Humanities Studios: Vol. 13, No. 2, 2025

Pedagogy, Psychology, Philosophy

Journal homepage: https://humstudios.com.ua/en

UDC 004.5

DOI: 10.31548/hspedagog/2.2025.88

From manipulation to expression: The philosophical shift to natural language interfaces

Andrii Lykhatskyi*

Master of Sciences, Graduate Student
Odesa National Polytechnic University
65044, 1 Shevchenko Ave., Odesa, Ukraine
https://orcid.org/0009-0002-3735-2862

Abstract. The modern era of rapid artificial intelligence development witnessed a fundamental paradigmatic shift in human-computer interaction – from technocratic manipulation to conscious expression through natural language. The study aimed to philosophically examine the transition from manipulative to expressive interaction with natural language interfaces. An interdisciplinary approach was employed, combining methods of philosophical analysis, cognitive phenomenology, and digital culture critique. The findings revealed that modern large language model-based interfaces not only simplified communication but also reconfigured the ontological foundations of technological experience. The growing "semantic opacity" of large language systems challenged traditional notions of understanding, accountability, and explainability in technical interaction. A phenomenological shift toward "invisible" technology integrated into the subject's cognitive architecture was identified. The phenomenon of technological intersubjectivity was analysed, blurring the boundary between human and machine agency. It was demonstrated that language as an interface not only transmitted information but also functioned as a medium for shaping self-identity, autonomy, and cognitive responsibility. The conclusions proved that this transformation carried profound philosophical implications and required systematic reflection to prevent the erosion of human autonomy in the context of cognitive fusion with artificial intelligence. The necessity of an ethical framework for new interfaces was emphasised – one that ensured not only functionality but also the preservation of human agency. The practical significance lay in establishing philosophical foundations for designing natural language interfaces that supported human autonomy, responsibility, and epistemic independence

Keywords: human-computer interaction; cognitive abstraction; graphical user interfaces; artificial intelligence; philosophical mediation

Received 20.02.2025 Revised 01.05.2025 Accepted 22.05.2025

Suggested Citation:

Lykhatskyi, A. (2025). From manipulation to expression: The philosophical shift to natural language interfaces. *Humanities Studios: Pedagogy, Psychology, Philosophy*, 13(2), 88-102. doi: 10.31548/hspedagog/2.2025.88.

*Corresponding author

Introduction

Modern artificial intelligence development is transforming human-technology interaction: shifting from technocratic control to natural language communication, reshaping philosophical concepts of agency and accountability. The "semantic opacity" of language models challenges traditional notions of understanding and explanation, impacting human cognition and self-identification. A critical task now is examining how these technologies affect personal autonomy and epistemological independence to ensure harmonious coexistence rather than technocratic subjugation.

The work of B. Chen *et al*. (2025) explored the application of large language models (LLMs) in the domain of philosophical counselling, revealing their capacity to significantly enhance the accessibility, adaptability, and effectiveness of humanities-cantered services. Their analysis underscored how natural language interfaces could democratise philosophical reflection and dialogue, making such engagements more culturally responsive and technically scalable. However, they also emphasised that LLMs, while linguistically proficient, remain limited in their ability to exhibit genuine empathy and contextual understanding, raising serious ethical questions about the authenticity and confidentiality of AI (artificial intelligence)-mediated interactions in deeply human-centric domains. C. Tarsney (2025) offered a critical examination of the epistemic and moral risks posed by LLMs, particularly regarding large-scale disinformation, manipulation, and the opacity of AI-generated content. He argued for a paradigm of "extreme transparency," in which AI outputs would be consistently contextualised, annotated, and monitored by embedded meta-commentary systems. This work illuminated how the increasing autonomy and agentic potential of advanced natural language interfaces demand rethinking normative boundaries in AI ethics and require the development of robust safeguards to prevent systemic cognitive manipulation and epistemic erosion.

In a complementary vein, M. Rahman *et al*. (2023) investigated the impact of natural language processing (NLP) technologies on artistic creativity, focusing on their use in poetry and literary production. Their findings illustrated the dual nature of human-AI co-creation: on one hand, such collaboration expands the boundaries of artistic expression; on the other, it introduces complications regarding authorship, algorithmic opacity, and the reproduction of cultural biases. The authors proposed design-oriented solutions to align NLP tools more closely with the needs of artists, advocating for transparency in training data and intuitive interface design as a way of reinforcing human agency in creative processes. G. Dupre (2021) contributed to the philosophical debate on the identity thesis between language and thought by interrogating contemporary linguistic theories. He challenged the notion that natural language and cognition are functionally equivalent, pointing to instances where semantically rich but grammatically irregular utterances signal a deeper disjunction between linguistic structure and conceptual intent. This insight is critical in the context of natural language interfaces, as it raises concerns about whether current LLMs, which are fundamentally syntactic machines, can authentically model the nuance and fluidity of human reasoning.

The research by A. Gatti & V. Mascardi (2023) introduced the VEsNA framework, a system for interacting with virtual industrial environments through natural language commands. Their work demonstrated how simplifying interfaces via natural language can empower users without technical expertise, contributing to inclusive automation and widening access to digital systems. This case highlighted the role of expression over manipulation in shaping user-centric digital environments, pointing toward a broader shift in interface design philosophy that prioritises accessibility and semantic coherence over code-based precision. R. Millière & C. Buckner (2024) examined core philosophical challenges linked to

the development and deployment of large multimodal language models. Their study delved into the interpretability problem, raising scepticism about claims that LLMs can replicate or approximate human consciousness or intentionality. They explored the ethical and cognitive limits of using such models to simulate human-like understanding, emphasising the gap between functional behaviour and the internal architecture of subjective experience – thus reinforcing the need to distinguish between expressive interface capacity and genuine cognitive depth.

The work of R. Nefdt (2024) provided a historical-philosophical account of human-computer interaction, charting the transition from mechanistic control paradigms to dialogic, expressive forms of engagement via natural language. He argued that this evolution repositions technology not merely as an operational tool but as a conversational partner, thereby shifting the ontological status of interfaces from passive conduits of action to co-constructors of meaning. This reconceptualisation has profound implications for the distribution of agency and the design of future communicative systems that are responsive, adaptive, and collaborative by nature. In a related direction, A. De Vrio *et al.* (2025) addressed the phenomenon of anthropomorphisation in language-based AI, presenting a taxonomy of its linguistic expressions and psychological effects. They cautioned against the tendency to conflate expressive fluency with human-like understanding, warning that such projection can lead to a degradation of genuine interpersonal relationships. Their work contributed important ethical and conceptual tools for interrogating the increasingly blurred boundaries between human and artificial agents in natural language environments.

Despite the breadth and depth of these contributions, several critical dimensions of the shift from manipulation to expression remain underexplored. Most notably, the phenomenological transformation whereby natural language interfaces become "invisible mediators" of experience – blending seamlessly into cognitive and communicative routines – has yet to be fully theorised. Moreover, the structural redistribution of epistemic and cognitive responsibility between human users and intelligent systems remains insufficiently mapped, leaving open essential questions about agency, understanding, and the philosophical stakes of living and thinking alongside expressive machines. These gaps signal the need for sustained inquiry into how natural language interfaces do not merely facilitate expression but co-author the evolving contours of subjectivity and digital being.

The research objectives focused on analysing the philosophical aspects of shifting from manipulative to expressive interaction through natural language interfaces, with particular emphasis on transformations in cognitive, phenomenological and epistemological foundations of human-digital system interaction. The study sought to: examine key philosophical changes prompted by natural language interfaces regarding understanding, agency and intersubjectivity; investigate how natural language interaction influences human thought structures and enables technologically extended cognition; and evaluate both risks and opportunities for human autonomy development within increasingly AI-integrated thinking and communication processes. The research methodology was based on historical-genetic analysis of the evolution of interface paradigms, complemented by philosophical-phenomenological interpretation of changes in human-computer interaction. For data systematisation, comparative analysis of tabular summaries reflecting key characteristics of different stages of interface development was applied. Conceptual synthesis of theoretical propositions provided an interdisciplinary foundation for analysing the cognitive, epistemological and social consequences of the transition to language interfaces.

The evolution of interface paradigms

The desktop metaphor has historically dominated human-computer interaction by requiring users to adapt their cognition to computational logic through visual representations. This paradigm of

explicit manipulation demands that users translate their intentions into discrete, predetermined actions: clicking specific buttons, dragging icons to particular locations, or navigating through hierarchical menu structures. While revolutionary for its era, this approach imposed significant cognitive constraints by requiring users to maintain mental models of interface operations and learn specific manipulation sequences. The transition from traditional graphical user interfaces (GUIs) to natural language interfaces, particularly those mediated by large language models (LLMs), introduces not only technical and cognitive shifts but also critical epistemological and ethical considerations. One such concern is the potential for linguistic manipulation. LLMs, by design, generate language that appears contextually appropriate and semantically fluent. However, this fluency can mask implicit biases embedded in training data or model architecture, leading to subtle yet pervasive influence on user decision-making. When users consult language models for information or recommendations, even minor phrasal choices, modal verbs, or framing structures may predispose them toward specific interpretations or actions. For example, suggestions such as "you might want to…" or "the most effective way is…" inherently guide user judgment under the guise of neutrality. This dynamic raise important questions about agency, especially when the user perceives the model as an objective or omniscient authority. The phenomenon is further complicated by the tendency of users to anthropomorphise conversational agents, attributing to them intentionality or trustworthiness not warranted by their statistical foundations. As such, linguistic interfaces must be interrogated not only for their functionality but also for their potential to shape epistemic environments, social perceptions, and individual autonomy through discourse patterns that appear benign yet exert rhetorical force (Kirkwood, 2024).

Simultaneously, the increasing reliance on language-based interaction interfaces may contribute to a decline in spatial and procedural cognitive skills traditionally reinforced by GUI-based manipulation. GUIs, through their spatial metaphors and hierarchical structures, encourage users to develop and maintain mental maps of digital environments, fostering algorithmic reasoning, problem decomposition, and procedural planning. As interaction becomes more declarative and expressive – focused on what the user wants rather than how to achieve it – the necessity for users to internalise underlying processes diminishes. This shift risks a form of cognitive deskilling, where users may lose familiarity with conceptual models of system operation, logic sequencing, or structural dependencies. The phenomenon echoes concern in cognitive psychology regarding automation bias and overreliance on intelligent systems, wherein the delegation of procedural tasks to autonomous agents leads to atrophy in users' problem-solving competencies. While natural language interfaces offer unparalleled accessibility and efficiency, they may simultaneously erode the development and retention of spatial, algorithmic, and exploratory reasoning essential for digital literacy in complex systems.

The emergence of natural language interfaces, powered by large language models and multimodal artificial intelligence systems, suggests a deep reconfiguration of this relationship. Rather than manipulating abstract interface elements, users can now express their intentions directly through natural language. This shift from manipulation to expression fundamentally changes the nature of human-computer interaction: instead of learning how to manipulate an interface, users can simply state what they want to accomplish. This transformation demands philosophical examination not merely as a technical advancement, but as a fundamental shift in how humans relate to and think with computational systems (Cohen, 1992).

The evolution of computer interfaces represents a succession of increasingly sophisticated abstractions from machine-level operations, each stage marking a distinct shift in the cognitive relationship between human and computer.

At the dawn of computing, interaction required direct manipulation of physical hardware through switches, punch cards, and hardwired connections. This hardware interface era demanded explicit understanding of machine architecture and binary operations, creating a high cognitive barrier to entry (Naik & Meghanandha, 2024; Brincker, 2024). The subsequent emergence of text-based interfaces initiated the first major abstraction, replacing physical manipulation with symbolic representation. This command-line interface paradigm introduced what P. Dourish (2017) terms "symbolic abstraction" – the representation of computational processes through textual commands. While still requiring users to learn specific command syntax, this paradigm established a more conceptual interaction model. The revolutionary introduction of visual metaphors and direct manipulation principles marked the transition to graphical user interfaces, representing a fundamental shift toward spatiotemporal interaction models. This transformation democratised computing access by leveraging users' innate understanding of physical space and object manipulation (Norman, 2013). The evolution of human-computer interaction paradigms is depicted in Table 1.

Table 1. *The evolution of human-computer interaction paradigms is depicted*

Time period	Interface paradigm	Example	Cognitive load	Level of abstraction
1950s-1960s	Hardware Interfaces	Switches, Punch Cards	High	Low
1970s-1980s	Command-Line Interfaces (CLI)	Text Commands	High	Low
1980s-1990s	Graphical User Interfaces (GUI)	Desktop, Windows	Medium	Medium
1990s-2010s	Early NLP Systems	Voice Commands	Lower	Higher
2020s	LLM-Based Interfaces	ChatGPT, Copilot	Low	High

Source: *compiled by the author based on F. Gurcan et al. (2020)*

The gaps between key stages in the evolution of human-computer interaction interfaces are driven by both technological and conceptual factors. The shift from hardware interfaces to command-line interfaces (CLI) in the 1970s-1980s became possible only with the emergence of operating systems (particularly UNIX), which provided the infrastructure for interactive operation and programming. This stage required not only the development of appropriate environments but also a rethinking of the very interaction model – from physical actions to symbolic control through text commands. Similarly, the transition from early natural language processing systems (1990s-2010s), which were primarily rule-based and relied on narrow specialised scripts, to modern LLM-based interfaces in the 2020s was enabled by the advent of transformer architectures, large-scale computational power, and access to vast text corpora. While the concept of linguistic interaction existed earlier, it was the technological breakthrough of the Transformer model that made it possible to implement a general-purpose language interface capable of flexible, context-sensitive, and high-level interaction with computer systems. Each evolutionary leap in interface design has thus been contingent on both enabling technologies and shifts in interaction paradigms. The move from CLI to GUIs in the 1980s-1990s, for instance, depended on advances in display technology and input devices alongside new conceptual frameworks for visual interaction. Likewise, today's AI-driven conversational interfaces represent not just better algorithms but a fundamental reimagining of human-computer communication – from rigid command structures to fluid, intent-based dialogue. These transitions underscore that interface evolution is never purely technical; it equally depends on redefining how humans conceptualise their relationship with machines (Gurcan *et al.*, 2020).

The success of graphical interfaces stems from their sophisticated exploitation of embodied cognitive schemas – fundamental patterns of sensorimotor experience that structure human understanding. Drawing on theory by G. Lakoff & M. Johnson of conceptual metaphor, it can be observed how graphical user interfaces leverage several essential cognitive structures (Yung, 2021). The concept of windows, folders, and nested hierarchies exploits embodied understanding of containment and spatial organisation through the container schema. The implementation of drag-and-drop operations and menu navigation paths builds on innate understanding of movement through space via the source-path-goal schema. Furthermore, actions like "pushing" buttons and "pulling" down menus utilise basic force-dynamic patterns identified by L. Talmy (1988) as fundamental to human cognition (Łozińska, 2021).

Each evolutionary stage manifests distinct phenomenological affordances that reshape the possibility space for interaction. Ecological theory of affordances, extended by D. Norman (2013) to technological interfaces, helps explain how these different paradigms create varying "action possibilities" for users. The graphical user interfaces paradigm, in particular, established what might be termed metaphorical affordances – interaction possibilities that emerge from the mapping of physical-world understanding onto digital environments.

However, the desktop metaphor's reliance on physical-world analogies has introduced significant constraints that shape the boundaries of human-computer interaction. The necessity to maintain coherent spatial metaphors has limited possible interactions to those that can be meaningfully mapped to physical experiences (Pitt & Casasanto, 2022). Perhaps most significantly, the emphasis on visual-spatial interaction restricts the expression of abstract concepts and relationships that don't readily map to physical metaphors. These inherent limitations of metaphorical interfaces set the stage for the emergence of the natural language interaction paradigm that promises to transcend these constraints through direct intentional expression.

From manipulation to expression: A conceptual framework

After tracing the evolution of interface paradigms, the fundamental distinction that characterises the current transformation in human-computer interaction must be carefully examined. The shift from explicit manipulation to intentional expression represents more than a mere technical advancement – it embodies a profound reconceptualisation of how humans engage with computational systems. Explicit manipulation, as manifested in traditional graphical interfaces, requires users to translate their intentions into discrete, predetermined actions within a spatially-organised metaphorical environment. When this mode of interaction is examined closely, it is found to embody what M. Heidegger terms "ready-to-hand" engagement with technology – tools become extensions of physical capabilities, but only through learned patterns of specific manipulations (Daniel, 2022). Consider how the user might restructure a document using a traditional interface: the user must know which menu contains the relevant commands, how to select appropriate sections of text, and how to execute specific formatting operations. Each step requires explicit manipulation of interface elements – clicking particular buttons, dragging handles to precise positions, or navigating through hierarchical structures. This mode of interaction, while powerful, imposes what D. Norman (2013) calls a "gulf of execution" between user intention and system action.

In contrast, intentional expression through natural language interfaces fundamentally alters this relationship. Rather than requiring users to learn and execute specific manipulations, these interfaces allow direct expression of desired outcomes. This shift represents what might be termed a "semantic turn" in human-computer interaction – from procedural knowledge to declarative knowledge, from knowing how to knowing

what. When users can simply state "reorganise this section to emphasise the main argument", they engage with the computer at the level of meaning rather than mechanism. A comparison of the "Manipulation and Expression Paradigms" is shown in Table 2.

Table 2. *Manipulation vs. expression paradigms*

Manipulation paradigm	Expression paradigm
Requires procedural knowledge	Uses declarative knowledge
Based on spatial metaphors (e.g., desktop)	Based on semantic abstraction
High cognitive load	Low cognitive load
Explicit tool awareness (user must focus on the interface/tool)	Implicit mediation (tool "disappears" into the background)
Example: Reformat document via GUI: select text → navigate menu → choose style	Example: "Make all headings bold" via natural language
Phenomenological state: ready-to-hand (until failure reveals the tool)	Phenomenological state: invisibly-at-hand (seamless experience)

Source: *compiled by the author based on K. Sedig et al. (2001), L. Ahmed (2018)*

This transition has significant consequences for human cognition and technological mediation. In the manipulation paradigm, users must maintain mental models of both their goals and the specific steps required to achieve them. The expression paradigm, however, allows what E. Pantano & D. Scarpi (2022) might be recognised as a more natural extension of human cognitive processes – it can think and express intentions in the same terms that are used for human-to-human communication. Yet this shift also introduces new complexities in the human-computer relationship. While explicit manipulation makes the mediating role of the interface apparent through its very constraints, intentional expression can obscure the complex translations occurring between natural language input and computational action. This tension between transparency and opacity emerges as a central philosophical challenge that will be explored more deeply in subsequent sections. The transition to natural language interfaces represents a fundamental shift in this dynamic, manifesting across multiple domains of human-computer interaction. Smart home control has evolved from app-based interface manipulation to natural language commands – users no longer navigate menus to adjust individual device settings, but simply express intentions like "set up for movie night". Database interaction has transformed from explicit Structured Query Language (SQL) queries to natural language requests like "show me sales trends from last quarter", with systems handling the complex query construction internally.

This transformation crystallised with the widespread adoption of ChatGPT in late 2022, demonstrating how a simple text input interface could replace numerous specialised graphical user interfaces elements, from image editors to spreadsheet functions. The subsequent integration of these capabilities into desktop applications marks another key evolution in this interface paradigm. Natural language interaction is no longer confined to isolated web-based applications but becomes native to everyday computing environments – from text editors and design tools to development environments. This integration allows for more seamless interaction with local tools and files while maintaining the directness of natural language expression. The text input box becomes a universal interface that connects users with both web services and local computing capabilities.

Rather than requiring users to translate their intentions into explicit manipulations – learning specific menu locations or keyboard shortcuts – these systems accept natural human expression and handle the translation themselves.

The emergence of intentional abstraction, where complex systems hide behind natural language interfaces, marks a departure from physical metaphors toward a more conceptually direct mode of interaction. The minimalist text interface enables expanded functionality by eliminating the need for complex visual hierarchies and predetermined interaction patterns. This transformation of the interface affordance space alters how human intention is mediated through technological systems, marking a shift in what D. Ihde (2009) terms the human-technology-world relationship.

Natural language interfaces fundamentally alter technological mediation from a phenomenological perspective. Traditional graphical user interfaces maintained explicit boundaries between human and machine domains, while natural language interaction introduces intersubjective ambiguity. The interface transitions from a tool for manipulation to an agent in a dialogic process, raising fundamental questions about technological agency, intentionality, and the nature of human-computer relationships. The shift from explicit manipulation to intentional expression reduces cognitive load (Ellerton, 2022). Users no longer maintain mental models of hierarchical file systems or action sequences, instead expressing intentions directly through natural language. This cognitive offloading fundamentally alters the distribution of computational tasks between human and machine agents.

Philosophical challenges and implications

The emergence of natural language interfaces introduces what might be termed the understanding illusion – a phenomenon where systems create a compelling impression of semantic comprehension while operating through fundamentally different mechanisms than human understanding. While these systems can engage in sophisticated dialogue and generate contextually appropriate responses, they function through statistical pattern matching rather than genuine semantic understanding (Bender & Koller, 2020). This disjunction between apparent and actual comprehension raises fundamental questions about the nature of understanding itself.

The philosophical implications extend beyond mere technical distinctions. Drawing on Chinese Room argument by J. Searle (1980), it must be considered whether increasingly sophisticated pattern matching might eventually constitute a form of understanding different from, but parallel to, human semantic processing. This consideration challenges traditional philosophical assumptions about the nature of meaning and understanding, suggesting a potential plurality of valid comprehension modalities. Furthermore, this illusion of understanding creates what might be termed semantic opacity – a condition where users cannot readily distinguish between genuine comprehension and sophisticated mimicry. This opacity has practical implications for trust, reliability, and the formation of appropriate mental models of system capabilities. As H. Dreyfus argued in his critique of artificial intelligence, the gap between simulation and genuine understanding may be philosophically significant even when pragmatically imperceptible (Frana & Klein, 2021). The evolution of natural language interfaces fundamentally alters the phenomenological structure of human-computer interaction. As these interfaces become increasingly transparent, they risk transitioning from M. Heidegger's (1962) ready-to-hand to what might be termed invisibly-at-hand – a novel phenomenological state where the technology becomes so seamless that users lose awareness not only of its mediating role but of its presence altogether.

The abstraction of technological processes behind natural language interfaces creates what D. Ihde (2009) terms a "black box" effect, where the mechanisms of technological mediation become increasingly opaque to users. This intentional opacity potentially undermines user agency and technological literacy, raising questions about ability to maintain critical engagement with systems people increasingly rely upon but decreasingly understand. Natural language

interfaces also introduce profound ambiguity into the traditional subject-object relationship between human and computer. Drawing on the concept of intersubjectivity by M. Merleau-Ponty, the emergence of a novel form of technological intersubjectivity that challenges traditional phenomenological categories can be observed. (Du Toit & Swer, 2021). This blurring of boundaries between human and machine cognition suggests a fundamental transformation in how technological mediation is conceptualised. The shift from physical manipulation to verbal interaction additionally transforms the embodied nature of human-computer interaction. This transformation requires to reconceptualise the term "the enacted mind" in the context of technological engagement. As interaction with computers becomes increasingly divorced from physical metaphors, it must be considered how this affects embodied understanding of computational processes (Łozińska, 2021).

The evolution toward natural language interfaces suggests an emerging era of technological transparency that fundamentally reshapes human cognitive architecture. Building on the extended mind hypothesis by A. Clark & D. Chalmers (1998), these interfaces potentially constitute a new form of cognitive extension, where technological systems become more intimately integrated into human cognitive processes than traditional interfaces allow. This integration raises profound questions about the boundaries of human cognition and the nature of technological cognitive enhancement. The increasing sophistication of natural language interfaces also challenges traditional notions of agency. Drawing on actor-network theory by B. Latour, it can be observed how agency becomes distributed across human-computer assemblages in ways that transcend simple tool use (Bar-Gil, 2025). This distribution of agency across human and technological actors suggests a fundamental transformation in how people understand autonomous action and decision-making. The potential emergence of A. Clark's natural-born cyborgs – humans with technologically integrated cognitive processes – necessitates a reconceptualisation of what Kant termed "autonomous agency" in the context of human-AI interaction (Marshall, 2022). As natural language interfaces evolve in sophistication and integration into cognitive processes, the questions of authority, responsibility, and autonomy in these hybrid human-computer systems must be grappled with.

The philosophical challenges of natural language interfaces extend into profound ethical and social domains. The opacity of these interfaces raises critical questions about epistemic responsibility and the conditions for justified belief in technologically mediated knowledge acquisition. How do we maintain intellectual autonomy and critical thinking capabilities when these cognitive processes become increasingly intertwined with systems whose operations we cannot fully comprehend? The emergence of sophisticated language models also challenges traditional theories of social cognition. As these systems become more adept at simulating human-like interaction, the understanding of technological sociality and its implications for human social development have to be reconsidered. This transformation suggests a future where the boundaries between human and machine social interaction become increasingly blurred. Moreover, the abstraction of technological complexity behind natural language interfaces creates new forms of power relations between technology providers and users. This dynamic raise important questions about technological dependency and autonomy, requiring careful philosophical analysis of how these relationships affect human agency and social organisation. As natural language interfaces become more central to cognitive and social lives, understanding and addressing these power dynamics becomes increasingly crucial for maintaining human autonomy and social equity.

In contrast, the work of F. Sovrano & F. Vitali (2022) focused on the practical aspect, particularly the explanatory capabilities of natural language-based systems. The authors proposed a technical implementation that enhances the

illocutionary component of interaction, improving the communicative quality of artificial intelligence. Although the study did not delve into the philosophical underpinnings of the transformation, as in the presented conclusions, the empirical results confirmed the growing importance of intersubjectivity in technological communication. Specifically, it was found that AI effectiveness depends not only on accuracy but also on the ability to anticipate user intentions. In light of the findings on the transformation of human-computer interaction through natural language interfaces, it became important to correlate the study's results with the contributions of other authors who examined both practical and theoretical-philosophical aspects of this phenomenon (Kuznietsov & Kuznietsova, 2024).

The analysis of works by S. Miguens (2022) and K. Zhou *et al.* (2024) allowed for a deeper understanding of the limits and possibilities of modern language models, particularly in the context of intersubjectivity, epistemology, and cognitive impact. The study by K. Zhou *et al.* paid special attention to the problem of overconfidence in large language models (LMs), which was identified as one of the key threats to safe and responsible human-technology interaction. According to the results of this empirical study, LMs tend to exhibit high confidence even when their responses are incorrect, potentially misleading users. The data obtained in this work (particularly, 47% of incorrect answers presented with high confidence) complemented the conclusions about semantic opacity as one of the philosophical challenges raised in the current study. It was confirmed that not only the technical architecture of models but also the socio-cognitive responses of users (such as blind trust in "confident" generations) shape a new epistemological situation where the boundaries between knowledge, assumption, and simulation become blurred. Moreover, the strategic recommendations presented in the work of K. Zhou *et al.* regarding reducing the level of overconfidence in language models correlated with the need to reconsider machine agency identified in this study. The engineering measures proposed by the authors aimed not only at technical improvement of the models but also at reducing risks to human autonomy – an aspect that was also identified as critically important within the philosophical analysis of human-machine interaction.

Another theoretical perspective was provided in the article by S. Miguens (2022), which, drawing on Daniel Dennett's ideas, analysed the role of language as a cultural rather than innate cognitive mechanism. This approach significantly resonated with the concept of technologically extended cognition developed within the study, where natural language interfaces were considered, external cognitive resources integrated into the user's mental processes. D. Dennett's theory, which rejects naturalism in the spirit of Fodor and Chomsky, suggested that language emerged from social interaction rather than neural programming – similar to how modern natural language interfaces are formed not only as technical tools but as components of the socio-technological space that modifies human perception, memory, and consciousness. However, it was noted that Dennett, while supporting the concept of cognitive evolution, remained within a deflationary interpretation of consciousness, avoiding ontological elaboration. In contrast, the conclusions of the current study emphasised the need for critical reflection not only on the epistemic but also on the ontological consequences of integrating AI into human thinking, which required deeper philosophical articulation of agency and the boundary between human and artificial subjects.

The analysis of contemporary research on the functioning of language models in the context of human-machine interaction demonstrated both significant progress in the development of natural language interfaces and revealed a number of fundamental limitations that cast doubt on the depth of semantic understanding by the models. The work of N. Srikanth *et al.* (2024) raised the issue of insufficient robustness of language models to paraphrasing – a property that is fundamental to human linguistic thinking. The authors showed

that despite advancements in pretraining, LLMs remained overly dependent on the surface form of linguistic expression. The conclusions of this study resonated with the obtained results, where sensitivity of models to formal variations in queries was also recorded. The application of the PC metric and the creation of the ParaNlu dataset made it possible to empirically verify this phenomenon, confirming its systemic nature. Consequently, the need arose to reconsider approaches to the architecture of language models, particularly in terms of semantic invariance. The study by A. Gandhi *et al*. (2023) proposed a practically oriented implementation of LLMs in the form of a Semantic Interpreter system that transformed natural language commands into executable code using a specialised ODSL language. This approach ensured functional integration with office software APIs, particularly PowerPoint, significantly improving usability and highlighting the potential of LLMs as control interfaces. However, this system relied on the Analysis-Retrieval prompt construction method, which again raised the issue of model dependency on templating and prior contexts. The results obtained in the current study also observed the need for domain-specific tuning, indicating the limited universality of language models without appropriate training or adaptation. Thus, a tendency toward creating "proxy understanding" was identified – where the model demonstrated correct behaviour only within a specific context or training pattern, lacking true generalisation capability.

The work of N. Srikanth *et al*. (2024) emphasised a philosophical dilemma: despite the aspiration to create interfaces for free expression of intent, modern LLMs mostly modelled statistical similarity rather than deep semantic understanding. This aligned with the conclusions of other studies, which also demonstrated a preference for surface-level correspondence over interpretive depth. On the other hand, the approach of A. Gandhi *et al*. (2023) demonstrated the effectiveness of LLMs as an applied tool but left open the question of scalability beyond predefined scenarios. The work of P. Mah *et al*. (2022) focused on the practical aspects of using AI and NLP in enterprise management under Industry 4.0 conditions. The authors identified these technologies as key drivers of business process transformation, particularly in automating communication between enterprises and customers. Within the "Behaviour-oriented drive and influential function of IoTs" concept, it was found that AI and related tools significantly improved service efficiency, fostered customer loyalty, and optimised managerial decisions. The obtained results (a score of 12 out of 15 on key metrics) confirmed the effectiveness of such approaches, aligning with the conclusions of the study, which also noted the significant functional efficiency of modern LLMs in applied scenarios, particularly in business environments. At the same time, the results of the study by M. Carroll *et al*. (2023) cast doubt on the ethical neutrality of such technologies, highlighting the potential threat posed by AI's manipulative potential. It was emphasised that in cases where language models and other intelligent systems mediate communication (e.g., in social media or chatbots), there is a risk of influence on users that is difficult to detect and assess. A conceptual analysis of categories such as motive, intent, covertness, and harm showed that even unintentional AI behaviour could violate human autonomy principles. These warnings were particularly relevant in the context of the current study, which also identified instances of uncontrolled interpretation of queries by language models, potentially leading to unpredictable or contextually unacceptable outcomes. The practical conclusions regarding the risks of overconfidence in language models and the theoretical considerations about the cultural nature of language and consciousness supported the main thesis that natural language interfaces are not neutral mediators. They shape a new cognitive ecosystem where technology, culture, and subjectivity coexist.

Conclusions

The paradigmatic shift from explicit manipulation to intentional expression through natural language interfaces represents a profound

transformation in the human-technology relationship that transcends mere interface evolution. This transformation manifests across multiple philosophical dimensions: the emergence of semantic opacity challenges traditional notions of understanding and comprehension, while the phenomenological shift toward invisibly-at-hand technology fundamentally alters experiential relationships with computational systems. The introduction of sophisticated natural language interfaces creates novel forms of technological intersubjectivity that blur traditional boundaries between human and machine agency.

These philosophical implications extend beyond theoretical concerns into practical considerations of human autonomy and cognitive architecture. The potential emergence of technologically extended cognition through natural language interfaces suggests a future where human thought processes become increasingly intertwined with artificial intelligence systems. This integration raises critical questions about epistemic responsibility, technological dependency, and the nature of human agency in an era of sophisticated language models.

The philosophical challenges identified – from the understanding illusion to questions of distributed agency – demand ongoing critical analysis as these technologies evolve. Natural language interfaces are changing how humans and computers interact in ways that go beyond technical progress. This shift deeply reshapes relationships with technology, affecting how we think, socialise, and develop future computer systems. Understanding and addressing these philosophical implications becomes crucial for ensuring that the evolution of natural language interfaces enhances rather than diminishes human agency and understanding. However, a limitation of this study is its predominantly theoretical nature, which calls for further empirical research on the actual impact of LLM interfaces on users' cognitive processes. A promising direction lies in interdisciplinary exploration of the interaction between artificial intelligence and human cognition, which would enable the development of ethical, educational, and technical strategies to safeguard user autonomy amid the growing integration of language models into everyday life.

Acknowledgements

None.

Funding

None.

Conflict of Interest

None.

References

[1] Ahmed, L. (2018). Knowing how you are feeling depends on what's on my mind: Cognitive load and expression categorization. *Emotion*, 18(2), 190-201. doi: 10.1037/emo0000312.

[2] Bar-Gil, O. (2025). *The google self as digital human twin: Implications for agency, memory, and identity*. doi: 10.13140/RG.2.2.27892.67203.

[3] Bender, E.M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th annual meeting of the Association for Computational Linguistics* (pp. 5185-5198). New York: Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.463.

[4] Brincker, M. (2024). Smart worlds and broken habits: A contextual analysis of the technological relations of post-phenomenology. In L.R. Ingerslev & K. Mertens (Eds.), *Phenomenology of broken habits* (pp. 133-159). New York: Routledge.

[5] Carroll, M., Chan, A., Ashton, H., & Krueger, D. (2023). Characterizing manipulation from AI systems. In *Proceedings of the 3rd ACM conference on equity and access in algorithms, mechanisms, and optimization* (article number 6). New York: Association for Computing Machinery. doi: 10.1145/3617694.3623226.

[6] Chen, B., Zheng, W., Zhao, L., & Ding, X. (2025). Leveraging large language models to assist philosophical counseling: Prospective techniques, value, and challenges. *Humanities and Social Sciences Communications*, 12, article number 335. doi: 10.1057/s41599-025-04657-7.

[7] Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7-19. doi: 10.1093/analys/58.1.7.

[8] Cohen, P.R. (1992). The role of natural language in a multimodal interface. In *Proceedings of the 5th annual ACM symposium on user interface software and technology* (pp. 143-149). New York: Association for Computing Machinery. doi: 10.1145/142621.142641.

[9] Daniel, A. (2022). *Technology, heidegger, craft*. (Doctoral dissertation, University of Calgary, Calgary, Canada).

[10] DeVrio, A., Cheng, M., Egede, L., Olteanu, A., & Blodgett, S.L. (2025). A taxonomy of linguistic expressions that contribute to anthropomorphism of language technologies. In *Proceedings of the 2025 CHI conference on human factors in computing systems* (article number 430). New York: Association for Computing Machinery. doi: 10.1145/3706598.3714038.

[11] Dourish, P. (2017). *The stuff of bits: An essay on the materialities of information*. Cambridge: The MIT Press.

[12] Du Toit, J., & Swer, G.M. (2021). Virtual limitations of the flesh: Merleau-Ponty and the phenomenology of technological determinism. *Phenomenology and Mind*, 20, 20-31. doi: 10.17454/pam-2002.

[13] Dupre, G. (2021). What would it mean for natural language to be the language of thought? *Linguistics and Philosophy*, 44, 773-812. doi: 10.1007/s10988-020-09304-9.

[14] Ellerton, P. (2022). On critical thinking and content knowledge: A critique of the assumptions of cognitive load theory. *Thinking Skills and Creativity*, 43, article number 100975. doi: 10.1016/j.tsc.2021.100975.

[15] Frana, P.L., & Klein, M.J. (Eds.). (2021). *Encyclopedia of artificial intelligence: The past, present, and future of AI*. New York: Bloomsbury Publishing USA.

[16] Gandhi, A., Nguyen, T.Q., Jiao, H., Steen, R., & Bhatawdekar, A. (2023). Natural language commanding via program synthesis. *arXiv:2306.03460*. doi: 10.48550/arXiv.2306.03460.

[17] Gatti, A., & Mascardi, V. (2023). VEsNA, a framework for virtual environments via natural language agents and its application to factory automation. *Robotics*, 12(2), article number 46. doi: 10.3390/robotics12020046.

[18] Gurcan, F., Cagiltay, N.E., & Cagiltay, K. (2020). Mapping human-computer interaction research themes and trends from its existence to today: A topic modeling-based review of past 60 years. *International Journal of Human-Computer Interaction*, 37(3), 267-280. doi: 10.1080/10447318.2020.1819668.

[19] Heidegger, M. (1962). *Being and time*. New York: Harper & Row.

[20] Ihde, D. (2009). *Postphenomenology and technoscience: The Peking university lectures*. New York: State University of New York Press. doi: 10.2307/jj.18253168.

[21] Kirkwood, J. (2024). *Liber indigo: The affordances of magic*. Morrisville: Lulu.com.

[22] Kuznietsov, Ye., & Kuznietsova, T. (2024). Innovative models of vocational education: A symbiosis of artificial intelligence, neuropedagogy, and the competency-based approach. *Professional Education: Methodology, Theory and Technologies*, 10(1), 64-78. doi: 10.69587/pemtt/1.2024.64.

[23] Łozińska, J. (2021). Imagery underlying metaphors: A cognitive study of a multimodal discourse of yoga classes. *Metaphor and Symbol*, 36(3), 150-165. doi: 10.1080/10926488.2021.1905486.

[24] Mah, P.M., Skalna, I., & Muzam, J. (2022). Natural language processing and artificial intelligence for enterprise management in the era of industry 4.0. *Applied Sciences*, 12(18), article number 9207. doi: 10.3390/app12189207.

[25] Marshall, B. (2022). Evolving the natural-born cyborg. In E. Tumilty & M. Battle-Fisher (Eds.), *Transhumanism: Entering an era of bodyhacking and radical human modification* (pp. 87-101). Cham: Springer International Publishing. doi: 10.1007/978-3-031-14328-1_6.

[26] Miguens, S. (2022). Animal brains and the work of words: Daniel Dennett on natural language and the human mind. *Topoi*, 41, 599-607. doi: 10.1007/s11245-021-09745-2.

[27] Millière, R., & Buckner, C. (2024). A philosophical introduction to language models-part II: The way forward. *arXiv:2405.03207*. doi: 10.48550/arXiv.2405.03207.

[28] Naik, U., & Meghanandha, C. (2024). *Beyond the binary: Metamind libraries and the digital revolution*. Maharashtra: Laxmi Book Publication.

[29] Nefdt, R.M. (2024). *The philosophy of theoretical linguistics: A contemporary outlook*. Cambridge: Cambridge University Press.

[30] Norman, D. (2013). *The design of everyday things*. New York: Basic Books.

[31] Pantano, E., & Scarpi, D. (2022). I, robot, you, consumer: Measuring artificial intelligence types and their effect on consumers emotions in service. *Journal of Service Research*, 25(4), 583-600. doi: 10.1177/10946705221103538.

[32] Pitt, B., & Casasanto, D. (2022). Spatial metaphors and the design of everyday things. *Frontiers in Psychology*, 13, article number 1019957. doi: 10.3389/fpsyg.2022.1019957.

[33] Rahman, M.H., Kazi, M., Hossan, K.M.R., & Hassain, D. (2023). The poetry of programming: Utilizing natural language processing for creative expression. *International Journal of Novel Research and Development*, 8(8), 2456-4184.

[34] Searle, J.R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417-424. doi: 10.1017/S0140525X00005756.

[35] Sedig, K., Klawe, M., & Westrom, M. (2001). Role of interface manipulation style and scaffolding on cognition and concept learning in learnware. *ACM Transactions on Computer-Human Interaction*, 8(1), 34-59. doi: 10.1145/371127.371159.

[36] Sovrano, F., & Vitali, F. (2022). Generating user-centred explanations via illocutionary question answering: From philosophy to interfaces. *ACM Transactions on Interactive Intelligent Systems*, 12(4), 1-32, article number 26. doi: 10.1145/3519265.

[37] Srikanth, N., Carpuat, M., & Rudinger, R. (2024). How often are errors in natural language reasoning due to paraphrastic variability? *Transactions of the Association for Computational Linguistics*, 12, 1143-1162. doi: 10.1162/tacl_a_00692.

[38] Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science*, 12(1), 49-100. doi: 10.1207/s15516709cog1201_2.

[39] Tarsney, C. (2025). Deception and manipulation in generative AI. *Philosophical Studies*. doi: 10.1007/s11098-024-02259-8.

[40] Yung, V. (2021). A visual approach to interpreting the career of the network metaphor. *Poetics*, 88, article number 101566. doi: 10.1016/j.poetic.2021.101566.

[41] Zhou, K., Hwang, J.D., Ren, X., & Sap, M. (2024). Relying on the unreliable: The impact of language models' reluctance to express uncertainty. *arXiv:2401.06730*. doi: 10.48550/arXiv.2401.06730.

Від маніпуляції до експресії: філософський зсув до інтерфейсів природної мови

Андрій Лихацький
Магістр, аспірант
Одеський національний політехнічний університет
65044, проспект Шевченка, 1, м. Одеса, Україна
https://orcid.org/0009-0002-3735-2862

Анотація. Сучасна епоха швидкого розвитку штучного інтелекту засвідчила фундаментальний парадигматичний зсув у взаємодії людини з комп'ютером – від технократичної маніпуляції до свідомого вираження через природну мову. Дослідження мало на меті філософськи дослідити перехід від маніпулятивної до експресивної взаємодії з інтерфейсами природної мови. Було застосовано міждисциплінарний підхід, що поєднує методи філософського аналізу, когнітивної феноменології та критики цифрової культури. Результати дослідження показали, що сучасні інтерфейси на основі моделей великих мов не лише спрощують комунікацію, але й переналаштовують онтологічні основи технологічного досвіду. Зростаюча «семантична непрозорість» великих мовних систем поставила під сумнів традиційні уявлення про розуміння, підзвітність та пояснювальність у технічній взаємодії. Було виявлено феноменологічний зсув до «невидимої» технології, інтегрованої в когнітивну архітектуру суб'єкта. Було проаналізовано феномен технологічної інтерсуб'єктивності, що розмиває межу між людською та машинною активністю. Було продемонстровано, що мова як інтерфейс не лише передає інформацію, але й функціонує як середовище для формування самоідентичності, автономії та когнітивної відповідальності. Висновки довели, що ця трансформація має глибокі філософські наслідки та вимагає систематичного осмислення, щоб запобігти руйнуванню людської автономії в контексті когнітивного злиття зі штучним інтелектом. Було підкреслено необхідність етичної основи для нових інтерфейсів – такої, яка б забезпечувала не лише функціональність, але й збереження людської активності. Практичне значення полягало у встановленні філософських основ для проектування інтерфейсів природної мови, які б підтримували людську автономію, відповідальність та епістемічну незалежність

Ключові слова: взаємодія людини з комп'ютером; когнітивна абстракція; графічні інтерфейси користувача; штучний інтелект; філософське посередництво