TEORIE VĚDY / THEORY OF SCIENCE / XLV / 2023 / 2
https://doi.org/10.46938/tv.2023.579

////// přehledová studie / survey article //////////////////////////

A HISTORICAL INTERACTION BETWEEN ARTIFICIAL INTELLIGENCE AND PHILOSOPHY

Abstract: *This paper delves into AI development's historical and philosophical dimensions while highlighting the symbiotic relationship between philosophy and AI from a technological perspective: philosophy furnishes foundational concepts, and AI supplies practical tools. The paper posits neurosymbolic AI as a solution to present challenges, sparking discussions encompassing both technical and philosophical considerations. Advocating a multidisciplinary approach calls for merging empirical AI insights with philosophy and cognition science to enrich our comprehension of intelligence and propel AI forward.*

Keywords: *history of AI; philosophy of AI; symbolism; connectivism*

Historická interakce mezi umělou inteligencí a filosofií

Abstrakt: *Tento článek se zabývá historickými a filosofickými rozměry vývoje umělé inteligence a zdůrazňuje symbiotický vztah mezi filosofií a umělou inteligencí z technologického hlediska: filosofie poskytuje základní koncepty a umělá inteligence praktické nástroje. Článek navrhuje neurosymbolickou umělou inteligenci jako řešení současných výzev a otevírá diskusi zahrnující technická i filosofická hlediska. Podpora tohoto multidisciplinárního přístupu vyžaduje spojení empirických poznatků z oblasti umělé inteligence s filosofií a kognitivní vědou. Tento přístup dokáže nejen obohatit naše porozumění inteligence jako takové, ale také posunout vývoj umělé inteligence vpřed.*

Klíčová slova: *historie umělé inteligence; filosofie umělé inteligence; symbolismus; konektivismus*

YOUHENG ZHANG
Department of Philosophy
Wuhan University of Technology
Luoshi Road 122, Wuhan, Hubei, China
email /zhangyouheng@whut.edu.cn
0000-0002-2363-487X

This study was supported by the "Fundamental Research Funds for the Central Universities (China)" (grant number: 2023VB067).

1. Introduction

It has been nearly 70 years since the emergence of AI technology. Alan Turing, widely regarded as the father of AI, could hardly have envisioned its profound impact on everyday life and academic research. In a thought-provoking contribution to a philosophy journal in 1950, Turing posed a question: "Where is the best place to start research? […] chess or hardware?"[1] This query shed light on two key components that have since become fundamental to AI research: algorithms (programming) and arithmetic (hardware). As a result, philosophical discussions on AI in the modern technical sense have flourished among both philosophers and the AI research community.

The pursuit of creating "intelligent" AI programs transcends the confines of the AI discipline, as it has become an interdisciplinary endeavor. Over the course of the past 70 years, two distinct research paths have emerged, known as symbolism and connectionism, representing different paradigms within the scientific community. These paradigms embody divergent approaches to AI research and have sparked lively debates and collaborations among researchers and scholars.

2. Brief History of Symbolism and Connectionism

In the relatively short history of AI, symbolic AI emerged as the dominant approach for an extended period. Its development can be characterized by three significant stages: Automatic Theorem Proving, Expert Systems, and Knowledge Representation. Symbolic AI aimed to emulate human reasoning by employing logic and manipulating symbols. It achieved notable advancements in formal reasoning and rule-based systems, contributing to applications such as automated theorem proving and expert systems.

However, in the 21st century, connectionist AI, also called Artificial Neural Networks (ANNs), gained prominence and eventually surpassed symbolic AI as the mainstream approach in modern AI research. Connectionist AI operates by simulating the interconnected structure of biological neural networks. Its ascendancy can be traced through three distinct stages: Perception machines, Backpropagation algorithms, and Deep Learning.

During the Perception machines stage, ANNs primarily focused on mimicking human sensory perception and pattern recognition. This in-

[1] Alan M. Turing, "Computing Machinery and Intelligence," *Mind* LIX, no. 236 (1950): 433–60.

volved the development of early neural network models and algorithms that enabled machines to process sensory inputs and make basic classifications.

The introduction of the Backpropagation algorithm marked a significant breakthrough in connectionist AI. This algorithm allowed ANNs to learn from examples by adjusting the strengths of connections between artificial neurons. It enabled the training of multi-layer neural networks, enhancing their capacity to handle complex tasks and learn hierarchical representations.

The most recent and revolutionary stage in connectionist AI is Deep Learning. Deep Learning leverages deep neural networks with many layers, enabling the models to automatically learn intricate patterns and representations from vast amounts of data. This approach has led to significant breakthroughs in various domains, including image and speech recognition, natural language processing, and even playing complex games.

In this section, we will delve deeper into this rich history's chronological progression and technical aspects, exploring the developments and breakthroughs that have shaped both symbolic AI and connectionist AI.

2.1 1950–1965 Start of Intelligence Dream

The origins of symbolic AI can be traced back to the field of Automatic Theorem Proving (ATP), which is closely intertwined with the study of logic.[2] By considering "formal" as synonymous with "mechanical," most of the methods employed in ATP can be categorized within the realm of logic.[3] In 1954, the logician Martin Davis created the first ATP program on a tube computer called JOHNNIAC at the Institute for Advanced Study in Princeton.[4] However, the research paper detailing this program was not publicly released until 1957. Davis's most notable contribution was his collaboration with Hilary Putnam in solving Hilbert's tenth problem, propelling their research partnership forward. Putnam, a renowned American philosopher known for his "brain in a vat" thought experiment exploring the nature of reality, worked extensively alongside Davis. Together, they introduced the

[2] Dominique Pastre, "Automated Theorem Proving in Mathematics," *Annals of Mathematics and Artificial Intelligence* 8, no. 3–4 (1993): 425–47.
[3] Donald W. Loveland, "Automated Theorem Proving: Mapping Logic into AI," in *Proceedings of the ACM SIGART International Symposium on Methodologies for Intelligent Systems*, eds. Zbigniew W. Ras and Maria Zemankova (Knoxville, TN: ACM, 1986), 214–29.
[4] Martin Davis, "A Computer Program for Presburger's Algorithm," *Symbolic Computation Automation of Reasoning* 1 (1957): 41–48.

Davis–Putnam (DP) procedure[5] and refined it into the Davis–Putnam–Logemann–Loveland (DPLL) algorithm.[6]

In 1957, "The Logic Theorist," an ATP program developed by Herbert Simon and Allen Newell, the founders of symbolic AI, gained significant attention for pioneering the use of heuristic procedures.[7] Another influential work during this period was Dag Prawitz's natural deduction algorithm in 1957, which not only marked the inception of natural deduction but also introduced the concept of "Unification."[8]Additionally, in 1958, logician Wang Hao implemented complete propositional logic and first-order logic programs on an IBM 704 computer. Subsequently, the latter program was improved to prove all 150 first-order logic and 200 propositional logic theorems in Russell's *Principia Mathematica*.[9] While numerous studies on ATP have followed, the fundamental works of the 1960s established a solid foundation for symbolic AI.

In contrast to symbolic AI, connectionist AI follows a different philosophical principle known as computationalism. Rather than deriving intelligent machine behavior from formal logic, Warren McCulloch and Walter Pitts drew inspiration from neuroscience and sought to formalize biological neural activity to imbue machines with intelligence. They posited that the "all-or-none" law of neural activity adequately represented the activity of any neuron as a proposition, setting a threshold θ to determine neuron activity based on its characteristics.[10] The original artificial neural network model, known as the M-P model, laid the groundwork for subsequent ANNs.

In 1957, Frank Rosenblatt, influenced by psychologist Donald Hebb and philosopher Friedrich Hayek, improved upon the M-P model with his theory of "Perception." This theory integrated inputs by adding weighted inputs

[5] Martin Davis and Hilary Putnam, "A Computing Procedure for Quantification Theory," *Journal of the ACM* 7, no. (1960): 201–15.
[6] Martin Davis, George Logemann, and Donald W. Loveland, "A Machine Program for Theorem-Proving," *Communications of the ACM* 5, no. 7 (1962): 394–97.
[7] Allen Newell and Herbert A. Simon, "The Logic Theory Machine. A Complex Information Processing System," *Journal of Symbolic Logic* 22, no. 3 (1957): 331–32.
[8] Dag Prawitz, Haå kan Prawitz, and Neri Voghera, "A Mechanical Proof Procedure and Its Realization in an Electronic Computer," *Journal of the ACM* 7, no. 2 (1960): 102–28.
[9] Hao Wang, "Toward Mechanical Mathematics," *IBM Journal of Research and Development* 4, no. 1 (1960): 2–22.
[10] Warren S. McCulloch and Walter Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," *The Bulletin of Mathematical Biophysics* 5, no. 4 (1943): 115–33.

with fixed weights obtained during the training phase.[11] If the sum of these weighted inputs exceeded a given threshold θ, the neuron would trigger. When triggered, the neuron's output was set to 1; otherwise, it was set to 0. While the Perceptron functioned similarly to the M-P model, Rosenblatt removed the absolute inhibition rule, granting the neuron greater adaptability. Rosenblatt implemented software for the perceptron model on an IBM 704 at Cornell Aeronautical Laboratory and subsequently developed the Mark I Perceptron, which his team used for image classification.[12] In 1962, Rosenblatt published "Principles of Neurodynamics," summarizing his work and delving into biology, psychology, and philosophy.[13] At the time, he became a highly sought-after scholar in the United States, attracting significant attention and receiving substantial funding from private and government sources. In his paper, he openly criticized many scholars in the symbolic AI field, arguing that their approach was incapable of achieving true AI.[14] However, his views were not well-received, and in Marvin Minsky and Seymour Papert's 1969 book *Perceptron: An Introduction to Computational Geometry*, they criticized the perceptron and the connectionist AI approach, pointing out limitations such as the inability to solve the linear separability problem.[15] According to Minsky and Papert, single-layer perceptrons could not perform certain functions, like the XOR function, even with multiple layers. Their critique aimed to discredit Rosenblatt's work, which ultimately led to the end of the initial boom in connectionist AI driven by the perceptron and tragically marked the end of Rosenblatt's life.

2.2 1965–1990 The Golden Age of Artificial Intelligence in the 20th Century

During the period when connectionist AI faced temporary suppression while symbolic AI dominated, an important collaboration took place between Feigenbaum, who had studied under Simon, and geneticist Lederberg. Their encounter occurred at a conference held at Stanford University's Center for

[11] Frank Rosenblatt, "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain," *Psychological Review* 65, no. 6 (1958): 386–408.
[12] John Cameron Hay, *Mark I Perceptron Operators' Manual (Project PARA)* (Buffalo: Cornell Aeronautical Laboratory, 1960).
[13] Frank Rosenblatt, *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms* (Washington, D.C.: Spartan Books, 1962).
[14] Rosenblatt, "Perceptron," 387.
[15] Marvin Minsky and Seymour A. Papert, *Perceptrons: An Introduction to Computational Geometry* (Cambridge, MA: MIT Press, 1969).

Advanced Behavioral Science Research in 1964.[16] United by their shared interest in the philosophy of science, they embarked on a fruitful interdisciplinary collaboration, with Lederberg's influence and leadership playing a central role. According to Bruce Buchanan, Lederberg's interest shifted after conceiving the project's philosophical framework, and it took Feigenbaum's computer team five years to actualize his vision.[17] The collaboration's first outcome, after the addition of Carl Djerassi, the inventor of the oral contraceptive pill, in 1965, was the Expert System DENDRAL.[18] DENDRAL utilized data from a mass spectrometer to generate the chemical structure of a given substance. The creators aimed for it to function as an "intelligent assistant" for chemists, capable of performing expert-level tasks without necessitating an expert level of theoretical knowledge.[19] Buchanan, as one of the participants in DENDRAL, noted that their early motivations included using heuristic search to tackle complex scientific problems and employing AI methods to gain a deeper understanding of fundamental issues in the philosophy of science.[20] Following the development of DENDRAL, Edward Shortlife, a doctoral student under Buchanan's guidance, designed the expert system MYCIN for medical diagnosis.[21] Although MYCIN was never put into use, it laid the foundation for its successor, EMYCIN.[22] Buchanan commented that MYCIN and the multitude of expert systems that followed it demonstrated the ability of a small amount of knowledge to facilitate intelligent decision-making processes in various significant domains. However, with the discontinuation of the fifth-generation computer program

[16] Gil Press, "History Of AI In 33 Breakthroughs: The First Expert System," *Forbes* (website), October 29, 2022, https://www.forbes.com/sites/gilpress/2022/10/29/history-of-ai-in-33-breakthroughs-the-first-expert-system/.

[17] Bruce G. Buchanan, "Oral History Interview with Bruce G. Buchana," interview by Arthur L. Norber, June 11–12, 1991, transcript (Pittsburgh, PA: Charles Babbage Institute).

[18] Joshua Lederberg, "How DENDRAL Was Conceived and Born," in *A History of Medical Informatics*, ed. Bruce I. Blum (New York, NY: Association for Computing Machinery, 1990), 14–44.

[19] Joshua Lederberg et al. "Applications of Artificial Intelligence for Chemical Inference. I. Number of Possible Organic Compounds. Acyclic Structures Containing Carbon, Hydrogen, Oxygen, and Nitrogen," *Journal of the American Chemical Society* 91, no. 11 (1969): 2973–76.

[20] Bruce G. Buchanan and Edward A. Feigenbaum, "Dendral and Meta-Dendral: Their Applications Dimension," *Artificial Intelligence* 11, no. 1–2 (1978): 5–24.

[21] Edward H. Shortlife, "The Computer as Clinical Consultant," *Archives of Internal Medicine* 140, no. 3 (1980): 313–14.

[22] William van Melle, "A Domain-Independent Production-Rule System for Consultation Programs," in *Proceedings of the 6th International Joint Conference on Artificial Intelligence – Volume 2* (San Francisco, CA: Morgan Kaufmann Publishers Inc., 1979), 923–25.

in Japan in 1988,[23] expert systems fell out of favor and became a term that people avoided mentioning. With the rise of E-commerce driven by the Internet, many Expert Systems, such as XCON, rebranded themselves as Rule Engines.[24] Today, there are few independent Expert System projects and R&D companies.

During this period, connectionist AI did not have as significant an impact as expert systems. However, it experienced a brief resurgence. In 1974, Paul Werbos demonstrated in his doctoral dissertation that the XOR problem could be solved by adding an additional layer to neural networks and utilizing Back Propagation.[25] Initially, Werbos's research received little attention, but he was later honored with the IEEE Neural Network Society's Pioneer.[26] The revival of neural networks in the 1980s is largely attributed to physicist Hopfield, who proposed neural networks capable of solving a wide range of pattern recognition problems and providing approximate solutions to combinatorial optimization problems.[27] This neural network model is now known as the Hopfield Network. In 1984, Hopfield implemented his proposed model using an analog integrated circuit,[28] which caught the attention of many physicists and invigorated the field of connectionism. The leaders of this movement were psychologists David Rumelhurt and James McLelland, along with computer scientist Geoffrey Hinton. Their collaboration led to the publication of *Parallel Distributed Processing*, a collection of papers that synthesized the perspectives of scholars from various disciplines and sparked a surge of research at the intersection of different fields and

[23] Hiroyuki Odagiri, Yoshiaki Nakamura, and Minoru Shibuya, "Research Consortia as a Vehicle for Basic Research: The Case of a Fifth Generation Computer Project in Japan," *Research Policy* 26, no. 2 (1997): 191–207.

[24] Wolfgang Runte, "Enhancing Business Process Management with a Constraint-Based Approach," in *New Trends in Software Methodologies, Tools and Techniques*, eds. Hamido Fujita and Roberto Revetria (Amsterdam: IOS Press, 2012), 215–37.

[25] Paul Werbos, "Beyond Regression: "New Tools for Prediction and Analysis in the Behavioral Sciences," PhD diss., Harvard University, 1974.

[26] IEEE Xplore, "Paul J. Werbos – Author Profile," accessed May 7, 2023. https://ieeexplore.ieee.org/author/37344537300.

[27] John J. Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational Abilities," *Proceedings of the National Academy of Sciences* 79, no. 8 (1982): 2554–58; John J. Hopfield and David W. Tank, "'Neural' Computation of Decisions in Optimization Problems," *Biological Cybernetics* 52, no. 3 (1985): 141–52.

[28] John J. Hopfield, "Neurons with Graded Response Have Collective Computational Properties like Those of Two-State Neurons," *Proceedings of the National Academy of Sciences* 81, no. 10 (1984): 3088–92.

neural networks.[29] Within the discipline, Hinton's research contributed to the optimization of Back Propagation algorithms.[30] However, hardware limitations hindered the effective use of Back Propagation in multilayer neural networks. At the end of the century, the Internet boom again overshadowed the brief resurgence of connectionism.

2.3 Early Studies of Symbolists and Connectivists in the 21st Century

Symbolic techniques such as frames, logic, word vectors, scripts, and semantic networks played a pivotal role in structuring knowledge and enabling AI to perform intelligent tasks. These techniques, combined with Expert Systems, laid the foundation for the field of Knowledge Representation and Reasoning. Key contributions in this domain include Marvin Minsky's introduction of "Frame Theory" in 1974, which influenced the design languages of related programs and fostered an object-oriented design philosophy.[31] In 1975, Roger Schank and Robert Abelson proposed "Script Theory," utilizing frames to organize knowledge.[32] John Sowa's 1976 proposal of the "Conceptual Graph" introduced a knowledge representation system based on Semantic Networks and Peirce's logic.[33] Additionally, Rudolf Wille's introduction of Formal Concept Analysis (FCA) in 1982 provided a method to derive conceptual hierarchies and formal ontologies from objects and their attributes.[34] Concurrently, the Cyc project was launched in 1981 to develop an encyclopedia-like Knowledge Graph. Led by D. Lenat and R. Guha, the project sought to build a comprehensive and robust AI system. Cyc aimed to encompass a vast collection of properly organized information, including

29 David E. Rumelhart, James L. McClelland, and PDP Research Group, *Parallel Distributed Processing*, vol. 2 (Cambridge, MA: MIT Press, 1986).
30 David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams, "Learning Representations by Back-Propagating Errors," *Nature* 323, no. 6088 (1986): 533–36.
31 A. J. Bayle, "Frames: A Heuristic Critical Review," in *Eighth Annual International Phoenix Conference on Computers and Communications. 1989 Conference Proceedings* (Scottsdale, AZ: IEEE Computer Society Press, 1989), 624–28.
32 Roger C. Schank and Robert P. Abelson, "Scripts, Plans, and Knowledge," in *Proceedings of the 4th International Joint Conference on Artificial Intelligence – Volume 1* (San Francisco, CA: Morgan Kaufmann Publishers Inc., 1975), 151–57.
33 John F. Sowa, "Conceptual Graphs for a Data Base Interface," *IBM Journal of Research and Development* 20, no. 4 (1976): 336–57.
34 Rudolf Wille, "Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts," in *Ordered Sets*, ed. Ivan Rival (Dordrecht: Springer, 1982), 445–70.

rules and fact.[35] While G. Miller's Wordnet project in 1995 created a machine-readable database linked by lexical and semantic associations, it lacked the rule-based capabilities of Cyc. In 2001, T. Berners-Lee et al. introduced the Semantic Web, which relied on a machine-readable database but diverged from the approach of the World Wide Web Consortium (W3C).[36] The acquisition of Metaweb by Google in 2010 led to the integration of its database into Freebase, ultimately resulting in the release of Google's Knowledge Graph in 2015.[37] These advancements marked significant milestones in the development of knowledge representation and the utilization of machine-readable databases for AI applications.

Simultaneously, the growth of the Internet and advancements in AI technology presented opportunities for neural networks. In 2006, Hinton's paper on Deep Learning unveiled a machine learning method that transformed high-dimensional data into a low-dimensional code, training it layer by layer.[38] This marked the resurgence of Deep Learning (DL) as a field, enabling machines to learn features from datasets without explicit labeling. AI's development drew inspiration from various disciplines and, in turn, started to influence them.[39] The renewed interest in DL sparked advancements in Artificial Neural Networks (ANNs), with specific types gaining prominence. Convolutional Neural Networks (CNNs), introduced by LeCun et al. in 1998, found applications in machine vision, utilizing convolutional, pooling, and fully connected layers.[40] Recurrent Neural Networks (RNNs), pioneered by Hochreiter and Schmidhuber in 1997, excelled in processing sequential data and natural language, employing hidden layers with dependent activation.[41]

[35] Charles Elkan and Russell Greiner, "Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project: D. B. Lenat and R. V. Guha," *Artificial Intelligence* 61, no. 1 (1993): 41–52.

[36] Tim Berners-Lee, James Hendler, and Ora Lassila, "The Semantic Web," *Scientific American* 284, no. 5 (2001): 34–43.

[37] Tanon Pellissier et al., "From Freebase to Wikidata: The Great Migration," in *Proceedings of the 25th International Conference on World Wide Web* (Republic and Canton of Geneva: International World Wide Web Conferences Steering Committee, 2016), 1419–28.

[38] Geoffrey E. Hinton and Ruslan Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," *Science* 313, no. 5786 (2006): 504–7.

[39] Bruce G. Buchanan, "A (Very) Brief History of Artificial Intelligence," *AI Magazine* 26, no. (2005): 53.

[40] Patrice Simard et al., "Boxlets: A Fast Convolution Algorithm for Signal Processing and Neural Networks," in *Advances in Neural Information Processing Systems*, volume 11, eds. M. Kearns, S. Solla, and D. Cohn (Cambridge, MA: MIT Press, 1998).

[41] Jürge Schmidhube and Sepp Hochreiter, "Long Short-Term Memory," *Neural Computation* 9, no. 8 (1997): 1735–80.

Graph Neural Networks (GNNs), introduced by Scarselli et al. in 2009, specialized in analyzing graphs representing entities and their relationships.[42] Quantum Neural Networks (QNNs), designed for quantum computers and proposed by Schuld et al. in 2014, aimed to overcome the limitations of traditional ANNs by applying quantum theory.[43] These diverse neural network models collectively strive to simulate brain intelligence, echoing the original vision of Rosenblatt.

3. Philosophical Thinking in AI Research

When examining the history of AI throughout the last century, it becomes evident that numerous philosophical terms were employed to label the technologies. AI researchers during this period displayed a notable inclination towards publishing in philosophical journals (as evidenced by references in the previous section). Nevertheless, it is crucial to acknowledge that disparities between disciplines persist. For instance, the term "ontology" carries distinct meanings in the realms of AI and philosophy.[44] Therefore, in this section, we will explore the philosophical differences existing between the two paradigms, encompassing their theoretical conceptions, practical implementation approaches, and the perspectives of professional philosophers. By delving into these aspects, a comprehensive understanding of the contrasting philosophical underpinnings between AI and philosophy will emerge.

3.1 Symbolic Versus Neural: A Difference in Theory

When we trace the history of AI, we discover that the theoretical foundations of connectionism emerged before symbolism. McCulloch, in collaboration with Pitts, amalgamated philosophical theory and scientific practice, culminating in their groundbreaking work "A Logical Calculus of the Ideas Immanent in Nervous Activity." This seminal piece serves as the cornerstone of connectionist philosophy in AI. The groundbreaking idea that "the law of 'all-or-none' neural activity can represent the activity of any neuron as a proposition" introduced a pioneering theory that reduced the brain and

[42] Franco Scarselli et al., "The Graph Neural Network Model," *IEEE Transactions on Neural Networks* 20, no. 1 (2009): 61–80.
[43] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione, "The Quest for a Quantum Neural Network," *Quantum Information Processing* 13, no. 11 (2014): 2567–86.
[44] Christopher Welty, "Ontology Research," *AI Magazine* 24, no. 3 (2003): 11.

mind to a mathematical model.[45] Importantly, it provided philosophers with a novel avenue to explicate the nature of the mind.

McCulloch and Pitts postulated that "the physiological relations between neural activities correspond to the relations between propositions."[46] By harnessing these logical relations, they sought to uncover the fixed processes of neural activity and present them in a formalized manner. Building upon this theory, they put forth two conjectures: the "nets without circle" theory and the "nets with circle" theory. The "nets without circle" theory gave rise to the M-P model, a neural network where synapses do not form loops. Here, first-order logic was referred to as "Temporal Propositional Expressions" (TPE).[47] This non-recursive neural network could implement TPE, effectively emulating the function of a Turing machine.

On the other hand, the "nets with circle" theory encompassed neural networks with recurrent connections. McCulloch and Pitts acknowledged the challenges involved in achieving this, stating that it was "much more difficult to achieve if one is not satisfied with the free assumption of acyclic nets."[48] In comparison to acyclic nets, the expressions for networks with cycles were more intricate due to the uncertain timing of neuronal activity within the loops, involving multiple quantifications. Stephen C. Kleene later enhanced the theory of acyclic nets by identifying patterns within them.[49] He described the set of all input neuron activation sequences through mathematical expressions, leading to a specific state for a given network with cyclic connections after complete processing.

In contrast to the neural activity perspective, the article "Computer Science as Empirical Inquiry: Symbols and Search" by the founders of the symbolic school, Simon and Newell, explicitly delves into philosophical considerations regarding symbolic AI. The article elucidates the nature of research within the symbolic AI framework, framing computer science as empirical research.[50] It emphasizes that AI is not a metaphysical fantasy but

[45] Warren S. McCulloch and Walter Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," *The Bulletin of Mathematical Biophysics* 5, no. 4 (1943): 117.
[46] Ibid.
[47] Ibid., 120.
[48] Ibid., 124.
[49] Stephen C. Kleene, "Representation of Events in Nerve Nets and Finite Automata," in *Automata Studies (AM-34)*, eds. Claude E. Shannon and John McCarthy (Princeton: Princeton University Press, 1956), 3–42.
[50] Allen Newell and Herbert A. Simon, "Computer Science as Empirical Inquiry: Symbols and Search," *Communications of the ACM* 19, no. 3 (1976): 114.

rather an empirical simulation and logical inference of human intelligence, aimed at discovering and analyzing new and known phenomena.[51]

The relationship between symbols and intelligence is a central question addressed in the article. It posits that there is no overarching "intelligence principle" to explain the fundamental manifestations of intelligence, similar to how there is no "life principle" that encapsulates the nature of life itself.[52] However, the ability to store and process symbols is deemed a necessary prerequisite for intelligence. Symbolism asserts that a physical symbol system has the potential for exhibiting intelligent behavior, thereby inferring that human beings, as systems with physical symbolic attributes, possess intelligence.[53]

Regarding the implementation of intelligence, the authors argue that symbol systems do not demonstrate intelligence when in a chaotic state.[54] They propose that symbol systems solve problems through heuristic search, employing structured heuristic algorithms as a systematic approach to problem-solving. The purpose-driven Problem Domain is considered an imitation of human purposeful intelligent activity and a defining aspect of artificial intelligence, delineating its scope.

3.2 Turing and Rosenblatt: Different Visions of Intelligent Machines

Throughout the real world, an array of logical symbols is prevalent, and within this landscape, Turing machines can be viewed as a symbolic endeavor. While the von Neumann structure constitutes the framework of a computer, the Turing machine is often regarded as its soul. It is intriguing to note that Turing machines were not originally devised for computers; rather, they emerged as an attempt to address Hilbert's fundamental mathematical question.[55] However, in contemporary applications, Turing machines hold far greater significance within the realm of computers than within mathematics. While alternative approaches, such as λ-evaluations, exist for resolving the decision problem, the Turing machine remains the definitive formulation for computers.

[51] Ibid.
[52] Ibid., 115.
[53] Ibid., 118.
[54] Ibid., 126.
[55] Alan M. Turing, "On Computable Numbers, with an Application to the Entscheidungsproblem," *Proceedings of the London Mathematical Society* 2, no. 42 (1936/37): 230–65.

In his seminal work on the philosophy of AI, *Computing Machinery, and Intelligence*, Turing posited the possibility that machines would eventually rival humans in all domains of pure intelligence.[56] His visionary ideas catalyzed subsequent AI research. This work made several noteworthy contributions to the field. First and foremost, it introduced the concept of the "imitation game" and provided a functional characterization of how electronic computers can simulate human-like intelligence. Turing's arguments, despite facing objections from various disciplines such as theology, mathematics, neuroscience, and computer science, laid the groundwork for machine intelligence.[57]

Secondly, Turing's concept of a "learning machine" served as a wellspring of inspiration for contemporary machine learning methodologies. His notion of a "child machine" presaged developments akin to the advancements seen with AlphaZero. The idea of "accidental deviation of computer intelligence behavior"[58] discussed in his paper continues to hold relevance and influence in the design of modern training and learning algorithms.

Moreover, Turing shed light on the "unpredictable" nature of intelligent systems, highlighting that intelligent behavior may exhibit slight deviations from fully self-consistent computational behavior, without resorting to randomness or meaningless repetitive loops.[59]

Lastly, Turing recognized the role of punishment in machine learning, emphasizing the need to construct machines in a manner that minimizes the occurrence of punishment signals.[60] He emphasized the use of "non-emotional" communication channels for punishment and reward, which could effectively teach machines to adhere to commands within specific languages, ultimately reducing the reliance on punishment and reward signals.[61]

While Turing's visionary perspective may have carried a touch of humor, it posed a critical question: "Where is the best place to initiate research? […] play chess or hardware?"[62] This question marked the advent of the AI era and served as an impetus for further exploration and advancements in the field.

Four years following Turing's passing, Rosenblatt emerged as a significant figure in the field. Distinguishing himself from the early pioneers of

[56] Turing, "Computing Machinery and Intelligence," 460.
[57] Ibid., 443–54.
[58] Ibid., 459.
[59] Ibid.
[60] Ibid., 457.
[61] Ibid., 459.
[62] Ibid., 460.

connectionism, Rosenblatt not only envisioned Artificial Neural Networks (ANNs) but also translated his theories into practical applications, leaving a profound impact on various domains of cognitive science. ANNs are machine learning models inspired by the functioning of the brain, and their development has been met with controversy due to the incomplete understanding of the human brain.

Rosenblatt's groundbreaking creation, the Perceptron, is often regarded as a composition of individual artificial neurons, serving as the foundational model for today's complex neural networks. He argued that the symbolic AI approach, which employs physical systems to replicate biological brain functions, fails to align with biological systems and is incapable of fully explaining biological intelligence.[63] While modern science has elucidated numerous principles, no single theory can comprehensively elucidate the intricate workings of the human brain and the operational principles of intelligence.

The success of symbolic AI in the previous century was inevitable. In an era marked by limited progress in biological science and technology, achieving what was deemed "artificial intelligence" necessitated creating machines or programs that exhibited at least a semblance of "intelligence." Symbolic AI pursued an external approach, relying on symbolic logic, digital computers, and switch theory, while assuming the physical and logical foundations of intelligence, including the existence of "intelligent agents" and "human intelligence." But according to Rosenblatt, "The models of symbolic AI are merely logical constructs designed to execute specific algorithms in response to stimulus sequences, but the language of symbolic logic and Boolean algebra is ill-suited for such investigations."[64]

Conversely, the perceptron embodies an internal approach, simulating biological neurons with digital counterparts, and posits that the fundamental basis for intelligence lies within the operation of the brain. This was Rosenblatt's ambition and aspiration, often referred to as "Rosenblatt's dream"[65] – not to create a contemporary notion of "artificial intelligence" but to construct an artificially intelligent brain. In the preface to *Principles of Neurodynamics*, Rosenblatt writes:

[63] Rosenblatt, "Perceptron," 388.
[64] Ibid., 387.
[65] Guang-Bin Huang, "What Are Extreme Learning Machines? Filling the Gap Between Frank Rosenblatt's Dream and John von Neumann's Puzzle," *Cognitive Computation* 7, no. 3 (2015): 263–78.

> It is only after much hesitation that the writer has reconciled himself to the addition of the term "neurodynamics" to the list of such recent linguistic artifacts as "cybernetics," "bionics," "autonomies."

For this writer, the perceptron program is not primarily concerned with the invention of devices for "Artificial Intelligence," but rather with investigating the physical structures and neurodynamic principles which underlie "natural intelligence." A perceptron is first and foremost a brain model, not an invention for pattern recognition.[66]

3.3 Dreyfus's Criticism

Hubert Dreyfus was a philosopher who played a significant role in criticizing symbolic AI. His influential report, *Alchemy and Artificial Intelligence*,[67] later published as *What Computers Can't Do*,[68] effectively bridged the gap between philosophy and AI. While some attribute his influence to social factors, the lasting impact of his work suggests otherwise.

Dreyfus's critique of symbolic AI primarily focused on three philosophical hypotheses: the psychological hypothesis, the epistemological hypothesis, and the ontological hypothesis.[69] The psychological hypothesis raises the question of whether the mind can be effectively modeled as a computer, particularly regarding the validity of using psychological computer models.[70] According to this view, human thinking operates hierarchically, and computers can simulate human thought by accessing information at the appropriate level. Dreyfus, however, argues that the specialized use of the term "information" within this theory differs from its ordinary meaning, and processing information heuristically does not equate to possessing genuine mental activity.[71]

The epistemological hypothesis critique centers on the belief among AI experts that all non-arbitrary behavior can be formalized using specific rules, which computers can replicate.[72] Dreyfus challenges this notion from

[66] Rosenblatt, *Principles of Neurodynamics*.
[67] Hubert L. Dreyfus, *Alchemy and Artificial Intelligence* (Santa Monica, CA: RAND Corporation, 1965).
[68] Hubert L. Dreyfus, *What Computers Can't Do: The Limits of Artificial Intelligence* (New York: Harper & Row, 1979).
[69] Ibid., 68.
[70] Ibid., 75.
[71] Ibid., 77.
[72] Ibid., 102.

a physical standpoint, emphasizing that behavior lacks universal laws and drawing a distinction between neurological and physicochemical laws. He asserts that meaningful human action transcends mere physical movements and questions the ability of machines to understand semantics, as they can generate arbitrary interpretations based on formal rules.[73]

Dreyfus's critique of the ontological hypothesis revolves around the understanding that intelligent acts should be comprehensible as independent elements.[74] He does not directly examine machines as ontological entities but criticizes the practice of confining intelligence within symbolic processing. He argues that the world must be expressed as a structured set of descriptions composed of initial elements, which serves as a crucial foundation for AI research.[75] His strongest criticism lies in the formalization and modeling of intelligence within the symbolic ontology.

However, when it comes to Artificial Neural Networks (ANNs), Dreyfus's position appears inconsistent in his earlier and subsequent studies. Initially, his critique of ANNs was focused on the biological level, contending that "the human brain may process information in a completely different way than a digital computer."[76] He argued that the M-P neural model only offered a partial explanation of how neurons functioned, and he considered ANNs to be an uncritically accepted explanation embraced by neuroscientists and AI experts due to their limited understanding of human experience. Dreyfus employed philosophical synthesis to challenge AI analysis, yet it is worth noting that philosophical synthesis is rooted in concrete analysis. Consequently, his simplistic analysis, devoid of formal education in computing, invited joint criticism from both AI and philosophy fields.

With the decline of symbolic AI and the resurgence of ANNs in the 1980s, Dreyfus revisited his critique in collaboration with his cousin,[77] narrowing the scope of his AI critique to symbolism while expressing some optimism toward ANNs. He suggested that valid methods discovered in the philosophy of natural sciences must also hold true in AI research.[78] However,

[73] Ibid., 106.
[74] Ibid., 118.
[75] Ibid., 123–24.
[76] Ibid., 71.
[77] Stuart Dreyfus S. E. is a cousin of Hubert Dreyfus H. L., whose research interests are in Artificial Neural Networks.
[78] Hubert L. Dreyfus and Stuart E. Dreyfus, "Making a Mind Versus Modelling the Brain: Artificial Intelligence Back at the Branchpoint," in *Understanding the Artificial: On the Future Shape of Artificial Intelligence*, ed. Massimo Negrotti (London: Springer, 1991), 33–54.

his arguments relied on vague references to traditional philosophy, attempting to draw connections between computer languages, AI programming, and long-standing ideas in philosophy. This far-reaching approach to proof drew similar criticisms as his earlier work.

4. Symbolism and Connectivism: What the Future Holds

In 2022, OpenAI introduced Chat GPT, an AI model that utilizes textual descriptions to generate content, reigniting discussions surrounding the development of "Artificial General Intelligence (AGI)." This breakthrough has sparked controversy and challenged many philosophical biases associated with AI. Amidst the media hype, GPT has even been hailed as a potential pathway to achieving AGI (Artificial General Intelligence).[79]

Within GPT, we can observe the presence of reasoning functions, albeit not as accurately as desired. This highlights the crucial role of reasoning abilities in future general-purpose models. Furthermore, it is important to acknowledge that language processing, while a significant aspect of intelligence, only encompasses certain functions and does not represent its entirety. To truly achieve intelligence, it begs the question: Do we require a more comprehensive and expansive model of intelligence?

4.1 The Reasoning Function Is Important

In the realm of AI, we have witnessed remarkable achievements through various approaches. IBM's Deep Blue, utilizing Expert Systems in the last century, employed exhaustive enumeration to defeat a chess grandmaster in 1997. Similarly, DeepMind's AlphaStar, based on Neural Networks, triumphed over a human professional in StarCraft 2 in 2019. Connectionist AI, trained through neural networks employing supervised and reinforcement learning, has proven its ability to outperform human players in numerous video games. However, it is essential to recognize that these successes are specific to particular environments. For instance, AlphaGo, designed for playing Go, cannot excel in StarCraft due not only to the uniqueness of the trained neural networks but also because these networks rely on decision-making rather than reasoning. While both decision-making and reasoning involve correct judgments within their respective environments, the ability

[79] Sam Altman, "Planning for AGI and Beyond," *Open AI* (blog), February 24, 2023, https://openai.com/blog/planning-for-agi-and-beyond.

to make judgments in settings with clear rules and objectives differs significantly from making judgments in open environments. Consequently, connectionist AI is inadequate as a trader in the stock market, which operates within clear rules but in an open environment.

One of the primary limitations lies in the fact that network models of connectionist AI struggle to effectively capture the macro-level effects of large-scale synchronized spiking activity and global connectivity resulting from the micro-level phenomena of synaptic reorganization, neurotransmitters, and hormonal neuromodulation.[80] Consequently, modeling the brain at multiple levels remains a challenge for connectionist AI, and to transcend the limitations of "specialized AI," we must develop systems that combine the expressive and programmatic diversity of symbolic systems with the ambiguity and adaptability of connectionist expressions.[81]

The concept of Neural-Symbolic AI (NeSy AI) represents a synthesis of connectionism and symbolism, and while this research has made theoretical progress, achieving breakthroughs in practical applications proves to be challenging. There are several reasons for this. Firstly, although the human brain can employ fully physical neural networks to solve highly abstract reasoning problems, our understanding of the underlying mechanisms remains partial. Secondly, the human brain operates through neurons exchanging chattering bioelectrical impulses rather than symbols like words, and the logical structures in mental thinking function differently from the workings of the brain.[82] Despite these significant differences, recent research by Paul J. Blazek and Milo M. Lin has made strides in addressing the neurosymbolic problem with Essence Neural Networks (ENNs).[83] ENNs alleviate the "uninterpretability" issues associated with backpropagation and stochastic gradient descent. Each neuron in ENNs possesses a differentiation function, with conceptual neurons distinguishing between "like A" and "not like A," and differentiation neurons distinguishing between different elements. ENNs simulate the function of a symbolic AI program and establish a computational framework that deciphers cognitive neural processes, as highlighted by Blazek. However, despite their advancements in logical reasoning, ENNs

[80] Tom Macpherson et al., "Natural and Artificial Intelligence: A Brief Introduction to the Interplay between AI and Neuroscience Research," *Neural Networks* 144 (2021): 603–13.

[81] Marvin L. Minsky, "Logical Versus Analogical or Symbolic Versus Connectionist or Neat Versus Scruffy," *AI Magazine* 12, no. 2 (1991): 34.

[82] Herbert Jaeger, "Deep Neural Reasoning," *Nature* 538, no. 7626 (2016): 467–68.

[83] Paul J. Blazek and Milo M. Lin, "Explainable Neural Networks That Simulate Reasoning," *Nature Computational Science* 1, no. 9 (2021): 607–18.

still require training, and the training set serves as an "innate condition" for their intelligence. Although ENNs excel in image recognition and text translation, achieving AGI capabilities remains a significant challenge.

One of the fundamental roles of AI is to introduce more structure into our understanding of human thinking. However, if we perceive thinking as a singular entity, we may struggle to delve deeper and gain a comprehensive understanding. It is imperative to explore various avenues and strive to grasp the complexities and nuances of human cognition.

4.2 Neurosymbolic Systems

We have mentioned one approach to implementing Neurosymbolic systems above and have described its rationale, but it is still necessary to understand the purpose of the research and the problems faced by this approach.

Neurosymbolic systems represent a fascinating area of research in the field of AI that aims to bridge the gap between the subsymbolic nature of neural networks and the explicit representation and reasoning capabilities of symbolic AI.[84] By integrating these two paradigms, Neurosymbolic systems strive to overcome the limitations of purely Connectionism or Symbolism approaches and provide a more comprehensive framework for understanding and replicating human intelligence.

The central idea of Neurosymbolic systems is that human minds exhibit both subsymbolic processing, characterized by neural activation and pattern recognition, and symbolic processing, which involves explicit representation, manipulation of abstract concepts, and logical reasoning.[85] By combining these two aspects, researchers aim to create artificial intelligence systems that capture the richness and complexity of human cognition.

A key challenge in Neurosymbolic systems is the problem of symbolic foundations.[86] Symbolic AI relies on symbols that have meaning and refer to objects, concepts, or events in the world. However, associating symbols with the objects they refer to in a meaningful way is difficult for purely symbolic approaches. Neuralsymbol systems attempt to address this challenge by lev-

[84] Pascal Hitzler and Md Kamruzzaman Sarker, *Neuro-Symbolic Artificial Intelligence: The State of the Art* (Amsterdam: IOS Press, 2022).
[85] Hugo Latapie et al., "Neurosymbolic Systems of Perception and Cognition: The Role of Attention," *Frontiers in Psychology* 13 (2022).
[86] Pascal Hitzler et al., "Neuro-Symbolic Approaches in Artificial Intelligence," *National Science Review* 9, no. 6 (2022): nwac035.

eraging the learning and pattern recognition capabilities of neural networks to ground sensory data with symbols from real-world experiences.

Another challenge lies in bridging the gap between subsymbolic and symbolic processing.[87] Neural networks excel at tasks such as image and speech recognition and can learn patterns from large datasets. Symbolic reasoning, on the other hand, makes possible logical inference and explicit representation of knowledge. Neurosymbolic systems are dedicated to combining the statistical learning capabilities of neural networks with symbolic reasoning to achieve a more comprehensive understanding of complex problems.

Current research on Neurosymbolic systems is focused on developing hybrid approaches that can exploit the complementary strengths of connectionism and symbolism.[88] Advances in Neurosymbolic reasoning and learning algorithms are also being actively pursued to enhance the capabilities of these systems. Ethical and social implications are being explored to ensure the responsible and beneficial deployment of Neurosymbolic AI.

In summary, Neurosymbolic systems represent a promising research direction that seeks to combine the power of neural networks and symbolic reasoning. By integrating these paradigms, researchers aim to create artificial intelligence systems capable of more comprehensive and human-like cognition. Addressing these challenges will require interdisciplinary research and collaboration across the fields of artificial intelligence, cognitive science, philosophy, and neuroscience.

Ongoing efforts are focused on developing new Neurosymbolic architectures, training algorithms, and evaluation frameworks to overcome these problems and improve the capabilities and understanding of Neurosymbolic systems.[89] While challenges and philosophical questions remain, ongoing research on Neurosymbolic systems has the potential to open up new possibilities for artificial intelligence and deepen our understanding of intelligence and thinking.

[87] Zenan Li et al., "Softened Symbol Grounding for Neuro-Symbolic Systems," in *Proceedings of the Eleventh International Conference on Learning Representations (ICLR 2023)*, Kigali, Rwanda, 2023.

[88] Amit Sheth, Kaushik Roy, and Manas Gaur, "Neurosymbolic Artificial Intelligence (Why, What, and How)," *IEEE Intelligent Systems* 38, no. 3 (2023): 56–62.

[89] Pascal Hitzler, "Some Advances Regarding Ontologies and Neuro-Symbolic Artificial Intelligence," in *ECMLPKDD Workshop on Meta-Knowledge Transfer*, eds. Pavel Brazdil et al. (Proceedings of Machine Learning Research, 2022), 8–10.

4.3 Cognitive Architecture & Foundational Models

In the above sections, we discussed the importance of combining symbolism and connectionism in the development of AI systems, emphasizing the significance of the reasoning function and the purpose of research on neurosymbolic systems. While this hybrid approach of Neurosymbolic AI suggests promising directions for future research, its limited technical implementation is hindered by the lack of a robust underlying theory. Cognitive architecture, on the other hand, is widely recognized in cognitive science as an approach to comprehending and elucidating the concept of "intelligence."

In the previous century, two prominent models of cognition, namely the SOAR model and the ACT-R model, have garnered significant attention in the field of artificial intelligence. The SOAR model, proposed by Paul Rosenbloom and others, provides a comprehensive framework that integrates reasoning, learning, perception, motor control, language, cognitive development, emotion, and potentially even consciousness within a cohesive structure.[90] Due to the multifaceted nature of human cognition, the SOAR model may not encompass all cognitive structures and may present solutions in applications that were unforeseen by its creators.[91]

On the other hand, the ACT-R model, proposed by John Anderson, is a representative model of artificial neural networks (ANNs). It offers a general framework for understanding the organization of the brain and how this organization generates thoughts.[92] Both the SOAR and ACT-R models transcend the traditional debates between connectionism and symbolic approaches. A comprehensive model must incorporate the strengths of both paradigms since symbolic approaches aim to describe thought as a manifestation of brain function utilizing specific types of entities and systems.

Recent theories in cognitive science have introduced the concept of dual processes, often referred to as System 1 and System 2, to explain human behavior. This theoretical framework provides insights into the coordination between symbolic and connectionist AI aspects. According to the dual-process theory of the mind, System 1 is characterized by associative,

[90] Paul S. Rosenbloom, Allen Newell, and John E. Laird, *Soar Papers: Research on Integrated Intelligence* (Cambridge, MA: MIT Press, 1993).
[91] M. Mitchell Waldrop, "Soar: A Unified Theory of Cognition?," *Science* 241, no. 4863 (1988): 296–98.
[92] Christian Lebiere and John R. Anderso, "A Connectionist Implementation of the ACT-R Production System,"in *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society* (Boulder: University of Colorado, 1993), 635–40.

implicit, imaginative, personalized, and rapid cognitive processes. In contrast, System 2 is characterized by analytic, episodic, verbal, generalized, and slower cognitive processes.[93] It is important to note that the relationship between System 1 and System 2 in human cognition is not a direct one-to-one correspondence. While System 1 may involve the use of symbols and abstract functions and algorithms that connect to AI, the functions and algorithms of System 2, although based on a symbolic approach, are implemented through the neural networks of the human brain.

Additionally, Global Workspace Theory (GWT) is a compelling theoretical framework in cognitive psychology and neuroscience that aims to elucidate the architecture underlying consciousness and cognitive processes. Developed by Bernard Baars, GWT proposes that conscious experience emerges from the coordinated activity of multiple specialized brain systems that exchange information within a global workspace.[94] According to GWT, the brain is comprised of various specialized processing modules, each responsible for specific tasks such as vision, language, motor control, and memory. These modules operate in parallel and independently, unconsciously processing information. However, when information becomes relevant to the entire cognitive system, it is broadcasted to a global workspace – a shared neural network accessible to multiple modules.

Despite its merits, some scholars have raised concerns about GWT. They argue that it lacks a detailed mechanism for the selection and dissemination of information in the global workspace.[95] Additionally, GWT falls short in addressing the subjective experience and the quality of consciousness. In 2023, the result of the bet between neuroscientist Christof Koch and philosopher David Chalmers brought renewed attention to GWT.[96] However, Koch's experiment highlighted that GWT did not provide a clear explana-

[93] Keith E. Stanovich, Richard F. West, and Ralph Hertwig, "Individual Differences in Reasoning: Implications for the Rationality Debate? – Open Peer Commentary – the Questionable Utility of Cognitive Ability in Explaining Cognitive Illusions," *Behavioral and Brain Sciences* 23, no. 5 (2000): 645–65.

[94] Bernard J. Baar and Stan Franklin, "An Architectural Model of Conscious and Unconscious Brain Functions: Global Workspace Theory and IDA," *Neural Networks* 20, no. 9 (2007): 955–61.

[95] David Kemmerer, "Are We Ever Aware of Concepts? A Critical Question for the Global Neuronal Workspace, Integrated Information, and Attended Intermediate-Level Representation Theories of Consciousness," *Neuroscience of Consciousness* 2015, no. 1 (2015): niv006.

[96] Mariana Lenharo, "Decades-Long Bet on Consciousness Ends – and It's Philosopher 1, Neuroscientist 0," *Nature* 619, no. 7968 (2023): 14–15.

tion for the mysteries of consciousness research within the specified time limit. This underscores the realization that we still have a long journey ahead in our quest to comprehensively unravel the cognitive theory of consciousness from a top-down perspective.

Moreover, the concept of a Foundational Model was introduced in a thought-provoking paper published in 2021 by Stanford University's Human-Centered Artificial Intelligence.[97] This paper provides a clear exposition of the limitations of the Large Language Model (LLM) in terms of implementing intelligence, and it advocates for the development of a more comprehensive model that encompasses a broader range of disciplines. The authors emphasize the need for collaboration and invite researchers in the humanities and social sciences to actively participate in this endeavor.

This notion of a Foundational Model suggests a shift in focus within the field of AI. While technical challenges in AI have progressively diminished, the scientific challenges have become more prominent, particularly those related to cognitive science. These challenges necessitate a concerted effort from the cognitive sciences as a whole, encompassing disciplines such as philosophy and AI research. By recognizing the interdisciplinary nature of intelligence and encouraging collaboration across these fields, we can address the complex scientific problems that arise in the pursuit of developing advanced AI systems.

5. Conclusion

When reflecting upon the entire trajectory of AI's historical development, we employ the designations of "symbolism' and "connectionism" as convenient tools to distinguish between the two distinct approaches toward research objectives. These classifications aid us in better discerning the outcomes derived from AI research. In truth, if we were to transport the researchers of the bygone century to our present era, they would likely be astonished by the evolution of their field. Many of these eminent pioneers aspired to create intelligent systems that could rival, or perhaps even surpass, human capabilities. Their visions extended far beyond the contemporary subdomains of AI, encompassing endeavors like speech recognition, image analysis, autonomous driving, and various other domains that have experienced remarkable breakthroughs within the expansive realm of artificial intelligence. It

[97] Rishi Bommasani and Liang Percy, "Reflections on Foundation Models," *Stanford HAI* (website), October 18, 2021, https://hai.stanford.edu/news/reflections-foundation-models.

is due to their great vision and deep interest in the study of philosophy that philosophy has developed a deep connection with AI.

Philosophical luminaries like René Descartes, John Locke, and David Hume laid the very bedrock upon which discussions concerning the nature of thought and knowledge were constructed, prompting AI researchers to embark on a quest to replicate these cognitive processes within machines. Similarly, concepts like the intriguing "trolley problem" and the persistent debates surrounding machine ethics have ignited fervent discussions concerning the ethical decision-making abilities of AI systems and the potential for AI to possess moral agency. In the realm of linguistics and language philosophy, figures like Ludwig Wittgenstein and Noam Chomsky have profoundly influenced AI research, particularly in the domains of natural language processing and comprehension. Yet, AI research reciprocates this influence by introducing fresh perspectives on traditional philosophical domains such as the mind, consciousness, and cognition. As AI research continues its inexorable march forward, it simultaneously begets new philosophical inquiries, particularly in the realms of ethics and sociology. Philosophy, therefore, serves as the very cornerstone upon which the edifice of AI inquiry is erected. It has been a guiding beacon in the development of AI technologies and an enduring catalyst for discussions concerning the societal implications of AI, its ethical underpinnings, and its capacity to emulate or even surpass human capabilities.

As the landscape of artificial intelligence research continues to unfold, our endeavors extend beyond the mere construction of computationally adept machines; they delve into the very essence of humanity itself. In our relentless pursuit of replicating intelligence, we find ourselves entangled in profound inquiries concerning the enigma of consciousness, the essence of the self, and the boundaries of our comprehension. As AI continues to evolve, our responsibility transcends the quest for technological innovation; it encompasses the preservation of those distinctive attributes that define our humanity. The future of AI research presents us with an intricate labyrinth, one woven with threads of ethics, morality, and existential contemplation. It serves as a poignant reminder that as we embark on the journey to engender intelligence, we may inadvertently unearth profound insights into the very core of our own existence.

Acknowledgments:

This article represents the culmination of my work, now in its final form after being initially published as preprint.[98] It stands as an enriched and refined iteration of our prior research, specifically building upon the foundations laid in our previous work titled "A Historical Review and Philosophical Examination of the Two Paradigms in Artificial Intelligence Research."[99] I extend our heartfelt gratitude to the editorial team at *Teorie vědy / Theory of Science*, as well as the anonymous reviewers, for their invaluable contributions and suggestions that have significantly enhanced the quality of this article.

Bibliography:

Altman, Sam. "Planning for AGI and Beyond." *Open AI* (blog), February 24, 2023. https://openai.com/blog/planning-for-agi-and-beyond.

Baars, Bernard J., and Stan Franklin. "An Architectural Model of Conscious and Unconscious Brain Functions: Global Workspace Theory and IDA." *Neural Networks* 20, no. 9 (2007): 955–61. https://doi.org/10.1016/j.neunet.2007.09.013.

Bayle, A. J. "Frames: A Heuristic Critical Review." In *Eighth Annual International Phoenix Conference on Computers and Communications. 1989 Conference Proceedings*, 624–28. Scottsdale, AZ: IEEE Computer Society Press, 1989. https://doi.org/10.1109/PCCC.1989.37457.

Berners-Lee, Tim, James Hendler, and Ora Lassila. "The Semantic Web." *Scientific American* 284, no. 5 (2001): 34–43. https://doi.org/10.1038/scientificamerican0501-34.

Blazek, Paul J., and Milo M. Lin. "Explainable Neural Networks That Simulate Reasoning." *Nature Computational Science* 1, no. 9 (2021): 607–18. https://doi.org/10.1038/s43588-021-00132-w.

Bommasani, Rishi, and Liang Percy. "Reflections on Foundation Models." *Stanford HAI* (website), October 18, 2021. https://hai.stanford.edu/news/reflections-foundation-models.

[98] Youheng Zhang, "A Historical Interaction between Artificial Intelligence and Philosophy," preprint, submitted July 23, 2022. https://arxiv.org/abs/2208.04148.
[99] Youheng Zhang, "A Historical Review and Philosophical Examination of the Two Paradigms in Artificial Intelligence Research," *European Journal of Artificial Intelligence and Machine Learning* 2, no. 2 (2023): 24–32.

Buchanan, Bruce G. “Oral History Interview with Bruce G. Buchanan.” Interview by Arthur L. Norber, June 11–12, 1991, transcript. Pittsburgh, PA: Charles Babbage Institute. http://conservancy.umn.edu/handle/11299/107165.

Buchanan, Bruce G. “A (Very) Brief History of Artificial Intelligence.” *AI Magazine* 26, no. 4 (2005): 53. https://doi.org/10.1609/aimag.v26i4.1848.

Buchanan, Bruce G., and Edward A. Feigenbaum. “Dendral and Meta-Dendral: Their Applications Dimension.” *Artificial Intelligence* 11, no. 1–2 (1978): 5–24. https://doi.org/10.1016/0004-3702(78)90010-3.

Davis, Martin. “A Computer Program for Presburger’s Algorithm.” *Symbolic Computation Automation of Reasoning* 1 (1957): 41–48.

Davis, Martin, George Logemann, and Donald W. Loveland. “A Machine Program for Theorem-Proving.” *Communications of the ACM* 5, no. 7 (1962): 394–97. https://doi.org/10.1145/368273.368557.

Davis, Martin, and Hilary Putnam. “A Computing Procedure for Quantification Theory.” *Journal of the ACM* 7, no. 3 (1960): 201–15. https://doi.org/10.1145/321033.321034.

Dreyfus, Hubert L. *What Computers Can’t Do: The Limits of Artificial Intelligence*. New York: Harper & Row, 1979. http://archive.org/details/whatcomputerscan00hube.

Dreyfus, Hubert L. *Alchemy and Artificial Intelligence*. Santa Monica, CA: RAND Corporation, 1965.

Dreyfus, Hubert L., and Stuart E. Dreyfus. “Making a Mind Versus Modelling the Brain: Artificial Intelligence Back at the Branchpoint.” In *Understanding the Artificial: On the Future Shape of Artificial Intelligence*, edited by Massimo Negrotti, 33–54. London: Springer, 1991. https://doi.org/10.1007/978-1-4471-1776-6_3.

Elkan, Charles, and Russell Greiner. “Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project: D. B. Lenat and R. V. Guha.” *Artificial Intelligence* 61, no. 1 (1993): 41–52. https://doi.org/10.1016/0004-3702(93)90092-P.

Hay, John Cameron. *Mark I Perceptron Operators’ Manual (Project PARA)*. Buffalo: Cornell Aeronautical Laboratory, 1960.

Hinton, Geoffrey E., and Ruslan Salakhutdinov. “Reducing the Dimensionality of Data with Neural Networks.” *Science* 313, no. 5786 (2006): 504–7. https://doi.org/10.1126/science.1127647.

Hitzler, Pascal, and Md Kamruzzaman Sarker. *Neuro-Symbolic Artificial Intelligence: The State of the Art*. Amsterdam: IOS Press, 2022.

Hitzler, Pascal. "Some Advances Regarding Ontologies and Neuro-Symbolic Artificial Intelligence." In *ECMLPKDD Workshop on Meta-Knowledge Transfer*, edited by Pavel Brazdil, Jan N. van Rijn, Henry Gouk, and Felix Moh, 8–10. Proceedings of Machine Learning Research, 2022. https://proceedings.mlr.press/v191/hitzler22a.html.

Hitzler, Pascal, Aaron Eberhart, Monireh Ebrahimi, Md Kamruzzaman Sarker, and Lu Zhou. "Neuro-Symbolic Approaches in Artificial Intelligence." *National Science Review* 9, no. 6 (2022): nwac035. https://doi.org/10.1093/nsr/nwac035.

Hopfield, John J. "Neural Networks and Physical Systems with Emergent Collective Computational Abilities." *Proceedings of the National Academy of Sciences* 79, no. 8 (1982): 2554–58. https://doi.org/10.1073/pnas.79.8.2554.

Hopfield, John J. "Neurons with Graded Response Have Collective Computational Properties like Those of Two-State Neurons." *Proceedings of the National Academy of Sciences* 81, no. 10 (1984): 3088–92. https://doi.org/10.1073/pnas.81.10.3088.

Hopfield, John J., and David W. Tank. "'Neural' Computation of Decisions in Optimization Problems." *Biological Cybernetics* 52, no. 3 (1985): 141–52. https://doi.org/10.1007/BF00339943.

Huang, Guang-Bin. "What Are Extreme Learning Machines? Filling the Gap Between Frank Rosenblatt's Dream and John von Neumann's Puzzle." *Cognitive Computation* 7, no. 3 (2015): 263–78. https://doi.org/10.1007/s12559-015-9333-0.

IEEE Xplore. "Paul J. Werbos – Author Profile." Accessed May 7, 2023. https://ieeexplore.ieee.org/author/37344537300.

Jaeger, Herbert. "Deep Neural Reasoning." *Nature* 538, no. 7626 (2016): 467–68. https://doi.org/10.1038/nature19477.

Kemmerer, David. "Are We Ever Aware of Concepts? A Critical Question for the Global Neuronal Workspace, Integrated Information, and Attended Intermediate-Level Representation Theories of Consciousness." *Neuroscience of Consciousness* 2015, no. 1 (2015): niv006. https://doi.org/10.1093/nc/niv006.

Kleene, Stephen C. "Representation of Events in Nerve Nets and Finite Automata." In *Automata Studies (AM-34)*, edited by Claude E. Shannon and John McCarthy, 3–42. Princeton: Princeton University Press, 1956. https://doi.org/10.1515/9781400882618-002.

Latapie, Hugo, Ozkan Kilic, Kristinn R. Thórisson, Pei Wang, and Patrick Hammer. "Neurosymbolic Systems of Perception and Cognition: The Role of Attention." *Frontiers in Psychology* 13 (2022): 806397. https://doi.org/10.3389/fpsyg.2022.806397.

Lebiere, Christian, and John R. Anderson. "A Connectionist Implementation of the ACT-R Production System." In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*, 635–40. Boulder: University of Colorado, 1993. https://doi.org/10.1184/R1/6613085.v1.

Lederberg, Joshua. "How DENDRAL Was Conceived and Born." In *A History of Medical Informatics*, edited by Bruce I. Blum, 14–44. New York, NY: Association for Computing Machinery, 1990. https://doi.org/10.1145/89482.89484.

Lederberg, Joshua, Georgia L. Sutherland, Bruce G. Buchanan, Edward A. Feigenbaum, Alexander V. Robertson, Alan M. Duffield, and Carl Djerassi. "Applications of Artificial Intelligence for Chemical Inference. I. Number of Possible Organic Compounds. Acyclic Structures Containing Carbon, Hydrogen, Oxygen, and Nitrogen." *Journal of the American Chemical Society* 91, no. 11 (1969): 2973–76. https://doi.org/10.1021/ja01039a025.

Lenharo, Mariana. "Decades-Long Bet on Consciousness Ends – and It's Philosopher 1, Neuroscientist 0." *Nature* 619, no. 7968 (2023): 14–15. https://doi.org/10.1038/d41586-023-02120-8.

Li, Zenan, Yuan Yao, Taolue Chen, Jingwei Xu, Chun Cao, Xiaoxing Ma, and Jian Lü. "Softened Symbol Grounding for Neuro-Symbolic Systems." In *Proceedings of the Eleventh International Conference on Learning Representations (ICLR 2023).* Kigali, Rwanda, 2023. https://openreview.net/forum?id=HTJE5Krui0g.

Loveland, Donald W. "Automated Theorem Proving: Mapping Logic into AI." In *Proceedings of the ACM SIGART International Symposium on Methodologies for Intelligent Systems*, edited by Zbigniew W. Ras and Maria Zemankova, 214–29. Knoxville, TN: ACM, 1986. https://doi.org/10.1145/12808.12833.

Macpherson, Tom, Anne Churchland, Terry Sejnowski, James DiCarlo, Yukiyasu Kamitani, Hidehiko Takahashi, and Takatoshi Hikida. "Natural and Artificial Intelligence: A Brief Introduction to the Interplay between AI and Neuroscience Research." *Neural Networks* 144 (2021): 603–13. https://doi.org/10.1016/j.neunet.2021.09.018.

Rumelhart, David E., James L. McClelland, and PDP Research Group. *Parallel Distributed Processing.* Volume 2. Cambridge, MA: MIT Press, 1986.

McCulloch, Warren S., and Walter Pitts. "A Logical Calculus of the Ideas Immanent in Nervous Activity." *The Bulletin of Mathematical Biophysics* 5, no. 4 (1943): 115–33. https://doi.org/10.1007/BF02478259.

Melle, William van. "A Domain-Independent Production-Rule System for Consultation Programs." In *Proceedings of the 6th International Joint Conference on Artificial Intelligence – Volume 2*, 923–25. San Francisco, CA: Morgan Kaufmann Publishers Inc., 1979.

Minsky, Marvin L. "Logical Versus Analogical or Symbolic Versus Connectionist or Neat Versus Scruffy." *AI Magazine* 12, no. 2 (1991): 34. https://doi.org/10.1609/aimag.v12i2.894.

Minsky, Marvin L., and Seymour A. Papert. *Perceptrons: An Introduction to Computational Geometry*. Cambridge, MA: MIT Press, 1969.

Newell, Allen, and Herbert A. Simon. "The Logic Theory Machine. A Complex Information Processing System." *Journal of Symbolic Logic* 22, no. 3 (1957): 331–32. https://doi.org/10.2307/2963663.

Newell, Allen, and Herbert A. Simon. "Computer Science as Empirical Inquiry: Symbols and Search." *Communications of the ACM* 19, no. 3 (1976): 113–26. https://doi.org/10.1145/360018.360022.

Odagiri, Hiroyuki, Yoshiaki Nakamura, and Minoru Shibuya. "Research Consortia as a Vehicle for Basic Research: The Case of a Fifth Generation Computer Project in Japan." *Research Policy* 26, no. 2 (1997): 191–207. https://doi.org/10.1016/S0048-7333(97)00008-5.

Pastre, Dominique. "Automated Theorem Proving in Mathematics." *Annals of Mathematics and Artificial Intelligence* 8, no. 3–4 (1993): 425–47. https://doi.org/10.1007/BF01530801.

Pellissier Tanon, Thomas, Denny Vrandečić, Sebastian Schaffert, Thomas Steiner, and Lydia Pintscher. "From Freebase to Wikidata: The Great Migration." In *Proceedings of the 25th International Conference on World Wide Web*, 1419–28. Republic and Canton of Geneva: International World Wide Web Conferences Steering Committee, 2016. https://doi.org/10.1145/2872427.2874809.

Prawitz, Dag, Haå kan Prawitz, and Neri Voghera. "A Mechanical Proof Procedure and Its Realization in an Electronic Computer." *Journal of the ACM* 7, no. 2 (1960): 102–28.

Press, Gil. "History Of AI In 33 Breakthroughs: The First Expert System." *Forbes* (website), October 29, 2022. https://www.forbes.com/sites/gilpress/2022/10/29/history-of-ai-in-33-breakthroughs-the-first-expert-system/.

Rosenblatt, Frank. "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain." *Psychological Review* 65, no. 6 (1958): 386–408. https://doi.org/10.1037/h0042519.

Rosenblatt, Frank. *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Washington, D.C.: Spartan Books, 1962.

Rosenbloom, Paul S., Allen Newell, and John E. Laird. *Soar Papers: Research on Integrated Intelligence*. Cambridge, MA: MIT Press, 1993.

Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. "Learning Representations by Back-Propagating Errors." *Nature* 323, no. 6088 (1986): 533–36. https://doi.org/10.1038/323533a0.

Runte, Wolfgang. "Enhancing Business Process Management with a Constraint-Based Approach." In *New Trends in Software Methodologies, Tools and Techniques*, edited by Hamido Fujita and Roberto Revetria, 215–37. Amsterdam: IOS Press, 2012.

Scarselli, Franco, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. "The Graph Neural Network Model." *IEEE Transactions on Neural Networks* 20, no. 1 (2009): 61–80. https://doi.org/10.1109/TNN.2008.2005605.

Schank, Roger C., and Robert P. Abelson. "Scripts, Plans, and Knowledge." In *Proceedings of the 4th International Joint Conference on Artificial Intelligence – Volume 1*, 151–57. San Francisco, CA: Morgan Kaufmann Publishers Inc., 1975. https://doi.org/10.5555/1624626.1624649.

Schmidhuber, Jürgen, and Sepp Hochreiter. "Long Short-Term Memory." *Neural Computation* 9, no. 8 (1997): 1735–80. https://doi.org/10.1162/neco.1997.9.8.1735.

Schuld, Maria, Ilya Sinayskiy, and Francesco Petruccione. "The Quest for a Quantum Neural Network." *Quantum Information Processing* 13, no. 11 (2014): 2567–86. https://doi.org/10.1007/s11128-014-0809-8.

Sheth, Amit, Kaushik Roy, and Manas Gaur. "Neurosymbolic Artificial Intelligence (Why, What, and How)." *IEEE Intelligent Systems* 38, no. 3 (2023): 56–62. https://doi.org/10.1109/MIS.2023.3268724.

Shortlife, Edward H. "The Computer as Clinical Consultant." *Archives of Internal Medicine* 140, no. 3 (1980): 313–14. https://doi.org/10.1001/archinte.1980.00330150027008.

Simard, Patrice, Léon Bottou, Patrick Haffner, and Yann LeCun. "Boxlets: A Fast Convolution Algorithm for Signal Processing and Neural Networks." In *Advances* *in Neural Information Processing Systems*, volume 11, edited by M. Kearns, S. Solla, and D. Cohn. Cambridge, MA: MIT Press, 1998.

Sowa, John F. "Conceptual Graphs for a Data Base Interface." *IBM Journal of Research and Development* 20, no. 4 (1976): 336–57. https://doi.org/10.1147/rd.204.0336.

Stanovich, Keith E., Richard F. West, and Ralph Hertwig. "Individual Differences in Reasoning: Implications for the Rationality Debate? – Open Peer Commentary – the Questionable Utility of Cognitive Ability in Explaining Cognitive Illusions." *Behavioral and Brain Sciences* 23, no. 5 (2000): 645–65.

Turing, Alan M. "Computing Machinery and Intelligence." *Mind* LIX, no. 236 (1950): 433–60. https://doi.org/10.1093/mind/LIX.236.433.

Turing, Alan M. "On Computable Numbers, with an Application to the Entscheidungsproblem." *Proceedings of the London Mathematical Society* 2, no. 42 (1936/37): 230–65.

Waldrop, M. Mitchell. "Soar: A Unified Theory of Cognition?" *Science* 241, no. 4863 (1988): 296–98. https://doi.org/10.1126/science.3291119.

Wang, Hao. "Toward Mechanical Mathematics." *IBM Journal of Research and Development* 4, no. 1 (1960): 2–22. https://doi.org/10.1147/rd.41.0002.

Welty, Christopher. "Ontology Research." *AI Magazine* 24, no. 3 (2003): 11. https://doi.org/10.1609/aimag.v24i3.1714.

Werbos, Paul. "Beyond Regression: "New Tools for Prediction and Analysis in the Behavioral Sciences." PhD diss., Harvard University, 1974.

Wille, Rudolf. "Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts." In *Ordered Sets*, edited by Ivan Rival, 445–70. Dordrecht: Springer, 1982. https://doi.org/10.1007/978-94-009-7798-3_15.

Youheng, Zhang. "A Historical Interaction between Artificial Intelligence and Philosophy." Preprint, submitted July 23, 2022. https://arxiv.org/abs/2208.04148.

Youheng, Zhang. "A Historical Review and Philosophical Examination of the Two Paradigms in Artificial Intelligence Research." *European Journal of Artificial Intelligence and Machine Learning* 2, no. 2 (2023): 24–32. https://doi.org/10.24018/ejai.2023.2.2.23.