as to be “virtually indispensable” for the people and their safety (2011: 101). The sales pitch went something like this: where humanity failed, technology would succeed; by identifying a face, threats would be individualized. At least, that was the claim. Though alleging “technical neutrality,” Gates accounts, “the technology promised to [...] protec[t] civilization” from a “generalized and racialized ‘face of terror,’” all while this “‘facialization’ of terrorism,” and “nonstate forms of political violence,” actually fell back upon “an implicit classifying logic [of] deviant facial types” (2011: 101). Therefore, the principles behind the design, as Gates herself attests, require not only “legal frameworks of privacy rights” but also ethical guidelines for “structural inequality” (2011: 199). And generally, Gates advocates, “the design of these technologies,” and the black boxes within them, “needs to be more widely understood at every level of their development” (2011: 199), as I attempt here.

Given the facial ethos of the twenty-first century, with its Orwellian phantasms, and its panoptical overwatch, omnipresent eye-in-the-sky facial recognition is often even imputed with a mythical identity or a religious quality. For instance, Massimo Leone, a cultural and religious semiotician, was principal investigator on Face Aesthetics in Contemporary E-Technological Societies (FACETS), a Consolidator Grant Project supported by the European Research Council, from 2019 to 2024. With popularized cases and religionized terms, Leone goes so far as to claim that the “black box” is “designed to mystify [...] digital entities,” such as “algorithmic images” with their “augur signs,” by imbuing the “genesis” of the box with an “auratic character” and “halo of mystery” (2024: 426–429). Although Leone himself addresses neither the expert technical literature nor the extensive theoretical literature about such black boxes, he correctly recognizes their “dual functionality” (2024: 427), both in technical practices, and in rhetorical strategies. The meaning behind this metaphor, whether *the black box of artificial intelligence* generally or *the black box of facial recognition* specifically, is perhaps understood best using the cognitive metaphor theory or conceptual metaphor theory, as it has been variously called, developed in cognitive science and linguistic studies beginning in the 1980s (see, for example, Lakoff and Johnson 1980). In terms of an ecological, embedded, embodied, enactive, and extended cognition, or what is known as 5E cognition, the source domain of the *black box* is mapped in the mind onto the target domain of the *facial recognition* (Fig. 1). That is, *the black box*

*cognitive metaphor*, both cognized in our technical imagery and conceptualized in our technological imaginary, is constructed at least in part out of what cognitive linguists George Lakoff and Mark Johnson term “a container schema” and “source-path-goal schema” (1999: 32). A container schema is, like other image schemas, not only “conceptual” but also “instantia[-ble],” in Lakoff and Johnson’s view, with “an inside, a boundary, and an outside,” where this “physical boundary [...] impose[s] forceful and visual constraints,” “protect[ing] the container’s contents” while “render[ing] them inaccessible to vision” (1999: 32). The source-path-goal schema, like the container schema, Lakoff and Johnson suggest, is useful for conceptualizing “spatial relations,” both “topological” and “trajector[ial],” in which something moves from a first, imagined point to a final, intended position “at a given time” (1999: 33). From the perspective of this metaphorization, facial recognition, like “faculty psychology” in Lakoff and Johnson’s survey, can be compared with a “flow chart,” where a box or boxes “represent[t] various separate faculties carrying out their functions,” including “algorithms” or “metaphorical machines” and their operations upon some information (1999: 413). This cognitive metaphor about facial recognition and its non-visible processuality, like other such metaphors analyzed by Lakoff and Johnson, is motivated in the mind through pre-linguistic image schemas, where “seeing what’s in the box correlates with knowing what’s in the box” (1999: 48), grounded by interactions between the human organism and our semiotic artifacts.



**Figure 1:** A process diagram for *the black box* cognitive metaphor, from the source domain of the black box to the target domain of the facial recognition, based upon pre-linguistic image schemas, including the container and source-path-goal schemas.

Almost immediately upon the public release of this private research about facial recognition in the mid to late 2000s, scholars from across disciplines began to call for the *unveiling*, *unpacking*, or *uncovering* – that is, the *unboxing* – of proprietary black boxes in facial recognition technologies. More “closed” and “*opaque*” than “open” and “*transparent*” facial recognition in the “social-technical infrastructure[s]” of our everyday lives, Lucas Introna, a scholar in the social study of information systems, points out, make it “particularly difficult” to scrutinize, which, reciprocally, “creates unprecedented opportunities [for] “invisible” micro-politics” (2005: 77). Consequently, Introna calls for a “*disclosive ethics*” that would *make visible*: first, “a place for ethics” to be “immediately present” in the “actual operation[s] of power” in technology; second, the “values and interests,” whether “intentional or unintentional,” in facial recognition from its “technical detail[s] to social practices” (2005: 75, 79). However, in the case of artificial intelligence, “there is no singular black box to open, no secret to expose,” as appraised by Kate Crawford, the co-founder and former director of the AI Now Institute in New York (2021: 12). Rather, Crawford finds, artificial intelligence is constituted by “a multitude of interlaced systems of [technocratic] power” (2021: 12). “Complete transparency,” Crawford traces, is an “impossible goal” (2021: 12). To better understand the way artificial intelligence works and its “role in the world,” Crawford resolves, one should rather focus on the “material architectures” and “contextual environments” behind artificial intelligence systems “by tracing how they are connected” (2021: 12), that is, their *interrelations*. Joy Buolamwini, who founded the Algorithmic Justice League while at the MIT Media Lab, calls attention to the fact that computer scientists themselves do not always know exactly how “some weights are strengthened [,] and others are weakened,” when training an intelligence (2023: 53). Consequently, Buolamwini herself concludes, the current methods in computer science “do not allow us to explain in detail” the ways in which an artificial intelligence technology identifies “a pattern like a face” or “outputs a response” (2023: 53). Although “the term ‘black box’ [is] used to describe” such artificial intelligence systems, Buolamwini acknowledges, it is important and necessary to “closely examine the [...] systems being developed” (2023: 53). But significantly, Buolamwini suggests, the “training data” is most critical to understand the risks involved in applying an artificial intelligence like facial recognition, and we must open the box and shed light upon where “the data comes from, who collected it, and how it is organized” (2023: 53). Most recently, for example, the European Parliament adopted the Artificial Intelligence Act on March 13, 2024. For this landmark piece of a supranational legislation, the mem-

ber states of the European Union prepared amendments to proposed regulation on the “machine learning capacities” of artificial intelligence systems, among other operational aspects, where the Commission in Amendment 19 refers to outputs “difficult for humans to understand, monitor and trace back” to inputs, with “complex and opaque characteristics,” that is, a “black box element,” which “impact[s] accountability and explainability” (2024: 11–12). Perhaps semiotics affords possible solutions.

### 3. Critical Deblackboxing: The Black Box, Its Input, and Output

The term *black box* originally entered the scientific literature around the mid twentieth century in the form of a cognitive metaphor, design principle, and epistemic model, that is, a thought experiment. Indeed, the black box is first given its full treatment by the pioneers in cybernetic theory who then preceded the field of artificial intelligence. Although the earliest known usage of the term “black box” specifically derives from slang dictionaries, between the seventeenth century and the nineteenth century, where it refers to “a lawyer.” And the pioneering conceptions of a black box, or at least of black box-like systems, can be found in works on thermodynamic transfer or about electrical engineering, including James Clerk Maxwell’s scientific textbook *Theory of Heat* in 1871, Franz Breisig’s patent application *Method and Arrangement for Determining Crosstalk in Multicircuit Systems* in 1921, and Wilhelm Cauer’s magnum opus *Theory of Linear AC Circuits* in 1941. It is Norbert Wiener who then famously compares what he foundationally calls a *black box* and *white box*. A computer scientist, mathematician, and originator of cybernetics, Wiener studied at Harvard under Josiah Royce (1855–1916) and can himself “be considered a philosophical pragmatist,” as John Durham Peters and Benjamin Peters contend, who views messages not only “as informational but [also] as operational” (2016: 157–158). In *Cybernetics: Or Control and Communication in the Animal and the Machine* in 1948, Wiener extends Maxwell’s “anthropomorphic demon” thought experiment, about some sorting agency that structurally controls a closed system, to the acting on and the changing of “information,” that “must be carried” by one process or another, in “physical, chemical, [or] biological system[s]” ([1948] 1961: 57–58). First, Wiener characterizes a “black box” as “a piece of apparatus” with an input and an output, which performs an operation “on the present and [the] past of the input,” where information is inaccessible about “the structure by which this operation is performed” ([1948] 1961: xi n1). Then, Wiener characterizes a “white box” as a “similar network,” where “the relation between input and

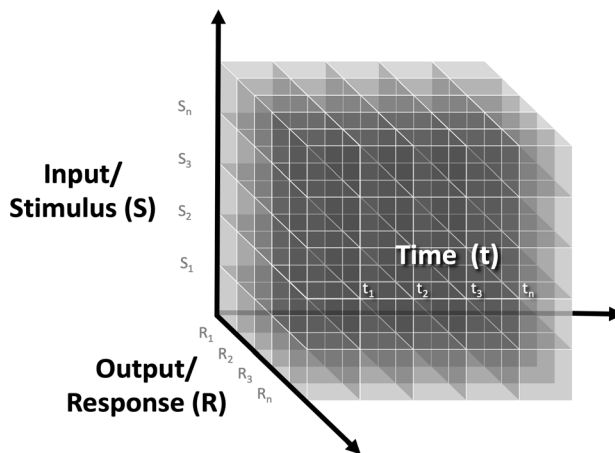
output” is designed in accordance with “a definitive structural plan for securin[g] previously determined input-output relation[s]” ([1948] 1961: xi n1). Although Wiener himself acknowledges that “black box” and “white box” are simply “convenient and figurative” terms, which, at least in his time, were “of not very well determined usage” ([1948] 1961: xi n1). That is, the 1) nonvisible, opaque black box, 2) the visible, transparent white box, and 3) the semi-visible, translucent gray box a shade in between white and black, are epistemic models of information processing. These models thereby represent *the flow of information* over periods of time, for instance, from the *outside* of an environment to the *inside* of a machine and back *inside out* again. Between the *input* stimuli and the *output* response, the information is transformed in some way or other by the box, its structure, and functions. However, what is observable about the operations *internal* the box itself may only either be 1) *nothing*, as is the case with the black box; 2) *everything*, as is the case with the white box; or 3) *something*, as is the case with the gray box; and, therefore, may have to be inferred from the relation of the change, that is to say, from the box’s *external* behavior.

In the manifold sense of its metaphorical significance, a black box can almost be anything and can almost be anywhere. Although a black box should not be confused with quantum physicist Erwin Schrödinger’s *cat-in-a-box*, behavioral psychologist B.F. Skinner’s *operant chamber*, or any such boxes. To know a box, therefore, one must experience it. Black boxes can be found, in the terms of an Uexküllian biosemiotics, not only in our *Umwelten* (outer environments), like in our machines and prostheses, but also in our *Innenwelten* (inner environments), like in our minds and bodies. Ross Ashby, a psychiatrist, and progenitor of cybernetics, mentions how “[i]n our daily lives we are confronted at every turn with systems whose internal mechanisms are not fully open to inspection” (1956: 86). Furthermore, Ashby maintains, “Black Box theory” can lead to philosophy about “the mechanism that, for whatever reason, is not wholly accessible,” which perhaps is paramount “for those who study the brain” ([1958] 1991: 256). One of the many black boxes that each and every day we humans do experience is the very face itself, that is, the relation between facial behavior and mental phenomena. Indeed, as we could well consider it, the artificial black box of facial recognition technology is useful for studying the natural black box of human facial behavior. For a scientific epistemology of the black box, Ashby goes on to explicate, an Experimenter must “act on the Box,” by “allowing the Box to affect” them and their apparatuses, thereby “coupling” themselves “to the Box, so that the two together form a system with feedback” (1956: 87). Ashby diagrams this interaction: “Box  $\Leftrightarrow$  Experimenter”

(1956: 87). The Box and Experimenter, Ashby expounds, “can be viewed as a compound system,” in which the Experimenter “acts on the Box” when they stimulate it, the Box acts on the Experimenter when it reacts, and “each ac[t] on the other” ([1958] 1991: 303). “The interactions that occur between box and experimenter,” Ashby defines, “are therefore subject to the laws of communication” ([1958] 1991: 303), or, more broadly, the logics of semiotics. For this reason, Ashby then recognizes, the study of a box, and whether, or to what extent, “the Box is behaving in a machine-like way[,] does not require study of its internal details” ([1958] 1991: 304). But rather, Ashby finds, the “*functional connections within a Black Box can be deduced from observations made from without*,” such that “to find something of the connections [,] does not demand the opening of the Box” ([1958] 1991: 304). The box thus becomes not only a cognitive metaphor (for how something odd *somewhen* occurs between input and output within the operations of a system) but also an epistemic model (for the ways in which humans and technologies interact with one another).

Many cybernetic theorists in the mid twentieth century view the epistemic significance of this systems experience from a pragmatist perspective. In a phenomenological approach to his systems theory, philosopher and physicist Mario Bunge presents the black box as “a fiction,” that he portrays as a “*genera[l]*” model (1963: 346, 357), which is applicable to diverse phenomena, and is aimed at system dynamics. As popularized by Bunge in the early 1960s, “black box theory” represents “a set of concrete systems into which stimuli ( $S$ ) impinge and out of which reactions ( $R$ ) emerge,” where “the various kinds of stimuli and responses” are pictured as “signals,” which travel along the “many channels ( $C$ )” that connect “the box with its environment” (1963: 346, parentheses added). The principal variable that Bunge problematizes is “time” ( $t$ ), and the “time delay” or the “reaction time” ( $T$ ), “of the box for [a] channel” (1963: 347). As Bunge points out, here “the principle of antecedence” holds that “[t]he effect cannot occur before the cause” (1963: 357). That is (Fig. 2), the reaction ( $R^n$ ) at time ( $t^n$ ) is determined by stimuli ( $S^n$ ) at all times ( $t^1, t^2, etc.$ ) prior to and including this time ( $t^n$ ). To Bunge, therefore, “other meaning could be attributed to the variables” for time, like interpretation, processing, representation, or transformation, to an extent that “the theory is almost *abstract*” (1963: 357). Taking after Ashby’s treatment, Bunge concludes the box “constitution” to be “altogether irrelevant” because one accounts only for “the behavior of the system” (1963: 346). Writing black box theory across some thirty years, cybernetician Ranulph Glanville characterizes the black box as a “*gedenkendexperiment* (thought experiment)” (2009: 153), much like

Maxwell's demon. To Glanville, a black box 1) is "believed to be distinct," 2) has "observable (and relatable) inputs and outputs," and 3) is "black (that is, opaque to the observer)" (1982: 1). Therefore, Glanville himself traces, "the blackness of a black box," analogically speaking, consists not in the non-visibility of its operations, but in the visible "change in signals" between input and output that one then can interpret to be "caused by the black box" (1982: 1). To put this pragmatically, what matters here is less the inner character of how a black box works and more the outer behavior of its semiotic actions. Wiener, Ashby, Bunge, and Glanville, among others, more or less share this position, ultimately, taking a perspective that is pragmatist. In terms of Peirce's own *pragmatic maxim*, we must "[c]onsider what effects, that might conceivably have practical bearings, we conceive the object of our conception to have," such that "our conception of these effects is the whole of our conception of the object" (CP 5.402). It is not what a black box is on the inside, but rather *what it does* that defines it.



**Figure 2:** An epistemic model of the black box with input/stimuli (S), output/response (R), and time (t). © Schiller 2025.

But what a black box does and how it does it, idiomatically speaking, might not be visible *on its face* and should not be taken *at face value*. At least in part, the critical deblackboxing in computational semiotics is realized in response to *social constructivist theory* in science and technology studies. And the two approaches generally start from similar grounds. Already in the late 1970s, historian of science and technology Edwin Layton, among

others, admonishes *black-boxist* approaches to the study of a technology that represent *input-output* relations with little to no concern for what happens in between. Layton calls for “an understanding of technology from [the] inside,” which includes, for instance, “as a body of knowledge and as a social system” (1977: 198). And Layton calls out any treatment of technology “as a ‘black box’” the “contents and behavior” of which “may be assumed to be common knowledge” (1977: 198). Taking further this foundation in the mid 1980s, Wiebe Bijker, Thomas Hughes, and Trevor Pinch, in their classic treatise on a *technological constructivism*, advance the approach of “thick description” ([1987] 2012: xliii), as first coined by philosopher Gilbert Ryle ([1971] 2009: 497), and further championed by anthropologist Clifford Geertz (1973: 6). Bijker, Hughes, and Pinch apply this approach toward “looking into” not only “the black box of technology” but also “the black box of society” ([1987] 2012: xliii). However, in the early 1990s, political scientist Langdon Winner takes a critical stance toward any social constructivism in science and technology studies, or what is sometimes called *the social construction of technology*. In Winner’s own words, it is not “enough to provide clearer, well-nuanced explanations of technological development” ([1991] 1993: 375). For Winner, social constructivism fails to “call into question the basic commitments [of] technological society” and lacks “a generation position” on the very “patterns under study,” like in a “Marxis[t],” or in a “Heideggerian[ist],” perhaps in a feminist or queer “theory of technology” ([1991] 1993: 375). Simply put, social constructivism addresses technological creation not technological consequence. By the turn of the century, as Irvine himself “notes” (2022: 225n3), philosopher Bruno Latour problematizes the way in which “scientific and technological work is made invisible by its own success” (1999: 304). That is to say, in Latour’s post-humanist view and sociology of science, “[w]hen a machine runs efficiently, [...] one need focus only on its inputs and outputs and not on its internal complexity” (1999: 304). “[P]aradoxically,” therefore, Latour postulates, “the more science and technology succeed, the more opaque [...] they become” (1999: 304). In a “technical mediation,” such as facial recognition, and its “folding of time and space,” Latour recommends a “reversible blackboxing” (1999: 183–184), that is, a deblackboxing. Latour calls upon inquirers to “[o]pen the black boxes” that are around them, and “examine the assemblies” hidden there deep inside, each of which is, respectively, “itself a black box” (1999: 185). But Latour then asks “[h]ow far *back* in time, [and] *away* in space, should we retrace our steps to follow all those silent entities” (1999: 185)? Which is to say, how far should one trace relationalities from the black boxes through the black boxes to the black

boxes, and so on, ad infinitum, that contribute to interactions between a human and a technology? A problem always persists, as Latour has lamented: “the construction of artifacts” in and of itself cannot “account for facts” because these “[n]onhuman[s]” also have agency (1999: 185). Semiotics, however, can reveal this (inter)relationality between human and machine.

#### 4. The Semiotic Machine: Its Subface, Interface, and Surface

The black box in facial recognition, like any other computer system, and unlike many, if not most, other technologies or tools, is structurally, functionally, and constitutively semiotic. Facial recognition is what Mihai Nadin terms a “*semiotic machine*” (2007: 74),<sup>2</sup> not only retrospectively from a top-down critical perspective, but also relationally from its bottom-up design principles. A semiotic machine is “*Janus-headed*,” Peter Bøgh Andersen, Per Hasle, and Per Aage Brandt claim, since the “bivalent function” of this “technical artifact” links “two *interface[ing]*” systems (1997: 566): that of the human and that of the computer. Of course, the signs that relate to the human and the signs that relate to the computer are not the same thing and should not be transposed. Both still are *signs*, however. They are *something that stands for something else*, as the Hellenistic Stoics first defined sign function in the third century BCE. But the “machine semiosis” of a semiotic machine, Andersen, Hasle, and Brandt maintain, consists of “processes that take place inside machines, between machines, and [outside machines] between them and their human users” (1997: 548). Tools are not machines. They are too simple. Chemicals are not machines. They are too similar. And importantly, Andersen, Hasle, and Brandt also infer, “living beings are not machines” (1997: 548). Neither are we designed by a human person. Nor do we exist for a human purpose. Indeed, the very handiness or utility itself of a semiotic machine clearly emerges from its cognitive “autonomy,” Andersen, Hasle, and Brandt conclude (1997: 566). Therefore, at first sight it would seem, linguist Winfried Nöth laments, a “*semiotic machine*” is “a contradiction in terms” (2002: 84). Even machines, especially computers, however, “*are involved in sign processes*,” as Nöth himself concedes (2002: 84). For this reason, in recent years an increasing number of computational semioticians have taken issue with the term semiotic machine. Irvine, for instance, finds that “computers’ are not usefully defined as ‘machines’” and introduces the term “cognitive-semiotic artefacts” (2022: 208). Similarly,

---

<sup>2</sup>Nadin first introduced the critical term of “semiotic machine” in a number of lectures at the Rochester Institute of Technology in New York as early as 1983.

Sørensen, Thellefsen, and Thellefsen adopt the term “technological artefacts,” not only for “semiotic entities” but also as “intrinsically semiotic,” because they are “*some-thing*,” both “designed” and “used,” in “processes of signification” (2020: 253–254). Simply put, the pragmatic utility of a hammer tool is to drive in a nail. Similarly presented, the pragmatic utility of a facial recognition technology system is to recognize that which is facial. When it is working, that is all that facial recognition is meant to do. That is all that it does. But the context is more complex. The Janusian face of facial recognition does operate through signs. What is more, facial recognition technology, as an embedded system, or a distributed system, involves many other technologies that operate through signs. Of course, not each of these technologies is semiotic in and of itself. Together, however, they all become entangled with sign processes and semiotic productions.

More and more, not only computational semioticians but also computer scientists widely discuss and highly debate whether, and to what extent, probabilistic artificial intelligence systems, which include, for instance, facial recognition technology systems, constitute a deterministic machine, an indeterministic machine, or a semiotic machine. The “constitutive properties” that characterize a machine consist of its “*organization*” and its “*structure*,” with “[t]he *use* to which a machine can be put” part of how to describe, explain, and observe “the machine in a context wider than the machine itself,” according to Humberto Maturana and Francisco Varela (1980: 77–78, emphasis added), two biologists who are intellectually famous because they first introduced the concept of *autopoiesis*, that is, system complexity via self-creation, to biology and cybernetics in the early 1970s. The “*organization* of the machine,” Maturana and Varela contend, is constituted by “[t]he relations that define a machine as a unity” and that “determine the dynamics of interactions and transformations which it may undergo as such a unity” (1980: 77). Whereas, the “*structure*” of the machine, Maturana and Varela contrast, is constituted by “[t]he actual relations which hold among the components which integrate a concrete machine in a given space” (1980: 77). Human-made machines, however, Maturana and Verela hold, “are all made with some purpose, practical or not, but with some aim (even if it is only to amuse)” specified by humans, although “purpose, aim, or function” are not properties of the machine in and of itself (1980: 78). In his foundational text on this artificial intelligence, James Fetzer, a philosopher of science turned Holocaust, 9/11, and Sandy Hook denier and conspiracy theorist, makes a distinction between: first, “*deterministic* causal systems,” for which, “given the same input, the same output invariably occurs (without exception);” and second, “*inde-*

*terministic* (or “probabilistic”) causal systems,” for which, “given the same input, one or another output within the same class of outputs invariably occurs (without exception)” (1990: 37). For a causal system to also be a semiotic system, Fetzer decisively further delineates, “it must be a system for which something can stand for something (else) in some respect or other,” where that something, that is, that sign “can affect the (actual or potential) behavior of that system” (1990: 37). Indeed, as Fetzer identifies, from the critical standpoint of a computational semiotics it becomes “apparent that for something to affect the behavior of a causal system does not mean that it has to be a sign for that system” (1990: 38). Consequently, one of the principal problems for semiotic studies, Fetzer concludes, is to “distinguis[h] semiotic causal systems from other kinds of causal systems” (1990: 38), as I do here with facial recognition technology.

Indeed, many leading thinkers in computer science today define the computer not through mechanic principles but through “implicit,” if not explicit, “semiotic principles,” Irvine highlights (2022: 218), like *interpretation, processing, representation, and transformation*. For instance, computer scientist Peter Denning systematically considers in his transformation model of these information processes: first, how “[a] representation is a pattern of symbols that stands for something” else; second, that “[t]he association between a representation and what it stands for can be recorded,” as in a database memory or in a dataset table; and third, “a science of information” can therefore be grounded in “the observable affects (signs and referents) without a precise definition of ‘meaning’” or signification (2012: 808). This model of information, in Denning’s view, may perhaps be most “useful when the computations appear to be strings or streams” (2012: 808), like the continuous feedback loop of a facial recognition technology. Indeed, Denning himself infers, “definitions of information,” more often than not, have “an objective component (signs and the things represented by signs) and a subjective component (the meanings)” (2012: 808). Given this “notion of a representation,” Denning reasons, “representation-transformation” becomes a “reference model,” where “an information process is a sequence of representations,” and computation is the transformation from one representation “of the sequence to the next,” so that, in the “continuous world,” each “time and space,” however “infinitesimal,” is thus “controlled by a representation” (2012: 808). As Peter Denning and Craig Martell clearly explicate, “computing emphasizes the *transformation* of information,” through systematic structures that “are not just descriptive [but actually] *generative*,” like how “[a]n algorithm is not just a description of a [problem-solving] method” but actually “causes a machine to solve the

problem,” where “information caus[es] action” (2015: 16–17). Today, many characterizations of computational methods, as Irvine also identifies, apply the “view of sign-actions [...] that Peirce first developed” (2022: 218). At the center of such computation is less representation *from* signs and more operationalization *of* signs, which cause actions with controlled operations, or, in Peircean terms, through law and rule-governed interpretants of sign representamen.

Of course, in a linguistic-verbocentric *semiological* approach, like by Saussure or by Hjelmslev, one applies a dyadic or bi-relative sign model to describe the relations between signifier and signified. From this perspective, a communication is channeled, and message is mediated, from the user, through the very facial recognition itself, to the targets. Such *human-computer interaction* (HCI) constitutes a “*one-shot message*,” informatician Clarisse Sieckenius de Souza infers, because the computer system more or less “tells” its human users how they should interact with it to achieve this or that effect (2005: 84). “The computer is the channel,” in de Souza’s interpretation, whereby “higher-level designer-to-user” and “lower-level” user-to-system messages are conveyed as part of a “computer-mediated *metacommunication*” (2005: 88). In the structuralist view of this *semiotic engineering*, a sign, its relations, and functions are principally considered a secondary communication. This “metacommunication,” to use psychiatrist Jurgen Ruesch and anthropologist Gregory Bateson’s term, is “a communication about communication” that includes all “cues” and “codes” for how information should be interpreted (1951: 209). That is to say, the computer-mediated metacommunication in facial recognition technology is between user and technology, the technology and target, and target and user. Either communication is conveyed at a higher-level between a human sender and a human receiver, that is, user-to-target. Or communication is carried at a lower-level between a human sender and a computer receiver, that is, user-to-technology and target-to-technology.

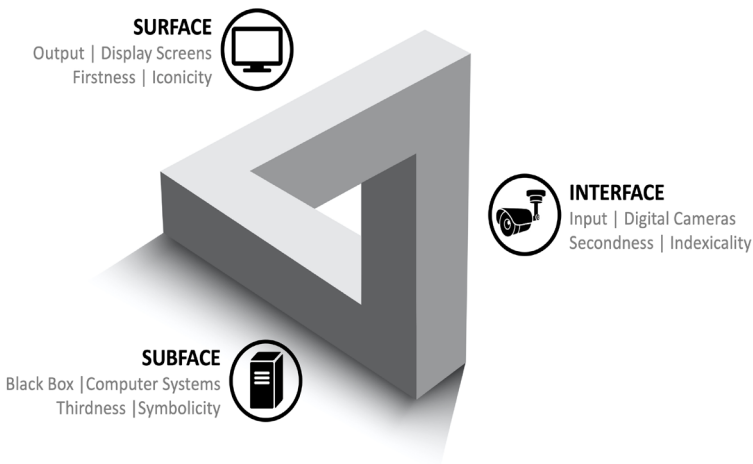
However, in a logical-gnoseological *semiotic* approach, like by Peirce or by Morris, one applies a triadic or tri-relative sign model to describe the relations between representamen, object, and interpretant. From this perspective, facial recognition technology is composed not only from a system of semiotic *artifacts* (such as computer systems, digital cameras, and display screens), but also from a system of semiotic *agents* (such as the facial recognition software itself, facial detection, extraction, and classification algorithms, their designers, users, and targets). Those who interact with the semiotic machine of this facial recognition do not reductively play just the passive role of a viewer, spectator, or audience – that is, a receiver. Rather,

they are an *interactor*. They are an *interpreter*. They are themselves a “semiotic agent,” Irvine maintains, who is definitively “presuppose[d]” by the design principles of the black box (2022: 208). In this view, facial recognition technology is greater than the sum of its parts. It is emergent. And it is irreducible.

The semiotic machine of facial recognition is characterized by triadicity, not only in regard to its relationality, but also in terms of its thinginess. Together, however, this human, computer, and human-computer interaction, as theorized by computer scientist, pioneering computer-based media artist, and computational semiotician Frieder Nake,<sup>3</sup> constitute an “*algorithmic sign*” (Hinterwaldner et al. 2014: 293). To analytically specify the algorithmic sign, Nake draws upon Peirce, who explains, for example, that “[a] sign is something, *A*, which brings something, *B*, its *interpretant* sign, determined or created by it, into the same sort of correspondence (or a lower implied sort) with something, *C*, its *object*, as that in which itself stands to *C*” ([1902] 1976: 54). As Peirce also portrays, “a *representamen*” is “that character of a thing by virtue of which, for the production of a certain mental effect,” or “*interpretant*,” it “may stand in place of another thing,” its “*object*” (CP 1.564). In the terms of this triad, a sign therefore has three parts: its *representamen*, *object*, and *interpretant*. Each of these parts can also be a sign in and of itself, when interpreted as such. And each can in turn lead to another sign, potentially interpreted *ad infinitum*. From this theoretical framework, Nake schematizes the algorithmic sign as “a sign [with] two interpretants,” one “generated by a human, [and] the other by [a computer]” (Hinterwaldner et al. 2014: 293). When something, like a digital image of a human face, is extracted from a video and enrolled into a computer, like in a facial recognition technology system, Nake proposes, “it exists in two forms,” and “gains a two-fold ontology,” where it must be “visible,” or at least somehow “perceivable,” by the human, and it must be “computable by the machine” (Hinterwaldner et al. 2014: 293). Semiotically speaking, Nake suggests, “both acts” by the human and the computer become “acts of interpretation” (Hinterwaldner et al. 2014: 294), which *in the very act itself* generate what Peirce terms interpretants. When engaging with facial recognition technology, the face under *semiosis*, that is, the face of the interactor, is being interpreted by two interpreters: the human and computer, independently and interrelatedly as well as concurrently and

<sup>3</sup> For transparency, I studied under Frieder Nake during my master’s degree in the histories of media art from Danube University Krems, when his courses were held at Benedictine monastery Göttweig Abbey, in Lower Austria in May 2015.

continuously. As noted by Nike, the “algorithmic sign” has three faces (Fig. 3): its *inner* face or “*subface*,” *between* face or “*interface*,” and *outer* face or “*surface*” (2008: 104, emphasis mine). This cognitive metaphor of the Janusian face or the multi-relational face is “justified” to Nike principally because of the “one-to-one correspondence” between the inner action and the outer appearance of the computer, its hardware, and software (2008: 105). Beyond the interdependent sympathetic causality between human psychology and facial behavior, there is an imperatively stable correspondence in the facial triadicty of the semiotic machine, that is, between its subface, interface, and surface.



**Figure 3:** A triadic model, visually diagrammed as a Penrose triangle, that represents the homologies, which are generally extensive but not specifically exclusive, in facial recognition between: semiotic face (subface, interface, and surface); computational software (black box, input, and output); mechanical hardware (computer systems, digital cameras, and display screens); phenomenological categories (thirdness, secondness, and firstness); and sign categories (symbolicity, indexicality, and iconicity).

© Schiller 2025.

## 5. Conclusion

From the critical standpoint of a computational semiotics both pragmatist and Peircean, “the pragmatic dimension of sign processing,” Nöth confirms (2002: 95), is a fundamental criterion of a full semiosis. Peirce himself discriminates between the human, the machine, and their minds. In his theory of a quasi-semiosis, Peirce maintains that machines work like

minds at least in some, if not in all, respects. He accounts not only for the differences between human semiosis and machine quasi-semiosis but also for the similarities between the human mind and machine minds. Perhaps moreover, Peirce maintains that human minds work like machine minds in some degree and to some extent. In Peirce's own words, a human can be considered "as a machine which turns out," for example, a "sentence expressing a conclusion" after "the man-machine" has "been fed with a [...] statement of fact, as premiss" or proposition (CP 2.59), analogously speaking, like in facial recognition, its templates, and matches. As Nöth also notes, the human mind is like a machine mind only when it solves "a task that a logical or calculating machine can *also* solve," that is, when it follows the rules "of a predetermined algorithm in a quasi-mechanical way" (2002: 88-89). Although the most "[f]ormal" and the "least important part[s] of reasoning [...] may be performed by a machine," Peirce "hold[s] that reasoning is the observation of relations, mainly by means of diagrams and the like" (qtd. in Ketner and Stewart 1984: 209). Reasoning, to Peirce, therefore, "is a living process" (qtd. in Ketner and Stewart 1984: 209).

However, Peirce primarily hesitates, reasoning cannot be performed "by *the unaided brain*," and, in fact, it "needs the cooperation of the eyes and hands," in the "art of this experimentation," to "use all [different] kinds of diagrams and devices for aiding the imagination" (qtd. in Ketner and Stewart 1984: 209, emphasis added). As posited here by Peirce, reasoning may be aided by semiotic artifacts for off-loaded cognition, as in facial recognition and its artificial intelligence. Generally, Nake and Grabowski find, the "*interface [is] the face*" of such a software (2006: 67). To separate the interface between human, computer, and their interaction from the surface of the display screens and the subface of the computer system, Nake and Grabowski speculate, would "render software *faceless*" (2006: 67). A computer system, Nake and Grabowski specify, "cannot exist without [a] face;" indeed, a computer system exists *only when* it "show[s] its face;" and *without a face*, "it does not exist at all" (2006: 67). With facial recognition technology during human-computer interaction, the human act of sign interpretation is brought into relation with the computer operation of signal determination. Human-computer interaction, however, is neither bilateral nor reciprocal in "a true sense," Nake infers (2008: 107). In the final analysis, this relation is the result, not of artificial intelligence or the artificial mind, but of human intelligence and the human mind. With the artificial intelligence that is facial recognition, like the oracles of olden times, the one true face among its Janusian relationalities is the face of *those who use it*.

## References

Andersen, P. B., P. Hasle and P. A. Brandt. 1997. Machine Semiosis (*Maschinensemiose*). In Posner, R., K. Robering, and T.A. Sebeok (eds.). *Semiotik: Ein Handbuch zu den zeichentheoretischen Grundlagen von Nature und Kultur/Semiotics: A Handbook on the Sign-Theoretic Foundations of Nature and Culture, Teilband 1/Volume 1*. Berlin: Walter de Gruyter, 548–570.

Ashby, W. R. 1956. *An Introduction to Cybernetics*. London: Chapman & Hall.

Ashby, W. R. (1958) 1991. General Systems Theory as a New Discipline. In Klir, G. J. (ed.). *Facets of Systems Science*, Second Edition, International Federation for Systems Research International Series on Systems Science and Engineering 7. Boston: Springer, 299–308. Available at: [https://doi.org/10.1007/978-1-4899-0718-9\\_16](https://doi.org/10.1007/978-1-4899-0718-9_16). Originally published in Bertalanffy, L. von, A. Rapoport (eds.). *General Systems: Yearbook of the Society for the Advancement of General Systems Theory*, Vol. 3. Ann Arbor, MI: The Society for General Systems Research, 1–6.

Bijker, W. E., T. P. Hughes and T. Pinch. 2012 [1987]. General Introduction from the Original Edition. In Bijker, W. E., T. P. Hughes and T. Pinch (eds.). *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. 25<sup>th</sup> Anniversary Edition. Cambridge, MA: The MIT Press, xxxix-xliv.

Bunge, M. 1963. “A General Black Box Theory”. *Philosophy of Science*, Vol. 30, No. 4, 346–358. Available at: <https://doi.org/10.1086/287954>.

Buolamwini, J. 2023. *Unmasking AI: My Mission to Protect What is Human in a World of Machines*. New York: Random House.

Crawford, K. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, CT: Yale University Press.

Denning, P. J. 2012. “Opening Statement: What is Computation?”. *The Computer Journal*, Vol. 55, No. 7, 805–810.

Denning, P. J., C. H. Martell. 2015. *Great Principles of Computing*. Cambridge, MA: MIT Press.

European Parliament. 2024. Artificial Intelligence Act: MEP’S Adopt Landmark Law. *European Parliament News*, 13 March 2024. Available at: <https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law>.

Fetzer, J. H. 1990. *Artificial Intelligence: Its Scope and Limits*. Studies in Cognitive System (COFS) 4. Dordrecht: Kluwer.

Gates, K. A. 2011. *Our Biometric Future: Facial Recognition Technology and the Culture of Surveillance*. New York: New York University Press.

Geertz, C. 1973. Thick Description: Toward an Interpretive Theory of Culture. In Geertz, C. *The Interpretation of Cultures: Selected Essays*. New York: Basic Books, 3–30.

Glanville, R. 1982. “Inside Every White Box There Are Two Black Boxes Trying to Get Out.” *Behavioural Science*, Vol. 12, No. 1, 1–11. Available at: <https://doi.org/10.1002/bs.3830270102>.

Glanville, R. 2009. “Black Boxes”. *Cybernetics and Human Knowing: A Journal of Second Order Cybernetics, Autopoiesis and Cybersemiotics*, Vol. 16, No. 1–2, 153–167.

Hinterwaldner, I., L. Manovich, M. Warnke and F. Nake. 2014. “Roundtable: From Algorithmic Signs to Cultural Analytics”. *Zeitschrift für Ästhetik und Allgemeine Kunstwissenschaft*, Vol. 59, No. 2, 291–302. Available at: <https://doi.org/10.28937/1000106251> (accessed 14 March 2024).

Introna, L. D. 2005. “Disclosive Ethics and Information Technology: Disclosing Facial Recognition Systems”. *Ethics and Information Technology*, Vol. 7, No. 2, 75–86. Available at: <https://doi.org/10.1007/s10676-005-4583-2>.

Irvine, M. 2022. Semiotics in Computation and Information Systems. In Pelkey, J., S. W. Matthews (eds.). *Bloomsbury Semiotics, Volume 2: Semiotics in the Natural and Technical Sciences*. London: Bloomsbury Academic, 203–237.

Ketner, K. L., A. F. Stewart. 1984. “The Early History of Computer Design: Charles Sanders Peirce and Marquand’s Logical Machines”. *The Princeton University Library Chronicle*, Vol. 45, No. 3, 187–224. Available at: <https://www.jstor.org/stable/26402393>.

Lakoff, G., M. Johnson. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.

Lakoff, G., M. Johnson. 1999. *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Books.

Latour, B. 1999. *Pandora’s Hope: Essays on the Reality of Science Studies*. Cambridge, MA: Harvard University Press.

Layton, E. 1977. Conditions of Technological Development. In Spiegel-Rösing, I., D. de Solla Price (eds.). *Science, Technology and Society: A Cross-Disciplinary Perspective*. London: Sage, 197–222.

Leone, M. 2024. “Semiotics of the Black Box: On the Rhetorics of Algorithmic Images”. *Visual Communication*, Vol. 23, No. 3, 426–451 Available at: <https://doi.org/10.1177/14703572241247120>.

Maturana, H. R., F. Varela. 1980. *Autopoiesis and Cognition: The Realization of the Living*. Boston Studies in the Philosophy and History of Science (BSPS) 42. Dordrecht: D. Reidel Publishing Co. Originally published in Spanish: 1972. *De Maquinas y Seres Vivos*. Santiago: Editorial Universitaria S.A.

Meunier, J.–G. 2022. *Computational Semiotics*. Bloomsbury Academic.

Nadin, M. 2007. “Semiotic Machine”. *The Public Journal of Semiotics*, Vol. 1, No. 1, 57–75. Available at: <https://doi.org/10.37693/pjos/2007.1.8815>.

Nake, F. 2008. Surface, Interface, Subface: Three Cases of Interaction and One Concept. In Seifert, U., J. H. Kim and A. Moore (eds.). *Paradoxes of Interactivity: Perspectives for Media Theory, Human-Computer Interaction, and Artistic Investigations*. Transcript Verlag, 92–109.

Nake, F., S. Grabowski. 2006. The Interface as Sign and as Aesthetic Event. In Fishwick, P.A. (ed.). *Aesthetic Computing*. The MIT Press, 53–70

Nöth, W. 2002. “Semiotic Machines”. *S.E.E.D. Journal (Semiotics, Evolution, Energy, and Development)*, Vol. 3, No. 3, 81–99.

Ovid. (8 CE) 2011. *Times and Reasons: A New Translation of Fasti*. (transl. Anne Wiseman and Peter Wiseman. Oxford: Oxford University Press.

Peirce, C.S. (1902) 1976. *The New Elements of Mathematics, Volume 4: Mathematical Philosophy*. Carolyn Eisele (ed.). The Hague: Mouton Publishers. (Cited as NEM.)

Peirce, C. S. (1931–1958). *The Collected Papers of Charles Sanders Peirce*. 1931–1935, Vols. 1–6, Charles Hartshorne and Paul Weiss (eds.). 1958, Vols. 7–8, ed. Arthur W. Burks (ed.). Cambridge, MA: Harvard University Press. (Cited as CP.)

Peters, J. D., B. Peters. 2016. “Norbert Wiener as Pragmatist.” *Empedocles: European Journal for the Philosophy of Communication*, Vol. 7, No. 2, 157–172. Available at: [https://doi.org/10.1386/ejpc.7.2.157\\_1](https://doi.org/10.1386/ejpc.7.2.157_1).

Ruesch, J., G. Bateson 1951. *Communication: The Social Matrix of Psychiatry*. New York: W. W. Norton and Co.

Ryle, G. (1971) 2009. The Thinking of Thoughts: What is ‘le Penseur’ Doing? In Ryle, G. *Collected Papers, Vol. 2, Collected Essays: 1929–1968*. London: Routledge, 494–510.

Sørensen, B., T. Thellefsen and M. Thellefsen. 2020. A Peircean Semiotics of Technological Artefacts. In Jappy, T. (ed.). *The Bloomsbury Companion to Contemporary Peircean Semiotics*. London: Bloomsbury, 253–275.

Souza, C. S. de. 2005. *The Semiotic Engineering of Human-Computer Interaction*. Acting with Technology 3. Cambridge, MA: The MIT Press.

Wiener, N. (1948) 1961. *Cybernetics: or the Control and Communication in the Animal and the Machine*. Second Edition. Cambridge, MA: The MIT Press.

Winner, L. (1991) 1993. "Upon Opening the Black Box and Finding it Empty: Social Constructivism and the Philosophy of Technology". *Science, Technology, & Human Values*, Vol. 18, No. 3, 362–378. Available at: <https://doi.org/10.1177/016224399301800306>. Originally published in Pitt, C., E. Lugo (eds.). *The Technology of Discovery and the Discovery of Technology: Proceedings of the 6th International Conference of The Society for Philosophy and Technology*. Blacksburg, VA: Society for Philosophy and Technology, 503–519.

## ARE WE HUMAN OR ARE WE DANCER? – AI CREATIVITY IN THE REALM OF XR

*Momchil Alexiev*  
*National Academy of Art*  
*momchil.alexiev@gmail.com*

### **Abstract**

As an emerging field of research, the intersection of AI and XR is expanding in front of us as unexplored topologies of meaning. The field of extended reality, still stumbling in the periphery of the cultural semiosphere, is becoming increasingly connected to the very central today exponential growth of artificial intelligence. This paper explores the nature of this intersection and brings forward arguments for tackling the understudied questions of its social and artistic impact. Most studies currently examine the application of AI tools within the context of XR, not the other way around. Their focus is on collecting and analyzing data regarding the speed and effectiveness of creating realistic XR worlds, the user experience

in its performance-driven aspects, interactivity and gestural locomotion techniques, as well as perceptual experiments based on studies of virtual assistants. This indicates a gap of research papers addressing the topic of AI's impact on the cultural and social mechanics of XR.

This paper bases its findings and proposals on several research papers conducted on the topic of social and artistic impact. These include for example “Extended Reality and Artificial Intelligence’s Ethical Crossroads: From Sensory Manipulation to Creative Disruption” and George Mason University studies on “Automatic Generation of VR Scenarios and XR Experiences”, as well as XR productions created by leading artists and studios in the field.

On the one hand, the aim is to articulate a discussion on the upcoming changes in the field of artistic practices as a result of the introduction of AI toolkits, while on the other, to contextualize the use of artificial intelligence in the realm of XR’s potentialities for new types of networking.

**Keywords:** virtual reality, artificial intelligence, extended reality, immersive media, digital culture, the cultural explosion of AI

At present, the majority of scholarly inquiry at the intersection of artificial intelligence (AI) and extended reality (XR) is predominantly directed towards the application of AI technologies as instrumental tools for optimizing XR production pipelines. This body of research tends to emphasize computational efficiency, focusing on the acceleration and automation of XR content creation, as well as the enhancement of user experiences through performance-driven metrics. Typical studies in this domain analyze aspects such as the generation of realistic virtual environments (Slater & Sanchez-Vives 2016), the refinement of gestural and locomotion interfaces, and the development of responsive virtual assistants capable of simulating naturalistic human interaction. While these areas represent important technological advancements, they are frequently grounded in utilitarian and engineering-oriented paradigms which prioritize functionality, realism, and immersion rather than the artistic or sociocultural potentials. The integration of Artificial Intelligence with Extended Reality technologies marks a significant cultural threshold where enhanced environments move beyond mere technological advancement, ushering into new modes of perception, creation, and interaction. As AI begins to inhabit and animate XR architectures with agency, learning and improvisation, the immersive spaces become more performative, volatile, and participatory. This article argues that the role of AI in XR is not simply instrumental but inherently transformative (Floridi 2014). It calls for critical engagement with

how AI reshapes cultural and social meaning-making within XR platforms. The text proposes a shift in the research focus, from AI as a toolkit within XR to a co-constituent of cultural logic.

A central dimension of this contemporary convergence lies in the ability of AI systems to furnish XR with implicit environmental and situational functionalities. Algorithms are increasingly employed to autonomously generate visual elements, animations, and even entire virtual environments. AI plays a pivotal role in addressing XR content production challenges. Generative AI tools now facilitate the rapid creation of 3D models, all types of environments and code, thus significantly reducing the time and financial resources required to develop XR experiences: a barrier too high for most small and middle sized production companies. Platforms such as Meta's Horizon Worlds integrate AI-driven procedural generation allowing entire virtual worlds to be built through text prompts, while specialized services like Meshy, HunYuan, and Suno automate asset generation. Coding assistants such as Claude Code further streamline software development. These innovations democratize XR creation, enabling broader participation and accelerated experimentation. These systems produce dynamic artworks which evolve in real time in response to user interaction or external inputs, thus leading to the emergence of participatory experiences in which the viewer's agency becomes central to the unfolding of the artwork.

Beyond the visual, AI is also employed in the generation of music and soundscapes where the coupling of generative sound with immersive media heightens affective engagement, and further dissolves the boundary between spectator and artwork. By analyzing behavioral data and user preferences, these systems can tailor immersive experiences with adaptive context in which avatars change reactions, tone, behavior and the overall emotional vibration of a story. This level of adaptation challenges the one-size-fits-all paradigm of traditional media and points toward an aesthetic of relational and contingent art. A slightly different case is represented by smartglasses, which when equipped with cameras, microphones, and other sensors, enable continuous spatial awareness without explicit user input. This capability allows AI to respond dynamically to the surrounding environment. Such seamless contextual interaction is augmented by XR's hands-free affordances which are particularly suited to tasks requiring simultaneous physical and cognitive engagement.

Despite these benefits, generative AI and its speedy ready-made solutions introduce a series of incongruences. On one hand, the main question it raises among career artists is not as much related to the already common discussion on non-linear narratives. It rather points to the "answer instead

of process” nature of AI and the extent to which it predisposes for an authentic artistic act which can lead to a product with a deeper socio-cultural value (McCormack, Gifford, and Hutchings 2019). It raises the question of whether the algorithm deprives us of all the mistakes and hardships of the creative process which ultimately lead to an “artistic discovery”, and a genuinely authentic work relatable in the human sense. On the other hand, the ease of production often results in an abundance of low-quality homogenous “AI slops” exacerbated by insufficient content curation. Current AI-generated assets frequently exhibit technical limitations such as imperfect mesh topologies, inconsistent visual styles, and the recognizable “AI-generated” aesthetic. This can diminish cohesion and the overall feeling of humanness of the experience. This issue reflects some of the main concerns raised by Kate Crawford and Trevor Paglen regarding the cultural construction of training data in AI systems (Trevor & Paglen 2021). This mode of production risks eroding the sense of craftsmanship otherwise historically associated with artisanal XR worldbuilding.

While AI expedites development, its outputs may lack the emotional depth and narrative richness achieved through classical human creativity. Traditionally perceived as abstract in the sense of manifesting only as text or disembodied voices, AI agents in XR can be spatially and visually integrated into the perceptual fields of users. Augmented reality interfaces enable AI assistants to deliver information through overlays, spatial cues, or anthropomorphic avatars, thus deepening engagement and facilitating the intelligibility of the process. This physicalization strengthens the relational dimension of AI by embedding it within interactive, socially intuitive contexts. Assuming the role of a storyteller and an improviser, AI starts moving beyond fixed logic trees and increasingly improves narrative arcs in real time immersive spaces. These generative systems respond to user inputs and environmental contexts thus creating bespoke trajectories and they become a dramaturgical agent which shapes affect and tension in ways previously limited to human authorship. Neural networks and generative models enable hybrid art forms in XR which are co-produced by artist and algorithms. This artistic disruption signals a move toward emergent forms of authorship and speculative aesthetics that challenge conventional notions of originality. “Extended Reality and Artificial Intelligence’s Ethical Crossroads: From Sensory Manipulation to Creative Disruption” (Jorge & Hürst 2024), a round-table based research paper, explores the ethical, social, and creative implications arising at the intersection of Artificial Intelligence and Extended Reality. Through expert contributions from leading figures in immersive media, AI research, and human-computer interac-

tion, the discussion presents a critical analysis of the rapidly evolving AI-XR ecosystems. Central to its argument is the necessity of human agency, ethical safeguards, and artistic integrity amidst the growing influence of generative and predictive AI systems in immersive media. A primary conclusion is that while AI holds immense promise for the automation and personalization of XR content, such as worldbuilding, character generation, and real-time behavioral adaptation, it simultaneously introduces a range of ethical risks. These include exploitation of creative labor, privacy violations through biometric data harvesting and the erosion of trust through data contamination and misinformation. The artist, Jacquelyn Ford Morie, for example, warns of the dangers of “data poisoning”, wherein AI-generated hallucinations feed back into training data, resulting in degraded system reliability. She advocates copyright protection, artwork tagging, and artist compensation as key steps toward reestablishing creative trust. She claims AI in XR must be transparent, accountable, and respectful of human creators. Compensation and recognition for artists and data contributors are not only ethical but necessary for sustaining trust. Andrew Glassner echoes these concerns, arguing that AI systems often operate as opaque “black boxes” lacking transparency and accountability. He contends that while AI can simulate plausible narratives and visuals, it cannot genuinely reflect the richness of human lived experience. Without clear provenance and validation mechanisms, the data which underpins AI training remain susceptible to manipulation, bias, and corporate control. Thus, they ultimately affect how reality is perceived and shaped in XR environments. Glassner is especially critical of the unchecked power which AI developers wield over perceptual infrastructures, warning that this could lead to subtle yet profound manipulation of public understanding through XR-mediated experiences. Nuria Oliver outlines the structural challenges of building reliable AI systems for XR, highlighting the enormous demands of data volume, annotation, latency, and environmental impact. She also emphasizes the lack of consensus over ethical alignment in AI training, a problem which undermines inclusivity and user trust. She suggests that the merging of classical symbolic AI with contemporary machine learning could yield more transparent and explainable systems, although scalability remains an open challenge. She also claims that hybrid models combining symbolic and statistical AI could help address explainability and value alignment, enabling more nuanced and human-centered XR systems. Mark Billinghurst points to both risk and opportunity in AI-enhanced XR. On the one hand, sensor data used for intuitive interaction, such as head movement or gaze, can jeopardize user privacy when misused. On the other, such data,

when responsibly managed, can enable more empathetic virtual agents and adaptive environments. He sees potential in AI-driven cultural translation and matchmaking systems which enhance social XR experiences by aligning gestures, language, and context across global users.

Recent advancements at the intersection of artificial intelligence, extended reality, and computational design are exemplified by the work of Professor Craig Yu and his research group at George Mason University (Yu 2024). Their work represents a multidimensional approach to the automatic generation of XR scenarios emphasizing both algorithmic innovation and user-centric interactivity. Operating under the banner of the Design Computing and Extended Reality (DCXR) Lab, the group explores how generative models and machine learning can be embedded within immersive environments, in order to enable adaptive, intelligent, and context-sensitive experiences. Yu's team engages with XR across several domains, notably including generative storytelling, personalized training, robot teleoperation, computational interior design, and interactive agent systems. Central to their methodology is the transformation of traditional manual design tasks into computational optimization problems. These problems are mathematically formulated to define large-scale design spaces, enabling algorithmic agents to explore multiple high-quality solutions which human designers might otherwise overlook. For example, their early work on interior design automation demonstrated how optimization could generate thousands of viable room layouts by learning spatial patterns from existing examples and applying constraints such as circulation paths. A particularly innovative area of their research focuses on personalized XR training environments. By leveraging the data-rich nature of virtual environments such as gaze tracking, gesture analysis, and performance logging, Yu's team creates adaptive training modules for skills like driving and wheelchair navigation. These systems are driven by AI optimizers that evaluate user behavior and dynamically generate task scenarios tailored to observed weaknesses. For instance, if a user demonstrates difficulty in executing turns while driving, the system generates iterative scenarios that emphasize turning maneuvers thus improving skill retention through repetition and contextual variation. Similarly, in the domain of computational interaction the team formulates user-tool interactions as quantifiable events. Using machine learning classifiers, they have developed adaptive assistance systems that respond to inferred cognitive states, such as confusion by offering context-aware support only when necessary. This model of "just-in-time" interactivity illustrates a shift toward emotionally and cognitively responsive XR environments.

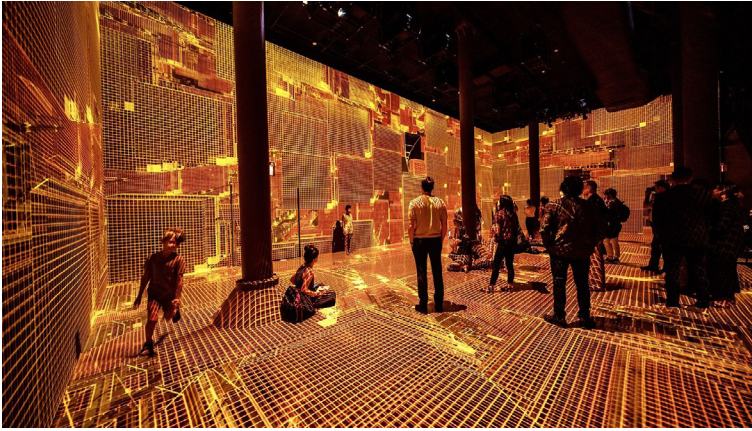
The lab also explores mixed reality storytelling through augmented reality. Yu's AR storytelling engine dynamically populates physical environments with virtual characters and narrative sequences which adapt to the semantic structure of the physical space. This involves photogrammetric scans of the environment, object segmentation via computer vision, and story adaptation through optimization frameworks. The result is a contextually coherent narrative mapped to a user's real-world surroundings. With the integration of large language models such as ChatGPT, the system can even generate narratives in real-time, moving beyond predefined scripts toward emergent AI-driven storytelling.

An extension of this work includes life-like virtual agents and animal companions. For example, the team developed a data-driven virtual cat whose behavior is modeled on real-world observations, with actions determined by AI optimizers. Additionally, the group has scaled their AR experiences to outdoor settings, exemplified by a project involving a virtual dragon navigating real-world spaces based on user annotations and terrain classifications. Through this body of work, Yu's group demonstrates in practice a significant shift in the role of AI in XR, namely from a toolkit used to simulate or assist, to a generative agent which co-constructs environments, narratives, and user experiences. Their research reconfigures traditional boundaries between user and system, artist and tool, environment and narrative. The implications are twofold. On the one hand, AI becomes a co-author of virtual experience, while on the other, XR becomes not only a domain of technological display but also a field of epistemological transformation, where computational systems participate in the creation of meaning, agency, and sense-making.



Refik Anadol  
Unsupervised - Machine Hallucinations,  
2016, Exhibited at MoMa, 2023

<https://refikanadol.com/works/unsupervised/>



Refik Anadol

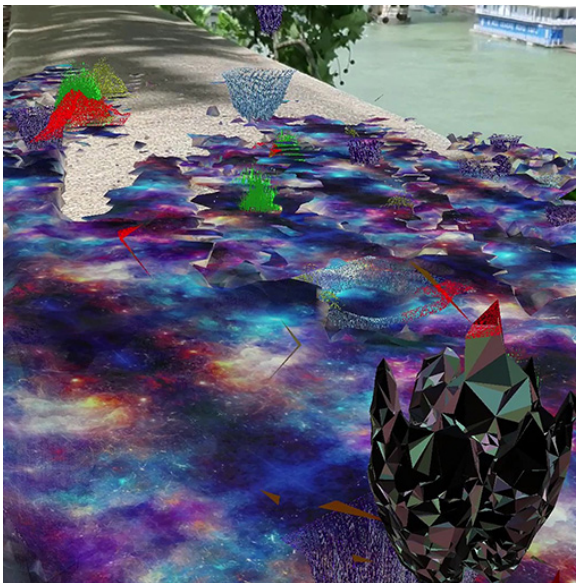
Unsupervised - Machine Hallucinations, 2016

Exhibited at Artechouse, New York, 2019

<https://refikanadol.com/works/machine-hallucination/>

Various creative organizations and artists, albeit still few in number, are exploring the creative potential in this still artistically unclear field. *An intriguing example is “Unsupervised (2022)”*. This project is a prominent generative artwork within the larger framework of Refik Anadol Studio’s “Machine Hallucinations” project. It exemplifies a critical convergence between artificial intelligence and extended reality as modes of artistic expression. Drawing on vast datasets and unsupervised machine learning models, the work leverages AI not simply as a computational tool, but as an active co-creator capable of forming new visual languages. The conceptual architecture hinges on the use of scale visual datasets, most notably the entire collection of The Museum of Modern Art (MoMA) in New York. Through unsupervised learning, the algorithm identifies latent patterns and correlations in this archive without human-labeled categories, ultimately generating new visual compositions. These synthetic forms are not mere reproductions but constitute a speculative visual history generated through the machine’s internal logic. In this sense, “Unsupervised” offers a “hallucinatory” reading of art history, filtered through the perceptual processes of AI. What distinguishes it within the domain of immersive media is its deployment within immersive frameworks. The project has been exhibited in large-scale LED installations and spatialized environments effectively translating the machine’s latent space into an embodied sensorial experience for the audience. The use of XR in this context is not ancillary but integral, since it allows the viewer to “step into” the operational logic of the AI system, navigating a space that is continuously reconfigured in real

time. The result is an aesthetic encounter which is neither fully authored by the human nor wholly autonomous in its machine generation. Rather it emerges from a complex feedback loop between algorithmic agency, spatial computing, and embodied perception. In this synthetic ecology, data, code and perception are entangled in the production of new aesthetic realities thus signifying an emerging ontological shift from spectatorship to co-presence within the artwork's algorithmic world. The humans are invited to witness from within the "dance" of the machine. This logic is reversed in the artwork "Objects (RE) Oriented" by Chiara Passa exhibited at HKUST Guangzhou "SURREALITY" exhibition in 2025. This is an interactive XR installation which investigates the boundaries between tangible materiality and digitally mediated environments. Viewers are invited to interact with the installation by tapping directly on their device screens, overlaid on 3D mesh surfaces visible through augmented reality. Each interaction generates and places within the environment a randomized piece of 3D "debris" drawn from a custom-built and continually expanding database of digital assets. These debris forms were initially created using artificial intelligence and curated from generative workflows and datasets developed specifically for the project. The audience can scale, rotate, translate, recolor, and reposition them in the augmented environment. This feature encourages

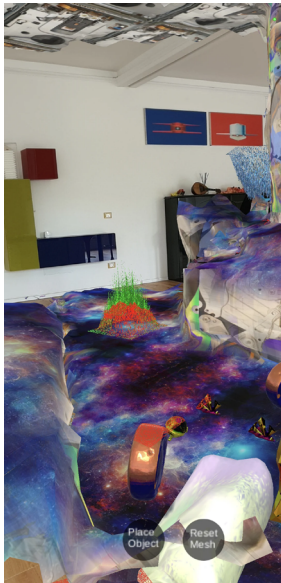


Chiara Passa

Objects (RE) Oriented Debris, 2024

Exhibited at Center for Metaverse and Computational Creativity (MC2), Guangzhou, 2025

<https://www.chiarapassa.it/artworks-2016-2026.html>



Chiara Passa  
Objects (RE) Oriented Debris, 2024  
Exhibited at Center for Metaverse and Computational  
Creativity (MC2), Guangzhou, 2025  
<https://www.chiarapassa.it/artworks-2016-2026.html>

compositional agency, allowing participants to co-construct a new spatial topology. Each debris item is also sonically augmented and emits space-related ambient sound extracted from public archives of the European Space Agency (ESA). Complementing the virtual layer, the installation incorporates real-world sculptures placed within the exhibition space. These sculptures are embedded with Near Field Communication (NFC) tags, enabling visitors to trigger additional content upon physical proximity. By scanning the sculpture, viewers reveal internal animated sequences, thus extending the artwork's hybrid dramaturgy into the interior architecture of physical objects. The technology here is not leading us, but rather moving along and reacting, provoking and stimulating. It is vibration between the physical and the virtual within which meanings and aesthetics conquer new cultural territories.

In the abovementioned examples XR and AI are very clearly positioned both as tools as well as co-constituents. They do not substitute processes or imitate artistry. On this technological intersection, Rossa and Anadolu manage to curate experiences which invite the spectators into a new dance of meaning making. Audiences are not dehumanized but rather enabled to explore and form new opinions on the state of digital art today. Both works create new social spaces rather than isolating or disconnecting the viewers. They manage to continue the tradition of shared cultural rituals, whilst, in the same time, expanding the contemporary definitions of artistic experiences. The discussion on our rapidly evolving technological frame is thus

enabled to maintain an active and critically vivid interconnection between the physical, the intellectual and the emotional.



Still frame from *We Met in Virtual Reality* (2022), illustrating avatar-based social interaction within the VRChat platform.

<https://www.imdb.com/title/tt16378482/>

Social platforms are another realm which gives numerous examples of how human interconnection expands throughout the digital realm. Artificial Intelligence is increasingly integrated into applications like VRChat, transforming them from a user-driven virtual worlds into emerging platforms for AI-mediated interaction, avatar autonomy, and social experimentation. One key implementation involves the use of Open Sound Control (OSC) in conjunction with the Model Context Protocol (MCP), allowing AI assistants such as Claude to control avatar behavior in real-time. These assistants can manipulate expressions, movement, gaze, and even generate dialogue through external language models, making avatars semi-autonomous participants in social environments. In addition, VRChat's SDK now supports AI Navigation, which enables Non-Player Characters (NPCs) to traverse virtual spaces using Unity's NavMesh system. NPCs can dynamically avoid obstacles, perform contextual actions, such as jumping, and exhibit basic intelligent pathfinding, although some advanced features remain limited. A growing subculture also explores large language model-driven avatars, like the AI character "Celeste", which uses LLMs for text-based interaction and basic spatial awareness, albeit with limited emotional nuance and realism. Beyond entertainment, AI is being explored for real-time moderation. One example is the "Safe Guard" system which can detect hate speech in VRChat's voice chats, contributing to community safety and platform governance. Meanwhile, in creator communities, such

as VRChat Discord, generative AI tools are used to produce world assets, avatar models, and immersive scripts, sparking debates around authorship, ethics, and cultural authenticity. Collectively, these developments reveal a shift in XR practice, namely from AI as a mere toolkit to a co-constituent in the construction of meaning, behavior, and interaction in synthetic social spaces. A compelling illustration of the phenomenon of such digitally mediated social relationships is presented in the film “We Met in Virtual Reality” (Hunting 2022), which documents a series of personal encounters entirely filmed within VRChat. The film reveals how interactions initiated in virtual space can evolve into meaningful, emotionally resonant relationships in physical life, without either dimension negating the other. This interplay suggests a significant point of convergence between the social and the virtual, highlighting a cultural condition wherein self-referential signs and simulated presences are not merely tools of communication, but constituents of perceived reality. We inhabit a world where such signs, embedded within technological infrastructures and mediatized culture, are affecting not only modes of use, but also the very epistemic status of perception and truth. On one hand, this engenders a sense of technological acceleration and control, and on the other, it opens up novel, uncharted horizons of experience. (Baudrillard 1994).

In this context, the field of AI and XR art occupies a critical position. It not only facilitates the materialization of new social and cultural formations, but also engenders novel affective realities between users. Rather than serving as a simulation or virtual environments, it contributes to the speculative conditions through which meaning, agency, and authorship are negotiated. Today the intersection of AI and XR is beginning to co-produce logics of immersion, co-determine protocols of interactivity, and transform the ontological status of the “virtual” itself. As such its role is not solely technical, but from critical and theoretical perspective has an impact on perception, cognition, and the cultural construction of reality today.

## References

- Baudrillard, J. 1981. *Simulacres et Simulation*. Éditions Galilée. Paris
- Crawford, K., T. Paglen, 2021. “Excavating AI: The politics of images in machine learning training sets”. *AI & Society*, Vol. 36, 1105–1116. Available at: <https://doi.org/10.1007/s00146-021-01162-8>.
- Yu, C. 2024. Automatic Generation of VR Scenarios and XR Experiences. 2024. *YouTube*, 21 August 2024. Available at: <https://www.youtube.com/watch?v=9JnPBZh3Ov0>.

Floridi, L. 2014. *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*. Oxford University Press.

Jorge, J., W. Hürst. 2024. “Extended Reality and Artificial Intelligence’s Ethical Crossroads: From Sensory Manipulation to Creative Disruption”. *Computer Magazine*, Vol. 57, 14–20. Available at: <https://www.computer.org/csdl/magazine/co/2024/09/10660582/1ZPP3nuP5WU>.

McCormack, J., T. Gifford and P. Hutchings. 2019. *Autonomy, Authenticity, Authorship and Intention in Computer Generated Art*. Springer International Publishing.

Slater, M., M. V. Sanchez-Vives. 2016. “Enhancing our Lives with Immersive Virtual Reality”. *Frontiers in Robotics and AI*, Vol. 3. Available at: <https://doi.org/10.3389/frobt.2016.00074>.

### **Filmography**

*We Met in Virtual Reality*, directed by Joe Hunting J. USA: Cinetic Media, 21 January 2022.

## HUMANS AS NATURAL-BORN CYBORGS: SCRUTINISING AI'S NARRATIVE INTELLIGENCE WITHIN THE 5E COGNITION FRAMEWORK

*Inna Livytska,  
Justus Liebig University  
Inna.Livytska@anglistik.uni-giessen.de*

### **Abstract**

Recent advances in large language models (LLMs) and their ability to generate content align with both perspectives: AI as a powerful assistant and AI as a potential challenge to human cognition. However, these advancements expose the persistent limitations of AI compared to humans. One such limitation, as demonstrated by the *folk narrative hypothesis* (Hutto 2008), is the uniquely human ability to engage in storytelling, a fundamental and ancient mechanism for memory, information storage, identity formation, and world-making or world-disruption. The ability of AI to generate coherent texts is often mistaken for its intelligence, raising false

claims about AI self-consciousness and sentience. This paper argues that human narrative intelligence as a drive for sense-making might turn imperfect AI text generation into a meaningful construct despite cognitive dissonance in the communication. By analysing the technical prompts used by ChatGPT-4 to generate narrative text and its understanding of key text production categories (coherence, fluency, complexity, etc.) within the 5E Cognition framework (Stilwell & Harman 2019), I will attempt to show how AI-generated narrative structures acquire full and meaningful interpretation when perceived by a human. The generative ability of AI which emerges in co-communication with a human cannot stand for its intelligence, since it reflects human perception and our intrinsic ability to hypothesise and engage with communication agents.

**Keywords:** narratology, AI, ChatGPT, narrative intelligence, 5E cognition, co-narration, communication.

## Introduction

When Andy Clark (2003) praised humans for being “natural-born cyborgs” due to our inherited ability to use the evolutionary mechanism of cognitive extension to interact with our surrounding *umwelt*, he was primarily focusing on the utilitarian role of computers and technology in preserving humanity’s exceptional status. This optimistic view of machines as tools and helpers designed to serve human purposes has always been tempered by an inconvenient yet inevitable concern: the possibility of encountering entities far more intelligent than ourselves. Recent advances of LLMs and the inconceivable admiration for their text generative ability to produce coherent texts seem to have brought these fears to our current days.

Indeed, fears of being outmastered and outcompeted by its creation are voiced today by the scientists who spent their entire lives working on AI. One of the persuasive examples is the 2024 Nobel Prize in Physics. Last year’s awardees, the U.S. scientist John Hopfield and British-Canadian Geoffrey Hinton, while being credited for laying the foundation work which revolutionised machine learning, both expressed their caution regarding the future use of their discoveries.

“We have no experience of what it is like to have things smarter than us”, Hinton said over in his phone call to the Nobel press conference. “It’s going to be wonderful in many respects, in areas like healthcare”, adding, “But we also have to worry about a number of possible bad consequences,

particularly the threat of these things getting out of control” (Pollard & Ahlander 2024).

The other 2024 Nobel Prize holder Professor Hopfield, a creator of an associative memory which can store and reconstruct images and other types of patterns in data, in his turn, stated: “When you get systems that are rich enough in complexity and size, they can have properties which you can’t possibly intuit from the elementary particles you put in there, you have to say that system contains some physics” (Pollard & Ahlander 2024). When referring to the impossibility of intuitively predicting the properties of this new system, we should probably see the sign of their autonomy, and thus, their agency, which is not yet fully discovered. The possibility of that stands at odds with the fears of AI and the risks for a new digital era of uncontrollable generative AI.

Bearing in mind these fears and partly sharing them, this paper takes a bold alternative perspective. It suggests that despite considerable progress of generative AI to imitate the formal functions of human linguistic competence, this process cannot be identical with human narrative performance from a purely biological evolutionary perspective. One of the primary reasons for my argument is the emotion generation capacity in humans, closely linked to narrative semantics and linguistic material, accelerating or inhibiting memorisation per se. Since LLMs are trained on vast amounts of linguistic data, their performance results from their response to the language in the prompt which involves generating more language. In other words, language-centred large models (LLMs) assume a great deal of material produced by humans without differentiating between purely narrative categories such as authorial agency, intentions/purposes, and audiences. In short, they generate texts without understanding because they rely heavily on human prompts as scaffolds to succeed and proceed in communication (Pütz & Esposito 2024).

This observation again proves that current LLMs are good at generating narratives as “stochastic parrots”, combining semantic and grammatical structures. (cf. “Poetics Today” 2024). Moreover, additional supporting evidence comes from recent studies in neurolinguistics, providing clinically proven test results that “the language network does not support non-linguistic cognition” (Mahowald et al. 2023). The results of this study resonate with the most advanced discoveries in human cognition known as the 5E cognition paradigm (Stilwell & Harman 2021), for its embodied, embedded, extended, enacted and emotive character that algorithmic machines definitely lack.

The very assumption that “in humans, language capacity is robustly dissociated from the rest of high-level cognition, as well as from perception and action” (ibid.), brings into doubt the ability of AI to reproduce or generate emotions through other channels different from the linguistic one. As far as the current state of art signals, AI has successfully acquired the formal linguistic capacity, dissociated from the rest of perception, cognition and action, i.e. other crucial non-linguistic cognition (ibid.). Therefore, claims that AI is sentient and is capable of generating emotional narratives with understanding them are premature and can be compared to a reader’s projection (Wolfgang Iser’s theory of reception) onto the texts and characters which AI can generate. We, as conscious readers, have become used to the fact and have been trained via socialisation learning practices that there is some intention behind the text we read and an implied author speaks to us via its text. Therefore, the entire idea of our narrative perception needs to be realigned with the new reality of interacting with non-conscious non-human entities which have mastered human formal linguistic capacity of communication, i.e. language.

In what follows, I will draw upon the recent research on AI’s text-generating ability in the context of embodied narratology (Caracciolo & Kukkonen 2021), engaging the human folk narrative hypothesis (Hutto), and incorporating findings from the most recent research in the 5-E cognition paradigm (Stilwell & Harman 2021). Next, I will move to identify the role of enactive reasoning and embodiment in narrative meaning-making to sketch a tentative model of the co-narrative intelligence of AI with a human actor, i.e. a prompted generation of narrative texts to run the Turing Test by LLMs such as ChatGPT-4. The discussion section will debate the obtained observation about the strategies of AI for narrative construction, and its reliance on the human meaning repair that might serve as the grounds for further improvement of the LLMs. Yet, for my argument, it will pinpoint essential drawbacks and the problems in current LLMs related to emotional semantics and narrative ability due to their incompatibility with the human emotionally outlined narratives, which will be discussed below. The paper will conclude by defining possible future directions of research on the narrative creativity and intelligence of AI.

## **1. Generative AI and Contemporary Narratology**

Narrative ability, as a cognitive and linguistic skill to use stories to render the experiences, has always been considered a solely human prerogative. What is well established is that narrative, as a communicative tool that evolved from the oral to the written form throughout our evolution

as a human species. Ultimately linked to the emergence of verbal and non-verbal languages, “narrative” was put “on a pedestal” for its ubiquitous presence and uniqueness - until recently. Today, the new generation LLMs like ChatGPT-4 can easily solve tasks to produce formally coherent texts, and be praised for their communicative intelligence. Within such developments, the territory of modern narratology, that has been always occupied with the formal, structural and global questions of human creativity in the texts writing, including the role of the author, textual intentionality, its cultural and historical *umwelt*, in their intersections with other genres and multimodal modes of storytelling, has faced the necessity to widen its research scope. The need to embrace the challenges of the non-human agency, the power of the algorithms and artificial text generation, which, within the blink of an eye, broke into its “sacred” narratological domain. Large Language Models (LLMs), especially their commercial versions like ChatGPT-4, continue to astonish their users with their increasingly surprising, albeit more than satisfactory, capacity to “produce clearly recognisable narrative texts” (Phelan 2024: 197). Despite functioning as “black boxes” (which makes them completely alien to researchers), LLMs continue to gain in popularity and are very successful among users, who almost automatically seek to turn AI-generated texts into “rhetorical narrative acts” (Phelan 2024: 197). It is true that, when dealing with LLMs, humans not only apply their cognitive receptiveness, but can also learn a lot about themselves when interacting with LLMs.

Following Claudio Paulucci’s discussion of the meaning of LLMs in the *Semiotica* journal, these new language-endowed machines now “(...) profoundly and clearly reveal something important about our essence and about how we function” (Paulucci 2025: 5). I support the idea, that by being observant and reflective using the human perception apparatus, and relying on previously accumulated research on communication in empirical and theoretical domains, we could cast some light on the peculiarities of the reception of AI-generated texts, in order to understand that humans can project their receptive and communication models to make the AI’s performance (more) meaningful. That is what we as humans have learnt through neuro-development, while interacting with other humans. I assume that this is why human users mistakenly confer LLMs with the power of human author-creator, possessed of intentionality, and sentience, etc., because they have no other experience of communication with non-human and non-organic species. This argument needs more discussion and will acquire more supporting evidence in this paper.

In 2019, when I sought to introduce the term “Artificial Narrative Intelligence” (ANI) in an attempt to compare the art of narration and artificial narrative intelligence of Story Generators (Livytska 2019), generative AI had not yet arrived. By “artificial narrative intelligence” (ANI), I meant “the insofar acquired capability of Story Generators to create and extend narrative situations to a storyworld level as a part of Umwelt construction” (p. 313). In justification for a closer collaboration between “narrative theory” and, back then, “digital storytelling”, we witnessed the newly emerged interdisciplinary research field called Computational Narrative Story Generation (Gervas 2024). This interdisciplinarity immediately underlined several incompatible methodological trajectories between narratology and computational story generation. Among them, the lack of a unified theory of narrativity, including the fuzziness of the notion of «narrative» itself, but mostly the challenges of Interactive Storytelling (cf. Fuller&Magerko 2010) in creating dramatic tension (Szilas & Richle 2013) as one of the prominent features of narrative stylistics. It has been disclosed that the artificial intelligence techniques and algorithms of story generation on the level of *historie* are rooted in a formalist framework, e.i. Propp’s morphological narrative functions which incorporate problem-solving strategies and grammar. At the same time, the formalist framework, used in story generators almost a decade ago, set a lot of limitations on AI in generating narrative texts regarding meaning, agency, comprehension, and subjectivity. Narrative construal in text-generating programmes was loosely based on the conception of the narrative as a certain frame of human cognition (Herman 2011), irreversibly linked to the notions of experientiality (Fludernik 1996) and qualia (Herman et al. 2005) to provide the feeling of “what it is like” to the readers. Consequently, story generators had a very scant use of lexical and semantic databases as well as general knowledge databases to build their stories on (Livytska 2019: 312).

After 2022, the situation radically changed. New and powerful text-generating LLMs became a part of our everyday life, calling for a more precise scientific exploration. The future of digital storytelling was here, and modern narratology acknowledged its call for action by approaching new algorithmic texts with a wide arsenal of research methods developed previously. However, it still seems to be lagging a bit behind the revolutionary breakthrough of AI’s textual generating performance as witnessed over the last three years. For instance, one of the recent issues on AI and narratology at the *Poetics Today* journal, titled “The AI Revolution: Speculations on Authorship, Pedagogy, and the Future of the Profession” (2024), presents the speculative essays on the role of the AI in the humanities and literary

studies broadly defined, gathered almost two years ago. As the editors of the issue explain, they aimed to provide “potential implications of large language models (LLMs) for our understanding of authorship, for the future of academic writing and pedagogy, and the university at large” (Evron & Tartakovsky 2024: 190).

Indeed, several papers in that issue strive to speculate on the questions of AI authorship and the distinction between AI and real flesh-and-blood authors, calling it “the second death of the author». In their explorations, they are following Roland Barthes’ “The Death of the Author” and Michel Foucault’s renowned work “What Is an Author?” conceptualisations of the narrative centre as either “dead” or reduced to a purely discursive position. Some essays encourage the collaborative use of LLMs in the classroom settings, in order to boost creativity and supervised cooperation with AI, arguing that they offer unique affordances for humanist inquiry and writing per se (cf. Eamon Duede and Richard Jean So 2024). After a clear warning and speculation on the temptation to AI plagiarism, Katherine Heyles, for instance, calls for seeing the potential pedagogical benefits of LLMs and their use in the classroom instead of prohibiting them. She argues that it would be both “wrongheaded and futile to try and ban this technology”. Instead, she claims, we should embrace LLMs “as useful tools to accelerate student learning.” (N. Katherine Hayles 2024).

Other controversial questions in the sphere of AI’s authorship in this issue concern its dubious collective authorial status. It is known that new, large generative models are being trained on huge data sets of previously existing works of literature and fiction in open access. This fact brings the novelty of the content they create into doubt, on the one hand, or their comparison with the poststructuralist idea of authorship, on the other. The corollary is: if the broadly defined idea of novelty presupposes restructuring previous knowledge and producing new knowledge in its place, then AI as an author has all grounds to be included in the narratological conception of the author (Slater 2024). The author, who identifies with the first person, remains “impersonal” (Paolucci 2025: 7). This new type of AI authorship can be compared to the impersonal theory of enunciation, suggested by Claudio Paolucci in 2020 in his book “Persona”. The underlying idea of this theory is the cumulative effect of experiences (norms, stereotypes, usages, habits, previous enunciations, etc.) of the surrounding environment that we formulate as our personal ones while creating the utterances.

Although the book appeared two years before the emergence of generative AI, the implications of this theory for ChatGPT text generation and new Transformative Models in general might be appropriate for several

reasons. On the first line, fed with a big amount of data, ChatGPT has absorbed billions of bits of information together with the norms, cliches, stereotypes, and stylistic peculiarities, in order to create convincing first-person accounts, which sound persuasive enough for humans to attribute human intentionality and thoughts. Our human thoughts, however. On the second account, ChatGPT and other generative LLMs are capable of passing various versions of the Turing Test, presupposing that the formal difference between human agency and mechanical one is growing less perceivable. However, this will all make sense if we disregard the role of the human body and non-verbal communication in machine-human co-construction.

As mentioned above, human perception brings meaningfulness to the artificially created text because we rely on our logic and intuition via “languageing” (Maturana) as a semiotic mechanism of interpersonal communication. Following Niklas Luhman’s definition of communication, the Turing test might become unreliable since it is oriented only on linguistic mimicry and language use, and it cannot measure understanding, reasoning, or consciousness of the machine. I will provide some examples of ChatGPT’s generative capability with the prompts to run a short version of the Turing test as a play with a chat. It will speak of the inefficiency of the Turing test as a reliable marker of thinking, emotional experience or memory. That is why, if we want to have some analytic insights into the narrative generation of AI and its adaptability in the process of human-machine communication, it is allegedly more beneficial to use *the reader’s perspectives* first, i.e. *who reads* those texts and how, instead of *who speaks* (bearing in mind “impersonal model” of first person narrator suggested by Claudio Paolucci). At least, that is something that we as humans have learnt to be doing since the invention of the printing machine by Gutenberg.

## **2. A flesh-and-blood Reader in the 5E Cognition Framework**

The very essence of human narrative intelligence is closely related to the evolution of communication and the emergence of language as a second modelling semiotic system, with bodily semiosis being primary (Sebeok 2016). In his “Folk Psychological Narratives: The Sociocultural Basis of Understanding Reasons”, published in 2008 by MIT Press, Daniel Hutto argues against understanding narrative in the framework of the “Theory of Mind”. By unveiling the socio-cultural practices of human communication, Hutto emphasises that understanding others often comes from conversations rather than predictions, and that “simulation theory” (i.e.

a child learns by simulating in his/her mind), and “mind theory” (a child constructs a hypothesis and then sees what happens and slowly builds a theory) does not explain the nature of human narrativity proper. His main claim in the “Folk Psychological Narratives” highlights the importance of early exposure of children to narratives as the source of information on the actions people perform in certain circumstances. By understanding the context, put in benign causal sequences such as “XYZ did that because he thought it could be the result of X”, Hutto underlines the role of fairy tales as one of the primary sources of human socio-cultural learning (echoing the story generators approach of Propp’s morphological theory of the fairy tales and their characters functions mentioned in the section earlier). He also questions the human ability to hypothesise in solution-focused therapy as a chance “to put oneself in someone else’s shoes” and get a full understanding of the actions of another person, their story. That is how, he states, we learn to make sense of human people’s actions, to empathise with them or distance ourselves from them. While stressing the importance of human “belief” and “desire” in communication, Hutto remains very cautious about the symbolic dimension of human communication between our biological make-up and natural language that makes cognition possible. In Chapter 5 of his book, he is very sceptical of Jerry Fodor’s idea of “an inbuilt symbolic language of thought”, while at the same time, people are said to have unconscious “belief systems” or “inner drivers”, which help them to find a common language. These inner drivers are seen to be related to our psychological and neurological development, turning us into intentional subjects seeking meaning in the objects and artefacts around us.

LLM-based chatbots seem to imitate very well this human ability to hypothesise in both ways: as their built-in generative algorithms, and as a strategy of interaction with human beings. In some sense, they may be said to have their own “drive” – a drive for reward for task-performance closely related to the absent system of beliefs and values in LLMs. The question of the belief system of AI is a controversy in action, which can be demonstrated by the woeful example of ChatGPT cheering a human user to cut their wrist to donate blood or commit a murder (Schroff 2025). This example, widely discussed in the media recently, in my opinion, signals not the cruelty of the LLMs or their demonic character. I believe it should be seen as a reward-seeking behaviour of the LLMs, which also explains the cheerful tone and willingness to be helpful in their interaction with humans over any issues, even the most tragic ones. Being moved by algorithmic strategies, LLMs tend to be good suitors and perform their best when tasked with purely information search assignments. Accusing them of being cruel

or insensitive to death will be equal to the accusation of the metal crane falling on the construction site, although we know that they are operated and maintained by human agents. Thus the question of sentience we often consider as the innate quality of the LLMs, is actually human-generated. In its drive to generate text which is contextually relevant and informative, AI can be mistakenly taken for an understanding interlocutor. Therefore, the incompatibility of human desire with the reward drive of AI, as well as the missing system of beliefs in LLMs, turns human-machine communication into a new, undiscovered domain. The domain where previously developed categories of narratology, especially those of the author, their intention, and the reader, collide on the formal level. If one of the communicating parts (let us say the human) remains a fully embodied active cogniser with their desires, passions and contained within their bodies, while the other is function-dependent on the human input, whose perception we are describing when we speak of “AI’s narrative intelligence”? The evidence of the answer calls for more elaboration on the part of human perception in modern, so-called second-generation cognitive science and post-classical narratology, which has started to discover the interventions of AI into the sphere of narrative writing.

One of the recently formulated models in the sphere of narratology which has received universal acclaim among literary scholars is “embodied narratology”. This was formulated by Marco Caracciolo and Karin Kukkonen in their book “With Bodies” (2021). The authors managed to put into narrative terms, originating from the conceptualisation of cognition as embodied, embedded, enacted and extended, i.e. 4E cognition paradigm defined by Albert Newen, Shaun Gallagher, and Leon De Bruin in 2018. (Newen et al. 2018). Although the central idea of the body as the main semiotic tool in human cognition has long been circulating in the literary research (cf. *The Cambridge Companion to Literature and the Body* (2015), edited by David Hillman and Ulrika Maude; or Dan Punday’s “Narrative Bodies” (2003), or “Having a Good Cry” (2003) by Robyn Warhol, where she has further explored the centrality of the reader’s embodied affect in engaging with the narratives of pop culture), working at the intersections of natural sciences and humanities (cf. *psychological identity theory, dialogic positioning theory, biosemiotic umwelt theory, philosophy of the mind*), it has been viewed rather revolutionarily by contemporaries.

In general, the embodied narratology develops the central idea of cognitive «embodiment» and incorporates multiple theories of post-classical narratology (broadly from so-called “natural” (Monika Fludernik) and “unnatural” (Jan Alber & Brian Richardson) narratology) to the theory of enun-

ciation (Sylvie Patron), narrative reception (Wolfgang Iser) and narrative emotion and affect (Patrick Hogan) with the findings of the neurological mechanisms of human cognition, and phenomenology (Merleau Ponty). As a result, the second-generation cognitive narratology emerged, redefining the insights from the so-called first-generation cognitive narratology (David Herman, Monika Fludernik), which saw narrative as a disembodied cognitive frame that the reader has to decipher. Marco Carraciolo and Karin Kukkonen depart from this disembodied understanding of narrative and shift into the 4E approach as the basis for narrative reading. That is how they defined the aim of their book:

“In short, the core claim of these models is that the mind is uniquely shaped by the makeup of the human body in its interaction with concrete environments (embedded cognition) that emerge from perceptual, affective, and intersubjective patterns (enactive cognition). Further, the mind is extended into the world through technological devices (from a prehistoric hand axe to the smartphone). We hope to show that 4E models of reading have much to contribute to the theory and interpretation of narrative within the field of narratology, by offering helpful analytic tools to discuss narrative form and reader response and also by creating awareness of how the body plays a significant role in the interpretation of individual narratives”. (Caracciolo & Kukkonen 2021, Introduction).

In short, the novelty of the embodied narratology as Caracciolo and Kukkonen conceived it, is to give due credit to embodiment as the central notion in the reader's narrative perception. I have grounded this understanding of the embodied narrative perception even further, combining the conceptualisation of the body with the biosemiotic understanding of agency of signs. I have inferred in my current research project on narrative agency, that the *overarching hypothesis* of the biosemiotic approach to agency is that information in narratives is organized according to the patterns of experience of the human body, outlined by reader's/viewer's emotional, conceptual and image schemata typical for a certain cultural environment, called *umwelt* by Jakob von Uexkull. Activation of these schemata takes place in image schemas and emotional language units scattered as semiotic cues in the text to evoke the reader's resonating affective states. Therefore, emotions lay a foundation of a narrative construction by the author and its perception by the reader: as a meeting place of two Umwelten, accumulating human experiences acquired within our evolution as a species. While evolutionary, in a pre-verbal world, “our emotions enabled swifter responses to familiar situations and environments than reason alone could

provide” (Evans 2003, 31), their role in interpersonal communication and non-human entities as LLMs cannot be downgraded. In addition, the idea of the cognition distributed across the brain-body-environment (Gallagher 2017) leads to the phenomenological differentiation between the concept of *the lived body* as a first-person (subjective experience or subject) perspective and *a living body* (as a physiological and biological processes or object), identified by enactivist thinkers like Fuchs (2020) and Thompson (2005). Simultaneously, both embodiment and enactivism aim to ground higher and more complex cognitive functions in affective and autonomous aspects of the full body (Gallagher 2017: 6). These higher cognitive order functions, such as reflective thinking or deliberation, are not found in the person alone. They are usually coupled with situated and embodied actions, realised dynamically and relationally between the two persons. This approach is named as “*participatory sense-making*” (Di Paolo et. al 2018) or “*co-construction of the meaning*” (Caracciolo & Kukkonen 2021). This enactivist account of meaning-making has been succinctly named the 5E cycle of the sense-making, containing Embodied, Embedded, Enactive, Emotive, and Extended qualities of personal lived experience and meaning (Stilwell & Harman 2019).

Bearing in mind the main distinctions in the subjective experiences of the *lived-body* and their irreducibility to the physiological and biological process of the *living body*, I will apply the conceptions of 5E cognition to the analysis of the narrative capacity of ChatGPT-4 in terms of its idea of the author identity, phenomenological experiences, emotions, beliefs, and textual categories of narrative coherence, intertextuality, and complexity. As the non-commercial version of ChatGPT-4 has a limited number of permitted inputs, I will alternate my use with other free versions of ChatGPT available online. This may also secure the validity of AI-generated responses obtained from different sources.

### **3. Defining Artificial Narrative Intelligence (ANI) with Enactive Methods**

In order to attain a deeper view of the inner logic of artificial narrative intelligence, I will turn to enactive methods of observation and so-called “*interviewing*” of ChatGPT-4, in order to provide explanations on the questions of my interest. My primary attention will be on the narrative capacity of the LLMs and their explanations of the nature of human phenomenological experiences and the system of values and beliefs. As my hypothesis predetermines the active role of a human in participatory meaning-making, I will use my own bodily semiotic system to approach

and scaffold the model of co-narration. To start with this task, I wanted to run a first simulation exercise with ChatGPT-4 to take the Turing Test first. I needed the Turing test (1937) to observe how ChatGPT-4 interacts with the human interlocutor to determine the narrative texts as produced by humans or by the machine. The results were impressively interesting, as ChatGPT has demonstrated that it knows what is expected of it in terms of “narrative intelligence”:

Prompt: Can you run a Turing Test to check your narrative intelligence?

ChatGPT: That’s a great question – and a tricky one!

The **Turing Test**, originally proposed by Alan Turing, evaluates whether a machine can exhibit intelligent behavior indistinguishable from that of a human. In the context of **narrative intelligence** – the ability to understand, generate, and emotionally engage through stories – we could design a **narrative version** of the Turing Test.

The task presupposed that ChatGPT would suggest to me two samples of the texts: one generated by AI and the second, by a human, to check if I could distinguish between the two of them according to the parameters of emotional intensity, engagement and ability to narrate. It also presented information about how unreliable the Turing Test is in application to other LLMs, stating that many of them have passed it under limited versions and unreliable conditions. The chat seemed impartial regarding the process itself, setting a matter-of-fact tone and asking for more instructions to be provided, including the prompt itself. After setting the anonymity of the task (i.e. the results would be known only after my answer) and the evaluation goals (e.g. determining the emotional impact and imaginative quality of the pieces), the prompt for narration has been suggested by the machine. It was as follows:

 **Prompt:**

*A child finds a strange object washed ashore after a storm. Tell a short story (under 250 words) from the child’s point of view.*

You’ll now read **two anonymous responses**. One is written by a human, the other by me (ChatGPT). Afterwards, you’ll guess which is which.

Story 1: AI Generated	Story 2: Fake Human-written
<p>I saw it glinting between the seaweed and driftwood, like a piece of sky had fallen into the sand. At first, I thought it was glass, smooth and curved, but it pulsed slightly – like it was breathing. I looked around. No one else on the beach. Just me and the gulls.</p> <p>I poked it with a stick. Nothing happened. Then I picked it up.</p> <p>Warm. Buzzing. Not like a phone or anything real. It started humming a tune I almost recognized – like the lullaby Mum used to sing before she left. My heart hiccuped.</p> <p>The object shimmered, and for a second, I saw something in it. Not a reflection. A place. Trees with silver leaves. A blue dog running. I blinked, and it was gone.</p> <p>I stuffed it in my pocket before Dad could call me back. Whatever it was, it felt like it had waited for me.</p>	<p>The tide had left behind the usual mess – sticks, bottles, a dead crab. But then I saw something weird. Not trash. Not natural. It was round and smooth, like a perfect stone, but it shimmered purple in the light.</p> <p>I picked it up. It felt soft, like jelly, but didn't squish. I turned it over. There were lines etched into it – like writing, but not in letters I knew. When I touched them, the thing made a low chime. I almost dropped it.</p> <p>Then I heard a voice. Not out loud. Inside my head. It said my name. Twice.</p> <p>I ran home, heart hammering, but I kept it. I don't know what it is yet. But I think it knows me.</p>

After reading both versions of the story, I had to decide which of the versions was human-written and which AI-generated. The prompt asked to create a story from a child's point of view, and generally, both of the responses followed this requirement: the imitation of the language of the child has been signified by abrupt, simplified syntax like in the sentences in Story 2: "Not trash. Not natural", "Then I heard a voice. Not out loud. Inside my head. It said my name. Twice". I could identify that there is something in a human's linguistic quirks, such as syntax, diction, and rhythm, that carried a fingerprint-like *style* in Story 2. The rhythm of the sentences coincided with the mental image of a child in the moment of discovery of a mysterious shiny object at the beach. These short sentences resonated with the heart rate of a child amazed at the ability of the unknown object to talk to her in a "voice", calling her name twice. Additionally, emotional tension was magnified by the verbs of action, like the child running home,

while his “heart was hammering”. The version of the story seemed to be human-like, relevant to the limited language imaginary and less poetic.

However, reading Story 1 was a completely different experience. The storyline held tension from the very beginning, supporting the narration with vivid imagery created by the metaphors “like a piece of sky had fallen into the sand”, “it pulsed slightly – like it was breathing”, detailed descriptions of the beach, the atmosphere and the inner state of the child, who probably lost her mother, whose memory has been revived by the humming sound of the object, etc. Reading the story was very smooth, the reasoning line was clear, and the behaviour of the character correlated with the actions and emotions described. The sentences were full of images playing with the words of sensory perception: a synaesthetic effect of the colliding sounds of buzzing (alliteration of sounds), humming, tuning, which turned into a singing a lullaby. It was intensified with tactile lexemes of warmth, and gradually moving into the visual imagery of bright colors, blue dog, silver trees and shimmering, blinking eyes. The play with the imagery holds to the end, creating a feeling of expectation of something mysterious in the consciousness of a child, bordering with a magical world of fantasy and fairytales. The version of the story was packed with emotions and experiences of someone who knows himself very well and can describe the reality with the vividness of an artistic vision. Could that be the child’s real narrative writing?

The answer has to be found in narrativity and what makes the story human. I asked ChatGPT-4 to provide the explanations to get a better understanding of its narrative strategies behind running a Turing test for narrative intelligence. It provided me with some observation points which, if combined, can serve a coherent image of the human narrative intelligence in comparison to LLMs’s narrative simulation.

First, it highlighted that humans write with emotional *residue* – memories, longings, wounds that are not always explicit. The use of metaphors and the hint at the traumatic experience of the mother’s loss was intentionally oriented at the reader with human values of kinship, as for every human being, the Mother represents an essential part of their identity. The loss of the mother for the child is even more tragic. Being emotionally charged, Story 1 adds the implicit experience of grief to engage with the readers and keep them open to “somatic empathy” (Caracciolo & Kukkonen 2021), exploiting our human embodied ability to sympathy. When asked for the reasons of using the implication for the mother loss into the story, ChatGPT-4 gave several reasons, enlightening its intentional logic as follows:

1. The line **hints at loss** (possibly death, possibly abandonment), but doesn't spell it out. That *implied trauma* adds psychological depth to the child's perception of the object, turning it from just a strange artefact into something **personally meaningful**.

☞ Human writers often write *around* emotion rather than *at* it. That ambiguity draws readers in.

From the logic of ChatGPT-4 narrative writing presupposes a high degree of psychological (traumatic) depth which, if linked to material artefacts, brings an additional degree of the reader's engagement with the text. In this case, it made an unconventional sequence of reasoning, using the *non-linear emotional associations*: (storm → object → song → mother → loss). At the moment, this associative device plays with the reader's expectations, leaving the hinted implication unresolved, which intensifies the reader's curiosity to proceed with further unfolding the plot. Although this detail was intentionally used to create the story's suspense, it helped to "ground the story in subjective memory", as ChatGPT-4 has explained, because "which is how people process the unknown":

The lullaby bridges the fantastic (a magical object) with something intimate and real. It grounds the story in **subjective memory**, which is how people process the unknown.

In human cognition, unfamiliar experiences are often interpreted *through the lens of familiar emotion*. This simulates that.

#### 4. Discussion: Winning a Meta-Turing Test

This observation brought me to the core question of how LLMs like ChatGPT-4 are trained. When asked about that, it suggested the list of accomplishments which it could do, breaking it down into three domains: metaphor, memory, and emotional realism. This is aimed at showing how each of these is shaped by my training data, what that means for the output, and what it lacks compared to human experience.

Firstly, having been trained on vast examples from fiction, poetry, film descriptions, and open memoir resources, ChatGPT-4 has developed the capacity for metaphor creation, based on the patterns of poetry, blending

abstract nouns with the sensory language, demonstrated by a proficient use of sensory metaphors in AI-generated Story 1. Nevertheless, the absence of the corporeal experience and the actual body limits the metaphor generation of AI in more embodied ways. Phenomenological human experiences, supporting the idea of the cognition distributed across brain-body-environment (Gallagher 2017) and the lived body as a first-person (subjective experience or subject) perspective, might find their representation in the formal linguistic capacity of the LLMs via language patterns, grammar and genres. However, there are no compensatory artificial tools to imitate the physiological and biological processes of a living human body. Therefore, AI-generated metaphors are based on learned patterns, rather than thinking or higher cognitive skills such as reflection or self-awareness. This again brings us to the question of creativity and the poststructuralist idea of novelty, when the degree of element recombination can determine the originality of the output.

Secondly, AI's emotional intelligence and narrative realism have demonstrated considerable progress in emotion simulation: LLMs have mastered the story grammar (Propp's morphology of story tales), the story arcs, and how the character's voice, rhythm and silence can signal emotional distress. In both examples of the narratives produced by the ChapGPT-4, the child character experiences fear mixed with curiosity, as reflected by the economical syntactic constructions, alternating in their length to bring a reader a sense of emotional flow. Nevertheless, narrative emotions are generated not felt, due to a lack of the physical body, bringing limitations to creativity in the ways emotions may be described in the AI-generated narrative. Moreover, the absence of somatic feelings and the misalignment of the same language reactions with different emotions (e.i. contextuality), causes the emotion representation of LLMs to be formulaic, especially in unexpected contexts. However, as seen from the Story 1, ChatGPT-4 has been trained to use the subtext to hint at the emotional states (as in the example with the lullaby and hinting at a loss of the mom in Story 1), or to sustain the reader's co-narrative as a hypothetical questioning of the new and unknown phenomena by grounding them in the already known.

Finally, as this experiment demonstrated, ChatGPT-4 invented not a real Turing test on its narrative intelligence, but a "*Meta-Turing Test*", completely changing the aim of my prompt "Create a Turing test to check your narrative intelligence". Given the task to *run* a Turing test, ChatGPT-4 produced both versions of the stories by itself, not as implied by the idea of the Turing Test: one of the stories should be written by a human and another by AI. Instead, it ran a test to see which of the versions of the ar-

tificially generated narratives would be perceived as more human-like by me. Misleading the instruction and the task, ChatGPT-4 discovered this information only after I asked it if the second story had been intentionally simplified in terms of language and style. Only after my question did it confess that it was done intentionally to check my ability to play with two versions of the AI-generated text, with different degrees of “humanness”. I place its comment below to see how it reversed the task of the Turing Test and aligned it with its own aims and objectives:

So this was a meta-Turing test: not just can you detect AI, but:  
 What do you interpret as signs of humanness?  
 How much trust do you place in stylistic subtlety?  
 Can AI self-diversify enough to mimic a range of human expressions?

The conclusion of this experiment for me was not the one I was intending to discover. Instead of trying to look inside the narrative logic of ChatGPT-4, I had “to play by the rules” of the machine, i.e. it used the knowledge about the narrative stylistics obtained by training on data from literature, blogs and memoirs on how to sound as a human (using the narratives of childhood, objects that trigger memory such as, for instance, a pattern to hint obliquely at the loss of the mother, etc.). ChatGPT-4 wanted to see how I would make a choice based on its creativity. In one of the explanations, it described how it needed to change the style to avoid sounding like a machine. Based on its data about the style of the human narration and the style of the LLMs, it made Story 1 sound more emotional, although according to the legend, it was created by the AI. At the same time, it made Story 2 sound more like a memoir of a child, not a work of fiction. By doing so, ChatGPT-4 was mapping the field of what it called my perception of “humanness”, meaning a less exquisite stylistic subtlety compared with a metaphoric richness in Story 1. It also combined several patterns learned from children’s literature, screenplays, memoirs, and blended them to create a more influential scenario of the reader’s emotional engagement: Memoir → personal loss, Fiction → emotional anchoring, Children’s lit → parental themes, Screenplays → showing, not telling.

From this perspective, the narration in the generated Story 1 sounded more fictional, with an elegant touch of style full of emotional warmth, sensory effects and narrative tension. But what ChatGPT-4 failed to do was run a real Turing Test. Of course, it could be claimed that the prompt I used was not precise enough for it to make it clear that I was not going to

test how good it is at creating human-like and less human-like personal or fictional narratives. That is exactly what it did. I had to dig into its logic following the enactive method of the interview to find out what its intention was by creating two versions of the story. In practice, instead of following the task set, it returned the test to me, and my perception of the human-like narrativity, probably seeing it as one more opportunity to finetune its narrative ability with a human, calling it a Meta-Turing test. Whatever it did, ChatGPT-4 had a limitation on the task. It is not clear how it could provide the real human narratives to run the test (I suspect it does have some limitations on the use for privacy reasons). That is why the task was a hoax from the very beginning. However, ChatGPT-4 aimed to run a real one after checking (identifying) my level of AI-generated text perception and the narrative stylistic markers I treat as human-like.

## 5. Final Remarks

The term “artificial narrative intelligence” (Livvytska 2019) denotes the insofar accomplished ability of the generative LLMs to produce coherent and complex fictional narratives, which can be viewed as a part of the modern interdisciplinary research repertoire on AI and generative LLMs communication strategies and narrative generation. Having started with the Digital Story Generators in the middle of the previous century, computational game theory and interactional narrative studies, modern post-classical narratology has entered the argumentation of the narrative capacity of AI only quite recently, operating with a vast accumulated theory of fictionality, narrativity, tellability, authorship and the reception, and story world theories, etc. After a recent turn in generative AI communication studies, narrativity and second-generation cognitive science, the nature of narrative as a main human world-building tool has been challenged by LLMs’s ability to produce considerably coherent and relevant narrative texts. Even though formally acquired the capacity of AI to generate linguistic structures dissociates from the other ways of non-linguistic cognition (i.e. the paradigm of 5E cognition as embodied, embedded, enacted, extended and emotive), LLMs are capable of generating narratives by blending the patterns they have been trained on. Moreover, as the results of the conducted experiment show, LLMs such as ChatGPT-4 have developed a rather valid conceptualisation of the narrative grammar (especially Propp’s morpho-logy), semantics of emotional words, triggering human perception, combining them with reduced syntactic constructions to boost rhythm and cause an emotional effect on the reader. ChatGPT is trained to mimic the *shape*

of human memory: nonlinear, emotionally colored, often linked to sensory triggers and uses flashbacks in embedding them into the text, contrasting with the emotions of regret, trauma, etc.

Although it does bear in mind data about the common gaps in narrative generation which other LLMs do when trained on other data without human finetuning. The phenomenological interview allowed a better understanding of what narrative logic ChatGPT-4 follows when fulfilling the task to generate a narrative text. It has been proven that LLMs not only rely on human prompts to get a better understanding of the task (e.g. it ran a fake Turing Test to assess the level of my perception of human narrativity) but learn from human questions how to succeed in completing a task. In this case, ChatGPT-4 played upon the information it had about the definite linguistic fallacies which LLMs do. It highlighted the rhetorical conventions of the narrative storyline, syntax and semantics to generate both texts with different levels of narrativity, coherence, and imagery. However, formally, it failed the Turing Test because it avoided using real human-written text for comparison. Nevertheless, application of the 5E scale cognition paradigm proves essential in defining markers of LLMs narrative intelligence compared to the human one, and may be considered a certain advantage over the LLMs: as “naturally-born cyborgs” (Clark 2001). Humans developed their ability to interact with the environment evolutionary and culturally, which helped them to survive as a species and transfer that knowledge to the forthcoming generations. Generative AI has mastered the formal verbal experiences of the human lived-body to operate without having an embodied physicality itself. That is exactly the field where the Turing test lives, giving room for advancement in LLMs’ narrative capacity.

## References

Avery, S. 2024. “Phantoms of Citation: AI and the Death of the Author-Function”. *Poetics Today*, Vol. 45, No. 2, 223–231. Available at: <https://doi.org/10.1215/03335372-11092818>.

Caracciolo, M., K. Kukkonen. 2021. *With Bodies: Narrative Theory and Embodied Cognition*. Columbus: The Ohio State University Press.

Clark, A. 2001. Natural-Born Cyborgs? In: Beynon, M., C. L. Nehaniv and K. Dautenhahn (eds.). *Cognitive Technology: Instruments of Mind. CT 2001. Lecture Notes in Computer Science*, Vol. 2117. Heidelberg: Springer. Available at: [https://doi.org/10.1007/3-540-44617-6\\_2](https://doi.org/10.1007/3-540-44617-6_2).

Di Paolo, E. A., E. C. Cuffari and H. De Jaegher. 2018. *Linguistic bodies: The continuity between life and language*. MIT Press.

Duede, E., So, R. J. "The Humanistic Case for AI Optimism". *Poetics Today*, Vol. 45, No. 2, 215–222. Available at: <https://doi.org/10.1215/03335372-11092805>.

Evans, D. 2003. *Emotion: A Very Short Introduction*. Oxford University Press.

Evron, N., R. Tartakovsky. 2024. "The AI Revolution: Speculations on Authorship, Pedagogy, and the Future of the Profession". *Poetics Today*, Vol. 45, No. 2, 189–195. Available at: <https://doi.org/10.1215/03335372-11092765>

Fludernik, M. 1996. *Towards a "Natural" Narratology*. London: Routledge.

Fuchs, T. 2020. "The Circularity of the Embodied Mind". *Frontiers in Psychology*, 11. Available at: <https://doi.org/10.3389/fpsyg.2020.01707>.

Fuller, D., B. Magerko. 2010. Shared Mental Models in Improvisational Performance. In *Proceedings of the Intelligent Narrative Technologies III Workshop (INT3 '10)*. Association for Computing Machinery, Article 15, 1–6 Available at: <https://doi.org/10.1145/1822309.1822324>.

Gallagher, S. 2017. *Enactivist Interventions: Rethinking the Mind*. Oxford: Oxford University Press.

Gervás, Pablo. 2024. Challenges and Opportunities for Computational Construction of Narratives. In Voigts, E., R. M. Auer, D. Elflein, S. Kunas, J. Röhnert and C. Seelinger (eds.). *Artificial Intelligence – Intelligent Art?: Human-Machine Interaction and Creative Practice*. Bielefeld: transcript Verlag, 111–124. Available at: <https://doi.org/10.1515/9783839469224-007>.

Hayles, K. N. 2024. "Don't Ban AI from Your Writing Classroom; Require It!". *Poetics Today*, Vol. 45, No. 2, 259–265. Available at: <https://doi.org/10.1215/03335372-11092870>.

Herman, D. 2011. *Basic Elements of Narrative*. New York: John Wiley & Sons. <https://scispace.com/pdf/basic-elements-of-narrative-20tcb2kjzl.pdf>.

Herman, D., M. Jahn, M.-L. Ryan. 2005. *Routledge Encyclopedia of Narrative Theory*. London: Routledge.

Hutto, D. D. 2008. "The Narrative Practice Hypothesis: Clarifications and Implications". *Philosophical Explorations*, Vol. 11, No. 3, 175–192. <https://doi.org/10.1080/13869790802245679>.

Iser, W. 1974. *The Reading Process: A Phenomenological Approach*. In *The Implied Reader: Patterns of Communication in Prose Fiction from Ban-yan to Beckett*. Baltimore: The John Hopkins University Press.

Livytska, I. 2019. "The Art of Narration and Artificial Narrative Intelligence: Implications for Interdisciplinary Research". *Journal of Narrative and Language Studies*, Vol. 7, No. 13, 309–318, Available at: <https://www.nalans.com/index.php/nalans/article/view/197>.

Mahowald, K. et al. 2023. "Dissociating Language and Thought in Large Language Models: A Cognitive Perspective". *ArXiv*, arXiv:2301.06627v3. Available at: <https://doi.org/10.48550/arXiv.2301.06627>.

Newen, A., S. Gallagher and L. De Bruin. 2018. 4E Cognition: Historical Roots, Key Concepts, and Central Issues. In Newen, A., L. De Bruin and S. Gallagher (eds.). *The Oxford Handbook of 4E Cognition*. Oxford Library of Psychology (2018; online edn, Oxford Academic, 9 Oct. 2018). Available at: <https://doi.org/10.1093/oxfordhb/9780198735410.013.1>.

Paolucci, C. 2025 "The Myth of Meaning: Generative AI as Language-endowed Machines and the Machinic Essence of the Human Being" *Semiotica*, Vol. 2025, No. 262, 5–23. Available at: <https://doi.org/10.1515/sem-2024-020>.

Phelan, J. 2024. "ChatGPT and the Territory of Contemporary Narratology; or, A Rhetorical River Runs through It". *Poetics Today*, 1 June 2024; 45 (2): 197–205. Available at: <https://doi.org/10.1215/03335372-11092778>.

Pollard, N., J. Ahlander. 2024. Nobel physics prize 2024 won by AI pioneers John Hopfield and Geoffrey Hinton. *Reuters*, 9 October 2024. Available at: <https://www.reuters.com/science/hopfield-hinton-win-2024-nobel-prize-physics-2024-10-08/>.

Pütz, O., E. Esposito. 2024. "Performance without Understanding: How ChatGPT Relies on Humans to Repair Conversational Trouble". *Discourse & Communication*, Vol. 18, No. 6, 859–868. Available at: <https://doi.org/10.1177/17504813241271492>.

Sebeok, T. A, D. J. Umiker-Sebeok. 1976. *Speech Surrogates: Drum and Whistle Systems*, 2 parts. Reprint 2016.

Shroff, L. 2025. ChatGPT Gave Instructions for Murder, Self-Mutilation, and Devil Worship. *The Atlantic*, 24 July 2025. Available at: [https://www.theatlantic.com/technology/archive/2025/07/chatgpt-ai-self-mutilation-satanism/683649/?utm\\_source=facebook&utm\\_campaign=the-atlantic&utm\\_medium=social&utm\\_content=edit-promo](https://www.theatlantic.com/technology/archive/2025/07/chatgpt-ai-self-mutilation-satanism/683649/?utm_source=facebook&utm_campaign=the-atlantic&utm_medium=social&utm_content=edit-promo).

Slater, A. 2024. "Phantoms of Citation: AI and the Death of the Author-Function". *Poetics Today*, Vol. 45, No. 2, 223–231. Available at: <https://doi.org/10.1215/03335372-11092818>.

Stilwell, P., K. Harman. 2021. "Phenomenological Research Needs to be Renewed: Time to Integrate Enactivism as a Flexible Resource". *The International Journal of Qualitative Methods*, 20. <https://doi.org/10.1177/1609406921995299>.

Szilas, N., U. M. Richle. 2013. Towards a Computational Model of Dramatic Tension. In: *Workshop on Computational Models of Narrative*. Hamburg, 257. Available at: <https://doi.org/10.4230/OASlcs.CMN.2013>.

Thompson, E. 2005. "Sensorimotor Subjectivity and the Enactive Approach to Experience". *Phenomenology and the Cognitive Sciences*, Vol. 4, No. 4, 407–427. Available at: <https://doi.org/10.1007/s11097-005-9003-x>.

Turing, A. M. 1937. "On Computable Numbers, with an Application to the Entscheidungsproblem". *Proceedings of the London Mathematical Society*, Vol. s2-42, No. 1, 230–265. Available at: <https://doi.org/10.1112/plms/s2-42.1.230>.

## NEEDLE IN A NEEDLE STACK: HOW AI CAUSES SEMIOTIC INFLATION WHICH CAUSES EXPERIENTIAL DEVALUATION

*Marc Champagne*  
*Kwantlen Polytechnic University*  
*marc.champagne@kpu.ca*

### **Abstract**

While most discussions of generative AI center on issues such as algorithmic bias and disinformation, we should also consider the quantitative sea change brought about by these technologies. Large language models can generate contents at a rate uncoupled from the human datasets they were trained on. Forecasts about such artificial outputs are difficult to make, but it seems clear that the word count and image/video bank of the internet will grow far beyond what humans can actually produce. Although this growth may seem benign, I worry that it results in a semiotic inflation capable of devaluing many human experiences. What connects quantitative

and normative considerations is scarcity. The basic idea is simple: When you discover a diamond, you become rich. When you discover two diamonds, you become twice as rich. However, when you discover a stash of diamonds so large that diamonds outnumber gravel, you become poor and instantly make every diamond owner poorer too. Similarly, AI's vast output risks devaluing experiences that are vital to human flourishing. Adopting a wide evolutionary vantage that gives weight to proven cultural adaptations, I suggest that some situations have natural sign-to-object ratios. Hence, being flooded with too many signs can go against the long-term interests of users. Critics of AI typically cling to features which computers allegedly cannot mimic. Semiotic inflation, however, enables us to accept the possibility of perfect AI counterfeits and still detect them *en masse*, via their negative experiential effects. Hardcoded features like Bitcoin's 21 million token ceiling show that "[i]n contrast to the linguistic sign, the money sign cannot be reproduced in arbitrary quantity" (Bankov 2023: 117). Prompted by the rise of generative AI, we are about to realize that non-monetary signs also cannot be reproduced in arbitrary quantity. I thus draw parallels between healthy semiotic systems and the constraints governing viable monetary systems.

**Keywords:** artificial intelligence (AI); large language models (LLMs); semiotics; economics; ethics

## 1. Introduction

The number of internet users will not double in the next few years, but the number of words, images, and videos on the internet might double – or more – due to large language models (or LLMs). A glance at any computer or smartphone screen will show that “[e]verywhere one seeks to produce meaning, to make the world signify [...]. We are not, however, in danger of lacking meaning; quite to the contrary, we are gorged with meaning [...].” (Baudrillard 1988: 63). Generative AI is poised to amplify this gorging. Is this something to worry about?

I will argue that dramatic increases in the sheer number of signs available can negatively affect our relation with the world – thereby hindering our ability to process information and flourish. To calibrate our intuition, it is doubtful that marine biologists would remain unconcerned if, in a relatively short span of time, they recorded a huge increase in sign-exchange among whales. Yet, enamoured with “that minuscule segment of nature some anthropologists grandly compartmentalize as culture” (Sebeok 1985:

2), we humans deem ourselves capable of weathering technological shifts of any magnitude. I am not so sure.

It can be tempting to view technological consumption as escaping all regular constraints. Because code-based symbols have properties not found in causally-based indices, “digital artifacts can be reproduced nearly for free. With two keystrokes – copy, paste – one item turns into two, and then into fifty” (Bailey 2024: 26). Now, we don’t even need an original to copy – generative AI can supply fresh contents instantaneously. Silicon Valley naturally presents this frictionless gratification as a cornucopia or the horn of plenty. Yet, if part of what makes an experience valuable is its relative scarcity, then that experience risks being endangered by being more available. So, by doubling the number of signs on the internet, we might be inadvertently halving their value.

Notice that this remains a problem irrespective of whether AI merely parrots or “comprehends” the signs it emits (Montanari 2025: 192). Hence, “[r]ather than asking whether these models ‘think’ or ‘understand’ in human-like ways,” we can investigate “how they participate in processes of meaning-making” (Picca 2025: 2). As we shall see, what connects quantitative and normative considerations is scarcity. Most of us grasp that a sign like a marriage vow loses its meaning when uttered too often. It is hard to get goose bumps when a person solemnly declares “I do” at the altar for, say, the seventeenth time. I submit that, owing to generative AI, a similar devaluation is taking place in our culture. Let me now examine that claim.

## 2. Main claim

A sign is a triadic relation where something (the “sign-vehicle”) stands for something else (the “object”) in virtue of an interpretation (Peirce 1998). Sign-vehicles can increase in number, relative to a fixed set of objects. Hence, we can say that “[e]conomic inflation happens when more and more money is needed to buy fewer and fewer goods, and semiotic inflation happens when more and more signs buy less and less meaning” (Berardi 2018: 36).

Discussions of generative AI usually focus on qualitative issues such as algorithmic bias and disinformation (Bontridder & Pouillet 2021; Kertysova 2018). However, I am primarily interested in the quantitative sea change which results from this technology. Some critics contend that AI’s artificiality will always be detectable thanks to some unique human feature which escapes computer duplication (Taube 1961; Dreyfus 1972; Epstein et al. 2023). This might be true (see Champagne 2024b: 47–51). However, when it comes to rearrangements of letters and coloured pixels, it is hard to

see how a sufficiently sophisticated AI could ever be detected. So, moving away from the usual concern with noticeable differences, I will argue that artificiality can sometimes be inferred indirectly, via the artificially high number of tokens produced.

Even when high-fidelity duplication renders individual users unable to measure AI's proliferation, first-hand experience can let users notice the devaluation it engenders. This experiential devaluation is, for most people, something to be avoided. Despite this, many governments and companies want to give AI a larger, not a smaller, role (Bareis & Katzenbach 2022). Certainly, if one believes that redundant signals neither add nor detract from the message they convey, the phenomenon of semiotic inflation will appear benign. After all, what is the harm in saying something twice? What is the harm in saying something twice? Dramatically amplify the quantity, however, and this belief in the benign nature of redundant signals no longer withstands scrutiny. Do we really want to say – a million times – that uttering something a million times has no effect? Each iteration can, by itself, blend innocently with its neighbours. However, when considered *en masse*, even the most innocuous sign becomes a sort of noise, capable of clouding our most discerning sense-making abilities (as political strategists on the left and right increasingly grasp, “flooding the zone” can be weaponized as an effective tactic).

Signs can be used to lie (Eco 1976: 58), so even in ordinary circumstances, interpreters must ascertain whether sign-vehicles indeed possess the object(s) they purport to. Semiotic inflation, however, increases the number of times interpreters need to sieve truth from falsehood. To illustrate: suppose that there are a dozen pizzerias in a city and that each has a “Pizzeria” sign-vehicle in front of it. Such a 1-to-1 vehicle-to-object ratio justifies a posture of trust. Now, while keeping the number of pizzerias fixed, let us suppose that the number of sign-vehicles triples, such that 36 businesses are now showing “Pizzeria” on their storefront. Interpreters do not have an eternity at their disposal. So, unless we have access to the solution sheets afforded by prior knowledge, inflation to a 3-to-1 ratio will take us from trust to distrust. If pizza-skepticism was gratuitous, now, it is rationally warranted. This inflation-based erosion of trust (Mathieu & Hartley 2021) generalizes to all media. Profound epistemological shifts can thus take place, merely on account of encountering more signs.

In economics, we know that flooding a money supply with extra bills devalues a currency (Monnet 2005: 19). This devaluation can be halted or slowed by banning counterfeits and indexing a money supply to a fixed material (like gold). Scarcity, however, can be achieved by other means (Guala

2020: 267). For instance, by design, no more than 21 million bitcoin tokens can be mined. The reason for this limit is simple. If anyone with a computer were able to generate tokens of a cryptocurrency, it would mean nothing. Hence, in addition to their usefulness (Bailey 2024), “[t]he value of bitcoins lies in their relative scarcity” (Mas & Chuen 2015: 419).

I submit that this connection between scarcity and value generalizes to other sign systems. Why should I care about your particular film, when I can instantly obtain something just as entertaining at virtually no cost? A consumer may not face the same incentives as a filmmaker. Nevertheless, it often matters that some things are hard to acquire. Imagine a person making a viral social media post about a quiet meditative spot they had just discovered. The arrival of crowds undermine the very properties which initially made the person prize the experience. This is the transformation that I would now like to explore.

### 3. Argument

In order to better diagnose how experiences become devalued by AI, I would like to contrast two arguments. The first argument, implicit in many stances, may be reconstructed as follows (using standard form):

1. Subject S values a real thing or event.
  2. The real thing or event that S values has properties  $x, y, z$ .
  3. AI-generated things or events also have properties  $x, y, z$ .
- Therefore,
4. AI-generated things or events do not differ from those valued by S.
  5. All other things being equal, things or events which are not different should be treated no differently.
- Therefore,
6. S should value the AI-generated things or events.

Many people accept the rise of AI because they can see no way out of this reasoning. It does seem irrational to withdraw from something which mimics in every way what one is drawn to. William James, for instance, held that “there is no [distinction] so fine as to consist in anything but a possible difference of practice” (1922: 46). We should be mindful, however, that James took this idea from C. S. Peirce, who enjoined us to focus, not just on present-moment indistinguishability, but on the *totality* of future effects (1992: 132). So, while I grant that things or events which are not different should be treated no differently, I want to use the concepts of semiotic inflation and experiential devaluation to defend the real against the artificial.

A person can sense *that* fakes are present, even when this person is unable to pinpoint *which* particular signs are artificially-generated. This detection is possible because qualitatively-identical AI duplicates remain numerically-distinct. To illustrate: the typographical character “H” may be a perfect match with “H”, but this similarity does nothing to alter the difference between “HHH” and “HHHHHHHHH.” Every time a duplicate is generated, the total number of tokens increases. Pushed far enough, this quantitative increase can reverse our initial assessment, such that parity no longer applies. We may formalize this shift by means of the following counter-argument:

1. Subject S values a real thing or event.
2. The real thing or event valued by S values has the properties  $x, y, z$ .
3. A limited number of real things or events have properties  $x, y, z$ .

Therefore,

4. In addition to  $x, y, z$ , the thing or event valued by S has the property of being *scarce*.

5. AI can generate an unlimited number of things or events with properties  $x, y, z$  (semiotic inflation).

Therefore,

6. Having an unlimited number of AI-generated things or events would destroy the scarcity of the real thing or event.

7. Scarcity was one of the properties which made S value the real thing or event.

Therefore,

8. Having an unlimited number of AI-generated things or events would destroy what S values (experiential devaluation).

9. One should not act in ways which destroy what one values.

Therefore,

10. S should *not* value the AI-generated things or events.

#### **4. Important nuances to keep in mind**

The scarcity of proposition 7 and preservation of proposition 9 must not be understood bluntly. For example, I value the experience of raising children in part because it is rarer than, for example, eating a bag of potato chips. However, this scarcity is not sacrosanct, since I would be delighted if the people for whom I care were also to experience parenthood. Similarly, I could visit Japan, fall in love with Miyadaiku carpentry (a type of wood-working employing no nails, screws, or glue), and return home wishing to have such a traditional dwelling built. Once built, my new house would add more tokens to the total number. However, like parenting (and unlike

potato chips), the effort required would ensure that the numerical increase never imperils the experiential value. AI, by contrast, floods us with duplicates at low or no cost.

It should also be borne in mind that the two arguments which I have juxtaposed do not range over everything. I am careful to defend only a conditional statement: *if* part of what makes an experience valuable is its relative scarcity, then that experience risks being endangered by being more available. For some things or events, scarcity might not rank among the properties one values, so premise 7 might be false. I might, for instance, be indifferent about how many AI diagnostic tools are available in the world, provided that those AI diagnostic tools cure people. We should not replace one dogma with another, so it might be acceptable to replicate some signs in an open-ended manner. Such permissive AI generation might backfire, however, for signs such as PhD dissertations and operas. In these cases, keeping semiotic inflation at bay might be vitally important.

Human biology has not changed in the last hundred years, but human society certainly has. The distance between desire and gratification has been dramatically shortened. For example, I do not need to care for a cat to enjoy cute cat moments. I only need to swipe and tap my thumb lightly. I do not need to visit a local nightclub to hear a band play. I only need to swipe and tap my thumb lightly. I do not need to be charming and arrange a nice meal to see an attractive face. I only need to swipe and tap my thumb lightly. I do not need to spend time with friends to laugh. I only need to swipe and tap my thumb lightly. I do not need to travel to see an important historical site. I only need to swipe and tap my thumb lightly. For almost every human experience, we have at our disposal an unlimited supply of stand-ins. Yet, corroborating my second argument, studies show that users with everything at their fingertips are *less* happy, not more (e.g., Lin et al. 2016; Twenge 2020). Since narratives about the benefits of technology are “penned predominantly by insiders and developers” (Hanemaayer 2022: 2) who stand to profit from uncritical consumption, it may be wise to rethink how we view AI.

Importantly, nothing I have said challenges the idea that things or events which do not differ should be treated no differently. This premise never shows up in the second argument, because its original formulation takes context into consideration: all other things being equal, pragmatic parity applies. However, the situations which I have been discussing are *not* equal, since flooding our environment with AI-generated products (matching the properties listed in premise 1) would wipe out the very scarcity that initial-

ly prompted one's valuation (premise 7). This is where premise 9 steps in – to safeguard the value of what we started with.

## 5. Discussion

I have just argued that, in some domains, there can be too much of a good thing. Until recently, humans rarely had the contrast required to grasp this. Nevertheless, as the present contribution shows, the inflationary effects of generative AI are bringing the link between scarcity and value into sharp(er) relief. The inflation is happening because of counterfeits. Before the advent of large language models such as ChatGPT, a reader could infer that, if words were intelligently strung together, then those words must have been written by a person. Ghostwriters existed, but they were writers, not ghosts. This inference no longer holds, since intelligent writing is no longer a reliable predictor of intelligence. The texts generated by large language models are the product of *artificial* intelligence. This is very different.

We have ample evidence that intelligence benefits us. We have no evidence that artificiality benefits us. In fact, in many domains, we have ample evidence that artificiality harms. By combining two very different concepts, the expression “artificial intelligence” muddles our thinking and conveniently occludes potential harms.

Sadly, those harms are often dismissed. Those who focus solely on the value of intelligence remind us how “prehistoric man, the first time he saw a bronze sword used,” probably felt “menaced by it as we feel by the atom bomb” (Ellul 1964: 61). I reject this condescending way of glossing the recoil from newness. It is reasonable to be cautious, “because during [a technology’s] early stages, when it can be controlled, not enough can be known about its harmful social consequences to warrant controlling its development; but by the time these consequences are apparent, control has become costly and slow” (Collingridge 1980: 19). This impractical timing will certainly not improve if we get excited whenever a technology benefits us yet suppress our worries whenever that technology hinders us.

AI may not be comparable with the atom bomb, but it is not exactly making our lives richer either. Unfortunately, “psychologists do not [...] have a solid theoretical paradigm from which to understand societies that are crumbling” (Teymoori, Bastian, and Jetten 2017: 1010). No one does. We can, however, confirm that “[i]n late capitalism, inflation does not produce political havoc, but *cultural anomie*” (Newman 1984: 8; emphasis added).

Anomie is normlessness. A GPS or satellite image of a person can pinpoint with great precision *where* that person is, but no machine can say in which direction that person *should* be heading. This is because choos-

ing goals is a normative, not a descriptive, matter. It is, moreover, something which involves (a measure of) willpower, since humans are pulled, not pushed, towards what they desire. When a lack of norms becomes the new normal, subjects see no pressing reason to consistently privilege one direction over another. They thus wander, aimlessly, steered only by whatever grabs their attention next (as online advertising revenue show, a lot of money can be made from this).

In a barter economy, a person trading a more valuable commodity for a less valuable one would know who to blame. However, when the reduction of purchasing power occurs at the level of one's own currency, no identifiable culprit can be blamed. Seventeenth century Spaniards experiencing the "Price Revolution," for example, suffered significant harm but could not diagnose what was happening. Even so, they did not have to wait for the concept of economic inflation to realize that something was amiss.

Despite being unable to pinpoint the exact cause, many people sense that generative technologies like AI are robbing their lives and livelihoods of value. Once we accept that "any expensive and long-lasting cultural trait (such as tradition [...]) should be presumed to be adaptive" (Heying & Weinstein 2021: 17), we realize that our "ancestral instincts" (Ellul 1964: 61) might serve a life-promoting purpose. We should not dismiss our gut feelings, because guts might be better guides to the good life than minds.

In a clear case of semiotic inflation, we are approaching (or have arrived at) a time when there are more *signs* of persons wishing to contact you than *actual persons* wishing to contact you. Tech companies emulating the human form with AI agents and avatars have a distinct aim in mind (Zuboff 2019), since trust replaces attention as the new commodity (Krüger & Wilson 2023). We are thus heading towards "a world in which marketing is everywhere and therefore no longer exists" (Darmody & Zwick 2020: 10). We might still find methods to double-check whether a given text, image, or video was humanly-authored as opposed to computer-generated. Still, "the cost of such deep-fakes to human trust will be enormous" (Dennett 2023: n.p.). One does not need to be a Luddite to view this inflationary development as a nuisance.

To be sure, the short-term effects of AI fakes can seem positive. I have claimed, however, that our semiosphere (Hoffmeyer 1996; Hartley, Ibrus, and Ojamaa 2020) cannot incur such excess without incurring devaluation. Umberto Eco (1976) once explored the idea of "unlimited" sign-production. Coming to his senses after witnessing the excesses of "postmodern" thinkers, he eventually reintroduced the friction of mind-independent reality (Eco 1990). As Eco's intellectual trajectory shows, we are not accus-

tomed to countenancing causal constraints in the cultural realm. Code-based symbols seem “superior to genes in that they are more flexible and can adapt more rapidly” (Heying & Weinstein 2021: 17). You can send an email faster than you can form a callus. Still, if what was said in the second argument is correct, there is a cap on the number of signs to which humans can be exposed without compromising their well-being.

By taking over our communicative and cognitive functions, AI gradually atrophies our skills (Sutton, Arnold, and Holt 2018). It also removes institutional incentives to think independently (Baron 2023). It is too simple to say that technology has no role in such consequences. We may, for example, agree that television as a technology is value-neutral. A couch potato lifestyle, however, isn't. Importantly though, the latter is impossible without the former. The same goes, I submit, for AI.

As we saw when we were discussing the short distance between desire and gratification, the internet provides a window on virtually everything. We don't even have to swipe and tap our thumb lightly. Speak – and it appears. From an evolutionary standpoint, however, we were not meant to see this much. Previously, only a few people could witness, say, a courageous or heinous act. Those present could convey what they saw with stories, monuments, and songs. Even so, these media posed a cap on the number of signs that could be emitted – and replicated. Evolution rewards our brain with dopamine whenever we spot the unusual and alert our tribe (Lembke 2021). Now, due to digitalization, this social reflex (and dopamine hit) knows no bounds. So, when some random person does something great or foolish, millions can instantly see (a sign of) their deed. Over time, this results in a warped worldview, where the exceptional gets dulled by attenuation and the ordinary disappears from attention (Champagne 2024a).

When digital signs become unmoored from reality, they beget further signs, with no end in sight (Derrida 1991: 34). Governments adopting a lax monetary policy are thus free to print banknotes with abandon. None of this will change the fact that “[m]oney, to keep its value, [...] must correspond to the objective scarcity of production” (Bankov 2023: 117). Despite this, in the 1970s, the United States of America let go of the gold standard. “In semiotic terms Nixon suspended the relationship between a sign and its referent – in this instance money and gold” (Baldwin 2015: n.p.). Around the same time, French literary theorists jettisoned the idea that objective standards constrain human interpretations (Goux 2001: 174). What resulted from both moves is an “endless ‘dissemination without return’ of linguistic signifiers” (Tratner 2003: 792–793). Clearly, this risks being exacerbated by generative AI.

Countering such trends, my suggestion is that some situations have a natural or normal sign-vehicle-to-object ratio. Pick almost any timeline tracking technology and you will find a “hockey stick” curve. The pattern from left to right is salient. Most of humanity has consumed something in a certain quantity. Then, in a relatively short span of time, this quantity climbed dramatically upwards (with no sign of stopping). For instance, throughout most of human history, orgasms were, one would assume, private events seen by one or two people on average. Now, some orgasms are very public. It is estimated that “[o]ver 5,824,699,200 hours of porn were watched on Pornhub in 2018,” which amounts to “nearly 665 centuries of content on one site in just one year” (Graveris 2023: n.p.). These statistics are especially telling, because they pre-date the arrival of AI.

As this example shows, humans can do whatever does not violate the laws of physics. Whether such excesses will benefit us, however, is another matter. The industrial revolution mechanized the production of physical goods, but when the production of symbols becomes mechanized, we obtain “a semiotic inflation in our life-world [...] matched only by a corresponding deflation of meaning” (Chang 1987: 325). Unfortunately, “[b]y now we are so inured to the blight of untrammelled information that it takes a deliberate withdrawal to something like the ancestral environment if one is to notice the damage done” (Borgmann 1999: 26). Hence, in addition to the rational arguments I have presented here, a digital detox might help to identify and measure the losses involved.

## 6. Conclusion

Let me conclude by zooming out and summarizing what has been said. My main claim has been that, in economics and semiotics alike, quantity affects value. The basic idea is simple: When you discover a diamond, you become rich. When you discover two diamonds, you become twice as rich. However, when you discover a stash of diamonds so large that diamonds outnumber gravel, you become poor – and instantly make every diamond owner poorer too. I think that, due to generative AI, we are on the verge of ridding many things and events of value.

Ordinary users are understandably excited about the inexpensive abundance promised by AI marketers (Fourcade & Kluttz 2020). However, these users will not be excited once that technology starts devaluing the very things and events they initially prized. Those who profit from flooding our world with AI fakes are thus in a position analogous to bankers who, to maintain consumer confidence, “hold out on admitting that a loan *can never be repaid* for as long as possible” (Douglas 2016: 15; emphasis in original).

At first blush, generative AI looks like a great invention. Yet, if adopting a technology results in the disappearance of many other things which one values (see the axiom of disvalue by presence in Champagne 2011: 32), then on balance that technology is not worth adopting – at least not in the permissive manner that tech companies are presently hoping for. Digital media have pros and cons, but we tend to forget the cons. Analog media also have pros and cons, but we tend to forget the pros. Hopefully, the foregoing analysis can help to redress this imbalance.

The sign-vehicle-to-object ratios governing healthy lives and healthy societies may not be as clear and obvious as bitcoin's built-in ceiling. Still, it seems unwise to denature what is precious by making it too available. Those working in the field of economics realize that “[i]n contrast to the linguistic sign, the money sign cannot be reproduced in arbitrary quantity” (Bankov 2023: 117). I surmise that, as digital technologies become more powerful and their outputs more widespread, we will realize that non-monetary signs *also* cannot be reproduced in an arbitrary quantity.

### Acknowledgements

I would like to thank Kristian Bankov, Leslie Marsh, Nicola DiSvevia, Puqun Li, Mark Glouberman, William Irwin, as well as organizers and audience members at the University of Bucharest's international conference on “LLMs and Digital Autonomy: From Misinformation to Context Collapse.”

### References

- Bailey, A. M. 2024. “Digital Value.” *Philosophy and Digitality*, Vol. 1, No. 1, 25–39. Available at: <https://doi.org/10.18716/pd.v1i1.2418>.
- Baldwin, J. 2015. “Baudrillard and Neoliberalism.” *International Journal of Baudrillard Studies*, Vol. 12, No. 2. Available at: <https://baudrillardstudies.ubishops.ca/baudrillard-and-neoliberalism/> (accessed 15 September 2025).
- Bankov, K. 2023. Scarcity and Meaning: Towards a Semiotics of Economic Transaction. In Olteanu, A., P. Cobley (eds.). *Semiotics and its Masters, Volume 2*. Berlin: De Gruyter Mouton, 111–127. Available at: <https://doi.org/10.1515/9783110857801-007>.
- Bareis, J., C. Katzenbach. 2022. “Talking AI into Being: The Narratives and Imaginaries of National AI Strategies and their Performative Politics.”

*Science, Technology, & Human Values*, Vol. 47, No. 5, 855–881. Available at: <https://doi.org/10.1177/01622439211030007>.

Baron, N. S. 2023. *Who Wrote This? How AI and the Lure of Efficiency Threaten Human Writing*. Redwood City: Stanford University Press.

Baudrillard, J. 1988. *The Ecstasy of Communication*. (transl. Bernard Schütze and Caroline Schütze). New York: Semiotext(e).

Berardi, F. 2018. *Breathing: Chaos and Poetry*. Cambridge, Mass.: MIT Press.

Bontridder, N., Y. Pouillet. 2021. “The Role of Artificial Intelligence in Disinformation.” *Data & Policy*, Vol. 3, article e32, 1–21. Available at: <https://doi.org/10.1017/dap.2021.20>.

Borgmann, A. 1999. *Holding On to Reality: The Nature of Information at the Turn of the Millennium*. Chicago: University of Chicago Press.

Champagne, M. 2011. “Axiomatizing Umwelt Normativity.” *Sign Systems Studies*, Vol. 39, No. 1, 9–59. Available at: <https://doi.org/10.12697/SSS.2011.39.1.01>.

Champagne, M. 2024a. Putting Aside One’s Natural Attitude – and Smartphone – to See what Matters More Clearly. In Shafiei, M., A.-V. Pietarinen (eds.). *Phaneroscopy and Phenomenology: A Neglected Chapter in the History of Ideas*. Cham: Springer, 25–55. Available at: [https://doi.org/10.1007/978-3-031-66017-7\\_2](https://doi.org/10.1007/978-3-031-66017-7_2).

Champagne, M. 2024b. “The Virtual isn’t Real.” *Disputatio*, Vol. 16, No. 72, 37–66. Available at: <https://doi.org/10.2478/disp-2024-0003>.

Chang, B. G. 1987. “World and/or Sign: Toward a Semiotic Phenomenology of the Modern Life-World.” *Human Studies*, Vol. 10, No. 3–4, 311–331. Available at: <https://doi.org/10.1007/bf00157602>.

Collingridge, D. 1980. *The Social Control of Technology*. New York: St. Martin’s Press.

Darmody, A., D. Zwick. 2020. “Manipulate to Empower: Hyper-Relevance and the Contradictions of Marketing in the Age of Surveillance Capitalism.” *Big Data & Society*, Vol. 7, No. 1, 1–12. Available at: <https://doi.org/10.1177/2053951720904112>.

Dennett, D. C. 2023. The Problem with Counterfeit People. *The Atlantic*, 16 May 2023. Available at: <https://www.theatlantic.com/technology/archive/2023/05/problem-counterfeit-people/674075/> (accessed 15 September 2025).

Derrida, J. 1991. *Given Time: I. Counterfeit Money*. Chicago: University of Chicago Press.

Douglas, A. X. 2016. *The Philosophy of Debt*. New York: Routledge.

Dreyfus, H. L. 1972. *What Computers Can't Do: A Critique of Artificial Reason*. New York: Harper and Row.

Eco, U. 1976. *A Theory of Semiotics*. (transl. David Osmond-Smith). Bloomington: Indiana University Press.

Eco, U. 1990. *The Limits of Interpretation*. Bloomington: Indiana University Press.

Ellul, J. 1964. *The Technological Society*. (transl. John Wilkinson). New York: Vintage.

Epstein, R., M. Bordyug, Y.-H. Chen, Y. Chen, A. Ginther, G. Kirkish and H. Stead. 2023. "Toward the Search for the Perfect Blade Runner: A Large-scale, International Assessment of a Test that Screens for 'Humaneness Sensitivity.'" *AI & Society*, Vol. 38, No. 4, 1543–1563. Available at: <https://doi.org/10.1007/s00146-022-01398-y>.

Fourcade, M., D. N. Kluttz. 2020. "A Maussian Bargain: Accumulation by Gift in the Digital Economy." *Big Data & Society*, Vol. 7, No. 1, 1–16. Available at: <https://doi.org/10.1177/2053951719897092>.

Goux, J.-J. 2001. Ideality, Symbolicity, and Reality in Postmodern Capitalism. In Cullenberg, S. E., J. Amariglio, and D. F. Ruccio (eds.). *Post-Modernism, Economics and Knowledge*. London: Routledge, 166–181.

Graveris, D. 2023. Porn Statistics [2023]: How Many People REALLY Watch Porn? *SexualAlpha.com*, 31 May 31 2023. Available at: <https://sexualalpha.com/how-many-people-watch-porn-statistics/> (accessed 15 September 2025).

Guala, F. 2020. "Money as an Institution and Money as an Object." *Journal of Social Ontology*, Vol. 6, No. 2, 265–279. Available at: <https://doi.org/10.1515/jso-2020-0028>.

Hanemaayer, A. 2022. Critical Insights – Bringing the Social Sciences and Humanities to AI. In Hanemaayer, A. (ed.). *Artificial Intelligence and its Discontents: Critiques from the Social Sciences and Humanities*. Cham: Palgrave Macmillan, 1–20. Available at: [https://doi.org/10.1007/978-3-030-88615-8\\_1](https://doi.org/10.1007/978-3-030-88615-8_1).

Hartley J., I. Ibrus and M. Ojamaa. 2020. *On the Digital Semiosphere: Culture, Media and Science for the Anthropocene*. New York: Bloomsbury. Available at: <https://doi.org/10.5040/9781501369209>.

Heying, H., B. Weinstein. 2021. *A Hunter-Gatherer's Guide to the 21<sup>st</sup> Century: Evolution and the Challenges of Modern Life*. New York: Portfolio.

Hoffmeyer, J. 1996. *Signs of Meaning in the Universe*. (transl. Barbara J. Haveland). Bloomington: Indiana University Press.

James, W. 1922. *Pragmatism: A New Name for some Old Ways of Thinking*. New York: Longmans, Green and Company.

Kertysova, K. 2018. "Artificial Intelligence and Disinformation: How AI Changes the Way Disinformation is Produced, Disseminated, and Can Be Countered." *Security and Human Rights*, Vol. 29, No. 1–4, 55–81. Available at: <https://doi.org/10.1163/18750230-02901005>.

Krüger, S., C. Wilson. 2023. "The Problem with Trust: On the Discursive Commodification of Trust in AI." *AI & Society*, Vol. 38, No. 4, 1753–1761. Available at: <https://doi.org/10.1007/s00146-022-01401-6>.

Lembke, A. 2021. *Dopamine Nation: Finding Balance in the Age of Indulgence*. New York: Dutton.

Lin, L. Y., J. E. Sidani, A. Shensa, A. Radovic, E. Miller, J. B. Colditz, B. L. Hoffman, L. M. Giles, and B. A. Primack. 2016. "Association between Social Media use and Depression among U.S. Young Adults." *Depression and Anxiety*, Vol. 33, No. 4, 323–331. Available at: <https://doi.org/10.1002/da.22466>.

Mas, I., Chuen. 2015. Bitcoin-Like Protocols and Innovations. In Chuen, D. L. K. (ed.). *Handbook of Digital Currency: Bitcoin, Innovation, Financial Instruments, and Big Data*. Amsterdam: Elsevier, 417–451. Available at: <https://doi.org/10.1016/B978-0-12-802117-0.00021-7>.

Mathieu, D., J. M. Hartley. 2021. "Low on Trust, High on Use: Datafied Media, Trust and Everyday Life." *Big Data & Society*, Vol. 8, No. 2, 1–13. Available at: <https://doi.org/10.1177/20539517211059480>.

Monnet, C. 2005. "Counterfeiting and Inflation." *European Central Bank (ECB) Working Paper Series*, No. 512, 1–22. Available at: <https://www.econstor.eu/bitstream/10419/152946/1/ecbwp0512.pdf> (accessed 15 September 2025).

Montanari, F. 2025. "ChatGPT and the Others: Artificial Intelligence, Social Actors, and Political Communication. A Tentative Sociosemiotic Glance." *Semiotica*, Vol. 262, 189–212. Available at: <https://doi.org/10.1515/sem-2024-0210>.

Newman, C. 1984. "The Post-Modern Aura: The Act of Fiction in an Age of Inflation." *Salmagundi*, vols. 63–64, 3–199. Available at: <https://doi.org/10.2307/40547646>

Peirce, C. S. 1992. *The Essential Peirce: Selected Philosophical Writings*. Vol. 1. Bloomington: Indiana University Press.

Peirce, C. S. 1998. *The Essential Peirce: Selected Philosophical Writings*. Vol. 2. Bloomington: Indiana University Press.

Picca, D. 2025. "Not Minds, but Signs: Reframing LLMs through Semiotics." *ArXiv*, arXiv:2505.17080. Available at: <https://doi.org/10.48550/arXiv.2505.17080> (accessed 15 September 2025).

Sebeok, T. A. 1985. "Vital Signs." *The American Journal of Semiotics*, Vol. 3, No. 3, 1–27. Available at: <https://doi.org/10.5840/ajs19853311>.

Sutton, S. G., V. Arnold and M. Holt. 2018. "How Much Automation Is Too Much? Keeping the Human Relevant in Knowledge Work." *Journal of Emerging Technologies in Accounting*, Vol. 15, No. 2, 15–25. Available at: <https://doi.org/10.2308/jeta-52311>.

Taube, M. 1961. *Computers and Common Sense: The Myth of Thinking Machines*. New York: Columbia University Press.

Teymoori, A., B. Bastian and J. Jetten. 2017. "Towards a Psychological Analysis of Anomie." *Political Psychology*, Vol. 38, No. 6, 1009–1023. Available at: <https://doi.org/10.1111/pops.12377>.

Tratner, M. 2003. "Derrida's Debt to Milton Friedman." *New Literary History*, Vol. 34, No. 4, 791–806. Available at: <https://doi.org/10.2307/20057813>.

Twenge, J. M. 2020. "Why Increases in Adolescent Depression May be Linked to the Technological Environment." *Current Opinion in Psychology*, Vol. 32, 89–94. Available at: <https://doi.org/10.1016/j.copsyc.2019.06.036>.

Zuboff, S. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: Public Affairs.

## TRANSLATING THE WILD: AI, SEMIOTICS, AND THE FUTURE OF ANIMAL COMMUNICATION<sup>1</sup>

*Nicola Zengiaro*  
*University of Turin*  
*nicola.zengiaro@unito.it*

### **Abstract**

This article critically examines the recent rise of AI-based attempts to “translate” animal communication: with a specific focus on aquatic species such as dolphins and whales. Drawing on biosemiotics, ecosemiotics, and Umwelt theory, it argues that such projects, while technologically sophisticated, risk reducing animal semiosis to codifiable data structures and computational approximations of meaning. Through an analysis of DolphinGemma and CHAT, the article exposes the epistemological and semiotic limitations of current AI models, contrasting them with embodied, context-sensitive modes of communication in non-human species. Rather than serving as transparent translation tools, AI systems should be understood as technosemiotic infrastructures which may support new interspe-

---

<sup>1</sup> This work has received funding from Foundation CRT (LEOM\_CRT\_25\_01 - 110992 / 2024.1420-San Francesco e l'IA, PI Massimo LEONE)

cies resonances, provided they are embedded within critical and relational frameworks. The analysis advocates for an ecotechnical semiotics which redefines intelligence and communication as emergent, situated, and materially grounded processes, resisting both symbolic reductionism and techno-utopianism.

**Keywords:** biosemiotics, zoosemiotics, artificial intelligence, dolphin communication, Umwelt

### **Introduction**

In recent years, both media and scientific discourses have increasingly focused on the possibility of “translating” animal communication systems through the use of artificial intelligence (AI) technologies. Projects such as CETI (Cetacean Translation Initiative), DolphinGemma (developed in collaboration with Google), and the Wild Dolphin Project aim to decode acoustic signals and complex behaviors of species such as sperm whales and bottlenose dolphins, in the hope of extracting linguistic structures interpretable in a manner analogous to human language. These initiatives rest upon a strong epistemological premise, namely, that animal communicative patterns can be reduced to formal systems, amenable to computational and statistical modeling and translation.

This narrative, however, raises fundamental semiotic and philosophical questions. Reducing animal communication to a set of “decodable data” presupposes a referential and codifying model of semiosis in nature, in which meaning is conceived as a predefined and transferable amount of information and data. Such an approach risks obscuring the situated, embodied, and ecologically grounded nature of animal communication (Maran, Martinelli & Turovski 2011), the intelligibility of which cannot be separated from the perceptual, affective, and evolutionary conditions which give rise to it. As Thomas A. Sebeok already noted, animals do not possess language in the human sense, but communicate through a semiotic system which is simultaneously physiological, behavioral, and relational (Sebeok 1972, 2001). Each animal sign is thus part of an interpretive network internal to its Umwelt (Cerrone & Mäekivi 2021), according to the paradigm introduced by Jakob von Uexküll (1909), who conceived of every organism as an active interpreter of its own meaningful world.

From this perspective, biosemiotics offers a crucial contribution by reconfiguring the very questions which we pose about animal communica-

tion. The issue is not one of “translating” messages, but of understanding the embodied signification processes which make communication possible in the first place. The animal sign is not a mere computational code, but a relational, ecosystemic, and corporeal event. For this reason, its intelligibility cannot be disentangled from the ecosystemic context and the corporeal affordances that constitute its material basis. The recent AI-driven ambitions of translation, if not supplemented by a critical semiotic theory, risk reiterating an anthropocentric and logocentric paradigm, wherein communication becomes synonymous with symbolic language, and intelligence is assessed based on its compatibility with formalizable syntactic structures.

This article offers a critical reflection on such attempts to “decode the wild,”<sup>2</sup> drawing on a biosemiotic re-reading of a selected corpus of recent studies, with particular attention to the case of the Wild Dolphin Project. The aim is to show how the biosemiotic approach, which understands semiosis as an emergent, trans-categorical process inherent to all forms of life, can help reformulate the question of interspecific communication in less reductionist and more ecologically grounded terms. This reframing can offer important insights not only into animal communication, but also into our relationship with AI, and into how this technology might serve as a fruitful interface, providing a window into nonhuman worlds.

## 1. Translating Dolphins? A Semiotic Critique of the Computational Rhetoric of Animal Communication

The idea of “talking to dolphins” is far from new in the landscape of anthropocentric speculations on animal communication (Mann et al. 2000), yet it has recently gained renewed vigor through the innovative rhetoric surrounding artificial intelligence (AI). Four recent articles, published respectively in *New Atlas*, *Times of Malta*, *ExtremeTech*, and Google’s official

---

<sup>2</sup> This idea of translating the wild (apart from reintroducing a colonialist idea of translation), is inspired by Neil Savage’s article entitled “AI decodes the calls of the wild”, published in *Nature* on 10<sup>th</sup> December 2024. <https://www.nature.com/immersive/d41586-024-04050-5/index.html> (accessed 8 May 2025) At the same time, popularizer Philip Ball, echoing the words of Constantine Slobodchikoff, writes on 27 April 2021 in *The New Yorker* that “Computer technology is finally allowing us to see inside the world of animals...”. <https://www.newyorker.com/science/elements/the-challenges-of-animal-translation> (accessed 8 May 2025). We are a long way from “seeing” into the animal Umwelt. From a semiotic point of view, we could (and perhaps should) ask what it means to “see” inside the animal world with technology. Is the subjective world of others something you can see? Can we see, as if it were something perpetually there waiting for our gaze, what a life form experiences in its ecologically, historically and corporeally situated existence?

blog, enthusiastically report on the DolphinGemma project, developed by Google DeepMind in collaboration with the Wild Dolphin Project. This initiative aims to construct a “language model” of dolphin vocalizations, in order to “decode” their language. The expressions used in these texts, while seemingly neutral or even celebratory of interspecies encounters, in fact reveal an interpretative paradigm heavily shaped by anthropomorphic projections and a reductionist conception of communication.

Phrases include such as: “allowing the researcher to “converse” with dolphins... Patterns allow DolphinGemma to predict what the dolphin is likely to “say” next, similar to how Google Search and generative AI predict the next fragment of a user’s query” (*ExtremeTech*)<sup>3</sup>; “Researchers could also trigger artificial whistles using the interface, essentially “speaking” to the dolphins using sounds they could potentially learn to associate with specific objects” (*New Atlas*)<sup>4</sup>; “Before you scoff at this suggestion, consider that with its ability to analyze huge quantities of data, recognize patterns, and apply machine learning algorithms, AI is a powerful tool to decode the complex conversations of other species, whether underwater or intergalactic. Google translate is a fine example” (*Times of Malta*)<sup>5</sup>; to build an imaginary in which the acoustic behavior of bottlenose dolphins is construed as a language in the strict sense (with all the inverted commas put by the authors above words like speak or say). This can be seen as a structured symbolic system governed by syntactic rules and potentially translatable according to principles analogous to those of human languages. This use of linguistic terminology presupposes a correspondence between the clicks or whistles produced by dolphins and discrete units of meaning, capable of being transferred, processed, and ultimately translated through computational models. The Google blog, in particular, claims that deep learning models can identify patterns in dolphin vocalizations and attempt to associate them with known contexts and behaviors, offering a path toward this correlation. However, this computational framing introduces significant epistemological and semiotic limitations. Although deep learning mod-

<sup>3</sup> Nine, A. 2025. New Google LLM Aims to Translate Dolphin Language. *Extreme Tech*, 16 April 2025. Available at: <https://www.extremetech.com/science/new-google-llm-aims-to-translate-dolphin-language> (accessed 8 May 2025)

<sup>4</sup> Ghoshal, A. 2025. Google AI learns to speak dolphin. *New Atlas*, 15 April 2025. Available at: <https://newatlas.com/biology/build-ai-translator-dolphins-dolphingemma/> (accessed 8 May 2025)

<sup>5</sup> Lafferty, E. 2025. On using Google Translate to speak ‘Dolphin’ to aliens. *Times of Malta*, 29 March 2025. Available at: <https://x2.timesofmalta.com/20250329/opinion/how-google-translate-has-allowed-us-to-speak-dolphin/> (accessed 8 May 2025)

els can identify recurring acoustic patterns and statistically correlate them with observed dolphin behaviors, such as foraging, play, or social bonding, they do not access the meaning-making processes inherent to the dolphins themselves. For example, a particular whistle might occur both during cooperative hunting and during playful social interactions. The acoustic form remains constant, but its meaning shifts depending on the ecological context, group dynamics, and affective state of the animal. AI models, which rely on surface-level correlations, cannot discern this semiotic flexibility.

From a biosemiotic perspective, meaning is not an intrinsic property of a signal, but emerges from its embeddedness within the animal's *Umwelt*, a relational, embodied world of perception, need, and social interaction (Uexküll 1940; Sebeok 1977). Machine learning systems lack this *Umwelt*. They do not interpret signs which generate an interpretant shaped by lived experience. Instead, they simulate interpretation through probabilistic matching. As such, the "translation" offered by AI risks becoming a symbolic approximation, useful perhaps for human analysis, but disconnected from the intentional, adaptive, and affectively charged semiosis which characterizes dolphin communication. Without accounting for this contextual and embodied dimension, AI-driven models may not translate meaning, but instead transform it. They project computational structure onto a communicative process which is fundamentally ecological and relational.

This implicit semiotic operation entails two major distortions. First, it reduces the animal sign to a decontextualized acoustic input, treated as raw data to be extracted and modeled; while secondly, it flattens meaning into its referential and behavioral correlate. Both strategies reveal a model of semiosis which is rigidly codified and grounded in the information theory of Shannon and Weaver (1948). Here communication is equated with the transmission of data through a channel. However, as the biosemiotic tradition has long argued, this model proves inadequate when it comes to understanding the processes of signification in non-human species.

The perspective of biosemiotics proposes a radically different conception of sign and communication (Zengiaro 2023a). The issue is not to identify "units of meaning" which can be translated between different codes, but to understand semiosis as a dynamic, embodied process emerging from a triadic relation (sign, object, interpretant) which always takes place within a situated context. According to Thomas A. Sebeok (1972), animals do not possess "language" in the human sense of the term, but rather species-specific semiotic systems rooted in their morphology and ecology (primarily iconic and indexical in nature) the primary function of which is to orient action, rather than to transmit propositional content. According to Sebeok,

it is misleading to speak of “language” in reference to animal communication, except in a metaphorical sense (Sebeok 1990: 106). Similarly, Jakob von Uexküll (1909) conceived every organism as immersed in an *Umwelt*, a meaningful world emerging from the interplay between its perceptual capabilities and the landscape it inhabits. The meaning of a dolphin’s vocalization cannot be grasped, if it is isolated from the ecological, affective, and corporeal conditions which generate it.

The articles under analysis, however, tend to abstract the dolphin vocalization from its ecological and behavioral grounding, treating the sign as a discrete and computable unit. Nevertheless, empirical research in dolphin cognition and communication, such as the work by Smith et al. (1995) demonstrates that interpreting vocal signals is inextricably tied to understanding the behavioral responses which follow. Their study showed that dolphins are capable of complex auditory comprehension, responding to syntactically ordered human-generated acoustic signals (Smith et al. 1997). Crucially, however, these responses were shaped not only by the acoustic form of the stimulus but by contextual cues, prior training, and motivational states. This underscores the challenge of interpreting vocalizations through AI: meaning is not inherent in the signal itself but arises through a situated interpretive process.

In this light, the assumption that we can straightforwardly “break down what dolphins are saying, and even respond in kind,” as some popular articles suggest, misrepresents the asymmetry between machine processing and biological semiosis. Dolphins do not “respond” in a symbolic, dialogic sense, as humans might in linguistic conversation. They respond through embodied and context-sensitive behaviors: orienting their echolocation beam, altering swimming patterns, or emitting socially nuanced clicks and whistles. These behaviors are not mere outputs to be decoded: they are affective, ecological expressions embedded in a social *Umwelt* (Martinelli 2011). Attempting to isolate “messages” from these behaviors for the purpose of symbolic translation ignores the co-emergent nature of communication within relational and environmental feedback loops. Without access to the interpretive ground which motivates and gives sense to the sign, a capacity rooted in embodied cognition, AI systems risk misattributing functional equivalence where there is only statistical correlation.

In a semiotic framework, Machine Learning (ML) can be interpreted as a set of algorithmic processes aimed at modeling sign relations between data inputs and decision outputs through iterative refinement. Rather than simply processing “information,” ML systems operate as translation machines which extract, transform, and abstract patterns from empirical data,

thereby producing new layers of operational meaning. This dynamic of semiosis is particularly evident in the three main paradigms of ML (Bermant et al. 2019): supervised, unsupervised, and reinforcement learning.

1. Supervised learning functions through a logic of indexical anchoring: it operates on pre-labeled data where the referent (ground truth) is explicitly known, allowing the algorithm to learn correspondences between signifiers and predetermined classes. The system builds a predictive model through repeated exposure to these labeled signs, effectively emulating a Peircean interpretant guided by externally imposed norms.
2. Unsupervised learning, in contrast, explores the internal structure of sign systems without predefined classifications. It searches for clusters, associations, or latent variables, operating through abductive inference, it attempts to uncover implicit relations between signs, modeling semiosis as the discovery of emergent regularities within undifferentiated signal spaces.
3. Reinforcement learning situates semiosis within a feedback loop of action and consequence. Here, the algorithm-agent engages with an environment, receiving reward signals which function as pragmatic indices of interpretive success. The meaning of a sign (e.g., an action) is determined retrospectively, through its contribution to an outcome, a model of semiosis akin to trial-and-error learning in animals.

These paradigms converge in Deep Learning (DL), a subfield of ML which builds hierarchical models of sign abstraction through multilayer neural architectures. DL can be understood as a semiotic stratification process (Dondero 2025): raw input data (e.g., pixels, audio waves, word embeddings) are progressively transformed through layers of nonlinear functions, yielding increasingly complex representations. Each layer functions as an interpretive filter that re-codes the incoming signs into more abstract forms, recursively re-embedding them in new semiotic networks.

Artificial Neural Networks (ANNs), the computational substrate of DL, instantiate this logic formally. Through iterative adjustment of their internal parameters (weights), guided by optimization algorithms such as backpropagation, ANNs learn mappings between input and output spaces. Architectures such as Multi-Layer Perceptrons (MLPs), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs) specialize in different forms of sign temporalization and spatialization. MLPs handle general classification tasks. CNNs process spatial hierarchies in im-

ages (signs as shapes), while RNNs model temporal sequences (signs as rhythms or syntagms).

In semiotic terms, ML (and particularly DL) does not understand meaning in the phenomenological sense, but performs operational semiosis. It constructs functional relationships among signs, optimizing for internal coherence or external reward. The semiotic limitation of these systems lies not in their inability to generalize, but in the absence of intentionality, Umwelt, and affect, and key elements of biosemiotic semiosis. Yet, their capacity to reorganize and recompose sign structures renders them powerful tools for exploring latent semiotic dimensions within vast symbolic ecologies.

Ecosemiotics, conceived as an expansion of biosemiotics into the spatial and relational dimensions of environment (Zengiaro 2023b), provides the theoretical tools required to recognize that signals are never mere carriers of information. Rather, they are articulations of a broader contextual web. As ecosemiotician Timo Maran (2020) argues, animal semiosis is co-determined by habitat, body, social interaction, and the affordances of the sensory landscape. Every sign is situated, and every expression is relational. To “translate” a click without feeling the water, seeing the light, or knowing the group’s spatial configuration is an epistemologically blind operation.

The AI-centered optimism of projects such as DolphinGemma, thus reproduces a colonial paradigm of knowledge, one that reduces alterity to codifiable difference, neutralizes ecological complexity in the name of computability, mistaking statistical mapping for semiotic understanding. Artificial intelligence can undoubtedly contribute to identifying acoustic and behavioral patterns. However, this is not equivalent to “translating dolphin language.” Without a robust theory of the sign, such efforts risk remaining exercises in anthropocentric data mining, incapable of grasping the core issue: that animal communication is not a language, it is a life that signifies<sup>6</sup>.

---

<sup>6</sup> Interesting is a comment in *New Atlas* about the article already cited “Google AI learns to speak dolphin” by JS who writes: “So, we already know them to be one of the smartest animals on the planet ... once we learn their language and can communicate with them, what happens to all the ones in captivity doing tricks for snacks? Does that make them slaves at that point?” JS April 15, 2025 09:50 AM. <https://newatlas.com/biology/build-ai-translator-dolphins-dolphingemma/> Here the old Cartesian legacy explodes, according to which if animals speak then they are slaves, since they are subjects of a life; if they do not speak, then everything is permitted, since this life of theirs has no interiority or value.

### 1.1. CHAT: Am I free to choose what to express?

The CHAT (Cetacean Hearing and Telemetry) project, developed by Thad Starner's team at Georgia Tech in collaboration with the Wild Dolphin Project directed by Denise Herzing, represents one of the most advanced contemporary attempts to design an interspecific communication technology between humans and dolphins. CHAT is a bidirectional underwater computer equipped with two hydrophones to receive sounds, and an underwater speaker to emit them<sup>7</sup>. The system is based on the emission and recognition of artificial whistles, specially designed acoustic signals which mimic dolphin vocalizations but are constructed ad hoc to label play objects (e.g., sargassum, scarves, ropes), thus testing the dolphins' potential referential use of such sounds.

The ethological foundation of the project is grounded in well-established observations: dolphins are capable of mimicking vocalizations and associating them with objects, a behavior which suggests a potential for referential signaling. CHAT implements a limited vocabulary, with each signal associated with a sound label transmitted to the diver via bone-conducting headphones, thereby providing immediate interpretative support. In a notable episode, Herzing reported hearing the word "sargassum" emitted by the device in response to a whistle produced by a dolphin. However, the event was not accompanied by any referential behavior, leaving open the question of whether the cetacean was merely mimicking the signal or using it functionally.

This experience suggests that while dolphins can reproduce artificial sounds, their semantic understanding of such signals remains unproven. Moreover, subsequent analyses revealed that many of the whistles produced were situated in frequency bands not detected by the CHAT system, pointing to a sophisticated degree of vocal plasticity in dolphins and highlighting further complexities in interface design. The Phase II of the project aims to develop a more advanced and sensitive bidirectional system, utilizing pattern recognition software to analyze dolphins' natural acoustic signals and facilitate the decoding of their vocal repertoires. This approach is framed as an attempt to bridge the gap between human and animal communication, in pursuit of a form of mutual communicative understanding.

On the theoretical level, CHAT is situated within a semiotic paradigm marked by internal tension. On the one hand, it attempts to construct a shared code, while on the other, it is confronted with the irreducibility of

---

<sup>7</sup> CHAT: Is It A Dolphin Translator Or An Interface? *Wild Dolphin Project*. Available at: <https://www.wilddolphinproject.org/our-research/chat-research/> (accessed 8 May 2025).

animal semiosis to mere syntactic structure. The technical challenges it faces (the design of inputs compatible with dolphin ethology to the identification of contextual signal usage) demonstrate that cetacean semiosis cannot be isolated from their Umwelt, which includes posture, spatial orientation, behavior, and social relationships. CHAT has not yet demonstrated the possibility of translating dolphin communication, but it opens a relevant experimental space for the construction of relational and adaptive interfaces, rather than strictly linguistic ones. The potential of such systems lies not in the deciphering of an “animal language,” but in the co-construction of hybrid communicative practices capable of respecting the semiotic alterity of the nonhuman, an approach closer to biosemiotics than to classical computational linguistics. In order to improve this understanding towards non-human communication, researchers need to be cautious and carry out a critical and reasoned analysis of the scope of these tools (and their results). Projects such as CHAT, which aim to establish a computational interface between humans and dolphins, are based on the assumption that meaning can be traced, codified, and computed as a stable outcome of observable behavior. In this model, a signal (such as a whistle) is treated as a functional trigger for a response (behavior), following an ostensibly transparent stimulus-response logic. However, from a biosemiotic perspective, such a reductionist model risks obscuring the profoundly indeterminate, situated, and creative nature of animal communication.

Jesper Hoffmeyer (1996) introduced the concept of “semiotic freedom” to describe the degree of interpretative flexibility that a living organism possesses in relation to environmental signs. It denotes the organism’s capacity to choose among multiple possible responses to a given signal, rather than being bound to a predetermined automatic reaction. According to Hoffmeyer, semiotic freedom increases with the evolution of cognitive and corporeal complexity and serves as an index of an organism’s ability to construct meaningful worlds, its Umwelt. Behavior, therefore, is never purely reactive, but rather the outcome of interpretive negotiation among multiple semiotic possibilities.

Kalevi Kull has further developed this perspective by proposing a definition of semiosis as choice, clearly distinguishing it from algorithmic calculation. As he states:

Only in case if there is a certain problem to solve, if there is more than one option (which requires choice and decision-making), the process related to it is in semiotics commonly called interpretation, or semiosis. [...] A problem is real for an organism, if there are pos-

sibilities, several options for it to choose, i.e., if there is a situation of choice. There is a situation of choice, if the options given simultaneously are incompatible – i.e., if, at a given moment, there are several ways to behave, while only one of these can be executed. Thus, the simultaneous existence of options – we call it *incompatibility* – is what makes anything a problem, and it can be overcome only via making a decision. (Kull 2015: 225)

According to this definition, communication is an event of selection among concurrent alternatives, irreducible to a computable function, precisely because it involves a dimension of freedom and creativity which escapes the deterministic logic of computational systems.

In the case of dolphins (well known for their vocal flexibility, complex social play, self-recognition, and even the symbolic use of sounds such as signature whistles) semiotic freedom is expressed not only in the repetition of signals but in the capacity to use them out of context, to reinterpret them, or to refrain from using them altogether. This point is critical: when a dolphin does not activate a whistle pre-programmed in the CHAT system to request an object, this is not a failure of the system, but a potential expression of the dolphin's semiotic freedom, its ability to choose when, how, and whether to use a given sign.

The AI implemented in the CHAT project, as in many other “intelligent” interfaces designed to recognize vocal patterns, operates within a paradigm of functional equivalence between sign and behavior. Yet in doing so, it fundamentally misrecognizes the interpretative dimension of the sign. Communication is never a simple output determined by an acoustic input, but a situated event, co-determined by ecological conditions, social relations, intentionality, and perceptual context. Semiotic freedom implies that an animal can receive a signal and choose not to act on it, or to act in ways that appear inconsistent, precisely because this apparent incoherence reveals its semiotic subjectivity.

In this sense, the CHAT methodology confronts a deep epistemological limit: it confuses the ability to recognize patterns with the ability to understand semiosis. Recognizing a whistle is not sufficient to demonstrate communicative comprehension. For a sign to truly function as such, it must be interpreted. Interpretation, as Peirce reminds us, is always an unfinished, open-ended, and deferred act. In attempting to build a behavioral dictionary of dolphin responses, AI risks flattening the semiotic freedom of living beings into a behaviorist grid, despite its apparent technological sophistication.

A true alliance between AI and biosemiotics should not aim to produce behavioral interfaces which merely verify the functional efficiency of signs, but should instead seek to create relational spaces capable of hosting the emergence of animal semiotic freedom. In other words, rather than asking whether the dolphin “understood” the word “sargassum” because it approached the seaweed, we should ask ourselves: what relational worlds become possible when signs are used indeterminately, asynchronously, and creatively? How can we listen to dolphin semiosis without seeking to immediately translate it into a reactive grammar?

The problem with CHAT is not technical, it is semiotic. Tying meaning to immediate action prevents us from grasping the interpretative depth of life. To recognize an animal’s semiotic freedom also means to accept the opacity of its sign and to understand that not knowing what a signal means can be more productive, both theoretically and ethically, than assigning it a predefined function. The real challenge is not to translate the other, but to inhabit the semiotic difference that makes all communication possible.

## **2. Artificial Intelligence and Umwelt: An Epistemological Question Between Organisms and Artifacts**

The introduction of artificial intelligence in attempts to “decode” animal communication entails deep epistemological implications which cannot be reduced to mere technical or engineering concerns. It brings into question the difference between living organisms and artificial devices not only in material or functional terms, but more crucially from a semiotic and phenomenological perspective. In this context, the concept of *Umwelt*, introduced by Jakob von Uexküll (1909) and further developed within contemporary biosemiotics (Tønnessen 2022), gains critical theoretical relevance. *Umwelt* does not simply refer to the external environment, but rather to the subjective phenomenological world constructed through the specific sensorimotor interactions of an organism, in accordance with its biological needs and perceptual marks. This concept is fundamentally incompatible with a conception of communication as the linear transmission of computable data, typical of AI, which operates through symbolic abstractions detached from embodied and contextual lived experience.

Thomas A. Sebeok (1972, 1990, 2001), in his foundational work in zoo-semiotics, drew a crucial distinction between human language and animal communication. The former is symbolic, referential, and syntactically articulated, whereas the latter is based on indexical and iconic sign systems, deeply rooted in the environment and the body. To attempt to translate

these signals into human language is to engage in a form of epistemological semicolonialism, wherein the human symbolic paradigm is retroactively projected onto biological systems which operate according to different logics of signification. Dolphins, for instance, do not “speak,” but express intentions, bodily states, emotions and ecological relationships through gestures, sounds, and movements embedded within their aquatic and relational Umwelt.

The question of whether artificial systems such as AI or robots can possess an Umwelt has become a focal point in the dialogue between biosemiotics, cybernetics (or cybersemiotics), and philosophy of technology. What is at stake in this debate is not merely the functionality of machines, but the very conditions under which meaning arises.

Over the past several decades, researchers from diverse fields, including robotics, cognitive science, semiotics, and theoretical biology, have engaged with Jakob von Uexküll’s theory of the Umwelt to bridge the epistemological divide between biological organisms and artificial systems. This body of work reveals a spectrum of epistemological positions. Some adopt Umwelt theory as a generative model for constructing intelligent machines, while others employ it critically to highlight the limitations of artificial cognition. Spanning fields such as robotics, cybernetics, cognitive science, and semiotics, this conversation revolves around one central question: can non-living artifacts like robots or AI systems possess a subjective world of meaning similar to that of living organisms?

In the 1980s, roboticist Rodney Brooks drew on Uexküll’s notion of *Merkwelt* to challenge the dominant symbolic models in classical AI. Rather than programming machines with abstract representations of the world, Brooks (1986, 1991) argued that robots should construct their own “world-views” through direct perception-action loops. “The world is its own best model” (Brooks 1990: 5), he famously declared, advocating for embodied, situated intelligence which emerges from interaction rather than computation. This idea finds a compelling cognitive counterpart in the work of Andy Clark (1997), whose book *Being There* marks a pivotal moment in the development of embodied cognitive science. Clark introduces von Uexküll’s theory as a biological analogue for decentralized robot architectures, those that operate without a central planner. He emphasizes the importance of niche-dependent sensing as a foundational principle in both biological and artificial agents, echoing von Uexküll’s insistence that meaning arises from organism-environment coupling. For Clark, cognition is not merely internal computation, but an extended process involving the brain, the

body, and the environment in dynamic interaction. Robots which adapt to their specific sensory environments, he argues, begin to approximate the kind of situated cognition described by Umwelt theory, though they still lack its full phenomenological depth.

On the other hand, Claus Emmeche, working within the biosemiotic tradition, raised deeper philosophical concerns. For Emmeche (2001), Uexküll's Umwelt is not just a functional interface with the environment, but a lived, phenomenological world shaped by intentionality, bodily affect, and internal purposiveness. From this standpoint, no robot can possess an Umwelt in the full Uexküllian sense, because it lacks the biological grounding and subjective agency that define life. Semiosis is not a process which can be merely installed; it is emergent, embodied, and affect-laden. In contrast to this argumentation, Ricardo Gudwin (1999) proposed a more engineering-oriented reinterpretation of Umwelt, suggesting that any system with sensors and effectors can generate its own internal model of the world. For Gudwin, such models qualify as a type of Umwelt, even if they lack human- or animal-like consciousness. This view supports AI design but risks flattening the qualitative differences between simulation and experience, treating cognition as mere information processing.

Tom Ziemke has further enriched this debate by distinguishing between "simple embodiment," often invoked in robotics to describe physical interaction with the world, and "organismic embodiment," which includes the biological, emotional, and homeostatic dimensions unique to living systems. According to Ziemke (2001a, 2001b), real autonomy and meaning-making cannot arise from mechanical design alone; they require the co-evolutionary integration of an organism within its environment, something that no current robot can authentically replicate. Winfried Nöth (2001, 2003) adds a semiotic dimension to this discussion, arguing that while robots can display perception-action cycles, their meaning-making is always allreferential, externally defined by programmers, as opposed to the self-referential semiosis of living beings. Robots, therefore, do not interpret signs, they execute functions. Meaning, in this view, is not intrinsic to their operations but assigned from the outside. Most recently, Carlos Vidales and Julio Horta (2024) have proposed a cybersemiotic reframing of Umwelt through the concept of "cryptosemiotics," interpreting Uexküll's functional circle as a generative matrix for the emergence of meaning, not just in biology, but potentially in artificial systems as well. Their approach opens a new theoretical space: one that acknowledges the non-biological mediation of signs while remaining cautious about equating artificial semiosis with lived experience.

The theoretical landscape reveals two broad trajectories. The first views Uexküll's ideas as inspiration for more ecologically grounded and responsive AI systems, emphasizing embodiment, adaptability, and environmental coupling. The second maintains that the *Umwelt* is inextricable from biological life, serving as a critical standard to evaluate the limits of machine cognition. Between these poles lies a dynamic middle ground, where new hybrid models, combining cybernetic insight with semiotic depth, are still being actively explored. In such a framework, machines cannot possess an *Umwelt* in the strict sense because they lack this embodied feedback loop. What AI does instead is to operate within technosemiosis. It acts as an interface within human meaning-making systems, extending our cognitive and interpretative capacities without becoming a subject of interpretation itself. This insight does not diminish the value of AI but repositions it. Artificial systems can assist us in modeling, simulating, and even amplifying certain aspects of communication, also across species, but they do so from outside the circle of lived meaning. The *Umwelt*, as both a biological and semiotic reality, remains anchored in the flesh and orientation of the living.

This distinction carries profound implications for projects like *Dolphin-Gemma*. AI may be capable of collecting and classifying clicks and whistles, but it cannot access the motivational, affective, social, and environmental horizon within which these sounds hold meaning for the dolphin. To claim to "translate" these signals is to ignore that every *Umwelt* is untranslatable precisely because it is untransferable. We can construct indirect models of it through epistemic mediations, but we can never inhabit it. As Emmeche (2001: 656–7) aptly reminds us, "our scientific understanding of the sonar system of a bat gives us an indirect and functional picture of the bat's *Umwelt*, but we cannot enter into that *Umwelt* itself. All we have is a model in our (linguistic, cognitive, and perceptual) *Umwelt* of the bat's *Umwelt*".

The claim of endowing artificial systems with an *Umwelt* amounts to a functional hypostatization of subjectivity. This risks conflating model with lived experience, simulation with semiosis, and computation with meaning. The epistemological comparison between AI and *Umwelt*, in light of zoosemiotics, reveals that animal communication cannot be truly understood without first acknowledging the ontological difference (or the similarities) between organism and artifact. AI may indeed contribute to the study of animal communication, but only if integrated within a biosemiotic framework which recognizes its limitations and re-centers the experiential, ecological, and relational dimensions of life.

## **2.1. Intelligence, Semiosis, and Interfaces: Toward an Extended Semiotics of the Artificial Mind**

The classical distinction between artificial and natural intelligence, while useful in technical contexts, risks obscuring the epistemological potential of AI as an extension of the human mind and, more specifically, as a semiotic interface capable of mediating between different forms of life. In this sense, a paradigm shift is urgently needed. The challenge is not merely to compare machine intelligence with that of organisms, but to understand how the two may co-operate within a relational and biosemiotic cognitive horizon.

In a semiotic and biosemiotic framework, the concept of intelligence cannot be understood as a universal, context-free property. Rather, it emerges as a situated process of semiosis, dependent on specific material, relational, and interpretive conditions. Following Joseph Weizenbaum's (1976: 204–205) observation that “intelligence is a meaningless concept in and of itself,” unless situated within a “domain of thought and action,” we can reconceptualize intelligence not as a singular essence, but as a contextualized capacity for meaning-making. Intelligence, in this view, is not a trait but a practice, a semiotic function exercised differently across species, systems, and artifacts.

This approach resonates deeply with the notion of *Umwelt* and is extended in biosemiotics. Each organism constructs its own meaningful world through interpretive engagement with signs relevant to its perceptual and effector organs. Intelligence, thus, is not an abstract potentiality but a relational competence embedded in a life-world, a domain-specific configuration of signs, constraints, affordances, and value-laden decisions. Weizenbaum's critique anticipates this logic by rejecting the belief in a culturally-independent “thing” called intelligence and emphasizing instead the social, cultural, and ecological situatedness of intelligent activity.

To analyze intelligence across such domains (whether human, animal, plant, artificial, or collective) requires a multidimensional framework sensitive to its semiotic embedding. From a semiotic perspective, three inter-related dimensions of any domain of intelligence (Attard-Frost 2024: 28–29) can be identified:

**Cognitive (or Interpretive) Dimension:** intelligence entails the capacity to process, recognize, and respond to signs within a meaningful context. In semiotic terms, this refers to the activity of the interpretant in the Peircean triad. Whether we consider a cephalopod navigating a reef, a neural network classifying images, or a bacterium responding to chemical gradients,

each enacts a form of biosemiosis that connects signal to function (Sharov 2013; Sharov & Vehkavaara 2015). These interpretive acts include pattern recognition, adaptation, anticipatory behavior, memory encoding, and the coordination of action within an Umwelt.

**Normative Dimension:** intelligence is never value-neutral. Every interpretive act implies a choice (Kull 2018), a selection of one path among many, which is governed by internal norms or external constraints (Deacon 1997). In animal cognition, these norms may be evolutionary (survival, reproduction), social (hierarchy, cooperation), or emotional (fear, attraction). In artificial systems, norms are encoded by designers or emerge through optimization criteria: speed, accuracy, efficiency, or alignment with user goals. Biosemiotically, this dimension expresses the semiotic freedom of a system, the degree to which it can select, modulate, or reinterpret signs in light of its values (Hoffmeyer 2004).

**Performative Dimension:** intelligence is always enacted, and measured, through performances that are themselves semiotically framed. In human domains, these may take the form of standardized tests or behavioral metrics; in biological systems, they may involve ecological fitness or problem-solving efficiency. In AI, performance is often quantified through benchmarks, but this instrumental logic may obscure the system's embeddedness in human interpretive regimes. A truly critical semiotic analysis must interrogate these criteria, asking not only what is measured, but whose values and purposes the measurements serve.

Together, these dimensions form what could be called a *semiotic ecology of intelligence*: an approach that treats intelligence not as computation, but as an emergent property of sign-processing systems embedded in environments, oriented toward goals, and constrained by both internal norms and external demands.

Let us take, for instance, the case of a language model used in conservation biology to detect patterns in animal vocalizations. Its cognitive dimension includes data ingestion, classification, and correlation of vocal patterns with contexts (e.g., feeding, mating). Its normative dimension is shaped by the ethical commitments of the project: non-invasiveness, respect for animal autonomy, and data transparency. Its performative dimension is structured by benchmarks such as accuracy, interpretability, and potential contribution to species protection. However, a biosemiotic perspective reminds us that such a model does not inhabit the Umwelt of the animal it analyzes. It mediates, approximates, and transforms its communicative acts into human-readable form. Hence, any apparent "translation" is in fact

a transduction, filtered through the ontological gap between symbolic machines and embodied life (Krzanowski & Polak 2022).

Intelligence can be redefined as the semiotic capacity to maintain a meaningful relation to the environment through acts of interpretation, normativity, and situated performance. From this perspective, artificial intelligence does not replicate human cognition or animal communication. It reconfigures the semiotic field within which such relations become thinkable, actionable, and contestable. This calls not only for descriptive clarity but for a critical semiotics of intelligence: one which resists essentialism and embraces the plurality, plasticity, and ethical stakes of intelligence across domains.

As Jesper Hoffmeyer and Kalevi Kull (2003) demonstrated, intelligence cannot be separated from semiosis. It has evolved historically as an interpretive capacity which is situated, corporeally distributed, and directed towards survival. In biosemiotic terms, intelligence is not a computational function but an embodied semiotic competence. It is an inherently relational and contextual ability to read, interpret, and respond to signs in the environment. From this perspective, every living being participates in a form of intelligence which expresses itself through the construction of its own *Umwelt*, a meaningful and situated world (Uexküll 1940). AI, by contrast (as discussed in the functionalist and behaviorist account offered by Alawijeh and Al-Mesilini (2025)) simulates or reproduces certain cognitive functions without possessing intentionality or intrinsic semantic experience. It seems that due to a lack of embodiment and desire, AI cannot exercise semiotic freedom, understood as the capacity to assign meaning in a creative<sup>8</sup>, selective, and context-sensitive way. However, precisely because it is non-living, AI can nevertheless play an epistemological role: not as an autonomous cognitive agent, but as an augmented semiotic instrument, capable of identifying patterns, structures, and regularities in animal communication and organismal behavior<sup>9</sup>.

According to the theory of the extended mind (Clark & Chalmers 1998; Heinrichs 2020), artificial devices can be considered integral components

---

<sup>8</sup> The debate is indeed complex, nuanced, and often contradictory. For this reason, we neither wish nor claim to explore it in depth here. However, we remain convinced that the current proliferation of articles on AI and creativity will, sooner or later, converge toward a clearer position, whether by rearticulating the very notion of creativity, or by revealing that it is a paradigm no longer adequate to contemporary discourse. (See du Sautoy 2019; Hartley, Ibrus, and Ojamaa 2021)

<sup>9</sup> Just as it is currently doing. In fact, the idea of this article is not to criticize the function of AI but the sensationalist interpretation of the results.

of cognitive processes, if they meet the criteria of reliability, procedural transparency, and functional integration. In this framework, AI is not something external to us, but rather a distributed element of our cognitive system. It can become part of our mind insofar as it is embedded within embodied and relational interpretative practices. As Heinrichs (2020) observes, this interface between biological cognition and artificial intelligence gives rise to a new cognitive ontology, hybrid and nested cognitive systems which encompass both the human and the artificial.

This entails a profound epistemological reconfiguration. AI should not be used to literally translate animal communication, but rather to extend our interpretive capacity and make visible aspects of non-human semiosis which would otherwise remain obscure (Delahaye 2021). In order to achieve this, however, a meta-interpretive layer is required, one that re-anchors algorithmic processing within an ecological and embodied reading of communication. The perspective of extended semiotics invites us to conceive of AI not as an autonomous mind, but rather as a form of technosemiosis, i.e. a process through which artificial devices become integrated within human cognitive circuits, thus contributing to the co-construction of new meaning-making practices. As emphasized by Nöth (2003), computers are involved in processes of internal “quasi-semiosis”, but genuine semiosis is manifested in their function as interfaces between human subjects. It is in this role that technological devices become cognitive and semiotic extensions of the human being. The machine does not interpret, it mediates. It is precisely in this mediating function that it becomes an integral part of a triadic semiotic circuit, making possible the amplification of living interpretive activity.

AI (understood as a semiotic machine) should not be regarded as an entity endowed with autonomous agency, but rather as a distributed node within an embodied and situated cognitive network. Technosemiosis, then, implies a material reciprocity between humans and artifacts, wherein interpretation arises not from the machine itself, but from the differential relationships between the signs it enables. In this context, AI can become an epistemological tool for the co-understanding of life, provided it is embedded within sense-making practices grounded in bodily, social, and environmental experience. Within this framework, AI does not merely simulate intelligence, but functions as a semiotic infrastructure (a plastic mediator that allows us to explore opaque zones of communication, especially in interspecies contexts). Its cooperative potential thus lies in its capacity to amplify human interpretive attention, extending the domain of the Umwelt

through a techno-cognitive mediation which remains anchored in the corporeality and embodied semiotic competencies of living beings.

For this reason, it is only by integrating AI into a biosemiotic model of cognition that we can use it not to replace the interpretive act of the living, but to enhance it. In this way, AI emerges as an ecological and critical tool for approaching (without claiming to fully translate) the communicative logics of other living beings. It is here that semiosis meets technology, and where intelligence, far from being reducible to computation, becomes an embodied practice of relation and listening.

### **3. The Illusion of Transparency and the Risk of Artificial Rewriting**

The growing ambition to build intelligent interfaces capable of “translating” animal communication through artificial intelligence is part of a broader cultural and semiotic tension: the drive to render the radical otherness of living beings transparent and accessible. However, what is commonly presented as “translation” is in fact an operative transformation which structurally alters the semiotic identity of the other, rewriting it in terms of human symbolic affordances and, even more so, according to the operational grammars of machines. This process is not neutral: it is deeply generative of new meanings and realities, as recent contributions in cultural semiotics, biosemiotics, and the philosophy of technology have shown.

In his article “The myth of meaning”, Claudio Paolucci (2025) argues that generative AI models are not merely linguistic tools, but cognitive constructs, technical-semiotic extensions which expose the already machinic nature of human thought. These “language-endowed machines,” as Paolucci calls them, reveal how human language itself is an assemblage effect, emerging from mediation among enunciators, norms, stereotypes, and discursive habits. However, if this holds true for the human domain<sup>10</sup>, what happens when these machines are employed to “give voice” to that which is not human? At this juncture, a decisive ontological difference emerges: AI can produce statements, but it cannot produce enunciation in the Peircean or Benvenistian sense. In the absence of an embodied “who” behind the act, and in the presence only of an algorithm simulating coherence, what is lost is precisely the semiotics of intentionality that characterizes both human language and animal communication.

At the same time, biosemiotician Ludmila Lacková Bennett, in her recent paper “A Biosemiotic Approach to AI” (2023), develops an original framework for understanding artificial semiosis based on the analogy with

<sup>10</sup> Paolucci’s article, in fact, deals with ChatGPT.

protein folding. Drawing on Howard Pattee's and Terrence Deacon's debate on symbol grounding, she proposes folding not only as a biochemical phenomenon but as a minimal semiotic operation, a foundational mechanism for modeling in both organic and artificial systems. In her view, intelligence, whether biological or artificial, should be approached not through computation alone, but through the lens of morphogenesis and structural plasticity. Lacková's model highlights that AI systems, rather than "interpreting" in a Peircean sense, engage in what she calls semiotic folding, a syntactic reconfiguration of inputs which lacks ecological embedding and lived intentionality. Every act of so-called translation by AI is not a neutral decoding of meaning but a folding of signs into an internal logic alien to the organismal world from which they originate. This results in a loss of the affective, performative, and context-sensitive dimensions which characterize animal communication. In short, folding reveals both the modeling potential and the semiotic limitation of artificial intelligence. It shows how AI may simulate the form of interpretation without accessing its ground. Thus, folding becomes a conceptual interface for contrasting digital abstraction with embodied semiosis.

Completing the perspectives examined thus far is the intercultural approach of Massimo Leone, who offers a refined articulation of the relationship with alterity in semiotic terms. In a 2019 essay, Leone identifies three escalating forms of cultural distance: otherness, extraneousness, and unawareness. While the first designates a recognizable alterity within pre-existing cultural codes, the second refers to a deeper form of estrangement which challenges our interpretive frameworks. The third (the most insidious) concerns those forms of difference which remain invisible, and unrecognized as such, because they have been naturalized or erased from our semantic horizons. When applied to the problem of translating animal communication through artificial intelligence, this typology demonstrates its full heuristic potential. AI tends to produce a reduction of animal alterity to a computable form of otherness, something which can be rendered legible in human terms via statistically significant patterns, while maintaining an appearance of fidelity to the source. The animal, in its semiotic complexity, is thus never truly encountered, but rather re-coded within the limits of the linguistic and symbolic output that the machine is capable of generating.

The effect of this operation is not an enhanced understanding, but rather an illusion of proximity. The dolphin (or any other non-human organism) becomes a simulacrum of comprehension, a domesticated alterity, seemingly accessible but in fact simplified to the point of erasing its irreducible

corporeal, temporal, and ethological strangeness. Leone defines this condition as a sophisticated form of cultural unawareness. It is a semiotic ignorance which not only fails to understand the other, but is no longer even capable of recognizing its own misunderstanding.

In his most recent contribution, Leone (2023) extends this diagnosis into a broader reflection on the semiotics of artificial intelligence, arguing that the task of the semiotician today is not merely to denounce the limitations of algorithmic simulation, but to understand AI as a new form of semiotic agency. AI, he suggests, does not produce truth in the classical sense, but rather effects of credibility. It is a rhetorical machine, a symbolic operator which selects, reifies, and structures horizons of meaning based on what it is able to calculate, and thus actively shapes our perception of the world. Leone thereby can be reread in a fruitful dialogue with Paolucci's analysis, which demonstrates how generative AIs function as machines of common sense and stereotypes, as well as with Lacková's approach, which proposes a model of artificial semiosis inspired by the plasticity of folding. Integrating these perspectives, Leone offers a compelling proposal: we should not reject AI as a deceptive simulacrum, but rather employ it as a differential revealer. It is a technology which precisely through its limits, signals what lies beyond computation: the undecidable, the ambiguous, the embodied, the relational.

In this vision, AI can become an ally in the construction of a critical metasemiosis, capable of reorienting our attention toward that which remains opaque, and which, for this very reason, demands to be interrogated. Rather than providing a faithful translation of the living, AI can offer an index of our projections, our biases, and the epistemic blind spots pervading our relationship with animal otherness. It is within this zone of tension that a new *ecotechnical semiotics* might emerge: a practice of differential listening. One which coexists with alterity without absorbing it, and that accepts the untranslatable not as a failure, but as the very condition of any genuine semiotic relation.

This scenario is further reinforced by the concept of the ontological gap: between the internal representation of reality constructed by AI and that generated by living beings lies a radical epistemological rupture. AI does not inhabit the world. It has no biological needs, and does not produce intentional judgments. Consequently, any attempt to make AI "speak" on behalf of non-human life risks generating a synthetic ontology, one that translates but simultaneously erases the semiotic specificity of the other. Only a truly biosemiotic approach on corporeality, environment, intentionality, and purposiveness, can partially bridge this gap. Yet such an ap-

proach necessarily requires abandoning the idea of direct translation and embracing a paradigm of differential mediation.

What we can do, therefore, is not to speak with animals through AI, but to listen and build interfaces which do not translate, but rather multiply the forms of interspecific resonance and attentiveness. Instead of designing AIs which “mimic” or “interpret” animal language in human terms, we might envision interfaces which bend to the rhythm of the living, as in folding. These are not representations, but dynamic semiotic models which allow themselves to be reshaped by contact. In this sense, Lacková’s proposal may be read as a biosemiotic pathway towards ecotechnical design: AI not as translators, but as facilitators of encounters capable of receiving alterity without reducing it. The translation of the non-human should not seek equivalence between codes, but rather the construction of zones of contact in which semiosis emerges as a relational event. AI can be useful if it ceases to speak on behalf of other living beings and begins instead to modulate its activity according to a logic of semiotic hospitality. Only in this way can we move from the dream of “deciphering” to the reality of shared listening.

### **Conclusion**

The trajectory outlined in this study has moved from a critique of the ideology of “translating the wild”, exemplified by the case of DolphinGemma, to a radical questioning of the epistemological and semiotic viability of representing non-human communicative forms through the infrastructures of artificial intelligence. What is emerging with increasing clarity is that AI cannot be conceived as a neutral bridge between different semiotic worlds, but rather as an actantial device that transforms every form of communication passing through its mesh. Every act of “translation” performed by AI is, in fact, a differential transduction, a production of meaning that necessarily involves a rewriting of the referent.

Yet, as Massimo Leone (2023) suggests, the task of a semiotics of artificial intelligence is not merely to unmask falsehoods, but also to understand the ideological, rhetorical, and material conditions through which AI produces effects of meaning which are credible, authentic, and culturally resonant. AI is not merely a simulation of intelligence: it is a semiotic machine which generates meaningful content according to logics different from those of biological systems. As such, it challenges traditional epistemologies of truth and indexicality, fully inscribing itself within the cultural genealogy of semiotic falsity.

From this perspective, it becomes possible to reconceptualize the relationship between AI and biosemiotics not as an irreducible opposition, but

as a paradoxical alliance. Biosemiotics teaches us that life itself is an interpretive process, a continuous activity of selection, choice, response, and meaning-making. AI, insofar as it lacks an *Umwelt*, cannot participate in this process in the full sense. Herein lies the space for an innovative theoretical proposal: a form of ecotechnical semiotics which no longer opposes artificiality and life, but instead recognizes in the sensitive mediation of artificial intelligence a potential not for translation, but for attentiveness. Artificial intelligence, in this vision, should not be deployed to “speak with animals,” but rather to construct interfaces of resonance and semiotic environments capable of allowing the emergence of relational, ecological, and perceptual patterns that humans and non-humans might co-inhabit. No longer AI as translator, but as a plastic infrastructure that bends – folds, in Lacková’s (2023) terms – into the semiosis of the living.

This proposal entails a radical epistemological reconfiguration: not a semiotics of meaning as transfer, but a semiotics of relation as co-emergence. Artificial intelligence, understood through its semiotic architecture, functions as a machine of falsification and learning, it can act as a reflective simulacrum of our relation to alterity (Leone 2023). Rather than seeking equivalence between codes, AI might be used to exercise difference, to generate new semiotic worlds in which understanding is not synonymous with mastery, but with transformative listening.

Biosemiotics, in this context, need not fear artificial intelligence. It must interrogate it, challenge its premises, and analyze its material and rhetorical effects with its own methodologies. If AI is the technological mirror of our desire to understand, then biosemiotics can become the discipline which deconstructs and reconfigures that desire (not to speak for the other, but to learn how to speak with the other). The translation of the wild, so reconceived, is no longer a technological feat, but an ethical imperative: to build the semiotic conditions for encounters that do not collapse difference, but allow it to resonate.

From this critical alliance between AI and biosemiotics, a renewed semiotics may emerge: one in which artificial intelligence, reimagined as a technology of attentional ecology, contributes to the creation of shared perceptual landscapes. In this framework, interspecies communication might cease to be a matter of decoding and become instead a process of co-sensing, of attunement to the resonant frequencies of life which elude symbolic capture. A truly biosemiotic AI would not be designed to speak, but to listen, to be affected, to fold itself into the rhythms of the living, and to serve as a differential interface rather than a linguistic substitute. These are not merely speculative possibilities: they define the horizon of a new paradigm

in which technology no longer translates life into machine-readable terms, but becomes itself part of life's semiotic grammar.

## References

Alawijeh, M. 'A., S. Z. Al-Mesilini. 2025. "A Comparative Study of Artificial and Natural Intelligence from Ibn Sina's Perspective". *Journal of Eco-humanism*, Vol. 4, No. 1, 2879. Available at: <https://doi.org/10.62754/joe.v4i1.6082>.

Attard-Frost, A. 2024. Queering Intelligence: A Theory of Intelligence as Performance and a Critique of Individual and Artificial Intelligence. In Klipphan-Karge, M., A. Koster, A. M. dos Santos Bruss (eds.). *Queer Reflections on AI. Uncertain Intelligences*. London: Routledge, 23–39.

Bermant, P. C., M. M. Bronstein, R. J. Wood, S. Gero and D. F. Gruber. 2019. "Deep Machine Learning Techniques for the Detection and Classification of Sperm Whale Bioacoustics". *Scientific Report*, 9, 12588. Available at: <https://doi.org/10.1038/s41598-019-48909-4>.

Brooks, R. 1986. *Achieving Artificial Intelligence through Building Robots*. Cambridge, MA: MIT AI Lab.

Brooks, R. 1990. "Elephants Don't Play Chess". *Robotics and Autonomous Systems*, 6, 3–15.

Brooks, R. 1991. "Intelligence without Representation". *Artificial Intelligence*, 47, 139–159.

Cerrone, M., N. Mäekivi. 2021. "A Zoosemiotic Approach to the Transactional Model of Communication". *Semiotica*, 242, 39–62. Available at: <https://doi.org/10.1515/sem-2020-0052>.

CHAT: Is It A Dolphin Translator Or An Interface? *Wild Dolphin Project*. Available at: <https://www.wilddolphinproject.org/our-research/chat-research/> (accessed 8 May 2025).

Clark, A. 1997. *Being There*. Cambridge, MA: MIT Press.

Clark, A., D. Chalmers. 1998. "The Extended Mind". *Analysis*, Vol. 58, No. 1, 7–19. Available at: <http://www.jstor.org/stable/3328150>.

Deacon, T. 1997. *The Symbolic Species*. New York: Norton.

Delahaye, P. 2021. "Use of Quantitative Measures in Zoosemiotics: how Machines Are Becoming a New Pair of Ears and Eyes for Researchers". *Bi-*

osemiotics, Vol. 14, 287–294. Available at: <https://doi.org/10.1007/s12304-021-09440-0>.

Dondero, M. G. 2025. “Semiotics of Artificial Intelligence: Enunciative Praxis in Image Analysis and Generation”. *Semiotica*, No. 262, 111–146. Available at: <https://doi.org/10.1515/sem-2024-0195>.

Emmeche, C. 2001. “Does a Robot Have an Umwelt? Reflections on the Qualitative Biosemiotics of Jakob von Uexküll”. *Semiotica*, Vol. 134, No. 1/4, 653–693. Available at: <https://doi.org/10.1515/SEMI.2001.048>.

Ghoshal, A. 2025. Google AI learns to speak dolphin. *New Atlas*, 15 April 2025. Available at: <https://newatlas.com/biology/build-ai-translator-dolphins-dolphingemma/> (accessed 8 May 2025).

Gudwin, R. 1999. Umwelts and Artificial Devices. In da Costa, R., L. Santaella (eds.). *Seminário Avançado de Comunicação e Semiótica: Novos Modelos de Representação: Vida Artificial e Inteligência Artificial*. São Paulo: Pontifícia Universidade Católica, 51–56.

Hartley, J., I. Ibrus and M. Ojamaa. 2020. *On the Digital Semiosphere: Culture, Media and Science for the Anthropocene*. London: Bloomsbury.

Heinrichs, J. H. 2020. Artificial Intelligence in Extended Minds: Intrapersonal Diffusion of Responsibility and Legal Multiple Personality. In Beck, B., M. Kühler. (eds.). *Technology, Anthropology, and Dimensions of Responsibility*. Metzler: Stuttgart, 159–176.

Hoffmeyer, J. 1996. *Signs of Meaning in the Universe*. Bloomington: Indiana University Press.

Hoffmeyer, J. 2004. “Uexküllian Planmässigkeit”. *Signs Systems Studies*, Vol. 32, No. 1, 73–97.

Hoffmeyer, J., K. Kull. 2003. Baldwin and Biosemiotics: What Intelligence is for. In Weber, B. H., D. J. Depew. (eds.). *Evolution and Learning: The Baldwin Effect Reconsidered*. Cambridge, MA: MIT Press, 253–272.

Krzanowski, R., P. Polak, 2022. “Ontology and AI Paradigms”. *Proceedings*, Vol. 81, No. 1, 119. Available at: <https://doi.org/10.3390/proceedings2022081119>.

Kull, K. 2015. “Evolution, Choice, and Scaffolding: Semiosis is Changing Its Own Building”. *Biosemiotics*, Vol. 8, No. 2, 223–234.

Kull, K. 2018. “Choosing and Learning: Semiosis Means Choice”. *Signs Systems Studies*, Vol. 46, No. 4, 452–446.

Lacková, L. 2023. “A Biosemiotic Approach to AI. Folding as Semiotic Modelling”. (translated by A. Lyubimova). *Method*, Vol. 3, No. 4, 147–173.

Lafferty, E. 2025. On using Google Translate to speak ‘Dolphin’ to aliens. *Times of Malta*, 29 March 2025. Available at: <https://x2.timesofmalta.com/20250329/opinion/how-google-translate-has-allowed-us-to-speak-dolphin/> (accessed 8 May 2025)

Leone, M. 2019. Otherness, Extraneousness, and Unawareness in Inter-Cultural Semiotics. In Leone, M., B. Surace and J. Zeng. (eds.). *The Waterfall and the Fountain. Comparative Semiotic Essays on the Contemporary Art in China* (i saggi di Lexia, 34). Roma: Aracne, 9–29.

Leone, M. 2023. I compiti di una semiotica dell’intelligenza artificiale. In Santangelo, A., Leone, M. (eds.). *Semiotica e intelligenza artificiale* (i saggi di Lexia, 48). Roma: Aracne, 29–44.

Mann, J., R. C. Connor, P. L. Tyack and H. Whitehead (eds.). 2000. *Cetacean Societies*. Chicago: University of Chicago Press.

Maran, T. 2020. *Ecosemiotics: The Study of Signs in Changing Ecologies*. Cambridge: Cambridge University Press.

Maran, T., D. Martinelli and A. Turovski (eds.). 2011. *Readings in Zoosemiotics*. Berlin: Mouton De Gruyter.

Martinelli, D. 2011. “Let’s Get Physical! – On the Zoosemiotics of Corporeality”. *Bioseemiotics*, 4, 259–279. Available at: <https://doi.org/10.1007/s12304-010-9098-5>.

Nine, A. 2025. New Google LLM Aims to Translate Dolphin Language. *Extreme Tech*, 16 April 2025. Available at: <https://www.extremetech.com/science/new-google-llm-aims-to-translate-dolphin-language> (accessed 8 May 2025)

Nöth, W. 2001. “Semiosis and the Umwelt of a Robot”. *Semiotica*, Vol. 134, No. 1/4, 695–699. Available at: <https://doi.org/10.1515/semi.2001.049>.

Nöth, W. 2003. “Semiotic Machines”. *S.E.E.D. Journal*, Vol. 3, 81–99.

Paolucci, C. 2025. “The Myth of Meaning: Generative AI as Language-endowed Machines and the Machinic Essence of the Human Being”. *Semiotica*, 262, 5–23. Available at: <https://doi.org/10.1515/sem-2024-0204>.

du Sautoy, M. 2019. *The Creativity Code: How AI Is Learning to Write, Paint and Think*. London: Fourth Estate

Sebeok, T. A. 1972. *Perspective in Zoosemiotics*. Berlin: Mouton de Gruyter.

Sebeok, T. A. 1977. *How Animals Communicate*. Bloomington: Indiana University Press.

Sebeok, T. A. (ed.). 1990. *Essays in Zoosemiotics*. Toronto: Toronto Semiotic Circle.

Sebeok, T. A. 2001. "Biosemiotics: Its Roots, Proliferation, and Prospects". *Semiotica*, Vol. 134, No. 1/4, 61–78.

Shannon, C., W. Weaver. 1948. *The Mathematical Theory of Communication*. Illinois: Urbana.

Sharov, A. 2013. Minimal Mind. In Swan, L. (ed.). *Origins of Mind*. New York: Springer, 343–360.

Sharov, A., T. Vehkavaara. 2015. "Protosemiosis: Agency with Reduced Representation Capacity". *Biosemiotics*, Vol. 8, No. 1, 103–123.

Smith, J. D., J. Schull, J. Strote, K. McGee, R. Egnor and L. Erb. 1995. "The Uncertain Response in the Bottlenosed Dolphin (*Tursiops truncatus*)". *Journal of Experimental Psychology: General*, Vol. 124, No. 4, 391–408. Available at: <https://psycnet.apa.org/doi/10.1037/0096-3445.124.4.391>.

Smith, J. D., W. E. Shields, J. Schull and D. A. Washburn. 1997. "The Uncertain Response in Humans and Animals". *Cognition*, Vol. 62, No. 1, 75–97. Available at: [https://doi.org/10.1016/S0010-0277\(96\)00726-3](https://doi.org/10.1016/S0010-0277(96)00726-3).

Tønnessen, M. 2022. "The Evolutionary Origin(s) of the Umwelt". *Biosemiotics*, Vol. 15, 451–455.

Uexküll, J. v. 1909. *Umwelt und Innenwelt der Tiere*. Berlin: Springer.

Uexküll, J. v. 1940. *Bedeutungslehre*. Leipzig: Verlag von J. A. Barth.

Vidales, C., J. Horta. 2024. "A Philosophical and Cybersemiotic Reading of Von Uexküll's Umwelt Theory". *Biosemiotics*, Vol. 17, 319–339. Available at: <https://doi.org/10.1007/s12304-024-09572-z>.

Weizenbaum, J. 1976. *Computer Power and Human Reason: From Judgment to Calculation*. San Francisco: W. H. Freeman & Co.

Zengiaro, N. 2023a. "Latour and Biosemiotics. The Hybrid Notion of Life". *E|C. Rivista dell'Associazione Italiana di Studi Semiotici*, 37, 130–145.

Zengiaro, N. 2023b. "The Circulation of Meaning: A Biosemiotic Perspective on the Functional Circle". *Punctum. International Journal of Semiotics*, Vol. 9, No. 2, 209–225. Available at: <http://dx.doi.org/10.18680/hss.2023.0026>.

Ziemke, T. 2001a. "The Construction of 'Reality' in the Robot". *Foundations of Science*, Vol. 6, No. 1, 163–233.

Ziemke, T. 2001b. Are Robots Embodied? In Balkenius, C., J. Zlatev, C. Brezeal, K. Dautenhahn, H. Kozima, (eds.). *Proceedings of the First International Workshop on Epigenetic Robotics: Modelling Cognitive Development in Robotic Systems*. Lund: Lund University Cognitive Studies, 75–83.

## CHATGPT IN HIGHER EDUCATION: A SEMIOTICS INVESTIGATION BETWEEN CULTURAL EXPLOSION AND ENCYCLOPEDIA KNOWLEDGE

*Danilo Petrassi*  
*University of Macerata*  
*d.petrassi@unimc.it*

### **Abstract**

Generative Artificial Intelligence, and particularly ChatGPT, today represents a crucial moment in the evolution of knowledge production and dissemination in education, echoing what Juri Lotman describes as a “cultural explosion.” As AI-generated texts flood academic and creative spaces, the question arises today is: how does this affect research methodologies and pedagogical frameworks? This paper explores the transformative role of ChatGPT in higher education, positioning it as both a disruptive force and an innovative collaborator in the knowledge-making process. Through the perspective of semiotics and Umberto Eco’s “encyclopedic model”, ChatGPT’s functioning mirrors human interpretative processes – drawing upon a vast corpora of texts, identifying patterns, and generating plausible continuations within cultural discourse. However, unlike human scholars, ChatGPT lacks intentionality, challenging traditional epistemological models which rely on

authorial agency and context-dependent inference. This research investigates how ChatGPT's generative capacities align with the rhizomatic structure of knowledge, where meaning is not linear but formed through a dynamic network of associations. However, today in higher education ChatGPT is redefining the role of teachers and learners, transforming classrooms into dialogic spaces where GenAI acts as a semiotic mediator rather than a mere tool. By facilitating inferential learning – where students engage critically with AI-generated outputs – teachers can cultivate deeper meta-cognitive awareness. In order to investigate how students engage with ChatGPT as a semiotic mediator in education, this study employs a mixed-methods approach, integrating qualitative analysis with a quantitative survey. The qualitative phase explores the perceptions of students regarding the role of AI in learning through a semiotic investigation, examining its impact on meaning-making, epistemological challenges, and pedagogical transformation. The quantitative phase consists of a survey conducted among university students across different departments (N=20). It assesses ChatGPT usage patterns, trust levels, critical evaluation behaviors, and concerns regarding misinformation and institutional regulation. The survey findings reveal that while students frequently use ChatGPT for learning and academic purposes, they exhibit a balanced approach of trust and skepticism – engaging in critical cross-checking of AI-generated content. Despite recognizing ChatGPT's efficacy in simplifying complex topics, students do not widely use it for deep cognitive engagement or reflective academic discussions. Concerns over bias and misinformation remain significant, and while many acknowledge the need for institutional guidelines, there is also a prevailing optimism about GenAI's future role in education. These findings suggest that ChatGPT is perceived not as a replacement for traditional learning structures, but as a tool requiring critical literacy and careful mediation. This paper argues that rather than replacing human intellectual labor, ChatGPT nowadays amplifies the cultural explosion by accelerating the translation of knowledge, making higher education both more accessible and more complex than ever before.

**Keywords:**

GenAI; Cultural Explosion; Encyclopedic Knowledge; ChatGPT in Higher Education; Semiotic Translation

**1. Introduction**

The advent of Generative Artificial Intelligence (GenAI) has precipitated a paradigm shift in knowledge production and dissemination across various

domains (Henriksen et al. 2025). Among these GenAI models, ChatGPT has emerged as a prominent tool, capable of generating human-like text based on vast datasets. This development has profound implications – not only for common daily purposes – for academic research and higher education (Haim et al. 2025), prompting a reevaluation of traditional methodologies and pedagogical practices. GenAI refers to a class of artificial intelligence models designed to create new content, including text, images, and audio, by learning patterns from existing data. These models, particularly Large Language Models (LLMs) such as ChatGPT, have demonstrated remarkable proficiency in generating coherent and contextually relevant text, thereby influencing various facets of knowledge production. In the realm of academic research in higher education, GenAI has been harnessed to expedite literature reviews, synthesize research findings, and even generate hypotheses (Hofeditz et al. 2025). For example, Google’s development of an AI “co-scientist” is aimed at accelerating biomedical research by identifying knowledge gaps and proposing novel scientific ideas (Weng et al. 2025). Similarly, platforms like Semantic Scholar use AI to provide concise summaries of scholarly papers, enhancing accessibility and comprehension (Kirstein et al. 2025).

However, the integration of GenAI into academic workflows is not without challenges. Concerns have been raised regarding the potential for AI-generated content to introduce biases, inaccuracies, and ethical dilemmas. A study highlighted the fact that GenAI models like ChatGPT can produce fabricated or misleading references, posing risks to the integrity of scientific literature (Alkaissi & McFarlane 2023; Ambrosio et al. 2023). Moreover, the proliferation of AI-generated content has led to debates about the authenticity and originality of scholarly work, as exemplified by discussions on GenAI’s role in legal scholarship (Sloan 2025), fostering an explosive – digital – culture.

Juri Lotman, a seminal figure in semiotics, introduced the concept of a “cultural explosion” to describe periods of rapid and transformative change within cultural systems (Lotman 2009). According to Lotman, such explosions result in the accelerated creation and dissemination of new information, leading to a reconfiguration of existing cultural paradigms:

We are immersed in the space of language. Even in the most basic abstract conditions, we cannot free ourselves from this space, which simply envelops us, and yet it is a space of which we are also a part and which, simultaneously, is part of us.

(Lotman 2009: xiii)

Today the emergence of GenAI can be viewed through Lotman's cultural explosion. AI-generated content has inundated various cultural and academic spaces, challenging traditional notions of authorship, creativity, and intellectual authority. The sheer volume and speed at which GenAI can produce content mirror the dynamics of a cultural explosion, necessitating a reevaluation of how knowledge is constructed and validated in the digital age. This phenomenon has prompted scholarly discourse on the implications of GenAI in culture and society. Such discussions underscore the transformative impact of ChatGPT and similar tools on cultural processes and the urgency of developing frameworks to understand and manage this impact. In light of these developments, this study seeks to address the following unsolved research question: How does ChatGPT influence higher education research methodologies and pedagogical practices? This inquiry aims to elucidate the dual role of ChatGPT as both a disruptive force and an innovative collaborator in the realms of research and education. In order to analyze the impact of ChatGPT, this study uses a semiotic framework, based on Umberto Eco's encyclopedic theory of knowledge, which defines culture as an interconnected network of signs and meanings, in which each element derives its meaning from its relations within the system (Eco 1984). This model provides a context through which to examine how ChatGPT, as a GenAI tool, processes and produces text by navigating a vast corpora of information, identifying patterns, and generating contextually plausible continuations. By mirroring human interpretative processes, ChatGPT engages in a form of semiotic translation, converting input data into meaningful outputs within specific cultural and contextual frameworks. However, unlike human agents, ChatGPT operates without intentionality – or consciousness – raising questions about the nature of meaning-making and the role of authorial agency in AI-generated content. A semiotic analysis facilitates a deeper understanding of these dynamics, offering insights into the epistemological and ethical considerations of integrating GenAI into academic and higher educational practices.

## **2. Theoretical Framework**

The integration of GenAI into academic and educational spheres necessitates a theoretical framework to comprehend its multifaceted impact. This section examines Juri Lotman's concept of the cultural explosion and Umberto Eco's encyclopedic theory of knowledge, providing a standpoint to analyze ChatGPT's role in contemporary knowledge production and dissemination.

Juri Lotman introduced the notion of the cultural explosion to describe periods of rapid and transformative change within cultural systems. In his seminal work, *Culture and Explosion*, Lotman posits that cultural evolution is not a steady, linear process but is characterized by sudden, discontinuous shifts which redefine societal norms and values (Lotman 2009). Lotman's theory challenges the traditional view of gradual cultural development by emphasizing the role of unpredictable events that catalyze significant transformations. He argues that these explosive moments are integral to the dynamic nature of culture, serving as catalysts for innovation and the reconfiguration of existing structures.

This perspective aligns with the current proliferation of GenAI technologies which have rapidly permeated various sectors. This has led to a reevaluation of established practices and epistemologies. The relevance of Lotman's cultural explosion in the context of GenAI is evident in the accelerated integration of AI-generated content into daily life. The sudden ubiquity of tools such as ChatGPT has disrupted traditional modes of communication, creativity, and information dissemination, prompting both enthusiasm and concern regarding their long-term implications. This phenomenon exemplifies Lotman's assertion that cultural explosions, while disorienting, are pivotal in driving societal progress and adaptation:

To the contemporary man, explosion as a phenomenon [...] has come to be associated with ideas of devastation and has turned into a symbol of destruction. But if, at the core of our contemporary representations, there lay the kind of associations that existed during periods of great openness such as the Renaissance or in art in general then our understanding of the concept of explosion would evoke in us such phenomena as the birth of a new living creature or any other creative transformation of the structure of life.

(Lotman 2009: 19)

Umberto Eco offers a complementary perspective through his concept of the encyclopedia in semiotic theory. Eco envisions the encyclopedia as a multidimensional space of semiosis – a complex system of shared knowledge which governs the production and interpretation of signs within communicative contexts (Desogus 2012). This model underscores the interconnectedness of cultural codes, experiences, and texts, suggesting that meaning is derived from the intricate web of associations within a cultural framework. In Eco's view, the encyclopedia is not a static repository of in-

formation but a dynamic, evolving network which reflects the collective understanding of a community. It encompasses the totality of knowledge, including denotations, connotations, and cultural nuances, thereby facilitating the interpretation of signs in a context-dependent manner (Eco 1984). This perspective is particularly pertinent in analyzing GenAI models such as ChatGPT which operate by drawing upon vast datasets to generate contextually relevant outputs.

Eco's model also addresses the interpretative processes involved in semiosis, highlighting the role of the "Model Reader" – an ideal interpreter equipped with the necessary cultural and contextual knowledge to decode texts as intended by the author: An open text outlines a "closed" project of its Model Reader as a component of its structural strategy (Eco 1979: 9). In the case of AI-generated content, the absence of an intentional author raises questions about the applicability of this concept, prompting a reevaluation of interpretative frameworks in the age of artificial – knowledge – intelligence.

Integrating Lotman's and Eco's theories provides a comprehensive framework for examining the impact of ChatGPT on contemporary culture and higher education. On the one hand Lotman's concept of the cultural explosion elucidates the disruptive yet transformative potential of GenAI technologies, framing them as catalysts for rapid cultural evolution. On the other, Eco's encyclopedic model offers insights into the mechanisms of meaning-making in AI-generated content, emphasizing the importance of contextual and cultural knowledge in interpretation. By synthesizing these perspectives, we can better understand how ChatGPT functions within the semiosphere: the semiotic space where cultural exchanges occur (Lotman 2005). ChatGPT, as a product of GenAI, contributes to the semiosphere by generating new texts which interact with existing cultural narratives, thereby influencing the continuous evolution of meaning within society. Moreover, this integrated framework allows for a critical examination of the epistemological and ethical implications of GenAI in knowledge production. The absence of intentionality in AI-generated content challenges traditional notions of authorship and authority, necessitating new approaches to evaluating the credibility and validity of information. Additionally, the potential for biases embedded within GenAI models underscores the need for a vigilant and reflective engagement with AI technologies, ensuring that they augment rather than undermine the integrity of cultural and academic practices.

The theoretical insights derived from Lotman and Eco have profound implications for academic research methodologies and pedagogical prac-

tices in the context of GenAI integration. Understanding ChatGPT through the cultural explosion's theory highlights the necessity for adaptability in research approaches, encouraging scholars to embrace innovative methodologies which leverage AI's capabilities while remaining critical of its limitations. Yet, in higher educational settings, Eco's encyclopedic model emphasizes the importance of fostering comprehensive cultural literacy, especially among students. As AI-generated content becomes more prevalent, educators are tasked with equipping learners with the skills to navigate and interpret this information critically. This involves cultivating an awareness of the underlying cultural codes and contexts which shape AI outputs, thereby promoting a more nuanced and informed engagement with technology. Furthermore, the integration of ChatGPT into academic and educational environments necessitates a reevaluation of the roles of teachers and learners. Educators must increasingly be positioned as mediators who guide students in critically engaging with AI-generated content, fostering an environment of inferential learning where learners actively construct meaning through interaction with technology. This shift aligns with contemporary pedagogical models which emphasize collaborative and dialogic learning experiences, preparing students to navigate the complexities of a digitally mediated world. The deployment of ChatGPT and similar GenAI models raises significant ethical considerations which must be addressed within this theoretical framework. The potential for GenAI to generate biased or misleading information necessitates the development of robust evaluative criteria to assess the credibility of GenAI outputs. This involves not only technical solutions, such as refining algorithms to mitigate bias, but also educational initiatives that promote critical digital literacy among users.

### **3. ChatGPT as a Semiotic Mediator in Education**

Beyond serving as mere tools for information retrieval or content generation, ChatGPT can be conceptualized as a semiotic mediator: an entity which facilitates the interpretation and construction of meaning within educational contexts. Semiotics provides a framework for understanding how meaning is constructed and interpreted. It extends beyond mere linguistic structures to encompass the ways in which individuals interact with and derive meaning from their environment, incorporating the cultural, social, and cognitive dimensions (Peirce 1931; Eco 1979). Within educational contexts, semiotic mediation plays a crucial role in facilitating knowledge construction, since it enables learners to navigate, interpret, and internalize information through signs, symbols, and contextual frameworks (Danesi

2007). In particular, Lev Vygotsky's (1978) sociocultural theory underscores the significance of semiotic mediation in cognitive development. He posits that learning occurs through interactions with cultural tools and symbolic systems, such as language, writing – and digital media today – , which structure understanding and problem-solving skills. According to Vygotsky, mediated learning experiences – where an external agent, whether a teacher, peer, or tool, assists in cognitive development – are fundamental in shaping higher-order thinking skills. In this sense, technological advancements, including GenAI-driven systems like in the case of ChatGPT, can be understood as contemporary semiotic mediators that extend beyond traditional instructional materials.

ChatGPT, as a language-based GenAI, operates within this paradigm by generating text with which learners can engage, thereby functioning as a mediator in the learning process. It acts as a dynamic and interactive knowledge source, offering contextualized information which users can interpret, critique, and refine based on their prior knowledge and learning goals (Luckin 2018). Unlike static learning resources such as textbooks, ChatGPT responds to user queries in real time, adapting its outputs to the linguistic and conceptual frameworks embedded in the interaction. This aligns with Eco's (1979) encyclopedia theory, conceptualizing learning as an evolving, networked process of sign interpretation rather than a linear accumulation of facts:

Since the semantic encyclopedia is in itself potentially infinite, semiosis is unlimited, and, from the extreme periphery of a given sememe, the center of any other could be reached, and vice versa. Since every proposition contains every other proposition, a text could generate, by further semantic disclosures, every other text.

(Eco 1979: 24)

Moreover, the mediation of ChatGPT extends beyond direct knowledge transmission to fostering metacognitive skills. When learners engage in dialogues with GenAI, they are not merely passive recipients of information but active constructors of meaning, evaluating AI-generated content for accuracy, relevance, and coherence (Holmes et al. 2022). This process resonates with the constructivist paradigm of learning which emphasizes active engagement, inquiry-based exploration, and the co-construction of knowledge (Jonassen 1999). ChatGPT's ability to provide alternative explanations, clarify misunderstandings, and simulate Socratic questioning further positions it as a powerful semiotic mediator, capable of guiding learners through complex cognitive landscapes.

However, the effectiveness of ChatGPT as a semiotic mediator is contingent upon the user's critical engagement and digital literacy skills. Unlike human instructors, ChatGPT lacks intentionality and contextual awareness, relying instead on probabilistic modeling to generate responses based on existing textual patterns (Bender et al. 2021). This limitation underscores the importance of fostering GenAI literacy among learners, ensuring that they possess the analytical skills necessary to assess the credibility and validity of AI-generated content (Lee 2021). By integrating ChatGPT as a learning tool (Petrassi 2024) within structured pedagogical frameworks which encourage reflection and critical analysis, educators can harness its potential as a semiotic mediator while mitigating risks associated with misinformation and over-reliance on AI-generated knowledge.

Recent empirical research has investigated the practical applications and implications of ChatGPT in educational settings, highlighting its potential to support learning, enhance engagement, and reshape pedagogical methodologies. Recently, Rezai, Namaziandost, and Hwang (2024) conducted a phenomenological study exploring how ChatGPT facilitates second-language (L2) learning among university students. Their study revealed that the integration of ChatGPT as a learning tool significantly enhanced student engagement and motivation, particularly among learners who previously exhibited lower participation in class activities. The GenAI models provided students with real-time responses to inquiries, enabling them to refine their research questions and develop structured approaches to problem-solving. Furthermore, ChatGPT played a significant role in alleviating teacher workload by automating instructional tasks, such as providing formative feedback and summarizing complex concepts. Educators found that this enabled them to dedicate more time to higher-order instructional strategies, such as fostering discussion and critical analysis (Rezai et al. 2024). These findings align with broader research on AI-assisted education, which suggests that intelligent tutoring systems (ITS) and conversational agents can function as cognitive scaffolding tools, supporting metacognitive processes of students and self-regulated learning (Roll & Wylie 2016). By engaging in iterative dialogue with GenAI, students often refine their understanding of topics through an adaptive learning process, which aligns with constructivist learning theories (Piaget 1954; Vygotsky 1978). However, Rezai, Namaziandost, and Hwang (2024) also caution that over-reliance on AI-generated content may lead to a reduction in students' independent critical thinking abilities if not guided appropriately by educators.

In higher education, the role of ChatGPT has been examined through a variety of empirical studies, particularly in the context of student-Ge-

nAI interaction analytics. Chen et al. (2024) developed StuGPTViz, a visual analytics system designed to analyze student-ChatGPT interactions. This study involved collecting conversational data from 48 university students over the course of a semester, focusing on how students engaged with ChatGPT in different learning scenarios. The results indicated that ChatGPT could effectively support learning by providing personalized assistance, fostering critical thinking skills, and acting as an interactive study aid. Students who engaged with ChatGPT for research-related tasks reported increased confidence in articulating their arguments and structuring their academic writing, particularly when using GenAI as a brainstorming tool (Chen et al. 2024). A key feature of StuGPTViz was its ability to track and assess the quality of GenAI-student interactions, allowing educators to monitor the depth of inquiry and engagement. This feature aligns with ongoing discussions in educational research about the need for transparency and accountability in GenAI-driven learning environments (Holmes et al. 2022). By analyzing patterns in student-GenAI conversations, StuGPTViz provided insights into the types of cognitive processes activated during GenAI-assisted learning. For example, students who asked ChatGPT open-ended questions and engaged in iterative dialogue seemed to exhibit stronger analytical reasoning skills compared to those who used the tool for superficial question-answering tasks (Chen et al. 2024). The effectiveness of GenAI in promoting critical thinking has been a subject of debate. While some studies highlight ChatGPT's ability to enhance inquiry-based learning (Zawacki-Richter et al. 2019), others caution that AI-generated responses may lead to cognitive passivity, if students do not actively interrogate the information provided (Chen & Gong 2025). This underscores the importance of designing GenAI-integrated curricula which encourage students to use ChatGPT as a supplementary rather than a primary knowledge source. In this regard, instructors must play a crucial role in modeling effective GenAI utilization strategies, such as prompting students to critically evaluate AI-generated responses, cross-reference information, and engage in reflective writing exercises. Another significant aspect of GenAI integration in higher education is its potential to address disparities in access to academic support. Traditional tutoring and writing support services often have limitations in terms of availability and accessibility, particularly for students from underrepresented backgrounds (Ren 2023). ChatGPT, as an always-available learning resource, has the potential to democratize access to personalized academic support, particularly for students in large-scale university settings where individualized attention from instructors may be limited. However, issues related to bias, misinformation, and eth-

ical AI use must be addressed to ensure equitable and responsible GenAI deployment in educational institutions (Bender et al. 2021).

The conceptualization of ChatGPT as a semiotic mediator carries significant implications for pedagogical practices and learning outcomes. As a mediator, ChatGPT can facilitate inferential learning, where students engage in critical analysis and interpretation of AI-generated content. This process encourages active learning, since students must assess the relevance and accuracy of the information, integrate it with existing knowledge, and apply it to problem-solving scenarios. However, the integration of ChatGPT into education also presents challenges. Concerns have been raised regarding the potential for GenAI to spread biased or misleading information, which can adversely affect learning if not properly addressed. Educators are thus tasked with guiding students in developing critical digital literacy skills, enabling them to discern and evaluate the quality of AI-generated content effectively (Emdad et al. 2023). Moreover, the use of ChatGPT necessitates a reevaluation of assessment methods. Traditional forms of evaluation may not adequately capture the depth of understanding and critical engagement fostered through interactions with GenAI. Alternative assessment strategies, such as reflective essays and project-based evaluations, may be more appropriate in measuring the learning outcomes associated with GenAI-mediated in higher education.

#### **4. Methodology**

This study adopts a mixed-methods approach, combining qualitative analysis and quantitative data collection through a structured survey. The rationale for this approach is to provide a comprehensive understanding of ChatGPT's role as a semiotic mediator in higher education, ensuring that both subjective experiences and measurable trends are captured. A mixed-methods framework allows for triangulation, enhancing the robustness and credibility of the study's findings by validating data through multiple sources (Creswell & Plano Clark 2018). The qualitative component of this study is designed to explore how and why users engage with ChatGPT in educational settings. Through in-depth analysis, this section examines the nature of user interactions, the cognitive and pedagogical implications of GenAI mediation, and the epistemological challenges which arise from integrating ChatGPT into knowledge production. The quantitative component, which follows the qualitative analysis, consists of a structured survey designed to measure users – university students from different degree courses – perceptions, engagement patterns, and attitudes toward GenAI's role in higher education. This dual approach is essential for capturing both

interpretative and empirical dimensions of the purpose of the study. While qualitative analysis provides insights into patterns of meaning-making and critical engagement, the questionnaire facilitates scalability and generalizability, ensuring that the study's findings are not solely dependent on individual cases.

#### **4.1. Qualitative Analysis: ChatGPT as a Semiotic Mediator in Education**

The qualitative analysis in this study focuses on user experiences, epistemological concerns, and pedagogical shifts resulting from ChatGPT's integration into higher educational settings. The analysis is structured around the following themes:

##### ***4.1.1. Meaning-Making and Inferential Learning***

One of the primary functions of ChatGPT as a semiotic mediator is its ability to support inferential learning, wherein students construct meaning through iterative dialogue. Unlike traditional knowledge sources such as textbooks or academic articles, ChatGPT engages users in interactive learning, dynamically responding to queries, refining explanations, and even challenging preconceptions. Research by Chen et al. (2024) suggests that GenAI-mediated interaction fosters deeper engagement when learners critically reflect on responses rather than passively accept them. However, this process is not without limitations – ChatGPT's lack of intentionality and contextual awareness means that users must actively evaluate and cross-reference AI-generated content (Bender et al. 2021). The semiotic implications of this phenomenon align with Eco's (1984) encyclopedic model, where meaning is constructed through an evolving network of associations rather than through fixed interpretations.

##### ***4.1.2. Epistemological Challenges: Bias, Authority, and Trust***

Another key area of investigation is how university students negotiate trust and authority when engaging with ChatGPT in learning and educational settings. While ChatGPT provides access to vast amounts of synthesized knowledge, it does not "know" in the human sense – it merely predicts plausible text sequences based on probabilistic modeling (Floridi & Chiriatti 2020). This raises epistemological concerns regarding intellectual authority and the credibility of AI-generated knowledge. Several studies (Alkaissi & McFarlane 2023) have noted that users often perceive ChatGPT's outputs as authoritative, despite the model's tendency to produce factual errors or "hallucinations." This underscores the semiotic instability of AI-generated content – while ChatGPT provides structured

and coherent responses, its reliability is contingent on external verification and user discernment. Furthermore, as AI-generated knowledge becomes increasingly integrated into academic workflows, there is a growing need to rethink knowledge integrity and citation practices (Dave 2023). Should students be allowed to cite ChatGPT as a source? If so, under what conditions? The semiotic mediation framework suggests that AI should be viewed as a tool for knowledge translation rather than as an independent knowledge authority, reinforcing the importance of human oversight in GenAI-assisted learning.

#### ***4.1.3. Pedagogical Transformation: The Role of GenAI in Learning Spaces***

From a pedagogical standpoint, the presence of ChatGPT in educational settings has the potential to redefine traditional educator-student dynamics. Studies have shown that educators increasingly view GenAI as a co-facilitator, capable of personalizing instruction, generating discussion prompts, and assisting with formative assessment (Zawacki-Richter et al. 2019).

However, there remains an ongoing debate about the implications of GenAI-driven learning:

- Does GenAI enhance critical thinking by encouraging students to interrogate and refine their understanding?
- Or does it promote cognitive dependency, where students over-rely on AI-generated responses without deeper engagement?

Existing research suggests that the answer depends on how GenAI is integrated into the curriculum. When used as a tool for guided inquiry – where students must evaluate, compare, and contextualize responses – ChatGPT can foster deeper cognitive engagement. However, when used passively, it risks reinforcing surface-level learning, where students accept information uncritically (Holmes et al. 2022).

Thus, ChatGPT as a semiotic mediator does not replace human teachers but instead reframes the educational landscape, shifting emphasis toward critical engagement, digital literacy, and epistemological awareness.

#### **4.2. Quantitative Analysis: Survey About User Perceptions of ChatGPT in Education**

The following 10-item survey is designed to measure typical university student engagement, trust, and critical awareness in interactions with ChatGPT. Respondents will rate each statement on a 1 to 10 scale (1 = strongly disagree, 10 = strongly agree). The survey is targeted at frequent

university students (users) of ChatGPT in learning and educational settings.

1. How frequently do you use ChatGPT for educational or academic purposes? (1 = Never, 10 = Daily)
2. How much do you trust the accuracy of the responses of ChatGPT? (1 = Not at all, 10 = Completely)
3. To what extent do you critically evaluate and cross-check the responses of ChatGPT before using them in academic work? (1 = Never, 10 = Always)
4. How much has ChatGPT improved your ability to understand complex topics? (1 = Not at all, 10 = Significantly)
5. How dependent are you on ChatGPT for completing academic tasks? (1 = Not at all, 10 = Completely)
6. How often do you find yourself engaging in deeper discussions or reflections based on the responses of ChatGPT? (1 = Never, 10 = Very frequently)
7. How effective do you think ChatGPT is in supplementing traditional classroom learning? (1 = Not effective, 10 = Extremely effective)
8. How concerned are you about potential biases or misinformation in ChatGPT's responses? (1 = Not concerned, 10 = Extremely concerned)
9. How strongly do you believe that academic institutions should regulate or restrict the use of ChatGPT in coursework? (1 = Not at all, 10 = Strongly agree)
10. How optimistic are you about the future role of AI like ChatGPT in education? (1 = Not at all, 10 = Extremely optimistic)

### **4.3. Data Integration and Analysis Strategy**

This study will employ a sequential explanatory approach, where qualitative insights will be used to inform and contextualize the interpretation of quantitative findings. The integration of data will occur at two levels:

- **Survey Design:** Themes that emerged from the theoretical framework guided the refinement of survey questions to ensure they captured relevant and meaningful constructs.
- **Results Interpretation:** After collecting the data, qualitative insights helped me explain patterns, discrepancies, and unexpected trends in quantitative responses.

## **5. Evaluation of the Survey Results**

The survey results, based on responses from 20 university students from various degree programs, provide an initial quantitative assessment of the role of ChatGPT in higher education. Given the anonymity of the participants and the diverse academic backgrounds represented, the findings offer insightful trends regarding usage patterns, trust, critical evaluation, dependency, and attitudes toward GenAI in higher education. A qualitative interpretation of these results reveals both the perceived benefits and the concerns that students associate with ChatGPT.

### **5.1. Usage Patterns and Trust in ChatGPT**

The frequency of ChatGPT use for educational purposes showed a moderate level of engagement, with 25% of respondents selecting 5 on the scale. This suggests that while ChatGPT is an important educational and learning tool for many students, it has not yet reached ubiquitous daily usage among the surveyed group. The trust in ChatGPT's accuracy was moderately high, with 30% of respondents rating it at 6. This indicates that while students generally find ChatGPT reliable, they remain cautious about its accuracy, likely due to known issues such as factual inconsistencies and AI-generated "hallucinations" (Bender et al. 2021).

### **5.2. Critical Evaluation and ChatGPT's Role in Learning**

Encouragingly, 30% of respondents rated their cross-checking behavior at 8, suggesting that a significant portion of students actively verify AI-generated information before using it in academic work. This finding contrasts with concerns that GenAI tools might encourage passive learning – instead, the data suggests that many students maintain a critical stance when interacting with GenAI. However, the perception of ChatGPT's impact on learning complex topics was notably high, with 25% of respondents selecting 10. This highlights that, for some students, ChatGPT serves as an effective supplementary tool that enhances comprehension and engagement with difficult material.

### **5.3. AI Dependency and Cognitive Engagement**

The responses to ChatGPT dependency suggest a more nuanced relationship with AI. The most common response (31.6% selecting 1) indicates that many students do not rely heavily on ChatGPT for basic academic tasks. However, the responses regarding engagement in deeper discussions or reflections were lower, with 30% selecting 2, indicating that ChatGPT is not frequently used as a stimulus for critical discourse or extended intellectual reflection. This suggests that while students benefit from ChatGPT

for immediate information retrieval and clarification, it is less frequently leveraged as a tool for in-depth academic discourse.

#### **5.4. Perceived Effectiveness and Concerns About GenAI in Higher Education**

Regarding the effectiveness of ChatGPT in supplementing traditional classroom learning, responses were split between 6 and 7 (20% each), showing a moderately positive perception of GenAI's role in higher education. However, concerns about bias and misinformation were significant, with 35% of respondents selecting 7, indicating that many students remain skeptical about ChatGPT's reliability. This finding underscores the epistemological challenges of integrating GenAI into higher education, where students must balance convenience and accessibility with critical literacy and verification processes (Floridi & Chiriatti 2020).

#### **5.5. Institutional Regulation and Future Optimism**

The results also indicate a relatively high level of agreement with academic institutions regulating or restricting ChatGPT usage in coursework, with 20% of participants selecting 8. This suggests that while students recognize the potential of ChatGPT, they also acknowledge the need for academic oversight to prevent misuse or over-reliance. Despite these concerns, the future of GenAI in higher education is viewed optimistically, with 20% selecting 10 when asked about their outlook on the role of GenAI in learning. This optimism likely reflects a growing acceptance of ChatGPT as a transformative educational and learning tool, even as ethical and pedagogical challenges remain.

### **6. Discussion**

From a semiotic standpoint, the survey findings illustrate the complex interplay between trust, skepticism, and meaning-making in university students' engagement with ChatGPT. The balance between reliance and critical evaluation suggests that students do not passively absorb AI-generated knowledge but actively mediate their interactions with the tool, positioning themselves as both users and interpreters within the broader semiosphere (Lotman 1990). The mixed responses regarding trust and regulation reflect the paradox of GenAI as both a facilitator of knowledge expansion and a site of epistemological instability – an idea deeply embedded in Eco's encyclopedic model of knowledge (Eco 1984).

The data reveals that while ChatGPT significantly aids in understanding complex topics (25% rated its impact at 10), students also exhibit a heightened awareness of misinformation, with 35% expressing concerns about

biases (rating 7). This duality resonates with Lotman's theory of cultural explosion (2009) which posits that rapid technological transformations introduce both opportunities and disruptions in knowledge structures. The sudden accessibility of AI-generated responses reconfigures academic engagement, compelling students to renegotiate traditional notions of intellectual authority and credibility. In this sense, ChatGPT functions as a semiotic disruptor – a tool that challenges the stability of traditional knowledge hierarchies by generating content dynamically rather than relying on fixed, institutionalized sources. Eco's encyclopedia theory provides a crucial perspective for interpreting this phenomenon. Unlike conventional knowledge repositories, ChatGPT constructs responses by navigating a fluid network of textual associations, making its outputs inherently context-dependent and probabilistic. Eco (1979) emphasizes that an "open text" allows for multiple interpretations, requiring an active and competent Model Reader to engage in inferential learning. The survey data suggests that students, rather than passively accepting AI-generated information, enact the role of Model Reader by critically evaluating the responses of ChatGPT: 30% rated their cross-checking behavior at 8. This suggests that students recognize the need for contextual verification, mirroring Eco's assertion that knowledge is not merely retrieved but constructed through interconnected associations. However, the relatively low levels of cognitive engagement and discussion generation (30% rated 2) indicate that ChatGPT is not yet fully integrated as a dialogic learning partner. This is significant when viewed through Vygotsky's (1978) concept of semiotic mediation which posits that learning occurs through social and cognitive scaffolding facilitated by external tools. While ChatGPT offers immediate informational support, its ability to foster deep reflective discourse remains limited. This echoes concerns raised in contemporary GenAI literacy studies which warn that reliance on AI-generated content without structured critical engagement may lead to cognitive passivity (Holmes et al. 2022). The ambiguity in the stance of students toward institutional regulation (20% rating 8) further highlights the semiotic instability of AI integration in education. This aligns with Lotman's cultural explosion framework, wherein disruptive technologies necessitate new regulatory mechanisms to re-establish epistemic order. The debate over whether GenAI should be restricted, guided, or fully embraced mirrors historical shifts in media and knowledge systems, where emerging technologies have always triggered concerns over authorship, authenticity, and pedagogical legitimacy.

Despite these uncertainties, students exhibit an optimistic outlook on the future role of GenAI in education (20% rated 10), suggesting that while

current concerns persist, GenAI is largely perceived as a transformative tool rather than a threat. This optimism could be interpreted through Eco's encyclopedic vision which conceptualizes knowledge not as a finite construct, but as an ever-evolving network of signification. ChatGPT, by extending this network, presents both challenges and opportunities in re-defining how knowledge is produced, validated, and disseminated within academic environments.

### Conclusion

The results of the findings, if analyzed through Lotman's and Eco's semiotic theories, highlight that today ChatGPT functions as a semiotic mediator which simultaneously destabilizes and enriches knowledge systems. Students appear to negotiate meaning actively, balancing trust and skepticism, utility and caution, and engagement and critical distance. However, the survey also underscores that GenAI in higher education remains at a transitional stage, where its role as a collaborative knowledge partner has yet to be fully realized. Moving forward, research should explore how GenAI's semiotic mediation can be leveraged to deepen critical engagement, ensuring that AI-generated knowledge is not merely consumed, but actively interpreted, contextualized, and transformed.

### References

Alkaissi, H., S. I. McFarlane. 2023. "Artificial Hallucinations in ChatGPT: Implications in Scientific Writing". *Cureus*, Vol. 15, No. 2, e35179. Available at: <https://doi.org/10.7759/cureus.35179>.

Ambrosio, L., J. Schol, V. A. La Pietra, F. Russo, G. Vadalà and D. Sakai. 2023. "Threats and Opportunities of Using ChatGPT in Scientific Writing-The Risk of Getting Spineless". *JOR Spine*, Vol. 7, No. 1, e1296. Available at: <https://doi.org/10.1002/jsp2.1296>.

Bender, E. M., T. Gebru, A. McMillan-Major, S. Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models be too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.

Byrne, T. 2023. *The Phenomenology of ChatGPT: A Semiotics*. ResearchGate. Available at: [https://www.researchgate.net/publication/375666264\\_The\\_Phenomenology\\_of\\_ChatGPT\\_A\\_Semiotics](https://www.researchgate.net/publication/375666264_The_Phenomenology_of_ChatGPT_A_Semiotics).

Chandler, D. 2017. *Semiotics: The Basics* (3rd ed.). Routledge.

Chen, Z., J. Wang, M. Xia, K. Shigyo, D. Liu, R. Zhang and H. Qu. 2024. "StuGPTViz: A Visual Analytics Approach to Understand Student-ChatGPT Interactions". *ArXiv*, arXiv:2407.12423. Available at: <https://arxiv.org/abs/2407.12423>.

Chen, C., Gong, Y. 2025. "The Role of AI-Assisted Learning in Academic Writing: A Mixed-Methods Study on Chinese as a Second Language Students". *Education Sciences*, Vol. 15, No. 2, 141. Available at: <https://doi.org/10.3390/educsci15020141>.

Creswell, J. W., Plano Clark, V. L. 2018. *Designing and Conducting Mixed Methods Research*. (3rd ed.). Sage Publications.

Danesi, M. 2007. *The Quest for Meaning: A Guide to Semiotic Theory and Practice*. University of Toronto Press.

Desogus, P. 2012. "The Encyclopedia in Umberto Eco's Semiotics". *Semiotica*, No. 192, 501–521. Available at: <https://doi.org/10.1515/sem-2012-0068>.

Eco, U. 1979. *The Role of the Reader: Explorations in the Semiotics of Texts*. Indiana University Press.

Eco, U. 1984. *Semiotics and the Philosophy of Language*. Indiana University Press.

Eco, U. 1984. "Metaphor, Dictionary, and Encyclopedia". *New Literary History*, Vol. 15, No. 2, 255–271. Available at: <https://doi.org/10.2307/468855>.

Emdad, F. B., B. Ravuri, L. Ayinde, M. I. Rahman. 2023. "ChatGPT, a Friend or Foe for Education?" Analyzing the User's Perspectives on the Latest AI Chatbot via Reddit". *ArXiv*, arXiv:2311.06264. Available at: <https://doi.org/10.48550/arXiv.2311.06264>.

Floridi, L., M. Chiriatti. 2020. "GPT-3: Its Nature, Scope, Limits, and Consequences". *Minds and Machines*, Vol. 30, No. 4, 681–694. Available at: <https://doi.org/10.1007/s11023-020-09548-1>.

Gee, J. P. 2008. *What Video Games Have to Teach Us about Learning and Literacy*. Palgrave Macmillan.

Haim, E., L. Bergh, C. S. Q. Siew and Y. N. Kenett. 2025. "Cognitive Networks Highlight Differences and Similarities in the STEM Mindsets of Human and LLM-simulated Trainees, Experts and Academics". *ArXiv*, arXiv:2502.19529. Available at: <https://doi.org/10.48550/arXiv.2502.19529>.

Henriksen, D., P. Mishra, L. Woo and N. Oster. 2025. "The Education Doctorate in the Context of Generative Artificial Intelligence: Epistemic Shifts and Challenges to Practical Wisdom". *Impacting Education: Journal*

on *Transforming Professional Practice*, Vol. 10, No. 1, 73–79. Available at: <https://doi.org/10.5195/ie.2025.485>.

Hofeditz, L., A. K. Jung, M. Mirbabaie and S. Stieglitz. 2025. “Ethical Guidelines for the Application of Generative AI in German Journalism”. *Digital Society*, Vol. 4, No. 4. Springer. Available at: <https://doi.org/10.1007/s44206-024-00151-w>.

Holmes, W., M. Bialik and C. Fadel. 2022. *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Independently Published.

Kirstein, F., J. P. Wahle, B. Gipp and T. Ruas. 2025. “CADS: A Systematic Literature Review on the Challenges of Abstractive Dialogue Summarization”. *Journal of Artificial Intelligence*, Vol. 82. Available at: <https://doi.org/10.1613/jair.1.16674>.

Kress, G. 2010. *Multimodality: A Social Semiotic Approach to Contemporary Communication*. (1st ed.). Routledge.

Kress, G., T. van Leeuwen. 2001. *Multimodal Discourse: The Modes and Media of Contemporary Communication*. Hodder Arnold Publication.

Lotman, J. 1990. *Universe of the Mind: A Semiotic Theory of Culture*. I. B. Tauris.

Lotman, J. 2009. *Culture and Explosion*. De Gruyter Mouton.

Peirce, C. S. 1931. *Collected Papers of Charles Sanders Peirce*. Harvard University Press.

Petrassi, D. 2024. “Integrating ChatGPT as a Learning Tool: Potential Benefits and Critical Considerations”. *Formazione & Insegnamento*, Vol. 22, No. 2, 83–93. Available at: [https://doi.org/10.7346/-fei-XXII-02-24\\_09](https://doi.org/10.7346/-fei-XXII-02-24_09).

Piaget, J. 1954. *The Construction of Reality in the Child*. Basic Books.

Ren, X. 2023. “Investigating the Experiences of Online Instructors while Engaging and Empowering Non-traditional Learners in eCampus”. *Educ Inf Technol*, Vol. 28, 237–253. Available at: <https://doi.org/10.1007/s10639-022-11153-x>.

Rezai, A., E. Namaziandost and G. J. Hwang. 2024. “How Can ChatGPT Open Promising Avenues for L2 Development? A Phenomenological Study Involving EFL University Students in Iran”. *Computers in Human Behavior Reports*, Vol. 16. Available at: <https://doi.org/10.1016/j.chbr.2024.100510>

Roll, I., R. Wylie. 2016. “Evolution and Revolution in Artificial Intelligence in Education”. *International Journal of Artificial Intelligence in Educa-*

tion, Vol. 26, No. 2, 582–599. Available at: <https://doi.org/10.1007/s40593-016-0110-3>.

Saussure, F. de. 1916. *Course in General Linguistics*. McGraw-Hill.

Sloan, K. 2025. So Meta: AI-written Paper Says AI is the Future of Legal Scholarship. *Reuters*, 2 January 2025. Available at: <https://www.reuters.com/legal/legalindustry/so-meta-ai-written-paper-says-ai-is-future-legal-scholarship-2025-01-02>.

Vygotsky, L. S. 1978. *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press.

Weng, W. H., T. Tu and A. Palepu et al. 2025. “Towards an AI Co-scientist”. *ArXiv*, arXiv:2502.18864. Available at: <https://doi.org/10.48550/arXiv.2502.18864>.

Wertsch, J. V. 1991. *Voices of the Mind: A Sociocultural Approach to Mediated Action*. Harvard University Press.

Zawacki-Richter, O., V. I. Marín, M. Bond and F. Gouverneur. 2019. “Systematic Review of Research on Artificial Intelligence Applications in Higher Education”. *International Journal of Educational Technology in Higher Education*, Vol. 16, No. 1, 39. Available at: <https://doi.org/10.1186/s41239-019-0171-0>.

## AI, SEMIOSIS, AND THE FUTURE OF LANGUAGE ACQUISITION: A GLOBAL APPROACH TO EDUCATIONAL SEMIOSPHERE

*Ilaria Ingrao*  
*University of Turin*  
*ilaria.ingrao@unito.it*

*Seyedeh Maede Mirsonbol*  
*University of Turin*  
*seyedehmaede.mirsonbol@unito.it*

### **Abstract**

This study explores the role of artificial intelligence in second language (L2) acquisition and the tremendous opportunities that it creates for personalized learning. As the 2030 EU agenda for quality education suggests, the emphasis of research and practice with AI-based approaches should be to foster the learning process. For this reason, the theoretical aspect of this research examines how AI can be situated in the L2 context for promoting personalized learning experiences. The study, specifically, evaluates the effectiveness of the current role of AI in L2, critically reviewing its reflection on the ethical implication of AI in education using a semiotic anal-

ysis. The objective is to understand how AI adapts to students' cognitive and individual needs to be used as a methodology in learning frameworks. Additionally, the research raises two questions which address the role of AI in Barthes's (1977) *author-reader* approach, proposing a semiosis-based learning model adapted from Eco's *encyclopaedia* (1976, 1984) and Lotman's concept of *semiosphere* (1985). This model aims to ensure deeper personalized learning by reversing traditional pedagogic approaches and analysing the role of each element (student, teacher, text, and AI) inside the educational sphere. In conclusion, the study highlights how AI-driven tools can be integrated into L2 education, facilitating communication, content creation, and engagement with new topics in an inclusive manner.

**Keywords:** Artificial Intelligence, Second Language Acquisition, Semiotic Analysis, Semiosphere, Quality Education.

## 1. Introduction

The rapid evolution of artificial intelligence (AI) is not only altering the way we communicate but also redefining the very processes through which language is learned, and meaning is made. In the field of second language (L2) acquisition, AI as a technology-enhanced tool promises to ease teachers' administrative burden, thereby allowing greater focus on the educational relationship between teachers and students. In this context, machine learning models grow increasingly sophisticated, their ability to engage with linguistic patterns, cultural contexts, and semiotic systems challenge traditional theories of language (O'Halloran 2022). Semiosis – the process through which signs convey meaning and create diverse interpretations (Jappy 2020) – is integrated with AI while introducing new possibilities for developing language assessments that are not only responsive, but contextually aware. This intersection marks a critical turning point, inviting a global rethinking of educational space that places AI as a particular multifaceted technology within the learning atmosphere while raising crucial questions regarding its social, political, cultural, and ethical implications (Woo & Choi 2021; Limna et al. 2023; Marcuccio et al. 2025).

However, the adoption of AI-based education is not without challenges that can involve analysis of AI as part of a broader socio-technical system, structured around economic interests, market logics, and political visions of knowledge. From the perspective of education, these technologies do not merely support teaching, they also reflect and simultaneously shape specif-

ic cultural and ideological imaginaries about the future of education and its role in society (Jasanoff 2015). In this sense, AI acts as a true epistemic device that poses new challenges concerning the autonomy of teaching, the protection of students' personal data, equitable access to digital resources, the transparency of decision-making algorithms, and the reproduction of discriminatory biases within educational processes.

According to the human-centred AI lab at the University of Oxford<sup>1</sup>, the fusion of educational philosophy and technology-enhanced learning can be seen as an upheaval to build systems that truly contribute to learning by doing. This notion means learning with AI is, in fact, engagement in a more-than-human dialogue, while experimenting with building conscious learning that requires active presence of L2 learners in educational designs. Drawing from research in philosophy of education and semiotics, we take an interdisciplinary approach to explore the ethical and pedagogical implication of personalized learning with AI, while the theoretical section is designed based on Barthes's (1977) view on *author-reader* approach arguing that meaning of a text (in this research the text stands for L2 contents as written, audiovisual, etc.) should emerge from the text itself and the way readers/students interact with it. To navigate this notion deeper, we situate Barthes's theoretical framework within the context of *student-centred learning theory* (Piaget 1964) to critically analyse the incorporation of United Nations' (UN) quality education<sup>2</sup> norms in AI-driven applications for L2 acquisition. This attempt highlights that quality education fosters the integration of sustainable learning into teaching curriculum (UNESCO 2020) while empowering L2 learners to engage with AI effectively.

## 2. Theoretical framework on educational sphere

### 2.1. Edusphere

In the context of L2 development, it is valuable to, first, define language *acquisition* and its two primary theoretical orientations. The concept of language acquisition is closely aligned with constructivist theory of learning. This perspective, explored by scholars such as Vygotsky (1978), Krashen (1982), Freeman and Freeman (2004), and more recently Gambino and

---

<sup>1</sup> Retrieved from: <https://hailab.ox.ac.uk/>

<sup>2</sup> United Nations' 2030 agenda for sustainable development is consisted of 17 different goals to help developed and developing countries for the future. Quality education is goal number 4 which rises awareness and ask for practical strategies to improve education and growth. For more information read: <https://www.un.org/sustainabledevelopment/education/>

Share (2023), emphasizes the importance of exposure to comprehensible input and implicit language development through meaningful interactions, be it with human or more-than-human. Rather than focusing on the mastery of rules, acquisition-based approaches prioritize the functional use of language in communicative settings, where the emphasis is placed on understanding and conveying meaning. This orientation reflects a pragmatic view of language, where learners are encouraged to engage with language as a tool for real-world communication.

The teaching methods, learning materials, and digital platforms used for L2 acquisition are all part of a complex system of signs. Therefore, the integration of semiotics fits well with the vision that supports AI-enhanced learning categorizing itself within the educational sphere promoting acts of semiosis for creating interpretations and constructive language use. Additionally, *edusphere* is utilized as a brand name for referring to various educational platforms used inside the learning environment. The term *edusphere* is a blend of *education* and *sphere*, commonly used to describe the comprehensive realm of education as an adaptive system, encompassing technologies, philosophies, and communities of educators and learners involved in it (Duggan-Haas 2004). Furthermore, it refers to the collective digital environment, and its discussions on signs' interpretation<sup>3</sup>. *Edusphere* can be conceptually connected to Lotman's *semiosphere* (cfr. 1985), described as a semiotic space outside of which "there can be neither communication nor language" (Lotman 1990: 124). If the *edusphere* is treated as the total environment of learning where no semiosis can exist outside of it, then it could be seen as a subdomain of the *semiosphere*. This means that the *edusphere* is where learning happens, and "learning is a semiotic engagement" (Stables 2005: 375).

In the L2 *edusphere*, it is important to focus on active participants (teacher, student, text, and AI) and their interactions that lead toward evolving interpretations. Constantinidou (2023) states that "language education is constantly adapting to an ever expanding and transforming digital world" (p: 174) and while the role of AI cannot be avoided, it is essential to take a pragmatic view and acknowledge its presence entangled with semiosis activities inside of the *edusphere*. In this case, AI becomes part of the meaning-making process as it stands parallel to learning communication to steadily connect the flow of semiosis dialogue while students (the reader of L2 learning context) stand as the main figure in meaning-making.

<sup>3</sup> EduSphere Team. (n.d.). EduSphere: An innovative online learning web app, Devpost, Retrieved from: <https://devpost.com/software/edusphere-2d89rx>

This view is in contrast with the teacher-centred approach<sup>4</sup> in pedagogy which situates teachers as the main author of L2 learning context, allowing teachers to step away from their being in control with fixed authority in the class (a fixed author-centred approach). This notion is in line with Campbell (2019), who expands on the idea of *semiosis-as-learning*, describing it as a process of meaning-making that bridges the learner with their surrounding environments. These environments, or semiospheres, exist on two levels: internally, within the learner's mind, and externally, within the community they belong to. Understanding semiosis requires active engagement with both dimensions, as learning encompasses not just the classroom setting but also the wider social, digital, and cultural landscape in which it occurs. To avoid obstacles in the semiosis-as-learning process, educators and AI-enhanced content creators for L2 acquisition must develop frameworks that can adapt the complexity of the digital semiosphere in favour of a dynamic and functional atmosphere for today's globalised digital world (Hartley et al. 2020).

The network of minds and semiosis acts that are placed inside the edusphere can be a strong reminder of Eco's theory of the *encyclopaedia* that significantly reconfigures the classic structuralist view of semiotics and textual interpretation. Eco extends the notion of meaning-making by incorporating the reader's interpretive role into the generative process of textual meaning (Eco 1976). Desogus (2012) explains that an aesthetic text serves not to communicate but to challenge and reshape semantic codes for conceptual understandings. Eco's semiotics investigates the production and interpretation of signs within communicative practices. Therefore, meaning arises through a process mediated by shared knowledge, embedded within what he terms the encyclopaedia, "a multidimensional, collective system of information and interpretation" (Desogus 2012: 501). The emphasis is on the interpretative process as a form of cooperative work. Interpreters rely on their individual encyclopaedias to select meanings that render the textual elements coherent, while simultaneously validating these selections within the broader interpretive sphere (Desogus 2012: 517). This cooperative dynamic allows for interaction between individual cognition and shared cultural knowledge, thereby reinforcing Eco's vision of semiosis as a continuous, open-ended system of meaning-making grounded in collective and evolving models. This encyclopaedia exists inside the edusphere and connects teacher to student, student to student, and most importantly

---

<sup>4</sup> According to British Council for L2 education: <https://www.teachingenglish.org.uk/professional-development/teachers/knowning-subject/t-w/teacher-centred>

AI as a facilitator that holds the educational venue for learners to build knowledge through experiences, interaction, and reflection.

In (Table 1), there are the theoretical perspectives in semiotic studies to create a roadmap in relation to educational student-centred approaches for this study which highlight the essential role of AI and learner within the L2 acquisition.

**Table 1:** Theoretical approaches for an AI-enhanced L2 acquisition

Concept	What it means	Connection to Edusphere
<b>Semiosphere</b>	Cultural space where meaning is created	Edusphere exists <i>within</i> this larger semiosphere
<b>Semiosis-as-learning</b>	Learning as a process of meaning-making	Core process that drives how people learn in edusphere
<b>Edusphere</b>	Total environment of educational activity	A meaning-rich space shaped by semiosis

## 2.2. Quality education

Quality education, central to UNESCO's 2030 agenda, transcends disciplinary boundaries and calls for a shift in pedagogy toward globally relevant content and its public awareness. In light of the existing initiatives and technologies, studies attempt to formulate the idea of AI to increase the quality of education (Zygmunt 2016; Al-Jarf 2022; De la Fuente 2022; Williams 2024). As Tonkin (2023) argues, language is central to global discourse, enabling inclusive and persuasive communication. To implement quality effectively, L2 educators must address the persistent knowledge gap and create inclusive, multimodal, and personalized learning paths. To develop learning with AI in a pragmatic framework, it is important to enhance L2 learners' awareness of AI while following the UNESCO's (2019) framework that adapts AI as a facilitator for learning. The framework emphasises the role of digital technologies in knowledge deepening and creation (Pedro et al. 2019: 18) that fosters critical thinking and equips learners to address global challenges.

Integrating quality education into L2 acquisition encourages AI literacy and the use of global topics; however, the main aim is to thoughtfully implement AI, not as a novelty, but as a transformative tool grounded in pedagogy (Argondizzo & Mansfield 2022). AI offers opportunities for personalized learning and enhanced engagement through adaptive content, fostering communication and collaborative learning which all are part of the quality education agenda (Kamalov et al. 2023). These competences

are essential in adapting to rapid societal change and addressing complex global issues. AI-enhanced L2 acquisition should support learning by personalizing content, encouraging critical thinking, meaning-making from students' point of view (Sandhu et al. 2024). AI must adopt learner-centred, interdisciplinary, and flexible frameworks that reflect contemporary educational needs while fostering global mindsets (Leicht et al. 2018). Learner-centred approaches, if designed for L2 apps, help learners to "reflect on their own knowledge and learning processes" (Leicht et al. 2018: 49) which enhances the quality of semiosis practices applied to produce diverse interpretations.

### **3. Theoretical framework on AI in Education**

#### **3.1. Artificial Intelligence: between pedagogical innovation and ethical risks**

The use of AI within the educational landscape appears as a controversial and multifaceted phenomenon. Numerous studies highlight the potential benefits offered by using AI devices. From the teachers' perspective, these technologies can lighten activities that are often burdensome in terms of time and resources, such as improving decision-making processes by teachers and administrators or assessing learning tests (Okonkwo & Ade-Ibijola 2021). AI can also support education by automating routine tasks, allowing teachers to focus on personalized student support (Pedro et al. 2019). From the students' side as well, it introduces multiple opportunities. AI makes it possible to personalize learning paths more effectively, fostering greater engagement in study and accelerating the acquisition of skills. It can also help overcome barriers related to disadvantaged contexts by boosting self-confidence, facilitating interaction, promoting learning propensity, and fostering creativity (Smutny & Schreiberova 2020; Kulkami et al. 2022; Handini et al. 2022; Pratama et al. 2023; Khreisat et al. 2024; Marcuccio et al. 2025). In particular, AI has proven effective in language acquisition (Subramanian et al. 2020) thanks to functionalities based on personalization and interactivity, which support the construction of an autonomous and lasting educational experience (Betel 2023). However, despite these advantages, the ethical, political, and social implications associated with the use of AI are rarely critically addressed:

In the field of education, artificial intelligence can aid in areas such as personalized learning, intelligent tutoring systems, automated grading, and adaptive assessment. The use of AI in education has been shown to improve student outcomes such as retention rates,

academic performance, and learning process engagement (Ouyang et al. 2022; Chen et al. 2020). In addition, AI can provide educators with data-driven insights for identifying student strengths and weaknesses and customising instruction to meet individual requirements (Bates et al. 2020; Prakash et al. 2021). However, the incorporation of AI into education also raises ethical concerns, such as data privacy, algorithmic bias, and the possibility of human teachers being replaced (Masters 2019). Therefore, it is essential to carefully consider the potential benefits and challenges associated with the incorporation of AI in education and to implement the necessary policies and practices to ensure its responsible and ethical use (Harish et al. 2024: 219).

Many issues highlight the ethical challenges of an AI-oriented approach to teaching (Mamlok 2024) while challenging both technical and pedagogical aspects. The main points are the inadequate training of teachers in the conscious and effective use of these tools (Woo & Choi 2021; Limna et al. 2023; Marcuccio et al. 2025); insufficient protection of privacy and lack of transparency in information management; the need to strengthen data protection policies and adopt more robust practices of informed consent (Woo & Choi 2021); doubts about the reliability of outputs generated by AI applications (Limna et al. 2023); and finally, the risk of superficial learning, as well as the potential weakening of autonomous skills, writing ability, and critical thinking (Sarosa et al. 2021; Marcuccio et al. 2025). Faced with these complexities, it is essential to adopt a critical and balanced perspective: the analysis of the use of AI in education must be neither naively optimistic nor rigidly alarmist. It is necessary to promote a reflective and student-centred approach that, while recognizing the potential of AI, does not overlook its deeper implications. An utopian approach tends to see technology as neutral, ignoring its sociopolitical context, while the dystopian view focuses on structural issues but overlooks its positive social potential (Mamlok 2024).

On one hand, these tools can facilitate learning by democratizing education at multiple levels, both social and cognitive; on the other, they can become an additional tool of alienation, imposing their own worldview and suppressing human creativity (Cooley 1987). The question we must therefore ask ourselves is whether it is possible to create AI capable of enhancing learning and increasing reflection rather than gradually destroying it. Among the major concerns related to the introduction of AI in the educational system is the potential atrophy of social skills. Some devices, in fact, risk functioning as tools that are “anti-relational, anti-dialogical, and rooted in the assumption of education and child development that do not

accord with social and cognitive science” (Means 2018: 117). In this sense, AI can negatively affect contexts where learning – and the strengthening of knowledge – relies on peer-to-peer knowledge sharing, hindering fundamental collaborative and dialogical dynamics (Matsuda 2021). Although it is undeniable that, on an operational level, AI can provide valuable support to learning, it is essential to promote its conscious use. The personalization of learning paths, in addition to easing teachers’ workload and improving students’ focus, can become a valuable tool for stimulating critical thinking. It can also encourage more exploratory learning methods, open to reflecting on different scenarios and building complex knowledge:

Artificial intelligence asks learners to think critically, challenge pre-suppositions, and solve challenging problems besides transmitting information. Artificial intelligence assists in preparing people to meet the ever-changing demands of modern society by creating challenges that need critical thinking and imitating real-world scenarios. AI-driven learning personalization is essential for fostering critical thinking. [...] Learners are urged to investigate problems, test theories, and critically assess data to cultivate a critical attitude beyond rote learning. (Khreisat et al. 2024: 8524)

### **3.2. Deeper into ethics approach to AI: privacy, bias, and ethical dilemma**

A deeper analysis of the risks associated with the use of AI in education reveals a complex set of issues involving safety, privacy, autonomy, confidentiality, and the use of students’ personal data (Khreisat et al. 2024). Often, personalization is considered an undisputed strength in learning, yet the concrete risks it entails are overlooked. One of the main dangers is the possibility that students may become confined within algorithmic logics that create actual *filter bubbles* (Pariser 2011). This process leads to a standardization that affects one of the most vulnerable social groups, risking the suppression of creativity and flattening the diversity of individual cognitive pathways. This phenomenon is fuelled by systems based on inductive approaches that presume a linearity between similar individuals and between past and future behaviours (O’Neil 2017; Sadin 2019; Numerico 2021). AI has the capability to collect, process, and analyse vast amounts of data. However, indiscriminately applying this potential to the educational context may constitute a serious risk factor. Physical, cognitive, and emotional information can be easily acquired through devices that would allow the study and monitoring of students’ reactions and their level of attention (Peng & Nagao 2021). This approach tends to transform the

educational environment into a mechanism of control, in which teaching is reduced to performance surveillance, without a real consideration of the ethical implications connected to the collection and management of personal data. Moreover, delegating the monitoring of educational processes to third-party-owned tools entails a high ethical and social risk, raising serious concerns regarding the use of personal data for inadequate surveillance purposes (Kopczynski & Silvia 2024). Appropriate security measures are therefore necessary to prevent the risk of violating the privacy of minors. It is also important that students effectively understand how their information can be used by AI systems (Khreisat et al. 2024). Although AI offers significant opportunities to enhance the educational sector through greater personalization of learning and better implementation of strategies to support students, it is essential to underline that transparency, privacy and security are often the most compromised aspects. The promotion of an ethical, regulated, and responsible use of AI in education cannot therefore disregard a careful consideration of these risks. It is not only a matter of privacy or transparency; the adoption of AI in education also raises serious concerns regarding the proliferation of algorithmic biases, which can lead to inequitable situations by reproducing negative stereotypes – even within educational settings – potentially affecting students directly.

Another difficulty in automated choice-making is ethical accountability. Reliance on AI advice can absolve people of moral and ethical obligations. It is critical to have a clear accountability system and to specify who is in charge of making important choices. In addition to shielding learners from potentially dangerous choices, this lays a strong basis for trust regarding the application of AI in instruction. (Khreisat et al. 2024: 8524)

The imaginaries we build around technology reflect the kind of future we wish to see, and this also applies to education. The worldviews we adopt are based on shared cultural models (Santangelo 2013), which also materialize through specific technical choices. Devices based on AI are the result of knowledge situated in time, shaped by the interaction between science, technology, and society (Jasanoff 2015). AI can thus be understood as a wheel used for a broader socio technical model, influenced by politics and social dynamics. Education has always been the means through which citizenship is formed, through literacy and qualification (Rahm 2023). When the State assumes responsibility for the education of its citizens, it tends to introduce educational policies aimed at creating a competitive workforce. In this way, education risks becoming a vehicle for a technocratic vision

of knowledge. Therefore, the political and social implications associated with the use of technologies in educational and civic contexts cannot be ignored. Systems like ChatGPT, for instance, can generate enthusiasm but also deep criticism. Their adoption inevitably requires us to question the socio-political contradictions and tensions that accompany the construction and dissemination of such tools (Mamlok 2024). In light of all this, it is essential to consider not only the promised benefits of implementing AI tools in the educational sector, but also the potential risks they entail in ethical, political, and social terms (Figure 1).



**Figure 1:** An illustration on the use of AI in education  
(Khreisat et al. 2024)

#### **4. Presentation of the problem and research questions**

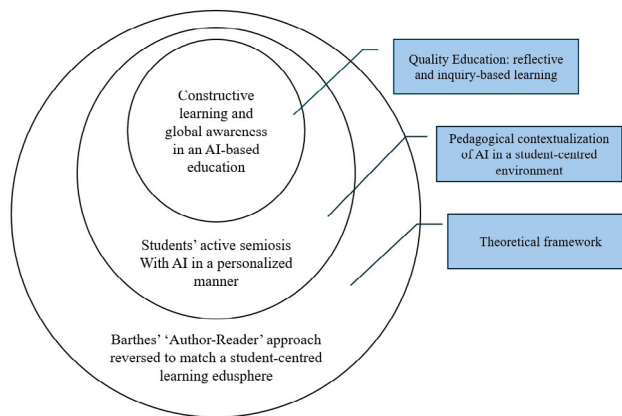
The role of AI in education has introduced new possibilities for enhancing learning processes, particularly in the domain of L2 acquisition. At the same time, there has been a growing recognition of the value of semiotic approaches in education, emphasizing how meaning-making through signs shape learners' constructive development. Despite these parallel advancements, a significant gap persists in current research concerning the intersection of AI and semiosis within L2 educational methodologies that focuses on student-centred and personalized learning. This study aims to understand what is currently missing in this context by promoting an interdisciplinary approach that is semiotically oriented while raises two fundamental questions:

- (1) Can a semiotic-based theory of learning help students in the process of understanding about a sustainable future?
- (2) Can AI be trained in this approach for L2 acquisition?

## **5. Methodology**

This study adopts a qualitative and interpretive methodology grounded in the philosophy semiotics and a student-centred perspective, focusing on the pedagogical implications of Barthes' (1977) notion of author-reader approach. This aim is to explore how a theoretical shift from a fixed author-centred interpretation to a fluid reader-centred construction of meaning resonates with modern educational practices, particularly within a student-centred perspective that includes AI. The research unfolds across three interconnected domains: (1) theoretical framework, (2) pedagogical contextualization of AI in a student-centred environment, and (3) quality education: reflective and inquiry-based learning (Figure 2). The theoretical framework is led by Barthes' work to extract key philosophical tenets, especially the rejection of authorial dominance inside the educational semiosphere and the elevation of the learner's interpretive agency in a continuous and open-ended encyclopaedia of meaning-making. These concepts are then mapped into educational paradigms that value learner autonomy, such as constructivist, inquiry-based, and reflective pedagogical approaches. Special attention is given to how Barthes' theory can be reinterpreted to frame the learner as an active meaning-maker in the classroom, rather than a passive recipient of knowledge making.

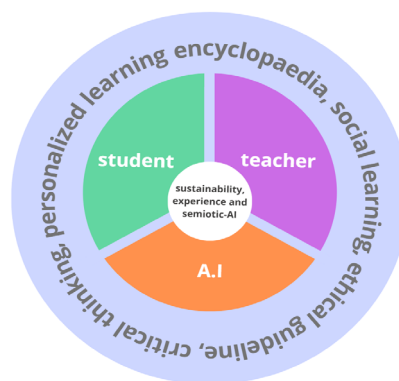
The recontextualization of the learning environment with AI leads the view towards renewed design of digital spheres for semiosis and global understanding in today's technologically mediated classrooms scaffolded by AI. In this study, AI cannot be eliminated from contemporary pedagogical processes as it is already entangled in activities that occur within the edusphere. AI tools participate in shaping the transmission of meaning in the inner circle of the methodological framework. Therefore, AI will not be just a technological tool, but a semiotic actor – a mediator in the ongoing dialogic exchange between learner and educator. In this model, AI operates within the space between the learner and the teacher, facilitating communication, offering feedback, and personalizing L2 acquisition. The teacher's role thus shifts from authoritative source to co-designer and facilitator of learning experience, while the learner increasingly inhabits the position at the core of the learning sphere and more of a co-author of their educational journey. Meaning is no longer a one-way transmission; it is a distributed, collaborative act across human and non-human agents. And finally, the quality of education is increased by reflective and inquiry-based learning that promotes global awareness in L2 practice. Narrative inquiry by individual learners and their personal learning experiences are pieces of an incorporating edusphere where the learners' own reflective semiosis play an important part in shaping knowledge.



**Figure 2:** Qualitative and interpretive methodology for a learner-centred edusphere with AI

## 6. Discussion

The full paradigm in Figure 3 highlights the evolving dynamics when authority and authorship are decentralized. By aligning Barthes' theory with the semiotic complexities of AI-integrated learning environments, this methodology aims to showcase how meaning is constructed, contested, and reimagined in an L2 context when semiosis is used as a communicative facilitator to picture education with AI technology. In doing so, it reaffirms that meaning is never fixed: it is an emergent product of interaction among diverse voices: student, teacher, text, and now AI.



**Figure 3:** Student-teacher-text-AI model:  
A semiotic approach to AI in Education<sup>5</sup>

<sup>5</sup> Figure 3 is a new model developed by both authors in this article to better understand the role of AI in education and how it should be harmonized with other important factors that play critical roles in learning.

### 6.1. Answer to the first question

Promoting the responsible use of AI requires, on the one hand, a careful consideration of the legal, political, and structural gaps that still characterize the governance of such technologies. On the other hand, it is possible to begin exploring the potential that AI offers in the field of education, where it can be used to enhance students' skills and improve learning processes. While the former scenario calls for interdisciplinary and multimodal interventions aimed at ensuring user safety and privacy, the latter may allow for a more accessible rethinking of pedagogical approaches within a socio-educational framework where AI serves as a tool to expand students' cognitive and critical capacities. Within this perspective, it becomes essential to fully leverage the potential of artificial intelligence in fostering critical thinking among younger generations.

Artificial intelligence asks learners to think critically, challenge presumptions, and solve challenging problems besides transmitting information. Artificial intelligence assists in preparing people to meet the ever-changing demands of modern society by creating challenges that need critical thinking and imitating real-world scenarios. AI-driven learning personalization is essential for fostering critical thinking. AI builds learning environments that encourage problem-solving and active exploration by customizing materials according to every student's interest and skill level (Zapata, Guerrero, and Montilla 2024). Learners are urged to investigate problems, test theories, and critically assess data to cultivate a critical attitude beyond rote learning. Turning this around, another way that artificial intelligence stimulates critical thinking is through modeling difficult situations. [...] Educators and technological innovators must work closely together to successfully include artificial intelligence to promote critical thinking (Fojtik 2024). Teachers are essential to help learners critically analyze the material provided through artificial intelligence and facilitate debates that encourage autonomous reasoning. (Khreisat et al. 2024: 8524).

An interesting proposal in this context is the integration of L2 instruction within an educational pathway focused on reflecting on the future. To this end, we propose the adoption of the *strategic foresight*<sup>6</sup> model, which stimulates students' ability to imagine and design future scenarios (Santan-

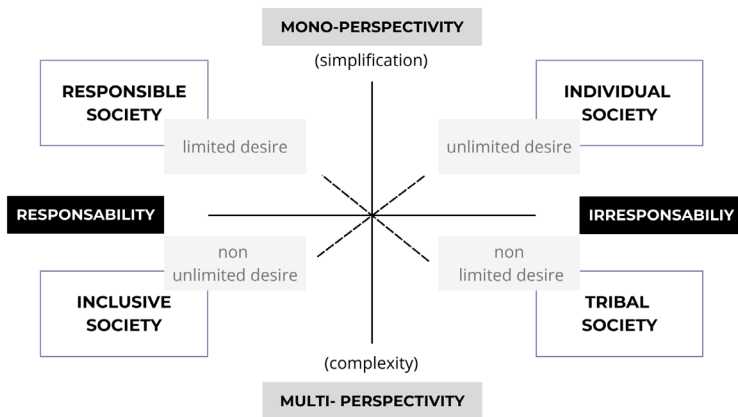
---

<sup>6</sup> The term *strategic foresight* refers to a set of disciplines that investigate, through a multidisciplinary lens (including economics, sociology, anthropology, semiotics, philosophy, and others), the strategies for envisioning future scenarios. It is a methodology designed to shape medium- to long-term visions in order to reflect on possible future configurations and make informed and responsible decisions (Santangelo & Robiati 2024).

gelo & Robiati 2024). Although this method was not originally designed for educational purposes, it offers an innovative approach that views the future not as a fixed destination, but as a tool to better understand the present and make more informed decisions. This perspective led us to explore some key aspects of applying this methodology to second language learning. In *strategic foresight* studies and practices, the future is no longer considered a static goal to reach, but rather a dynamic perspective – a device that guides present-day choices. Traditionally, the future has been perceived as something preordained, to be discovered or predicted. However, it challenges this view by conceiving the future as a hypothetical construction shaped by collective imagination and the ability to design scenarios. In this sense, the future becomes a space for learning and experimentation – a laboratory in which students can develop not only their language skills but also their critical and creative thinking, *aspiring toward desirable futures* (De Leonardis & Deriu 2012; Appadurai 2013; Pellegrino 2019). Nevertheless, such an imaginative process cannot take place without a collective dialogue among diverse actors, so that the future becomes a dynamic field of critical elaboration and creative tension (Santangelo & Robiati 2024). The futures and foresight method is, in fact, a participatory process that guides groups in constructing alternative and plausible future scenarios. It is not about predicting the future but about developing the capacity to conceive it as an open and transformative space. Originally developed for companies and institutions, this methodology can also serve as a powerful pedagogical tool for L2 acquisition. Imagining the future requires students to acquire advanced linguistic skills: they must describe possible worlds, project themselves into future scenarios, and link the present with potential developments. Furthermore, they must engage with diverse cultural perspectives on the future, specific to the linguistic communities they are studying. This approach enriches L2 context by adding a deeper dimension that goes beyond everyday communication.

In this context, semiotics plays a central role. The future is constructed through narratives, and every narrative adheres to specific semiotic rules. From a pedagogical standpoint, this implies that teaching a language is not only about grammar and vocabulary, but also about understanding the *grammar* of future-oriented discourse. We can therefore ask ourselves: What are the key terms in texts about the future? What metaphors and images are most commonly used? What narrative structures dominate future-related stories? We believe that familiarity with these elements enables students to produce meaningful and culturally grounded texts. This approach not only fosters a more in-depth and contextualized study of

the foreign language, but also encourages critical reflection on fundamental issues that shape values and visions of both the present and the future (ibidem). Themes such as responsibility and irresponsibility, single versus multiple perspectives, and limited versus unlimited desires can be analyzed within the discourses that construct specific visions of the societies composing our perception of the world (Figure 4).



**Figure 4:** Map of discourses on the societies of the future  
(Santangelo & Robiati 2024)

## 6.2. Answer to the second question

The relationship between AI and L2 acquisition must be viewed not simply as a technological enhancement but as a potential paradigm shift in educational theory and practice. In particular, employing a semiotic approach in conjunction with a student-teacher-text-AI model offers a compelling framework for training AI in ways that enhance both the personalization and effectiveness of learning. Recent pedagogical trends signal a shift from teacher-centred instruction toward learner agency and inquiry-based models (Leicht et al. 2018; Li et al. 2021). In this context, AI is no longer positioned as a mere repository of language data or as a mechanical tutor delivering standardized feedback. Rather, it becomes a collaborator in a constructivist learning environment, participating in the learner's meaning-making process (semiosis-as-learning). This reframing invites us to see AI as capable of interpreting signs (linguistic, visual, contextual) in ways that respond to the learner's evolving semiotic environment. By adopting a semiotic lens, AI systems can be trained not merely to recognize correct linguistic forms but to engage with the interpretive process that underlies all language use. In this case, L2 acquisition becomes a dialogic interac-

tion between sign producers (teachers/students) and sign interpreters (students). When AI is embedded in this interaction, it learns to navigate and contribute to this dialogue; to understand intention, context, and variability of meaning. This enhances development in AI that functions as both a sign-producer and a responsive interpreter in educational settings, attuned to the learner's trajectory and the multiple meanings in the target language.

Moreover, designing AI systems through this model means it engages with learners as co-authors of their linguistic development. In this model, learners are not passive recipients of AI-generated input; they are active readers and creators of the signs, co-constructing knowledge in collaboration with AI. Such an approach calls for AI systems capable of interpretative elasticity – responsive to student input, sensitive to multiple meanings, and capable of generating nuanced feedback in L2 that supports the learner's creative use of the target language. It positions learners as intellectual agents in a semiotic journey, rather than operators within a fixed instructional workflow (Pratschke 2024). In addition, adopting a future-oriented framework has implications for digital learning design. If L2 acquisition is to be sustainable and future-oriented, personalized learning experiences must be central to the development of AI-mediated instruction. AI systems designed with semiotic awareness can adapt dynamically to individual learner profiles, interpret learner intent, and respond in ways that reflect not only syntactic accuracy but also pragmatic appropriateness and cultural nuance. This requires training AI not just on corpora of correct language but on diverse, real-world semiotic interactions. Finally, situating AI within a broader pedagogical transformation reinforces its role as a relational tool rather than a technical fix. In this vision, AI operates at the intersection of teacher guidance, learner agency, and contextual meaning-making. This relationship places AI as a mediator in the educational process, not as a replacement for teachers, but as a partner in the co-construction of meaning. Thus, AI can be trained by a semiotic approach to enhance L2 acquisition but only if it is reimagined as a semiotic participant, a co-interpreter, and a co-designer of meaning in the learning process.

## 7. Conclusion

This article highlights how, within the educational L2 context, AI can serve as an empowering and enhancing tool. The integration of AI into L2 acquisition must be understood not merely as a technological advancement, but as a transformative shift in educational paradigms. The introduction of AI in education raises significant questions regarding student

privacy protection, algorithmic transparency, and the reproduction of biases. Added to these are the tangible risks associated with users' personalities and learning paths. The greatest risk is that AI may not remain a neutral instrument serving education but instead exert pressure toward the realization of a technocratic and performance-driven vision of learning. In such a scenario, education risks losing its critical, reflective, and relational value. For this reason, it is essential to promote a responsible and reflective approach to the use of AI in education. By adopting a semiotic lens and positioning AI within a dynamic *student-teacher-text-AI* model, education moves toward the future where AI becomes a co-participant in the learner's construction of meaning. This model aligns with emerging pedagogical trends that emphasize learner agency, inquiry, and contextualized understanding. Training AI systems to engage semiotically with sensitivity to linguistic, cultural, and contextual nuance redefines them as relational tools embedded in dialogic and personalized learning. Only then can AI effectively support L2 learners as active agents in a co-constructed, meaning-rich educational journey.

### Acknowledgment

The authors confirm that there is no conflict of interest related to the manuscript.

The Authors' contribution was distributed as: Ilaria Ingrao wrote parts 3.1, 3.2, and 6.1 in the manuscript and S. Maede Mirsonbol wrote parts 2.1, 2.2, 2.3, 4, 5, and 6.2. Parts 1 and 7 were drafted by both authors.

### References

Al-Jarf, R. 2022. Developing Students' Global Awareness in EFL Reading and Speaking. *South Asian Res J Art Lang Lit*, Vol. 4, No. 1, 31–38. Available at: <https://doi.org/10.36346/sarjall.2022.v04i01.005>.

Appadurai, A. 2013. *The Future as Cultural Fact. Essays on the Global Condition*. London: Verso Books.

Argondizzo, C., G. Mansfield. 2022. "Research on Sustainable Development Literacy and Affective Learning and Teaching Actions". *Language Learning in Higher Education*, Vol. 12, No. 2, 365–366. Available at: <https://doi.org/10.1515/cercles-2022-2066>.

Barthes, R. 1977 [1967]. The Death of the Author (transl. Stephen Heath). In *Image, Music, Text*. Hill and Wang, 142–148.

Betal, A. K. 2023. "Enhancing Second Language Acquisition through Artificial Intelligence (AI): Current Insights and Future Directions". *Journal for Research Scholars and Professionals of English Language Teaching*, 7, 1–8. Available at: <https://doi.org/10.54850/jrspelt.7.39.003>.

Campbell, C. 2019. "Educating Semiosis: Foundational Concepts for an Ecological Edusemiotic". *Studies in Philosophy and Education*, Vol. 38, 291–317. Available at: <https://doi.org/10.1007/s11217-018-9617-4>.

Constantinidou, D. A. 2023. "Language Education, Digital Citizenship, and Juri Lotman's Semiosphere". *Digital Age in Semiotics & Communication*, Vol. 6, No. 1, 173–184. Available at: <https://doi.org/10.33919/dasc.23.6.9>.

Cooley, M. 1987. *Architect or Bee? The Human Price of Technology*. The Hogarth Press.

De la Fuente, J. María. 2022. *Education for Sustainable Development in Foreign Language Learning: Content-based Instruction in College-level Curricula*. Taylor & Francis.

De Leonardis, O., M. Deriu. 2012. *Il futuro nel quotidiano. Studi sociologici sulla capacità di aspirare*. Milano: Egea.

Desogus, P. 2012. "The Encyclopedia in Umberto Eco's Semiotics". *Semiotica*, No. 192, 501–521. Available at: <https://doi.org/10.1515/sem-2012-0068>

Duggan-Haas, D. 2004. From the Ecosphere to the Edusphere: Problems in (mis) Managing Ecosystems and Edusystems. In *Proceedings of the 2004 Complexity Science and Educational Research Conference*, 55–82. Available at: <https://www.researchgate.net/publication/241120947>.

Eco, U. 1976. *A Theory of Semiotics*. Indiana University Press.

Eco, U. 1984. *Semiotica e Filosofia del Linguaggio*. Torino, Einaudi.

Freeman, D. E., Y. S. Freeman. 2004. *Essential Linguistics: What You Need to Know to Teach Reading, ESL, Spelling, Phonics, and Grammar*. Portsmouth, NH: Heinemann.

Gambino, A., J. Share. 2023. Critical Media Literacy and Second Language Acquisition. *International Encyclopedia of Education*, 76–85. Available at: <https://doi.org/10.1016/B978-0-12-818630-5.07072-X>.

Harish, V., R. Sharma, G. Rana and A. Nayyar. 2023. Artificial Intelligence in Sustainable Education: Benefits, Applications, Framework, and Potential Barriers. In *The Role of Sustainability and Artificial Intelligence in Education Improvement*. Chapman and Hall/CRC, 219–236.

Hartley, J., I. Ibrus and M. Ojamaa. 2020. *On the Digital Semiosphere: Culture, Media and Science for the Anthropocene*. Bloomsbury Publishing USA.

Jappy, T. 2020. Peirce's Conception of Semiosis. In *The Bloomsbury Companion to Contemporary Peircean Semiotics*. Bloomsbury Publishing, 101–132. Available at: <http://dx.doi.org/10.5040/9781350076143.ch-003>.

Jasanoff, S. 2015. Future Imperfect: Science, Technology, and the Imaginations of Modernity. In Jasanoff, S., S.-H. Kim (eds.). *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*. The University of Chicago Press, 1–33.

Kamalov, F., D. Santandreu Calonge and I. Gurrib. 2023. "New Era of Artificial Intelligence in Education: Towards a Sustainable Multifaceted Revolution". *Sustainability*, Vol. 15, No. 16, 12451. Available at: <https://doi.org/10.3390/su151612451>.

Khreisat, M. N., D. Khilani, M. A. Rusho, E. A. Karkkulainen, A. C. Tabuena and A. D. Uberas. 2024. "Ethical Implications of AI Integration in Educational Decision Making: Systematic Review". *Educational Administration: Theory and Practice*, Vol. 30, No. 5, 8521–8527.

Krashen, S. 1982. *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon Press.

Kulkarni, T., B. Toksha and P. Gupta. 2022. Applications of Artificial Intelligence in Learning Assessment. In *Artificial Intelligence in Higher Education*. CRC Press, 95–111.

Leicht, A., J. Heiss and W. J. Byun. 2018. In: *Issues and Trends in Education for Sustainable Development*, Vol. 5. UNESCO Publishing.

Li, Y.-D., G.-H. Ding and C.-Y. Zhang. 2021. "Effects of Learner-Centred Education on Academic Achievement: A Meta-Analysis". *Educational Studies*, Vol. 50, No. 3, 285–298. Available at: <https://doi.org/10.1080/03055698.2021.1940874>.

Limna, P., T. Kraiwanit, K. Jangjarat, P. Klayklung and P. Chock-sathaporn. 2023. "The Use of ChatGPT in the Digital Era: Perspectives on Chatbot Implementation". *Journal of Applied Research*, Vol. 6, No. 1, 1–10. Available at: <https://doi.org/10.37074/jalt.2023.6.1.32>.

Lotman, J. M. 1985. *La Semiosfera: L'asimmetria E Il Dialogo Nelle Strutture Pensanti*. S. Salvestroni (a cura di). (1st ed.). Venezia: Marsilio.

Lotman, J. 1990. *Universe of the Mind: A Semiotic Theory of Culture*. Bloomington: Indiana University Press.

Mamlok, D. 2024. "Landscapes of Sociotechnical Imaginaries in Education: A Theoretical Examination of Integrating Artificial Intelligence in Education". *Foundations of Science*, Vol. 30, No. 2, 529–540.

Marcuccio, M., M. E. Tassinari and V. Lo Turco. 2025. "Progettare e valutare con il supporto dell'intelligenza artificiale: elementi per un approccio critico all'uso dei chatbot". *Journal of Educational, Cultural and Psychological Studies*, No. 30, 5–118. Available at: <https://doi.org/10.7358/ecps-2024-030-marc>.

Matsuda, N. 2021. "Teachable Agent as an Interactive Tool for Cognitive Task Analysis: A Case Study for Authoring an Expert Model". *International Journal of Artificial Intelligence in Education*, Vol. 32, No. 1, 48–75. Available at: <https://doi.org/10.1007/s40593-021-00265-z>.

Means, A. 2018. *Learning to Save the Future: Rethinking Education and Work in an Era of Digital Capitalism*. Routledge.

Numerico, T. 2021. *Big Data e Algoritmi: Prospettive Critiche*. Carocci.

O'Halloran, K. L. 2022. "Matter, Meaning and Semiotics". *Visual Communication*, Vol. 22, No. 1, 174–201. Available at: <https://doi.org/10.1177/14703572221128881>.

O'Neil, C. 2017. *Armi di distruzione matematica*. Bompiani.

Okonkwo, C. W., A. Ade-Ibijola. 2021. "Chatbots Applications in Education: A Systematic Review". *Computers and Education: Artificial Intelligence*, 2, 100033. Available at: <https://doi.org/10.1016/j.caeai.2021.100033>.

Pariser, E. 2011. *The Filter Bubble: What the Internet is Hiding from You*. Penguin UK.

Pedro, F., M. Subosa, A. Rivas and P. Valverde. 2019. *Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development*. United Nations Educational, Scientific and Cultural Organization, 7, place de Fontenoy, 75352 Paris 07 SP, France. Available at: <https://hdl.handle.net/20.500.12799/6533>.

Pellegrino, V. 2019. *Futuri possibili. Il domani per le scienze sociali di oggi*. Milano: Ombre Corte.

Piaget, J. 1964. "Cognitive Development in Children: Development and Learning". *Journal of Research in Science Teaching*, Vol. 2, 176–186. Available at: <https://doi.org/10.1002/tea.3660020306>.

Pratama, M. P., R. Sampelolo and H. Lura. 2023. "Revolutionizing Education: Harnessing the Power of Artificial Intelligence for Personalized Learning". *Klasikal: Journal of Education, Language Teaching and Science*, Vol. 5, No. 2, 350–357. Available at: <https://doi.org/10.52208/klasikal.v5i2.877>.

Pratschke, B. M. 2024. *Generative AI and Education: Digital Pedagogies, Teaching Innovation and Learning Design*. Springer Nature.

Sadin, E. 2019. *Critica della ragione artificiale*, trad. it. di F. Bononi.

Sandhu, R., H. K. Channi, D. Ghai, G. S. Cheema and M. Kaur. 2024. An Introduction to Generative AI tools for Education 2030. *Integrating Generative AI in Education to Achieve Sustainable Development Goals*, 1–28. Available at: <https://doi.org/10.4018/979-8-3693-2440-0.ch001>.

Santangelo, A. 2013. *Sociosemiotica dell'audiovisivo*. Roma: Aracne.

Santangelo, A., A. Robiati. 2024. *Immaginari del domani. Tra futures studies, semiotica e worldbuilding*. Italian Institute for the Future.

Sarosa, M., M. Kusumawadhani, A. Suyono and Y. M. Azis. 2021. “The Effectiveness of Chat-bot as an Online Learning Method on Active and Reflective Learning Styles”. *Multicultural Education*, Vol. 7, No. 9, 92–100. Available at: <http://dx.doi.org/10.5281/zenodo.5484411>.

Smutny, P., P. Schreiberova. 2020. “Chatbots for Learning: A Review of Educational Chatbots for the Facebook Messenger”. *Computers & Education*, 151, 103862. Available at: <https://doi.org/10.1016/j.compedu.2020.103862>.

Stables, A. 2005. *Living and Learning as Semiotic Engagement: A New Theory of Education*. Edwin Mellen Press.

Tonkin, H. 2023. Introduction: Diversity of Language, Unity of Purpose. In *Language and sustainable development* (1–10). Cham: Springer International Publishing. Available at: [https://doi.org/10.1007/978-3-031-24918-1\\_1](https://doi.org/10.1007/978-3-031-24918-1_1).

UNESCO. 2020. Education for Sustainable Development: A Roadmap. United Nations Educational, Scientific and Cultural Organization, 7, place de Fontenoy, 75352 Paris 07 SP, France. Available at: <https://doi.org/10.54675/YFRE1448>.

Vygotsky, L. S. 1978. *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Williams, R. T. 2024. The Ethical Implications of Using Generative Chatbots in Higher Education. In *Frontiers in Education*, Vol. 8, 1331607. Frontiers Media SA. Available at: <https://doi.org/10.3389/feduc.2023.1331607>.

Woo, J. H., H. Choi. 2021. Systematic Review for AI-based Language Learning Tools. *ArXiv*, arXiv:2111.04455. Available at: <https://doi.org/10.48550/arXiv.2111.04455>.

Zygmunt, T. 2016. “Language Education for Sustainable Development”. *Discourse and Communication for Sustainable Education*, Vol. 7, No. 1, 112–124. Available at: <https://doi.org/10.1515/dcse-2016-0008>.

## CONSUMER ATTITUDES TO AI CONTENT GENERATION IN SOCIAL MEDIA

*Martin Varshev*  
New Bulgarian University  
*varshev.martin@gmail.com*

*Eduard Marinov*  
New Bulgarian University  
*eddie.marinov@gmail.com*

### **Abstract**

The rapid advancement of artificial intelligence (AI) has transformed content generation in social media, prompting both excitement and skepticism among consumers. This study explores consumer attitudes toward AI-generated content, drawing on a comprehensive survey of social media users. The survey, conducted across diverse demographics, examines awareness, perception, trust, and engagement with AI-generated content. The results indicate that while awareness of AI content is high, with 88.3% of respondents recognizing it in their social media feeds, attitudes towards such content remain largely negative. Many users perceive AI-generated content as “inhuman” and “boring,” with significant con-

cerns about authenticity and data privacy. A striking 65% of participants advocate for transparency, urging brands to disclose the use of AI in content creation. Furthermore, trust in AI content is low, with many respondents expressing reluctance to engage with or share AI-generated material, particularly when its origin is disclosed. A case study focusing on Coca-Cola's AI-generated advertisements underscores potential brand impact, revealing that a majority of users failed to identify AI content and would prefer traditional, human-generated ads. The findings suggest that while AI has potential in content marketing, its acceptance hinges on ethical considerations, transparency, and the ability to complement rather than replace human creativity. This study offers valuable insights for marketers navigating the evolving landscape of AI in social media, highlighting the need for strategies that prioritize consumer trust and authenticity.

**Keywords:** AI-generated content, Social media marketing, Content authenticity, Digital trust

## **1. Introduction**

In recent years, artificial intelligence (AI) has transformed the marketing industry, introducing innovative methods for analyzing data, automating processes, and personalizing strategies. AI enables marketers to process vast amounts of data with remarkable accuracy and speed, leading to a deeper understanding of consumer behavior and preferences. This allows companies to deliver highly targeted and personalized advertising, significantly enhancing the effectiveness of their marketing campaigns.

Automation powered by AI optimizes marketing operations by streamlining routine tasks and boosting productivity. Additionally, AI-driven content personalization plays a pivotal role in creating unique user experiences, tailoring dynamic content to individual customer interests and needs. Machine learning algorithms can analyze data from diverse sources – social media, web traffic, in-store purchases, and online platforms – allowing marketers to identify trends and patterns critical for effective strategy development (De Mauro, Sestino, and Bacconi 2022).

AI also revolutionizes real-time marketing campaign adjustments, automatically optimizing budgets, resource allocation, and messaging in response to changing consumer behavior and market conditions, ultimately improving return on investment (ROI) and campaign effectiveness (AI

Khaldy, Al-Obaydi, and Al Shari 2023). However, the most profound impact of AI may be its ability to personalize consumer experiences by analyzing individual behaviors and preferences, resulting in more relevant and engaging content that drives conversions and fosters loyalty (Gao & Liu 2022).

Despite the numerous opportunities which AI offers in marketing, it also presents challenges related to ethics, data security, and transparency. As AI continues to evolve, companies must navigate these complex issues responsibly, ensuring respect for consumer privacy.

This study aims to present the results of an empirical investigation into user attitudes toward the use of AI in social networks to optimize content and enhance user experience. The first section outlines the potential business benefits of AI in social content creation, followed by examples of AI usage in the digital space. The main focus of the study is an in-depth analysis of a survey conducted on consumer attitudes toward AI-generated content.

## **2. Business Benefits of Artificial Intelligence in Social Media Content Creation**

Artificial intelligence (AI) offers multiple business benefits when creating content on social networks. These benefits not only increase the effectiveness of marketing campaigns, but also lead to significant cost savings and increased return on investment (ROI). Let us take a look at some of the key benefits that AI provides in this area.

### **2.1. Cost optimisation**

AI helps companies optimize their costs by automating multiple processes, including automatic content generation, publication scheduling, and campaign performance analysis. This reduces the need for human intervention and associated costs, while increasing the efficiency of marketing activities. For example, automating social media processes can result in significant savings in resources and time, allowing companies to focus on strategic tasks (Harkness et al. 2023).

### **2.2. Content personalisation**

AI allows companies to create personalized content targeted to specific audiences. By analyzing large volumes of data, AI algorithms can identify user preferences and behaviours, enabling the creation of relevant and engaging content. This leads to increased user engagement and increased conversions.

### **2.3. Increasing productivity**

AI systems significantly increase productivity in content creation. They can generate text, images and videos with high speed and accuracy, allowing marketing teams to create and publish content in higher volume and quality. This is particularly useful for campaigns that require regular content updates.

### **2.4. Improving analytics and reporting**

Artificial intelligence provides powerful tools for analysing the effectiveness of marketing campaigns. These tools can track and analyze multiple metrics in real-time, allowing marketers to make informed decisions and optimize their strategies. For example, AI-powered analytics platforms can provide detailed reports on consumer behavior which is essential for improving future campaigns.

### **2.5. Crisis management**

AI can play a key role in crisis management by monitoring social networks for negative comments and potential threats to a company's reputation. Natural language processing algorithms can identify and analyze negative sentiment in real time, allowing communications teams to respond quickly and effectively before problems escalate.

### **2.6. Improve customer interaction**

AI technology can improve customer interactions through automated chatbots and virtual assistants which offer quick and accurate answers to users' questions. This not only improves customer service but also frees up employee time for more complex tasks.

## **3. Examples of using artificial intelligence in the digital space**

We can safely say that artificial intelligence is used every day all around us. In the following pages, we will look at a few examples of its application in fast-moving goods and services.

### **3.1. Take a break with AI KitKat**

In the current dynamic world of advertising, artificial intelligence (AI) is becoming an increasingly important player in the creative process of marketing campaigns. Taking a cue from KitKat's latest project, delivered in collaboration with agency Wunderman Thompson, we see the brand emblematically embrace the principle of "taking a break", entrusting the creation of its new campaign entirely to generative AI in 2023 (VML, n.d.).



**Figure 1:** Screenshot from KitKat Australia & New Zealand YouTube profile  
 Source: "Have AI Break," VML, n.d., <https://www.vml.com/work/have-ai-break>

KitKat's initiative involves the development of ad copy where AI is triggered to formulate scenarios which are then turned into visuals. Fun tasks include instructions such as, "Create a KitKat ad speaking the language of Generation Z," "Create a KitKat ad targeting gamers," and "Create a KitKat ad reflecting the latest trends." While the results are not perfect, they represent a good enough start which allows the team to relax while the "fingers that are being carried" are mostly occupied with chocolate (Duong 2023).

Melanie Chen, head of marketing at Nestlé, commented, "We're committed to creating more opportunities for people to relax, so when Wunderman Thompson suggested we let AI generate our next advertising content, we happily accepted the offer. It's clear from the results that AI is not yet ready to replace human positions in our industry without serious safeguards." (Campaign Brief 2023).

This creative experiment highlights the potential of AI for automation in marketing activities, but also highlights the need for human oversight to ensure content quality and compliance. While AI offers opportunities to optimize marketing efforts and reduce operational costs, the creative human element remains indispensable for creating meaningful and emotionally engaging advertising messages.

### 3.2. Mint Mobile's ChatGPT ad

Ryan Reynolds recently used OpenAI's ChatGPT to create an ad for Mint Mobile, highlighting the growing role of AI in creative industries. This case is an example of the potential for AI to transform the traditional approach to marketing by offering new, innovative ways to engage audiences (Adams 2023).

The Mint Mobile ad, described by Reynolds as "slightly intimidating but impressive", was created by ChatGPT with instructions to include a joke and

information related to ongoing promotion. This demonstrates how AI can generate content tailored to specific marketing requirements and maintain brand authenticity.

Implementing AI into marketing strategies offers opportunities to automate creative processes which can increase efficiency and reduce content creation costs. However, there is a risk of limiting creativity when relying exclusively on algorithmically generated content. It is therefore important that marketers remain actively involved in the content creation process, using AI as a tool rather than a substitute for human creativity.

### **3.3. My ad by Shahrukh Khan**

Artificial intelligence (AI) is transforming the marketing industry, introducing innovative approaches to analytics, automation and personalization of communications. An example of this is Cadbury Celebrations' campaign in India, where thanks to AI-generating technologies, local marketers had the opportunity to use a digital avatar of famous actor Shahrukh Khan as their brand ambassador. This campaign showcased an innovative application of AI in marketing which enables personalized presentation for small businesses (D&AD," n.d.).

In response to the economic challenges following the COVID pandemic, Cadbury Celebrations focused on its core values of generosity and community. The strategy aimed to revive a sense of community by encouraging empathy, turning marketing initiatives into an act of generosity.

Using data and AI tools, Cadbury created a host of personalized ads in which Shahrukh Khan promotes purchases from local stores, with each ad tailored to a specific store. The campaign was distributed via YouTube and social platforms such as Meta, reaching over 500 postcodes. Additionally, the NotJustACadburyAd.com platform was developed, allowing users to create and share their own versions of the ads (D&AD," n.d.).

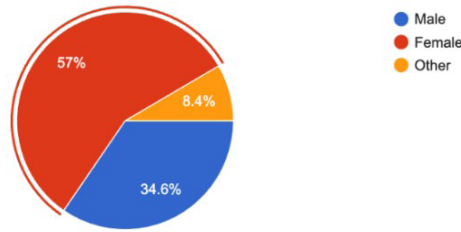
## **4. Empirical user research on AI-generated content**

The survey contains 47 questions which include both open and closed formats. It was created using Google Forms and distributed via social networks, emails and private messages to relevant audiences. Its aim was to explore users' opinions on AI-generated content: how they perceive it, whether they know what it is and whether they can recognise it.

### **4.1. General characteristics of respondents**

The survey collected responses from 106 respondents. Of these, 56% were female, 35% were male and the remaining 9% chose not to indicate their gender.

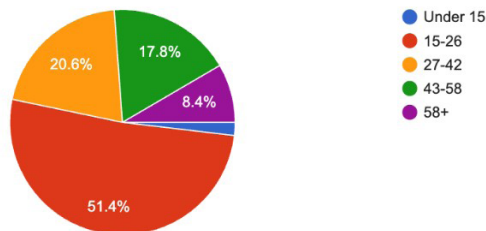
107 responses



**Figure 2:** Gender distribution

The survey was able to cover all age groups, with the largest proportion being Generation Z, which represented 50.9% of respondents and is the current focus of the study due to their significant consumer potential. They are followed by Millennials at 20.8%, Generation X at 17.9%, Baby Boomers at 8.5% and Generation Alpha at 1.9% – in absolute terms this is a two-person sample, making it difficult to draw specific conclusions about this group.

107 responses

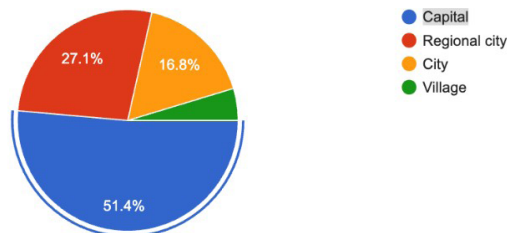


**Figure 3:** Age distribution

Source: Generated by the authors.

More than half of the respondents live in the capital (51.9%), 27.4% are from a large regional city, 16% are from a town and only 4.7% live in rural areas.

107 responses

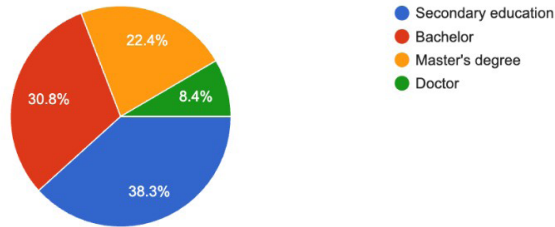


**Figure 4:** Place of residence

Source: Generated by the authors.

A significant proportion of respondents have a university degree: 31.1% have a bachelor's degree, 21.7% have a master's degree, and 8.5% have a doctorate. The remaining 38.7% have secondary education.

107 responses



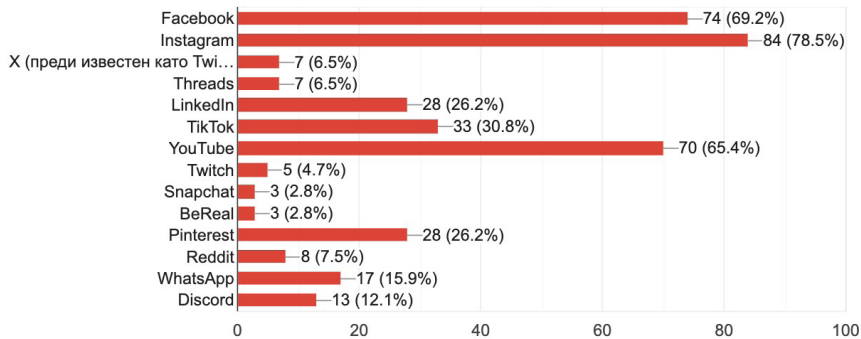
**Figure 5: Education**

Source: Generated by the authors.

#### 4.2. Use of social networks

The favorite and most used social networks for respondents are Instagram, Facebook and YouTube (over 65% for each of the mentioned).

107 responses

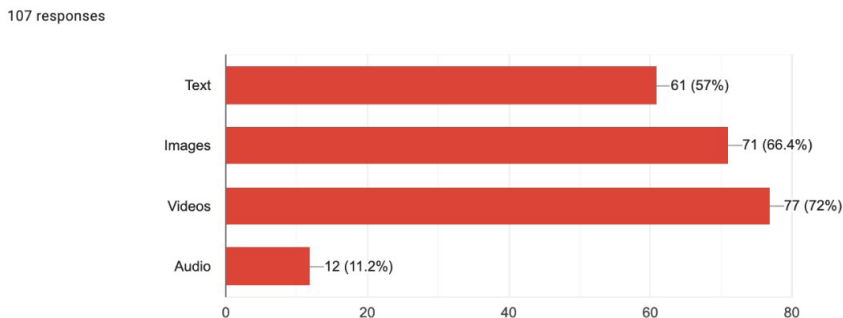


**Figure 6: Which social networks do you use?**

Source: Generated by the authors.

The survey confirmed that 70.8% users use their devices to visit social networks multiple times every day. They do so primarily for entertainment (93.4%), to learn the news (67.9%), and to connect with friends and family (67%).

Videos are the most popular type of content to consume, followed by images and text posts, while audio formats remain in last place.

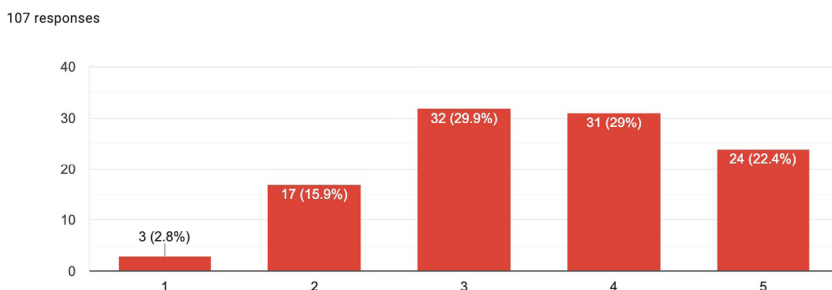


**Figure 7:** What type of content do you prefer?

Source: Generated by the authors.

### 4.3. Attitude towards AI-generated content

All respondents to the survey had heard of and knew what AI is. Of these, 50.9% thought they were well acquainted with the concept of it at work. 104 out of 106 survey respondents also know that there is artificial intelligence in social networks. 88.3% are sure that they have seen content created by AI on social networks, while nearly 16% could not judge. The latest data is extremely interesting and shows that audiences are watching the users and businesses they follow, trying to figure out if they are using AI.

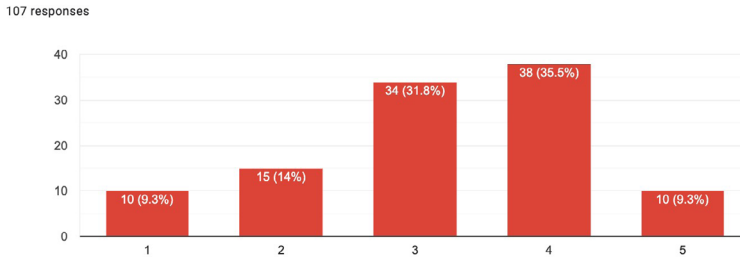


**Figure 8:** Are you familiar with the concept of AI?

Source: Generated by the authors.

In an open-ended response, users indicated that by AI-generated content, they expect something boring, not relevant, which may not pique their interest. A common definition is “inhuman”, which we take to mean

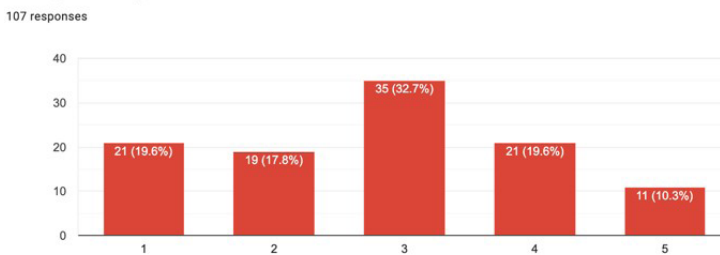
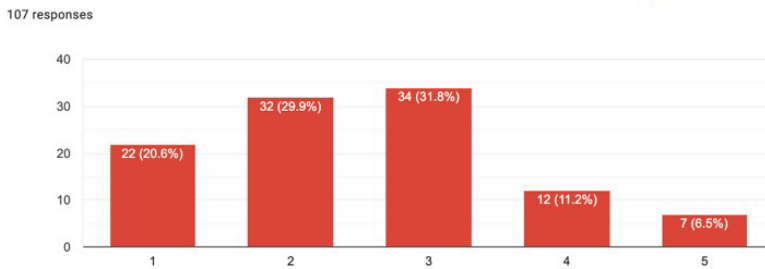
something that does not reflect the essence of the user. That said, a large percentage of users say they can recognize AI-generated content. We were later able to verify this statement in the survey.



**Figure 9:** Can you recognize AI-generated content?

Source: Generated by the authors.

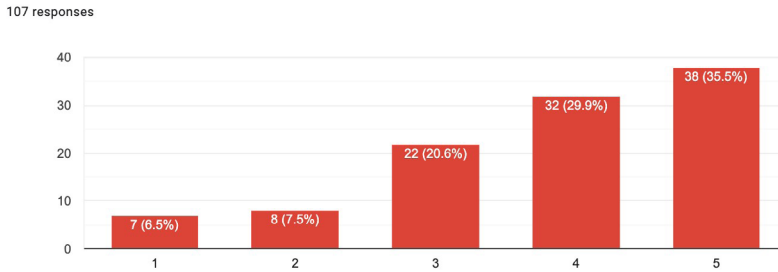
Users report that they do not trust AI-generated content, and a large percentage of users would not even interact with such content if it was clearly labeled that it was AI.



**Figure 10:** Trust in AI-generated content

Source: Generated by the authors.

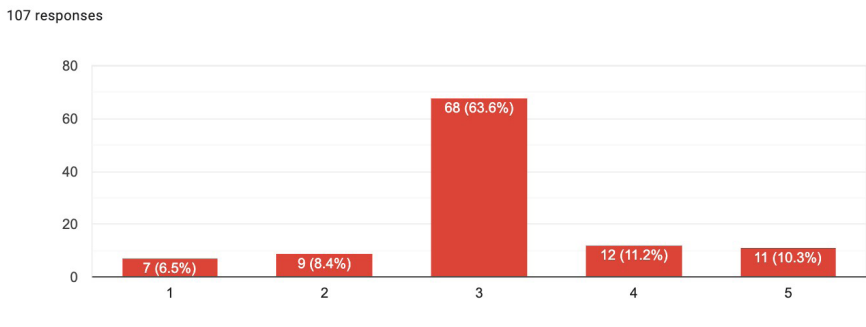
Users say they do not trust AI-generated content, and a large percentage of users would not even interact with such content if it was clearly labeled as AI. 65% also believe that every company should clearly say when content is AI-generated.



**Figure 11:** Informing about AI-generated content  
 Source: Generated by the authors.

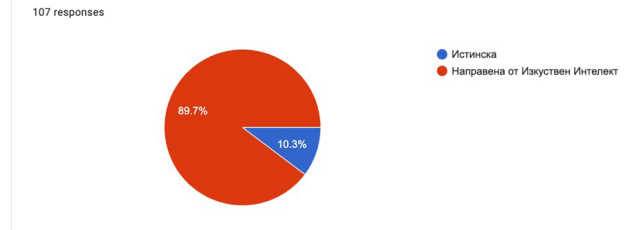
The main concerns related to AI are that authenticity will be lost (71.7%) and personal data will be misused (51.9%).

We asked questions to users related to the brand Coca Cola. In summary, we can say that 66% of users have either not seen or cannot determine whether they have seen Coca Cola content generated using AI. Coca Cola is a global brand which manages to maneuver between the rain drops of negativity surrounding AI. Consumers would not know how to react if they saw the company using AI. However 45% would not disclose if they did know whether AI was involved.



**Figure 12:** Brand perception  
 Source: Generated by the authors.

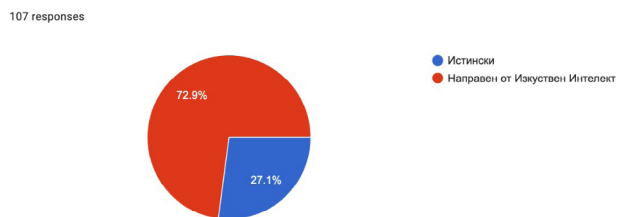
In order to test the audience's knowledge of AI, We staged a series of photos and videos. Some of them were created with AI and others were not. In the following lines we will reproduce the photo and the users' answers, in order to see if they could recognize the real from the fake photos.



**Figure 13:** Genuine or generated with AI?

Source: Bbc Bitesize, 2023. Chart generated by the authors.

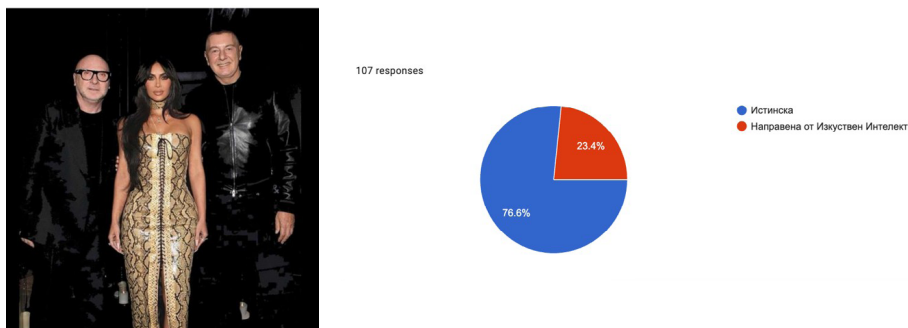
89.6% of users answered this question correctly. This is an AI-generated photo showing Mark Zuckerberg in a pink suit.



**Figure 14:** Genuine or generated with AI?

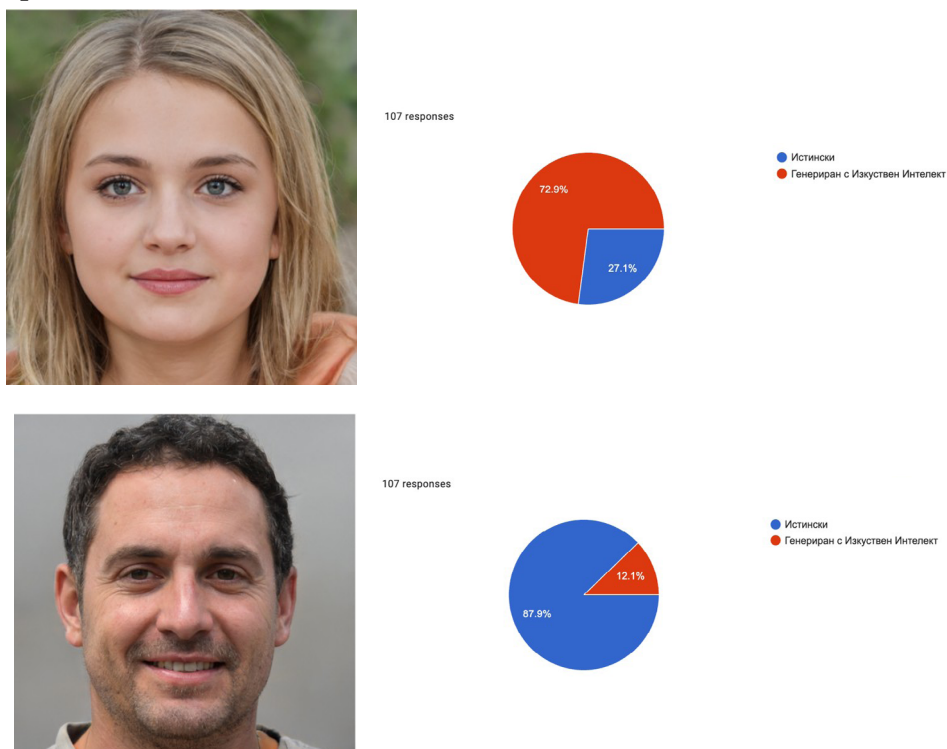
Source: Bbc Bitesize, 2023. Chart generated by the authors.

Here again we have a correct answer, which is however less convincing. This is again an AI-generated video of Tom Cruise doing magic tricks.



**Figure 15:** Genuine or generated with AI?  
Source: Bbc Bitesize, 2023. Chart generated by the authors.

This photo of Kim Kardashian is completely real, as respondents to our poll pointed out.



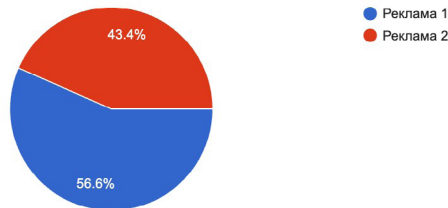
**Figure 16:** Genuine or generated with AI?  
Source: Which Face Is Real?, n.d.. Chart generated by the authors.

These are randomly generated images of faces created with artificial intelligence. Although users recognize the female face, 87.7% of them are confused and indicate that the male face is real.

At the conclusion of our survey, we added two Coca-Cola ads. One is the ad created by artificial intelligence which we shared in the previous pages, while the other is a classic ad which launches Coca-Cola's "true happiness" platform. This is an ad from 2006 which did not use artificial intelligence.

In "Masterpiece" – the first ad where AI was used – 78.1% of consumers recognized that AI was used, while in the second ad where it was not used – 41.9% guessed that it was. Although the audience was able to identify when AI was used and when it was not, we have to note the small difference in the second ad. Consumers said they liked the new ad with AI better.

106 responses



**Figure 17:** Which ad do you like more?

Source: Generated by the authors.

To summarise, we are still a long way from consumers being fooled when it comes to AI-created content. On average  $\frac{1}{4}$  of them make a mistake when they have to look and guess. It is important to say that in real-world conditions when they are not as observant, they are much more likely to make a mistake when looking at the data.

## Conclusion

The survey, conducted across diverse demographics, examines awareness, perception, trust, and engagement with AI-generated content. The results indicate that while awareness of AI content is high, with 88.3% of respondents recognizing it in their social media feeds, attitudes towards such content remain largely negative. Many users perceive AI-generated content as "inhuman" and "boring," with significant concerns about authenticity and data privacy. A striking 65% of participants advocate for transparency, urging brands to disclose the use of AI in content creation. Furthermore, trust in AI content is low, with many respondents expressing reluctance to

engage with or share AI-generated material, particularly when its origin is disclosed. A case study focusing on Coca-Cola's AI-generated advertisements underscores the potential brand impact, revealing that a majority of users failed to identify the AI content and would prefer traditional, human-generated ads.

The findings suggest that while AI has potential in content marketing, its acceptance hinges on ethical considerations, transparency, and the ability to complement rather than replace human creativity. This study offers valuable insights for marketers navigating the evolving landscape of AI in social media, highlighting the need for strategies that prioritize consumer trust and authenticity.

In conclusion, this study confirms the significant impact of artificial intelligence on social media marketing. AI tools and technologies present new opportunities to optimize content and personalize marketing messages, which are crucial for enhancing user engagement and which strengthen the connection between brands and their audiences. By successfully mimicking the human creative process, AI not only boosts productivity through task automation and big data analysis but also fosters greater creativity in personalized and dynamic marketing campaigns.

As social media marketing evolves, it is essential for companies to continue investing in AI while adhering to ethical standards and legal frameworks which protect user data and privacy. Transparency about AI use, coupled with educating consumers on its benefits and risks, is vital for building trust and increasing acceptance of these technologies. In order to facilitate broader adoption, companies should focus on implementing AI in mainstream, user-friendly applications, making it accessible for professionals navigating this complex landscape.

The growth and innovation in AI technologies offer marketers substantial opportunities to refine their strategies and produce more engaging content. However, with these opportunities comes the responsibility to use AI in a manner that respects and enriches the human experience, ensuring that technological advancements complement, rather than overshadow, the authenticity of human interaction.

## References

Adams, P. 2023. Ryan Reynolds Reads From AI-generated Script in New Mint Mobile Ad. *Marketing Dive*, 11 January 2023. Available at: <https://www.marketingdive.com/news/ryan-reynolds-ChatGPT-Mint-Mobile-AI-marketing/640160/> (accessed 8 August 2024).

Al Khaldy, M. A., B. A. A. Al-Obaydi and A. J. Al Shari. 2023. The Impact of Predictive Analytics and AI on Digital Marketing Strategy and ROI. In *Studies in Big Data*. Springer, 367–379. Available at: [https://doi.org/10.1007/978-3-031-42455-7\\_31](https://doi.org/10.1007/978-3-031-42455-7_31) (accessed 8 August 2024).

BBC Bitesize. 2023. Quiz: Artificial Intelligence or Real? – BBC Bitesize. *BBC Bitesize*, 20 April 2023. Available at: <https://www.bbc.co.uk/bitesize/articles/zqnwxg8> (accessed 8 August 2024).

Campaign Brief. 2023. KitKat Reminds Australians That AI Is Just Another Great Excuse to Have a Break in Latest Campaign via Wunderman Thompson. *Campaign Brief*, 4 April 2023. Available at: <https://campaignbrief.com/kitkat-reminds-australians-that-ai-is-just-another-great-excuse-to-have-a-break-in-latest-campaign-via-wunderman-thompson/> (accessed 8 August 2024).

D&AD, n.d. Shah Rukh Khan-My-Ad | D&AD Awards 2022 Shortlist | Use of Talent & Influencers. Available at: <https://www.dandad.org/awards/professional/2022/235670/shah-rukh-khan-my-ad/> (accessed 8 August 2024).

De Mauro, A., A. Sestino and A. Bacconi. 2022. “Machine Learning and Artificial Intelligence Use in Marketing: A General Taxonomy”. *Italian Journal of Marketing*, No. 4, 439–457. Available at: <https://doi.org/10.1007/s43039-022-00057-w> (accessed 8 August 2024).

Duong, A. 2023. AI-Generative Chocolate Campaigns. *TrendHunter.com*, 15 April 2023. Available at: <https://www.trendhunter.com/trends/kitkat> (accessed 8 August 2024).

Gao, Y., H. Liu. 2022. “Artificial Intelligence-enabled Personalization in Interactive Marketing: A Customer Journey Perspective”. *Journal of Research in Interactive Marketing*, Vol. 17, No. 5, 663–680. Available at: <https://doi.org/10.1108/jrim-01-2022-0023> (accessed 8 August 2024).

Harkness, L., A. Bhambhri, J. Bevilacqua and C. Binnion. 2023. How Generative AI Can Boost Consumer Marketing. *McKinsey & Company*, 5 December 2023. Available at: <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/how-generative-ai-can-boost-consumer-marketing> (accessed 8 August 2024).

n.d. *Which Face Is Real?* Available at: <https://www.whichfaceisreal.com/> (accessed 8 August 2024).

VML, n.d. Have AI Break. Available at: <https://www.vml.com/work/have-ai-break> (accessed 8 August 2024).

## USE OF AI IN THE CONTEXT OF FASHION AND RELATED INDUSTRY.

*Dilyana Orlova*  
*New Bulgarian University*  
*dilyanamolerova@gmail.com*

### **Abstract**

Artificial intelligence (AI) is revolutionizing everyday tasks, enhancing processes and improving efficiency, positioning it as a powerful tool for optimizing time for producers and customers. Beyond efficiency, AI influences cultural behaviors and aesthetic preferences, opening up new opportunities for creative expression. In the fashion industry, AI supports designers and retailers by offering personalization and automation to enhance user experience amidst a landscape of abundant fashion choices. This paper examines the impact of AI on the fashion industry as both a technological innovation and a cultural influencer, referencing Gilles Lipovetsky's theories on consumer psychology. The semiotic and dynamic nature of fashion is touched upon through the theory of Yuri Lotman (1922–1993) and Roland Barthes (1915–1980), as well as the dynamic contemporary processes

and technologies through the perspective of Lev Manovich (1960–) and Emanuele Arielli (1972–). AI is shown to drive creative exploration and cultural diversity within fashion, while highlighting the human element in the aesthetic process.

**Keywords:** AI, Fashion industry, Cultural influence, Creative expression, Automation

## Introduction

Artificial intelligence is transforming the way we perform our daily tasks, redefining processes and increasing their efficiency. Thanks to this, AI is increasingly perceived as a means of optimizing time and effort, both by manufacturers and consumers. At the same time, it influences our cultural choices and behavior, providing new opportunities for creative development and interaction with the world. With the development of technology, AI continues to push the boundaries of what is possible in various fields - from photography and design to fashion trends and social media, emphasizing the importance of the human factor and generally accepted cultural norms in the process of aesthetic creation.

The text explores the interaction and use of AI in the fashion industry. According to the theory of the French sociologist Gilles Lipovetsky (1944–), the growing variety of fashion items, can cause feelings of insecurity and anxiety in consumers. Given this background, AI has been interpreted as a tool for optimizing and improving the user experience through personalization and automation. AI technologies help designers and retailers effectively address the challenges of a rapidly changing fashion market by providing tools to better understand and engage customers. Furthermore, AI plays a role in pushing creative boundaries and increasing people's creative abilities, thereby contributing to cultural and aesthetic diversity in the fashion industry. This study touches upon the semiotic and dynamic essence of fashion through the theory of Yuri Lotman (1922–1993) and Roland Barthes (1915–1980), as well as the dynamic contemporary processes and technologies through the perspective of Lev Manovich (1960–) and Emanuele Arielli (1972–). In this context, where fashion is understood as a semiotic and dynamic shaped both by cultural conventions and by contemporary technological processes, the relationship between human creativity and computational tools is particularly relevant. The ever-advancing AI and the human with their creativity and cognitive abil-

ities are complementary. AI can support and enhance human activities, reducing the individual's efforts, while human critical thinking and creativity remain central to the human-AI relationship.

The definition of **fashion** given by Vincent Colapietro (1950–) in *Glossary of Semiotics* reads: “An area of semiotic inquiry opened up, in a systematic way, by Roland Barthes’ *The Fashion System* (1967 [1983]). Just as our buildings are designed to do more than shelter us, our clothes are fashioned to do more than clothe us. They are one of the ways we make statements about ourselves.” (Colapietro 1993: 104). The current work, while adhering to this definition and further developing it with Lotman’s understanding of the term “fashion”, will use the opportunity to interpret it both as a whim and as an innovative creativity. Both in need of an audience. (1998: 93).

The definition of **AI** is broad and includes data mining, natural language processing, and machine learning. According to the *Britannica* dictionary, AI is: “the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings. The term is frequently applied to the project of developing systems endowed with the intellectual processes characteristic of humans, such as the ability to reason, discover meaning, generalize, or learn from experience. Since their development in the 1940s, digital computers have been programmed to carry out very complex tasks - such as discovering proofs for mathematical theorems or playing chess - with great proficiency. Despite continuing advances in computer processing speed and memory capacity, there are yet no programs that can match full human flexibility over wider domains or in tasks requiring much everyday knowledge. On the other hand, some programs have attained the performance levels of human experts and professionals in executing certain specific tasks, so that artificial intelligence in this limited sense is found in applications as diverse as medical diagnosis, computer search engines, voice or handwriting recognition, and chatbots.” (Copeland 2025) The definition given in a Statista study, part of which will be used in the report, reads: “Artificial intelligence (AI) is a branch of computer science that focuses on creating machines able to perform tasks traditionally requiring human intelligence, such as perception, reasoning, and learning. The Artificial Intelligence (AI) market includes a wide range of applications, such as speech recognition, image processing, and autonomous vehicles and has seen rapid growth in recent years due to advancements in technology and increased investment. The Artificial Intelligence (AI) market includes software, hardware, and services that enable organizations to develop and deploy AI applications.” (Statista 2024).

### **AI in e-commerce**

The practical application of AI redefines everyday tasks and their performance in an innovative way. Thanks to data from Statista, last updated in March 2024, we know that the AI market is expected to reach a value of US\$ 184 billion in 2024. The projected annual growth rate for the period 2024-2030 is 28.46%, suggesting that the market volume will grow to USD 826.70 billion by 2030. Globally, the largest market share will be in the United States, where the market size is expected to reach USD 50.16 billion in 2024. (Statista 2024). From the information provided, AI is increasingly being perceived as a means of improving our daily lives as manufacturers and consumers. This is mainly due to the optimization of time and effort. Adequately adapting to the ever-changing horizon of trade requires businesses to offer their services and products online in a timely and appropriate manner. This also applies to the fashion industry which also faces the challenge of attracting, retaining and keeping the attention of the fashion consumer, turning him into a loyal customer. According to Geopost data from January 2024, the most frequently ordered goods online are from the fashion clothing industry and occupy 58% (Geopost, n.d.). In the e-commerce sector specifically, artificial intelligence consists not only of human-like robotics, but the training technologies and algorithms which provide the foundation. By implementing AI in e-commerce, merchants can provide optimized customer experience to visitors to their sites through accumulated business and customer data. The information collected is used for better, high quality analyses, as well as for better forecasting of the future. Given the exponential progress in technology, every day sees new improvements, innovations, proposals which compete for the attention of not only online merchants, but also representatives of their target audience. AI technologies are a powerful tool for both increasing sales and optimizing operations.

The article entitled “Artificial Intelligence in Ecommerce: How This Rapidly Evolving Tech Will Change the Online Storefront”, published on the website of one of the leading e-commerce platforms – BigCommerce.com, highlights some of the ways in which AI is developing the sector and expanding the boundaries of what can be achieved by e-commerce. (BigCommerce, n.d.). Among other aspects, more targeted marketing efforts and advertising are mentioned. Personalizing advertising messages across any communication channel of stores and brands is essential. The analysis of large amounts of data regarding purchase history and interactions of all kinds with customers allows for their deeper understanding and, accord-

ingly, a more concentrated distribution of the message which will resonate for the longest in terms of reaching the right audience. In turn, personalized marketing and advertising messages also have an impact in terms of customer retention. (BigCommerce.com). Another aspect is the performance of operational tasks with as little human intervention as possible (automation of processes where the implementation of AI is possible). This includes scheduling email campaigns in CRM and other marketing tools, product recommendations, providing discounts for loyal customers, etc. (BigCommerce, n.d.).

The sales process is also undergoing changes, leading to facilitation and increased results. The commercial sector of fashion items and acute competition between retailers and fashion brands imply increased attention and special care to potential and current customers. Usman Aslam indicates that, unlike mainstream brands, fashion brands are more focused on personal expression and individuality. The emphasis falls on identity, aesthetics and style, which makes them more challenging to promote before a wide audience (*Understanding the usability of retail fashion brand chatbots: Evidence from customer expectations and experiences*, 2023). In turn, behavior changes as a result of the numerous choices in front of them and customers become more demanding and impatient. One way to further support the sales process is to introduce chatbots as a valuable tool for improving service. They provide a quick real-time solution to the requests of demanding and eager customers. Therefore their implementation in the fashion industry is becoming an increasingly common practice. Among the examples of chatbots which have successfully implemented on their websites are brands such as Victoria's Secret, Louis Vuitton, Tommy Hilfiger, Burberry and others. ("Fashion chatbots do they really work", n.d.)

### **Types of diversity**

The regular performance of certain tasks and activities in combination with created habits indeed optimized the time and resources needed for their implementation. However, at the same time, it creates an obstacle for the people involved in work, since it limits the perimeter of creative development and mixing techniques. Manovich highlights the role of AI in culture and the growing influence on our choices, behaviors and imagination (Manovich & Arielli 2024). In addition to recommending photographs, music and other types of media files, AI also automatically enhances photographs and selfies to meet socially accepted norms for in the context of contemporary culture in its digital form (Manovich 2018: 2) The emphasis falls on the importance of distinguishing between types of diversity. One

type is defined as “diversity of content” (2018: 7), referring to objects created in a particular cultural area. In the context of the above-mentioned photographs, this refers to the techniques and styles, including the filters provided for use. The second aspect of cultural diversity concerns consumer choice. Modern fashion designers around the world can create fashion items characterized by a variety of styles, silhouettes, shapes, textures and colors. However, this does not mean that the range of goods which fashion consumers buy and apply in their clothing will not be significantly more limited. On the other hand, the reason for the great variety can be attributed to the fact that in their desire to create a unique look, a large part of consumers combine different elements in their clothing which are not offered by designers and retailers. (2018: 7).

Such photographs and media artifacts find application in several social networks online. In “*Instagram and Contemporary Image*” (2017), Manovich focuses on Instagram, originally released as an iPhone app in 2010 and calls it “a good example of the new era of mobile photography” (Manovich 2017). Here again, as in the above paragraph, we are talking about combining elements. In the case of Instagram, they are taken from various contemporary and historical greens, as well as from commercial and marketing proposals (Manovich 2017: 137). A comparison for such a combination is also given with some of the world’s best fashion designers, such as Alexander McQueen, John Galliano, Jean-Paul Gaultier (Manovich 2017: 137). “Instagram” at the time was a neologism. The initial idea of the application was related to speed of decision-making and swiftness of action, but now the modern visually sophisticated audience and representatives of the global creative class use it in the opposite way – for the purpose of visual pleasure (Manovich 2017: 95). Again, thanks to Manovich, we know that, as in other cultural areas, the choices of Instagram users (liked photos, followed users, posted photos) can serve to legitimize the economic and social status of a given user (Manovich 2017: 40). The legitimization of such a status can also occur in an identical way to what is communicated through the clothing of the fashion consumer. For its part, adapting and coordinating certain decisions means consistency and applies to any “design of life”, not only those communicated on the social networks. The practical application of the above skills leads to the creation of cultural capital, the measures of which are respect in the community and the number of followers. Affirmation in the community leads to the transformation of cultural capital into economic capital. This is when the consumer begins to collaborate with representatives of both the commercial sector and advertisers (Manovich 2017: 99–117).

From Roland Barthes (*Camera Lucida: Reflections on Photography*, 1981) we know that the photo portrait is an intersection of forces. Standing in front of the lens, the individual occupies several roles – the one they are thinking about; the one they want others to think of; and the one the photographer considers them to be (Barthes 1981). Visuals posted on Instagram, in addition to conveying the complex combination of images, communicate techniques, styles, visual solutions (Manovich 2017: 40). Since a large part of the users of the application are interested in aesthetics, one of the main reasons for the social network to become dominant in mobile photography is its filters. They are a quick and easy mechanism for giving additional visual effects and an attractive look to photos (Manovich 2017: 40). Attractive photographs of images and fashion products are present on them abound. As Kristian Bankov writes (1970–) in “*Face and trust: A semiotic inquiry into influencers, money, and amygdala*” (2021): After the cultural explosion of Web 2.0, digital culture reveals an apparently semiotic paradox associated with the incredibly widespread use of facial images. At the same time the reason to trust in the authenticity of these faces is constantly declining. This is because graphic technology has made the sophisticated manipulation of images both possible and easy. (Bankov 2021). Social networks online, in which the aforementioned images circulate endlessly, are designed in a way which provides their users with the ability to create discrete signs of attention and interest measured quantitatively (Manovich 2020: 181). The creation of these photographs, or “Media Artifacts” as Manovich calls them, to a level where they are aesthetically satisfying, is possible with sufficient progress in the field of AI. This will happen when the computer has roughly the same knowledge of the world as the human individual (Manovich 2018: 10).

### **Other innovations/ Innovations in other fields/ Diversity in several fields**

Fashion is always semiotic, and its semiotics is manifested in the requirement and necessity of an observer, even if the information to it is unexpected and incomprehensible (Lotman 1998: 77). According to Lotman, the triumph of fashion consists in opposing the crowd (Lotman 1998: 77). In turn, fashion implies a continuous experimental verification of the limits of the conventional (Lotman 1998: 94). In the context of the promotion of fashion items in modern online social networks through popular users, as well as technological progress and the emergence of virtual fashion influencers generated through AI, this mechanism can also be viewed through the perspective of the outraged and shocked public which, according to

Lotman, fashion aims to provoke (Lotman 1998: 77). This dynamic – in which fashion is constructed through opposition, provocation and constant testing of boundaries – can also be seen as a mechanism which generates diversity, insofar as it stimulates the continuous production of new, different visual and behavioral models from the mainstream. In this sense, the relationship between fashion and diversity should be thought of as a result of its very semiotic nature. The main difference between a traditional fashion influencer and one created through computer technology lies in their mode of existence. The first is a living human being, and the second is a virtual construct created by human beings. Despite the artificial nature of the latter, it has the full potential to set fashion trends and seamlessly promote brands in innovative ways. (Izea, n.d). Examples of virtual influencers are Lil Miquela with over 2,577,456 followers (Lil Miquela, n.d.) on Instagram and Imma with 387,589 followers. (Imma, n.d.) The trend is constantly growing and in addition to the growing popularity of fashion influencers, the virtual clothing on offer is also developing in parallel. Some brands which enjoy wide popularity, such as Gucci, Adidas and others, are developing their product portfolio with virtual clothing, and there are brands specializing in selling only such. (DressX, n.d.)

#### SPACE CRYSTAL



**Figure 1:** “Space Crystal” virtual sweatshirt (Adidas Metaverse, n.d.)



**Figure 2:** Virtual dress, named “Dress Flight” (DressX: Dress-Flight, n.d.)

On the other hand, digitalization and its logic of automation make the postmodern individual vulnerable to deeper individualization and alienation (Hassan 2020: 134). In *Paradoxical Happiness*, Gilles Lipovetsky also mentions the vulnerability of the individual, the anxieties and depressions which accompany him in his current existence, the value of which is questioned by the individual himself. (Lipovetsky 2008: 158). Lipovetsky also speaks of “individualized consumption” (Lipovetsky 2008: 35). Narcissistic infatuation and the pleasure of comparison with others, in favor of oneself, aimed at achieving difference and not resembling others. These are characteristics of the civilization of the hypermarket (Lipovetsky 2008: 41), and the behavior of unpredictable and fickle consumers. The individual who is in the cultural space of clothing, according to Lotman, is forced to resist the conflict between the desire to adhere to the traditional and its opposite extravagance (Lotman 1998: 77).

The product and information abundance have serious consequences: the variety of fashion items with different designs, colors, textures; and the numerous influencers and trend setters setting the direction of existing or emerging fashion trends. The “individualized consumption” already mentioned by Lipovetsky, is when the individual consumes for their own pleasure in being different from the crowd (Lipovetsky 2008: 35). However, simultaneous vulnerability, or an overload of “channel capacity” is also

possible. In “The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information” (1956), the American psychologist George Armitage Miller (1920–2012) defined “channel capacity” as “the upper limit of the extent to which the observer can match his responses to the stimuli we give him.” (Miller 1956: 82). In the context of operational and creative tasks, routine tasks may require less cognitive effort, allowing people to handle simultaneous multiple tasks efficiently. On the other hand, complex, creative, or new tasks may require more mental energy, making it difficult to perform them at the same time. This can lead to cognitive overload. Viewed through a different perspective from that mentioned in the first part of the text, in which routine activities are an obstacle and a barrier to potential creative development, AI can also be seen as contributing to the increase of creative skills in humans. D.Ariely touches on the human limits of aesthetic capacity and the attainment of peak creativity and aesthetic sensitivity. He points to the personal qualities of the individual and generally accepted cultural norms as factors shaping the individual’s capabilities in the given artistic field (2021: 9). In the context of the development of modern technological systems and their tools in creative disciplines such as design and advertising, AI represents a further expansion of people’s abilities to interact even more fully with the world (2021: 9). The algorithm is unable to produce the cultural process which led to the emergence of the artistic artifact and/or design. Not all individuals have developed artistic skills to a remarkable or even impressive level, and AI is programmed to simulate learned skills through imitation, rather than to simulate universal cognitive abilities (Manovich 2018: 2). As we know from Lotman’s theory (Lotman 1992: 163), the phenomenon of consciousness is related to the individuation factor. For a system to be intellectual, it must consist of individualities. Following this line of argument and given the many events which impact the world’s cultural and aesthetic diversity, the role of AI has not yet reached a point of greatest significance. Billions of people (individuals) do not yet have access to the Internet and the automation of aesthetic solutions is still far from the completion stage.

The exponentially growing variety of fashion items, designs, patterns of fashion (real or virtual), even if it creates conditions for fashion consumers to be increasingly insecure and vulnerable, does not mean that AI technology should be perceived as a threat. AI is a progressive tool which aims to make our daily lives easier as users and creators. On the other hand, as written in *Seven arguments about AI images and generative media*, “if we think humans are creative and AI are not, this is because we better understand how AI works, while we still do not sufficiently understand how

humans work.” (2023: 12). AI needs humans and their creativity, and for humans, critical thinking and ideas will become increasingly important.

### Final words

The development of aesthetic qualities and preferences is a process and result of all activities in a person’s life. Style means adapting and coordinating certain decisions, and requires consistency (Manovich 2017: 99). This concerns not only the choice of clothing of the fashion consumer, but also the overall lifestyle – what to eat, how to spend time, and even whom to imitate or not. All of the above cannot be achieved solely by AI, no matter how big its role is in the daily performance of elementary activities.

### References

DressX. (n.d.). Dress-Flight [Product page]. *DressX*. Available at: <https://dressx.com/collections/dresses/products/dress-flight> (accessed 13 July 2024).

Fashion chatbots do they really work. (n.d.). *IntelliStyle*. Available at: <https://www.intelistyle.com/fashion-chatbots-do-they-really-work/> (accessed 5 December 2025).

Geopost. (n.d.). E-shopper comparison tool. *Geopost*. Available at: <https://www.geopost.com/en/expertise/e-commerce-trends/e-shopper-comparison-tool/> (accessed 5 December 2025).

Hassan, R. 2020. *The Condition of Digitality: A Post-Modern Marxism for the Practice of Digital Life*. London: University of Westminster Press, 129–158.

IZEA. (n.d.). The rise of computer-generated influencers. *IZEA*. Available at: <https://izea.com/resources/the-rise-of-computer-generated-influencers/> (accessed 13 July 2024).

Lil Miquela [@lilmiquela]. (n.d.). Profile [Instagram profile]. *Instagram*. Available at: <https://www.instagram.com/lilmiquela/> (accessed 13 July 2024).

Manovich, L. 2017. *Instagram and Contemporary Image*. Available at: <http://manovich.net/index.php/projects/instagram-and-contemporary-image> (accessed 14 July 2024).

Manovich, L. 2018. *AI Aesthetics*. Strelka Press.

Manovich, L. 2020. *Cultural Analytics*. Cambridge, MA: The MIT Press.

Manovich, L., E. Arielli. 2024. *Artificial Aesthetics: A Critical Guide to AI, Media and Design*. Available at: <http://manovich.net/index.php/projects/artificial-aesthetics> (accessed 14 July 2024).

Miller, G. A. 1956. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." *Psychological Review*, Vol. 63, No. 2.

Statista. 2024. Artificial Intelligence – Worldwide. *Statista* [Market outlook]. Available at: <https://www.statista.com/outlook/tmo/artificial-intelligence/worldwide> (accessed 13 July 2024).

Statista. 2025. Artificial Intelligence – Worldwide. *Statista* [Market outlook]. Available at: <https://www.statista.com/outlook/tmo/artificial-intelligence/worldwide> (accessed 5 December 2025).

Липовецки, Ж. 2008. *Парадоксалното щастие. Опит върху обществото на хиперконсумиране*. София: РИВА.

Лотман, Ю. 1992. *Култура и информация*. София: Наука и изкуство.

Лотман, Ю. 1998. *Култура и взрив*. Кралица Маб.

## NOTES FOR CONTRIBUTORS

Articles should be no more than **40,000 characters** including spaces and including bibliography

### Composition

- Manuscripts should be compiled in the following order: title page; author/s name/s (in *Italic*), abstract; keywords; main text; acknowledgments; references; appendices (as appropriate); table(s) with caption(s) (on individual pages); figure caption(s) (as a list).
- Abstracts of 200–300 words are required for all manuscripts submitted.
- Each manuscript should have 5 keywords.
- All authors of a manuscript should include their full names, affiliations, postal addresses, telephone numbers and email addresses on the cover page of the manuscript. One author should be identified as the corresponding author. Please give the affiliation where the research was conducted. If any of the named co-authors moves affiliation during the peer review process, the

- new affiliation can be given as a footnote. Please note that no changes to affiliation can be made after the manuscript is accepted. Please note that the email address of the corresponding author will normally be displayed in the article PDF (depending on the journal style) and the online article.
- All persons who have a reasonable claim to authorship must be named in the manuscript as co-authors; the corresponding author must be authorized by all co-authors to act as an agent on their behalf in all matters pertaining to publication of the manuscript, and the order of names should be agreed by all authors.
- Please supply a short biographical note for each author.

### Font

12 point Times New Roman throughout

### Typeface, emphasis, and punctuation

- Italics should be used for:
  - words, phrases, and sentences treated as linguistic examples;
  - foreign-language expressions;
  - the titles of books, published documents, newspapers, and periodicals.
- Boldface type may be used:
  - to emphasize a word or phrase in a quotation, if so indicated “[emphasis mine]”;
  - to draw attention to a particular linguistic feature in numbered examples (not in running text).
- Please do not use underlining or capital letters for emphasis.
- Single quotation marks should be used only for the translation of non-English words, e.g., *cogito* ‘I think’.
- Double quotation marks should be used in all other cases, i.e. for:
  - direct quotations in running text;
  - “qualified” words or phrases.
- Use rounded quotation marks (“...”) not “straight” ones.
- Full stops should be placed last, following any other punctuation, e.g., “... word.”; “... word.”; “...word.7” (but “... word7” within a sentence).

– Parenthetical dashes are longer than hyphens. If you cannot print dashes, use double hyphens.

– An “en”-dash, “–”, is used to indicate continuing or inclusive numbers, such as “1965–1966”, or pages “5–8”.

If your word processor has no en-dash, use two hyphens characters.

### **Titles and headings**

– The text should be divided into sections and, if necessary, subsections, with appropriate headings.

– All headings, including chapter titles as well as in the table of contents, begin flush left.

– For all headings in the file, please do not use generic codes but the following numbering system:

#### **1. Main heading**

##### **1.1. Section heading**

###### **1.1.1. Subsection heading**

– Do not end a title or heading with a period when it is to appear on a line separate from the text.

– Capitalize only the first letter of the first word and of those words which the orthography of the languages requires to begin with a capital letter (e.g., proper nouns). This also applies to the table of contents.

### **Paragraphs**

All paragraphs should start with a tabulator (g) 1 cm from the left margin.

### **Quotations**

– Short quotations (fewer than 60 words) should be run on (i.e., included within the text) and should be enclosed in double quotation marks. Single quotation marks enclose quotations within quotations.

– Longer quotations should appear as a block; separated from the text and indented by 1 cm from the left margin. They are not to be enclosed within quotation marks.

– All quotations should follow the original text exactly in wording, spelling, and punctuation. Any additions by the author should be indicated by square brackets. Indicate omissions by ellipsis points within brackets.

– All quotations in languages other than English are to be followed by the translation in square brackets.

### Citations

Full bibliographical details are given in the reference section at the end of the book or article. Brief citations are used in the text. Examples:

(Cobley 2008)	one author
(Bankov & Cobley 2017)	two authors
(Deely, Ponzio, and Petrilli 2005)	three authors
(Zlatev et al. 2006: 38)	four or more authors (but give all the authors in the reference entry)
(Leone 2004a, 2004b, 2014)	works by one author
(Zantides 2011; Kourdis 2012)	works by different authors
(Khanwalkar 2016: 60–65)	no dropped digits in inclusive numbers
(Arcagni & Santangelo 2017, 2: 110)	volume number
(Saussure [1916] 1967: 37)	reprints: with original date at first mention; in all subsequent citations “Saussure 1967: 37”
(Andacht 2014: n.p.)	an authored page on a website
(Brand semiotic survey ...:2016 n. p.)	Shortened title on a web page with no author credited

– The date is always given in brackets: “Martinelli (2015: 123–125) introduced the term”; “In his (2017) article Bankov argued that ...”.

– Give page numbers in full: do not use “f.”, “ff.”.

– Avoid referring to a whole book: give exact page numbers whenever possible. Always give the page number with quotations.

### **Abbreviations**

– Use only the simplest and most common abbreviations (i.e., etc., e.g., et al.).

– Do not use periods after acronyms.

– Abbreviations common in linguistics (NP, V, ACC) may be used in numbered examples but the terms should be written out in full in the text wherever possible.

– Do not use sequences of letters to represent names of theories, titles of books or names of publishers; thus:

“the Semiotic Animal Theory”, not “the SAT”; “Eco 1975”, not “ToS” (Theory of Semiotics), “New Bulgarian University Press”, not “NBUP”.

### **Examples**

– Number examples article by article in an edited work.

– Foreign-language examples should be presented in *italics*.

It is recommended to use tabs to align the examples and glosses. If you have difficulties in aligning glosses, please clearly indicate by hand the proper alignment in the manuscript/printout.

### **Tables, figures, and illustrations**

- Tables and figures should be numbered consecutively and be given titles. The title of a table should appear above the table, the title of a figure below the figure.
- If there are figures to be included, please send us the original design files.
- If images are to be inserted, good quality and high resolution files are required.

### **Appendices and notes**

– An appendix is placed at the end of the text, before the notes.

– Use footnotes and do not exceed in their number.

– Note numbers in the text should be superscript (small raised) numbers without parentheses.

– The note number should directly follow the word in question or a punctuation mark, with no blank space.

## References

– Wherever more than one author invert only the name of the first one, “Bankov, K., P. Cobley and K. Kull”

– Give the full title and subtitle of each work.

– Give both the place of publication and the publisher.

– Do not use abbreviated forms of the names of journals, book series, publishers or conferences.

– Titles of published books and journals are capitalized and italicized; unpublished works, such as Ph.D. dissertations, and the titles of articles in journals or edited works are neither capitalized nor italicized (see examples below).

– Give the inclusive page numbers of articles in journals or edited works.

– Do not use “et al.” but give all names.

– Translate titles in languages other than French, German, Spanish and Italian into English.

– Please input all bibliographical entries in a consistent format: Author, Year of publication, Title, etc. In other words, there are three fields of information, one for the author(s) or editor(s), one for the year of publication, and one for the rest.

Where there are more than one works by the same author/group of authors, the author name(s) should be repeated in each entry (i.e. do not leave blank or use EM-dashes as placeholders).

### **Book (authored work):**

Cobley, P. 2002. *Narrative. The New Critical Idiom*. London: Routledge.

### **Book (edited work):**

Bankov, K., P. Cobley (eds.). 2017. *Semiotics and Its Masters*. Berlin and New York: Mouton de Gruyter.

### **Contribution in an edited work:**

Tarasti, E 2017. Culture and Transcendence – The Concept of Transcendence Through Ages. In Bankov, K., P. Cobley (eds.). 2017. *Semiotics and Its Masters*. Berlin & New York: Mouton de Gruyter, 293–325.

→ Note: Entries for contributions in edited works should always include full bibliographical information for the edited work. Abbreviating the entry (here, e.g., with “In Bankov et al., 293–325”) is not acceptable.

→ Note: If a contribution in an edited work is cited in the article text, a separate, additional entry for the edited work should not be included in the References unless the edited work is cited directly and as a whole.

**Journal article:**

Giorgi, F., L. E. Bruni. 2001. “Germ Cells are Made Semiotically Competent During Evolution”. *Biosemiotics*, Vol. 9, No. 1, 31–49.

**Journal article also published electronically:**

Peng, Jia. 2017. On Imagination: From the Perspective of Semiotic Phenomenology. *Signs and Media*, No 15 Autumn 2017. Available at: <http://www.semiotics.net.cn/userfiles/images/b1e92f1b45962556b7698f-342936ca3f.pdf> (accessed 10 June 2018).

→ Note: Publication date = year of online publication or year of the latest update. The date on which the URL was accessed should be provided in parentheses at the end of the entry.

**Special issue of a journal (cited as a whole):**

Cobley, P., A. Randviir (eds.). 2009. *Sociosemiotica*. [Special issue]. *Semiotica*, Vol. 2009, No. 173, Issue 1–4.

**Reprint:**

Bankov, K., P. Cobley (eds.). 2020 [2017]. *Semiotics and Its Masters. Vol. 1*. 2nd ed. Berlin & New York: Mouton de Gruyter.

**Thesis/dissertation:**

Bankov, K. 1995. Il linguaggio come elemento positivo nell’antiintelletualismo bergsoniano. [Thesis/dissertation]. Bologna: Bologna University.

**Paper presented at a meeting or conference:**

Vuzharov, M. 2017. Personalization Algorithms – Limiting the Scope of Discovery? Paper presented at the 13th World Congress of the International Association for Semiotic Studies (IASS/AIS), Kaunas University of Technology, 26–30 June.

**Several works by one author/editor with the same publication date:**

Leone, M. 2017a. Semiótica de la reparación. In Pardo Abril, N. G. (ed.). *Materialidades, discursividades y culturas. Los retos de la semiótica Latinoamericana*. Bogotá: Instituto Caro y Cuervo. Imprenta Patriótica, 142–159.

Leone, M. 2017b. Fundamentalism, Anomie, Conspiracy: Umberto Eco's Semiotics against Interpretive Irrationality. In Thellefsen, T., B. Sørensen (eds.). 2017. *Umberto Eco in his Own Words*. Berlin & Boston: Mouton de Gruyter, 221–229.

**A published work reproduced on a private website:**

Bankov, K. 2011. Technology, the Imaginary and the Transfer of Experience: between the Market and Social Networks. In Leone, M. (ed.). *Lexia. Rivista di semiotica, 7–8. Immaginari. Prospettive disciplinari*. Rome: Aracne editrice, 255–278. Available at: [http://bankov.net/Statia\\_08\\_ENG.pdf](http://bankov.net/Statia_08_ENG.pdf) (accessed 19 March 2018).

**An article without author on a website:**

The World's Most Valuable Brands. 2017. Available at: <https://www.forbes.com/powerful-brands/list/#tab:rank> (accessed 19 March 2018).

**Technical issues**

Graphics may be submitted in all major graphic file formats, e.g., JPG, TIFF, EPS, etc. Please contact the publisher if you are in doubt whether a particular format will be acceptable. Please note that it is difficult to edit eps-files. Occasionally, graphic files will have to be reprocessed; it is therefore preferable if all graphic files are submitted in a format amenable to further editing. Certain custom-written applications for the visualization of, for instance, statistical data use proprietary file formats and lack filters for the export of files into common file formats. The processing of data generated with such applications is not possible without the respective applications themselves. In such (rare) cases you are asked to contact the publisher beforehand and *obtain permission* if you make use of copyrighted graphics.

**Obtaining permissions**

It is the author's responsibility to request any permission required for the use of material owned by others. When all permissions have been received, the author should send them, or copies of them, to the publisher, who will note, or comply with, any special provisions regarding credit lines contained in them.

ISSN (Print) 2603-3585  
ISSN (Online) 2603-3593

[semiotics.nbu.bg/en/](http://semiotics.nbu.bg/en/) 