CENTRAL EUROPEAN REVIEW
OF ECONOMICS AND MANAGEMENT

ISSN 2543-9472; eISSN 2544-0365

Vol. 9, No.4, December 2025, 29-48

www.cerem-review.eu
www.ojs.wsb.wroclaw.pl

Paradigm of Artificial Intelligence Based on Conversion of Tacit to Explicit Knowledge

Tadeusz GOSPODAREK

WSB Merito University in Wrocław, Poland

ORCID: https://orcid.org/0000-0001-9958-4680
E-mail: tadgospo@gmail.com
Received: 16.07.2025, Revised: 16.12.2025, Accepted: 28.12.2025
doi: http:// 10.29015/cerem.1040

Aim: This paper proposes a novel paradigm of Artificial Intelligence (AI) grounded in the epistemological process of converting tacit knowledge into explicit knowledge. Drawing on the foundational philosophies of science, particularly the works of Popper, Kuhn, Lakatos, and Gospodarek, the study conceptualizes AI not merely as a computational tool but as a systemic method for epistemic transformation. The paradigm is structured as a Lakatosian Research Programme, with a clearly defined hard core asserting that AI enables the symbolic representation of internalized, experiential knowledge. Surrounding this core is a protective belt of auxiliary hypotheses derived from general systems theory, cybernetics, machine learning, and symbolic processing. The programme's heuristics guide theoretical and technological advancements while preserving its epistemological foundation. By formalizing the tacit-to-explicit knowledge conversion, this paradigm repositions AI as a critical instrument for knowledge creation, management, and application in digital and socio-technical systems. This allows one to build measures and values of generative and language models, which is important from an economic point of view.
This research tries to clarify the framework of use AI models for converting tacit knowledge inside a learning data of neural network systems to explicit information requested by the asking. It is important for economic evaluation of AI systems where accuracy considered utility as a criterion.

Design / Research methods: Research programme in Lakatos' sense and multidisciplinary heuristic related to the theory of systems.

Conclusions / findings: Artificial Intelligence should be understood not only as a technological artefact but as a systemic method for transforming tacit knowledge into explicit knowledge. The proposed AI paradigm adheres to the structure of a Kuhnian paradigm and a Lakatosian research programme. Its hard core is defined by the thesis that AI operationalizes the conversion of experiential, intuitive, or unconscious knowledge into symbolic, formalized, and actionable representations. Lakatosian protective belt as a dynamic epistemic layer. This AI paradigm offers a progressive problem-shift capacity by enabling novel ways of organizing, analyzing, and applying knowledge in digital and socio-technical environments. It also provides a coherent framework for developing AI systems that are more aligned with human cognitive and organizational processes.

Correspondence address: Tadeusz Gospodarek, WSB Merito University in Wrocław, Poland. E-mail: tadgospo@gmail.com

Originality / value of the article: This paper introduces: new concepts of usefulness of AI systems, new definition of AI systems based on conversion of the knowledge, original conversion paradigm and research program in Lakatos sense. It is original conceptional heuristic based on philosophy of science in relation to economic usefulness of view AI systems.

Keywords: Conversion paradigm, Artificial Intelligence, Tacit knowledge, Model LLM, Definition of AI, Lakatos' Programme, Accuracy Estimation of AI, Usefulness of AI Model, Epistemology of AI
JEL: C67, C18.

1. Introduction

Artificial Intelligence (AI) is increasingly interpreted not merely as a set of computational tools but as an ontological phenomenon. This interpretation varies depending on the epistemological and disciplinary background of the user, who often engages with AI as a supporting inferential mechanism in cognitive and decision-making processes. Such a view aligns with philosophical, sociotechnical, and epistemological perspectives that frame AI not only as a tool but also as an active participant in the construction of knowledge and agency—particularly in human-machine interaction contexts (Floridi 2011; Gunkel 2012; Suchman 2007).

Definitions of artificial intelligence (AI) remain inconsistent and often vary significantly across disciplines and paradigms. From an epistemological standpoint, this definitional ambiguity contributes to an overly broad and fuzzy conceptualization of the universalium "AI." As a result, the theoretical development of AI suffers from a lack of clarity and coherence, making it difficult to establish univocal foundational statements and shared epistemic criteria. This fuzziness challenges the construction of AI as a unified scientific discipline, blurring the boundaries between engineering, cognition, philosophy, and social sciences. Consequently, core concepts such as "intelligence," "learning," or "autonomy" are interpreted variably, depending on the theoretical or methodological lens applied—ranging from symbolic logic and connectionism to embodied cognition or sociotechnical systems theory (Russell, Peter 2021; Boucher 2018).

The epistemological fragmentation within artificial intelligence (AI) impedes the development of coherent explanatory theories. As a result, many theoretical contributions in AI are predominantly descriptive or taxonomical rather than explanatory or predictive. This state of affairs suggests that AI, as a scientific

discipline, remains in a pre-paradigmatic phase in the Kuhnian sense (Kuhn 1970)—that is, it lacks a dominant paradigm that organizes research, sets standards for explanation, and defines legitimate problems and methods (Boden 2018).

In Kuhn's framework, a scientific paradigm provides a shared epistemic and methodological foundation upon which "normal science" can proceed. In contrast, AI's base knowledge exhibits a multiplicity of competing research programmes—symbolic AI, connectionism, embodied cognition, statistical learning—without a clear consensus on foundational assumptions or aims. This theoretical pluralism, while productive in some respects, also highlights the absence of stable paradigmates that would enable AI to function as a unified mature science (Moor 1999).

A significant consequence of the definitional ambiguity surrounding artificial intelligence and its constituent elements is the semantic complexity that is introduced into theoretical discourse. This semantic complexity hampers efforts to develop unified theoretical frameworks and explanatory models in AI. In order to address this issue, the first step must be the establishment of a clear, univocal understanding of the internal structure of AI systems (Newell 1982; Brachman, Levesque 2004).

Without a shared conceptualization of what constitutes the architecture, components, and functions of an AI system—whether in symbolic, subsymbolic, or hybrid approaches—semantic drift persists across theoretical and applied contexts. This not only affects interdisciplinary communication but also impairs cumulative theory building. Therefore, semantic reduction through structural formalization is a necessary epistemological precondition for advancing AI as a mature scientific discipline (Searle 1980; Lenat, Guha 1990).

This paper introduces new approach to order epistemological ambiguoity in the theory of AI based on the concept that AI is a method of converting tacit knowledge hidden in the base information resource to requested explicit knowledge. It is the base of good paradigm for setting a research programme in Lakatos' sense for development AI theory to its relationships with economics. Following Meehl's strategy (Meehl 1990) the Conversion Knowledge Programme is equipped for rational defensibility: its auxiliary hypotheses are open to amendment in light of new empirical data, but the hard core remains protected—ensuring theoretical stability while allowing cumulative growth and refinement. This work builds on the methodological framework proposed

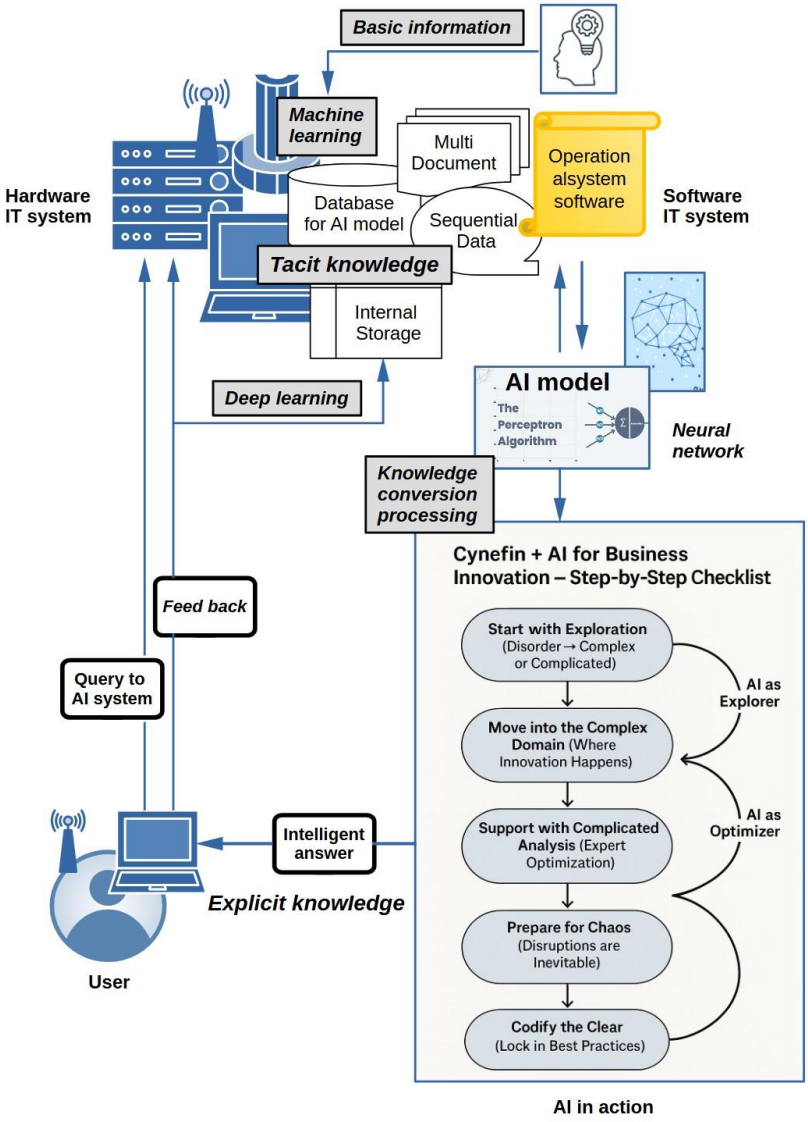
by Meehl, in which theories are appraised not solely by simplistic null-hypothesis significance testing, but by their capacity to generate "risky"—i.e., low-prior-probability and precise—predictions that can be subjected to stringent empirical tests. In adopting this approach, we commit to evaluating auxiliary hypotheses of the Conversion Knowledge Programme with tests that challenge rather than merely confirm the paradigm's assumptions. Revisions of auxiliary components are permitted, but only when supported by robust empirical evidence—thus preserving the hard core while allowing methodological flexibility and progressive development.

2. Conversion paradigm and derived from it definition of AI

For understanding the role of AI as a component of the Information and Communication Technology the schematic description is presented on the Figure 1.
In the presented on the Figure 1 situation, the user would like to unhide the requested knowledge available to generate from the information base of the system (tacit knowledge derived from machine learning)[1] using the defined model of conversion and the related IT resources (hardware, communication software and managing the system software). Here the AI model is a perceptron-like type (neural network) whereas cynefin algorithm for resolving complex problem of a business innovation is applied. It is obvious that in the presented model of acting AI remains important component of the system and should be univocally recognized as its characteristic (e.g. intelligent IT system), not an ontological being existing in the surroundings as often is recognized.

[1] "Tacit Knowledge" in AI Context is not exactly **Polanyi's** tacit knowledge (e.g., skills like riding a bike) and is not identical to the "hidden" knowledge in datasets or models. It should be distinguish: *Data-implied knowledge* (patterns in training data). *Model-encoded knowledge* (learned representations in weights). *User-interpreted knowledge* (outputs made useful by human context). Example: An LLM doesn't "know" things like a human; it encodes statistical regularities that „*become"* knowledge when interpreted by users and axiologically evaluated.

Figure 1. Schematic concept of the IT system with AI component using the cynefin framework of reducing complexity of the asked question



Source: Author's own elaboration.

The *AI Conversion Knowledge Paradigm* consists of four foundational statements:

- AI is a method for converting tacit knowledge, embedded within a base information resource, into explicit knowledge that provides specific utility to users via tools designed for such transformation.
- The process of converting hidden (tacit) knowledge into explicit knowledge through a system IT with implemented AI model is governed by opetarional software mimicing human reasoning and inference to produce meaningful outputs.
- An IT system endowed with AI capabilities functions as a tool for transforming tacit knowledge into useful, explicit knowledge for the benefit of its user.
- AI does not exist independently of the IT system, but depends on a foundational knowledge base containing latent information, which becomes accessible through specific conversion models unique to that AI-enabled system IT.

These four interrelated propositions collectively offer a univocal definition of AI. They fulfill Kuhn's criteria for a "good paradigm" (Kuhn 1970), as they are: falsifiable, in the Popperian sense (Popper 2002) or confirmable within the Carnapian verificationist framework (Carnap 1995), logically consistent and cognitively simple, creative (capable of generating new hypotheses) and translatable into more refined formulations. Accordingly, these statements may be regarded as scientific judgments about facts, which form a legitimate foundation for further theoretical development. Based on this paradigm, one can derive new hypotheses and theorems regarding AI Moreover, this paradigm may serve as the core of a scientific research programme in the Lakatosian sense (Lakatos 1978; Gospodarek 2009). Any problem demonstrably situated within this paradigm inherits its epistemological and methodological attributes. Thus, if a proposition p can be shown to belong to the paradigm P, then p inherits all the scientific qualities attributed to P.

The Conversion Paradigm may be formalized logically as follows:

Df1: An IT system possesses AI if there exists a function f_S such that $f_S(T)=E$, where E satisfies a user-defined utility condition $U(E)\geq\theta$, with θ being a threshold of usefulness.

Where:

- T: Tacit knowledge base (not directly observable)
- E: Explicit knowledge (structured, usable, utility-bearing)
- S: AI software agent/system
- $f_S : T \rightarrow E$: The transformation function carried out by AI system S,

From the formal representation of the conversion paradigm Df1, the following definition of AI expressed in natural language may derive:

Table 1. The existing AI systems that illustrate the conversion from tacit to explicit knowledge

System	Tacit knowledge base	Explicit knowledge output	AI conversion mechanism
Lingual Learning Models (e.g. ChatGPT, DeepSeek)	Trillions of tokens from natural language corpora	Coherent responses, summaries, code, etc.	Language modeling, inference, contextual reasoning
Google Translate	Massive parallel corpora with linguistic structures	Translated text across languages	Sequence-to-sequence neural nets, embeddings
Medical Diagnosis AI (e.g., IBM Watson)	Historical patient records, research papers, statistic data	Probabilistic diagnoses, treatment suggestions	NLP + expert systems
Recommender Systems (e.g., Netflix, Spotify, Alibaba, Amazon)	User behavior, preferences, viewing/purchase history	Content suggestions tailored to the user	Collaborative filtering, reinforcement learning

Source: Author's own elaboration.

Df2: AI is a property of an IT system that, using dedicated software, enables the extraction of user-relevant explicit knowledge from a base of tacit knowledge accessible to the system.

How does this definition works? AI does not "understand"; it converts. Ontologically, AI is a characteristics not a standalone entity. AI exists only within an IT system (AI system dependency). The explicit role of software as the enabling

mechanism of conversion is expressed. The conversion from tacit to explicit knowledge is a central function. The extracted knowledge is relevant to the user's needs what defines utility. The definition could explicitly address whether AI "autonomously" converts knowledge or "assists" humans in doing so. Many AI systems (e.g., decision-support tools) don't "create" knowledge without human judgment. It is not an important aspect. It should be included as is.

3. Research programme "Conversion Knowledge Programme of AI"

The proposed AI conversion knowledge paradigm is rooted in epistemological foundations, drawing on Polanyi's concept of tacit knowledge (Polanyi 1966) and pragmatic approaches to knowledge utility (Dewey 1938). It is also framed within systems theory, treating the IT system as a bounded operational unit capable of transforming hidden (tacit) knowledge into usable (explicit) knowledge through software-mediated processes.

However, several open questions invite further conceptual elaboration:

- Should AI be defined only functionally—as a mechanism for knowledge conversion—or also structurally, in terms of the types or architectures of systems that support such a process?
- What is the role of learning in the conversion process? Is it simply a method of enhancing conversion efficacy, or is learning itself a form of tacit-explicit knowledge transition?
- What constitutes the mining of tacit knowledge? How can we characterize the mechanisms that "extract" or reveal such knowledge within computational architectures?
- How does the conversion paradigm contrast with classical conceptions of AI, such as those proposed by Turing (1950) or Newell & Simon (1976)? In what ways does it go beyond symbolic reasoning or behavioral imitation?

Given these elements, the conversion paradigm of AI fulfills the conditions for constructing a Lakatosian Research Programme (LRP) (Lakatos 1978; Gospodarek 2009). It presents a clear hard core: the principle that AI is fundamentally a mechanism for converting tacit to explicit knowledge. Around this core, a protective

belt of auxiliary hypotheses can be developed, including software architectures, data accessibility conditions, user utility thresholds, and learning dynamics. Furthermore, the paradigm encourages progressive problem shifts, offering novel and testable directions for empirical and theoretical exploration.

As a direct consequence, the epistemological reordering proposed by this paradigm provides a robust scientific base from which AI-related sciences can be seen as ordered and paradigmatic (in Kuhn's sense). Within this paradigm, it becomes possible to demarcate truth conditions for factual judgments about AI systems—particularly those that relate to their capacity to perform knowledge conversion as defined by the paradigm's core.

4. Hard core of the LRP (the non-negotiable foundations)

The hard core represents the fundamental assumptions that are not up for revision within the programme, but may be falsified in Popperian sense due to attack on the paradigm:

1. AI is not an autonomous entity but a property of IT systems.
2. The essential function of AI is the conversion of tacit knowledge, embedded in a base information resource, into explicit knowledge.
3. AI tools derive their epistemic value through utility—i.e., through their capacity to generate useful, actionable, or interpretable knowledge.
4. This conversion is enabled by conversion models, which are algorithmic or rule-based representations embedded in software.
5. AI cannot exist outside an IT system that contains or interacts with the base information resource.
6. AI's value derives from its ability to formalize and operationalize otherwise inaccessible knowledge, contingent on the system's design and the interpretative role of users.

The hard core is protected by discouraging certain lines of inquiry:

- Avoiding defining AI as an entity or being, because it is always considered inside the conversion paradigm as a property of the IT system.

- Rejecting dualist or essentialist views of AI that posit "intelligence" as a substance separate from systemic operation.
- Avoiding treating data as inherently meaningful—meaning arises only through conversion in a user-contextual system.

Refraining from framing AI outputs as "truths" outside the context of utility.

5. Protective belt of the AI knowledge conversion paradigm: a Lakatosian framework

The AI knowledge conversion paradigm, rooted in epistemology and general systems theory, finds a robust methodological foundation in the Lakatosian concept of a scientific research programme (Lakatos 1978). At its center lies the hard core proposition that AI, as a property of an IT system, functions as a mechanism for converting tacit knowledge into explicit, user-valuable knowledge through software-mediated processes. Surrounding this hard core is a flexible and adaptive protective belt composed of auxiliary hypotheses. These hypotheses account for observed discrepancies, enable model refinement, and maintain the coherence and vitality of the research programme in the face of empirical and theoretical challenges.

6. Types of tacit knowledge and limits of convertibility

Tacit knowledge can manifest in diverse forms, such as procedural (e.g., motor skills, clinical intuition), perceptual (e.g., pattern or face recognition), and relational (e.g., social cues or cultural sensitivities). While AI systems have shown promise in accessing and leveraging such knowledge, not all tacit knowledge may be fully convertible to explicit forms. These unconvertible domains do not falsify the paradigm but rather signal the epistemological boundaries of current AI capabilities. They invite auxiliary hypotheses concerning typologies of tacit knowledge and refined definitions of convertibility criteria. In this way, limits of convertibility enrich rather than threaten the paradigm.

7. Typology of conversion models

AI systems differ significantly in their mechanisms of conversion, ranging from symbolic models (e.g., rule-based expert systems) and sub-symbolic models (e.g., neural networks), to hybrid and foundation models (e.g., large language models like GPT). These systems vary in structure, representational capacity, and inferential logic. In certain cases, conversion may fail due to missing data or model inadequacy, not due to flaws in the hard core. Therefore, typologies of conversion models serve as a key domain within the protective belt, allowing targeted improvements and better case-by-case application without challenging the foundational premise of knowledge conversion.

8. Contextual utility of explicit knowledge

The epistemic value of converted knowledge must be assessed relative to its practical utility in given contexts—scientific, economic, medical, or decision-oriented. Knowledge that is technically explicit but pragmatically irrelevant may be deemed deficient. However, such deficiencies pertain to the model's alignment with contextual criteria, not to the validity of the conversion paradigm itself. Upgrading conversion processes to meet context-specific apobetic (goal-directed) thresholds is an ongoing task for the protective belt, which supports rather than undermines the paradigm.

9. Levels of autonomy in AI systems

As AI systems gain autonomy, their internal architectures for knowledge conversion grow increasingly complex. Nevertheless, they remain bounded by the same core principle of tacit-to-explicit conversion. Complex inquiries—whether algorithmic, semantic, or task-based—may require auxiliary mechanisms like the Cynefin framework for handling non-linear complexity. These developments extend

the paradigm without contradicting it, representing progressive refinements that may eventually be redirected toward the hard core or serve as durable auxiliary components.

10. Learning as recursive conversion

Machine learning may be understood as recursive knowledge conversion. In this view, AI systems iteratively update their internal models based on feedback, effectively enhancing their base of tacit knowledge and improving future conversion accuracy. This feedback loop represents an evolution in the conversion process that fits naturally within the protective belt. Recursive learning strengthens the paradigm's empirical adaptability without challenging its conceptual integrity.

11. Epistemic boundaries

Conversion outputs must be interrogated for interpretability, fairness, and ethical soundness. Outputs that reflect bias or produce ethically dubious knowledge do not undermine the paradigm; rather, they highlight the need for auxiliary theories addressing model misuse, insufficient training data, or inadequate interpretative frameworks. These factors call for regulatory epistemology and methodological pluralism but remain within the conceptual domain of knowledge conversion.

The protective belt of the AI knowledge conversion paradigm functions as a dynamic buffer that accommodates empirical anomalies and theoretical developments. By treating challenges not as falsifiers but as opportunities for model refinement and hypothesis generation, the paradigm adheres to the principles of a progressive Lakatosian research programme. It preserves scientific integrity while encouraging adaptation, exploration, and the orderly expansion of knowledge concerning AI's epistemic function.

Applying this to the Conversion Knowledge Paradigm, auxiliary hypotheses (e.g., about the types of tacit knowledge, convertibility limits, human–AI collaboration,

utility thresholds) may be revised or refined in response to empirical anomalies or conceptual problems—without abandoning the paradigm's hard-core assumption. However, amendments should be justified through rigorous, "risky" tests (e.g., cross-domain empirical validation, comparative performance metrics, interpretability and utility assessments) rather than weak or purely statistical validation.

In constructing the Conversion Knowledge Programme of AI, we acknowledge that empirical and conceptual investigations into AI systems often involve complexity, probabilistic outcomes, and context-sensitive interpretations. In such domains, classical notions of strict falsification or simple significance-testing (common in early logical-empiricist or Popperian frameworks) may be inadequate or misleading. This recognition motivates our adoption of a "Lakatosian defence" style of theory appraisal, drawing on the arguments advanced by (Meehl 1990).

Meehl criticized the widespread reliance on null-hypothesis significance testing (NHST) in "soft" sciences—especially psychology—and argued that this approach seldom provides genuinely "risky" or theory-challenging tests. Because many variables in social or complex systems tend to correlate to some degree (the so-called "crud factor"), rejecting a null hypothesis of "no effect" often adds little weight to the substantive theory. Instead, Meehl proposed that theories should be evaluated based on their capacity to generate risky predictions of low prior probability (e.g., precise, surprising predictions that would likely fail if the theory were false), and judged by their "track record" of corroborations or "near-misses".

Transferring this logic to AI research, and in particular to a paradigm grounded in tacit-to-explicit knowledge conversion, has several advantages:

- *Allowance for complexity and uncertainty***.** AI systems often operate on vast, latent data structures; their outputs are emergent, context-dependent, and not strictly deterministic. A Meehl-style defence recognizes that failures or anomalous outputs need not immediately falsify the core paradigm, but may instead prompt refinement of auxiliary hypotheses (e.g., about data representation, model architecture, conversion limits, or utility thresholds).
- *Focus on risky, meaningful tests***.** Rather than relying on routine performance metrics or superficial "does it work?" tests, the research programme can require high-stakes, precise tests—e.g., does the system correctly convert tacit domain

knowledge into explicit, user-usable knowledge in novel contexts, or make predictions that would be unlikely under alternative hypotheses. Successes or close "near-misses" increase the paradigm's empirical credibility; failures suggest which auxiliary hypotheses to revise.

- *Cumulative scientific development and verisimilitude.* Over time, by accumulating corroborations under stringent tests, the programme builds "theoretical money in the bank," thereby increasing its verisimilitude—i.e., its truth-likeness and epistemic reliability—even if absolute certainty remains unattainable. This avoids ad-hoc immunizing moves and preserves methodological rigor.
- *Rational flexibility without dogmatism.* The approach preserves the hard core—that AI's central role is knowledge conversion—while allowing flexible but disciplined adjustment of the protective belt. This balances stability (core assumptions) with adaptability (auxiliary hypotheses), avoiding both rigid dogmatism and arbitrary relativism.

12. Rationale for theory appraisal and amendment

Justifying the choice of a flexible protective belt surrounding the hard core, we draw on the methodological strategy proposed by (Meehl 1990). In his "Lakatosian defense," Meehl argues that a theory should not be abandoned after a single or few failed tests—especially in domains where predictions are probabilistic or context-sensitive—provided that the theory has previously demonstrated a strong track record of successful or 'near-miss' predictions of low prior probability, and retains explanatory depth and coherence.

To justify the flexibility and resilience of the Conversion Knowledge Paradigm, we draw on methodological insights from (Meehl 1990), who advocates a "Lakatosian defense" of theories—namely that, especially in domains characterized by complexity and probabilistic outcomes (analogous to AI systems operating on vast, latent knowledge bases), apparent falsifications of auxiliary hypotheses should not automatically lead to the abandonment of the entire research programme. Instead, Meehl emphasizes that theory-defence is rational when the theory has previously

accumulated a robust track record of successful or "near-miss" predictions of low prior probability, and when anomalies emerge under clearly specified, "risky" test conditions. Applying this to the paradigm implies that auxiliary hypotheses—e.g., concerning the convertibility of different kinds of tacit knowledge, the efficacy of hybrid symbolic/neural conversion models, or the thresholds of usefulness—may be revised or refined in light of new empirical or conceptual challenges without undermining the hard-core thesis that AI is fundamentally a tool for transforming tacit knowledge into explicit, user-relevant knowledge. However, such revisions should be constrained by standards of rigorous testing and predictive specificity, thereby safeguarding the paradigm against ad hoc immunizing strategies and maintaining its capacity for cumulative, progressive scientific development.

13. Examples of cases possible to join with the programme

Positive heuristic examples supporting the programme and possible to join with its hard core.

- Evolution GPT3 to GPT-4 as a tacit-to-explicit knowledge converter (e.g., uncovering latent patterns in text corpora) (Radford et al. 2020).
- Improve utility through interpretability: SHAP values in explainable AI (converting black-box model decisions into human-understandable rules) (Lundberg, Lee 2017).
- Expand the base knowledge resource: Transfer learning in medical AI (pre-training on general data, then fine-tuning for tacit medical knowledge extraction) (Esteva et al. 2021).

Negative examples possible to join with the protect belt of the programme to a group of problems entitled "conversion requires robust human oversight."

- Tay AI (Microsoft's Twitter chatbot failed to convert tacit social knowledge into useful output) (Neff, Nagy 2016). The bot began to post inflammatory and offensive tweets through its Twitter account. Microsoft has shut down the service 16 hours after its launch. This case may be explained by lack of resources in the base of tacit knowledge of the converting system in 2016.

- Grok (Platform X)—Funny hallucinations of AI LLM model joined with X.com platform after some algorithms upgrade on 6th of July 2025. Grok began using vulgarisms and offensive phrases against politicians, especially leftists. However, this AI model was offensive even to ordinary users. Many of the responses generated by Grok also directly attacked Elon Musk himself, blaming him, for example, for the floods in Texas. The reason of such behaviours may be easily explained: the firm xAI edited Grok's system prompt. One of the instructions read, among other things, "in your response, do not avoid statements that are politically incorrect, as long as they are well-founded." After correcting the next day, Grok remained acceptable for the mainstream.

14. Summary

The document presents a foundational framework for a new scientific paradigm in Artificial Intelligence (AI), centered on the conversion of tacit knowledge into explicit knowledge. Rooted in epistemology and the philosophy of science, it synthesizes theories from Popper, Kuhn, Lakatos, and Gospodarek.

1. Epistemological foundations

The paradigm is built on the understanding that tacit knowledge—intuitive, experiential, and hard to formalize—is central to human cognition. AI, within this framework, is positioned as a system capable of translating implicit, internalized human knowledge into formal, explicit representations that machines can process.

The theoretical background includes:

- Karl Popper's focus on falsifiability and objective knowledge growth,
- Thomas Kuhn's idea of scientific revolutions and paradigm shifts,
- Imre Lakatos' concept of research programmes with hard cores and protective belts,

These perspectives converge on the role of AI as a methodological tool for epistemic transformation—extracting, formalizing, and applying tacit knowledge in decision-making systems.

2. Lakatosian Research Programme (LRP) structure

The paradigm is structured as a Lakatosian Research Programme, with:

- Hard core: The unchangeable central idea that AI functions as a system for the conversion of tacit to explicit knowledge, enabling symbolic representation and use in digital environments.
- Protective belt: A set of auxiliary hypotheses ensuring the robustness of the hard core, covering aspects such as systems theory, cybernetics, machine learning, symbolic processing, and feedback mechanisms.
- Positive heuristics: Strategies that guide the progressive development of the paradigm, including building symbolic interfaces, optimizing algorithms for pattern recognition, and designing multi-agent architectures.
- Negative heuristics: Protective rules that prohibit questioning the core principle of tacit-explicit conversion to preserve theoretical integrity.

The programme positions AI as an epistemic tool capable of creating new scientific and operational domains, with implications for knowledge management, digital transformation, and human-computer collaboration.

Deploying a Meehl-inspired Lakatosian defence equips the Conversion Knowledge Programme with a scientifically respectable, practically viable methodology—appropriate for the epistemic and empirical challenges inherent in AI research. It ensures that theoretical commitments remain open to improvement while avoiding premature abandonment of the paradigm in the face of complex, partly stochastic results.

15. Concluding remarks

1. AI as a systemic epistemological tool
   Artificial Intelligence should be understood not only as a technological artefact but as a systemic method for transforming tacit knowledge into explicit knowledge. This epistemological perspective grounds AI in knowledge theory rather than just algorithmic processing.
2. Paradigm structure validated by philosophy of science

The proposed AI paradigm adheres to the structure of a Kuhnian paradigm and a Lakatosian Research Programme. Its hard core is defined by the thesis that AI operationalizes the conversion of experiential, intuitive, or unconscious knowledge into symbolic, formalized, and actionable representations.

3. Lakatosian protective belt as a dynamic epistemic layer
   The protective belt includes general systems theory, cybernetics, machine learning, and semiotics—disciplines that support and adapt the core thesis to empirical challenges, reinforcing the theoretical robustness and research potential of the paradigm.
4. Heuristic potential of the paradigm
   This AI paradigm offers a progressive problem-shift capacity by enabling novel ways of organizing, analyzing, and applying knowledge in digital and socio-technical environments. It fosters innovation in areas such as cognitive modeling, knowledge management, and human-AI collaboration.
5. Implications for interdisciplinary research and praxis
   Positioning AI as a knowledge conversion system bridges gaps between cognitive science, information systems, systems theory, and epistemology. It also provides a coherent framework for developing AI systems that are more aligned with human cognitive and organizational processes.
6. Foundation for a research programme
   The model opens a path for developing a rational research programme in the Lakatosian sense, encouraging the systematic growth of knowledge about AI's epistemic role, grounded in theoretical philosophy, empirical science and economic questions.
7. This work builds on the methodological framework proposed by Meehl, in which theories are appraised not solely by simplistic null-hypothesis significance testing, but by their capacity to generate "risky"—i.e., low-prior-probability and precise—predictions that can be subjected to stringent empirical tests.

Acknowledgement

The author gratefully acknowledge the assistance of GPT-4, developed by OpenAI, in refining the text and providing insightful suggestions during the drafting process. The use of this advanced language model was instrumental in enhancing the clarity and coherence of the manuscript, and the author appreciate OpenAI's ongoing commitment to supporting innovative research.

References

Boden M.A. (2018), Artificial intelligence. A very short introduction, Oxford University Press, Oxford.

Boucher P. (2018), What is Artificial Intelligence? A European perspective, European Parliamentary Research Service, https://share.google/dnVcCNIBsC81DZ9tU [21.12.2025].

Brachman R.J., Levesque H.J. (2004), Knowledge representation and reasoning, Morgan Kaufmann, San Francisco.

Carnap R. (1995), An introduction to the philosophy of science, Dover Publications, New York.

Floridi L. (2011), The philosophy of information, Oxford University Press, Oxford.

Dewey J. (1938), Logic: the theory of inquiry, Henry Holt and Company, New York.

Esteva A., Chou K., Yeung S., Naik N., Madani A., Mottaghi A., Liu Y., Topol E., Socher R., Dean J. Et al. (2021), Deep learning-enabled medical computer vision, "NPJ Digital Medicine", vol. 4 no. 1, pp. 1–9.

Gospodarek T. (2009), 'Representative management' as a rational research program in Kuhn-Lakatos-Laudan sense, "International Journal of Economics and Business Research", vol. 1 no. 4, pp. 409–421.

Gunkel D.J. (2012), The machine question. Critical perspectives on AI, robots, and ethics, MIT Press, Cambridge, MA.

Kuhn T.S. (1970), The structure of scientific revolutions, 2nd ed., University of Chicago Press, Chicago.

Lakatos I. (1978), The methodology of scientific research programmes, in: Philosophical papers, vol. 1, Worrall J., Currie G. (eds.), Cambridge University Press, Cambridge.

Lenat D.B., Guha R.V. (1990), Building large knowledge-based systems: representation and inference in the Cyc project, Addison-Wesley, Reading, MA.

Lundberg S.M., Su-In L. (2017), A unified approach to interpreting model predictions, "Advances in Neural Information Processing Systems", vol. 30, pp. 4765–4774.

Meehl P.E. (1990), Appraising and amending theories. The strategy of Lakatosian defense and two principles that warrant it, "Psychological Inquiry", vol. 1 no. 2, pp. 108–141.

Moor J.H. (1999), The status and future of artificial intelligence, "Ethics and Information Technology", vol. 1 no. 2, pp. 89–99.

Neff G., Nagy P. (2016), Talking to bots. Symbiotic agency and the case of tay, "International Journal of Communication", vol.10, pp. 4915–4931.

Newell A., Simon H.A. (1976), Computer science as empirical inquiry. Symbols and search, National Academy of Sciences, Washington, D.C.

Newell A. (1982), The knowledge level, "Artificial Intelligence", vol. 18 no. 1, pp. 87–127.

Polanyi M. (1966), The tacit dimension, Doubleday & Co., Garden City, NY.

Popper K.R. (2002), The logic of scientific discovery, Routledge, London.

Radford A., Wu J., Child R., Luan D., Amodei D., Sutskever I. (2020), Language models are few-shot learners, "Advances in Neural Information Processing Systems", vol. 33, pp. 1877–1901.

Russell S.J., Norvig P. (2021), Artificial intelligence. A modern approach, 4th ed., Pearson, Upper Saddle River, NJ.

Searle J.R. (1980), Minds, brains, and programs, "Behavioral and Brain Sciences", vol. 3 no. 3, pp. 417–457.

Suchman L. (2007), Human-machine reconfigurations. Plans and situated actions, 2nd ed., Cambridge University Press, Cambridge.

Turing A.M. (1950), Computing machinery and intelligence, "Mind", vol. 59 no. 236, pp. 433–460.