Humanology, *Vol. 1*, n. 2, 2025
ISSN 3104-2570 (online)
DOI: 10.46793/HumanologyI-2.215M

Received: 25/01/2025
Accepted: 25/05/2025
Published online: 07/09/2026

Filip Mladenović[1]
PhD Student
Faculty of Philosophy
University of Belgrade

THE COGNITIVE REVOLUTION AS THE INTELLECTUAL FRAMEWORK FOR ESTABLISHING THE THEORETICAL FOUNDATIONS OF ARTIFICIAL INTELLIGENCE

Abstract*: This paper examines the cognitive revolution as the intellectual framework that enabled the theoretical foundations of AI. Tracing the decline of behaviorism and the rise of internalist models in psychology and philosophy of mind, the study highlights how the information-processing metaphor and the formalization of mental representations provided the conceptual tools for early symbolic AI. The discussion follows the transition from symbolic to connectionist models, emphasizing the significance of parallel distributed processing and deep learning architectures in overcoming the limitations of classical AI. Special attention is given to the role of linguistics, functionalism, and the symbol-grounding problem in shaping both the possibilities and boundaries of machine intelligence. The paper also addresses the ethical and conceptual challenges posed by contemporary AI, including issues of transparency, autonomy, and the „black box" problem.*

[1] dahfica@gmail.com

Ultimately, the cognitive revolution is presented not merely as a historical context but as the enduring intellectual framework that continues to define and challenge our understanding of AI, cognition, and the nature of intelligent systems

Keywords: *cognitive revolution, artificial intelligence, behaviorism, philosophy of mind, functionalism, deep learning.*

1. Introduction

The emergence of the cognitive revolution in the mid-twentieth century redefined the question of what scientific psychology and the philosophy of mind can, and ought to, explain. This shift involved a reorientation from the study of exclusively external behavior toward the acceptance of internal mental processes as a legitimate and experimentally accessible subject matter. Instead of treating the mind as a closed „black box“, the new approach conceived cognition as a system that processes, transforms, and interprets information. During the same period, the first wave of research in artificial intelligence (AI) appeared. Its theoretical foundations were drawn precisely from this newly formed information-processing metaphor of the mind, from computational models, the „formalization of thinking“, and from the emerging theories of functionalism. Naturally, the cognitive revolution developed as a broad interdisciplinary endeavor spanning philosophy, linguistics, anthropology, neuroscience, computer science, and psychology (Miller, 2003: 143). Computer science itself developed by taking human cognition as

both its model and its inspiration, grounding its core assumptions directly in scientific conceptions of cognition.

2. The Decline of Behaviorism and the Need for Internal Structure

Behaviorism was the dominant paradigm preceding the cognitive revolution and the theoretical framework against which the revolution first articulated itself. Throughout the first half of the twentieth century, behaviorism dominated psychology by insisting that scientific investigation must remain restricted to S-R (stimulus–response) models. This restriction was motivated by the assumption that only S-R relations could be empirically verified. The S-R framework offered predictability, relied on observable data, and allowed experimental control over variables, thereby enabling a form of experimental design modeled on the natural sciences. For this reason, John Watson emphasized that psychology is an objective and experimental science that does not require introspection (Watson, 1913: 176). Thus, within behaviorism, there was no conceptual space for anything that did not fit into the S-R framework.

Noam Chomsky's critique demonstrated that the behavioristmodelcouldnotexplaintheproductivity of language, the generation of entirely novel sentence constructions, or the systematic patterns of errors children make when acquiring their native language (Chomsky, 1959: 55–59). The counterexamples Chomsky presented were consistent with a broader model of rules intrinsic to internal linguistic structures, which become visible precisely when observing a child

acquiring language. Learning a language is not a matter of chaining words or sentences together, it is a process that presupposes, and requires, internal rules and structures. According to Miller, Chomsky's critique represented the single most decisive turning point toward the cognitive revolution (Miller, 2003: 142). It delivered a fatal blow to behaviorism and opened the way for a new conception of the mind, one that treats representations as constitutive and necessary for explaining behavior.

This development was also crucial for the emergence of AI, since machine intelligence cannot plausibly be described as a sequence of reactions to stimuli. It necessarily presupposes operations over symbols and rules, precisely the kinds of processes behaviorism could neither recognize nor account for.

3. The Information Metaphor and the Mind as an Information-Processing System

As noted in the introduction, the first wave of AI research grounded its theoretical foundations in the information-processing metaphor of the mind. This metaphor lies at the very core of the cognitive revolution. It presupposes that the mind is no longer to be understood as an abstract entity whose inner content is inaccessible, but as a system that receives, processes, stores, and reproduces information. This conception of the mind made it possible for mental processes to be formalized, simulated, and experimentally tested, thereby establishing the legitimacy of mental representations as explanatory constructs for behavior.

Three aspects of the information-processing metaphor are especially relevant for the development of AI. The first concerns the idea that cognitive functions can be analyzed as sequences of discrete steps, a prerequisite for algorithmic modeling. Mental phenomena are thus treated as theoretical constructs that can be manipulated at a formal level, while, more importantly, questions about their operation can be formulated as empirically testable and falsifiable hypotheses (Gardner, 1985: 29–30).

The second aspect concerns the existence of distinct levels of analysis within the new framework. It introduces the possibility of asking what a system, as a functional whole, is doing, how it performs its operations, and upon what mechanisms these operations are based. David Marr identified three levels at which any information-processing system must be understood: the computational theory, the level of representations and algorithms, and the hardware implementation (Marr, 2010: 25). The separation of these levels enables the maintenance of different kinds of explanatory validity. A model may be predictive without mirroring its physical realization. This structure of levels opened space for methodological cooperation between abstract theory and empirical data, producing a level of precision and productivity previously unattainable. It also prevented the imposition of a single worldview or discourse as the "one and only true one", truth could now be approached as a convergence among different levels of explanation.

The third aspect is grounded in the previous two, it provides a theoretical and empirically supported justification for treating mental representations as legitimate scientific constructs. Without stable mental

representations, phenomena such as memory or reasoning cannot be explained, and representations would soon become the foundation upon which early symbolic AI was built.

4. Computer Science and the Early Formation of AI

These developments enabled the emerging field of computer science to articulate the problem of intelligence in terms of algorithms and symbolic representations. Alan Turing introduced the concept of a universal machine in 1936, demonstrating that any effectively specified procedure can be carried out by general algorithmic mechanisms. A single machine can simulate any other machine, provided that it is supplied with a description of that machine as input (Turing, 1936: 241–242). This established the understanding of computation as the formal transformation of symbols.

Building on this foundation, Turing's 1950 paper introduced the imitation game as an operational criterion for empirically examining intelligent behavior. His proposal did not rest on any essential mental substance distinguishing humans from machines, but rather on the claim that if a machine can successfully reproduce intelligent behavior, there is no principled reason to deny it the status of a thinking system. Turing explicitly noted that machines could one day compete with humans in all purely intellectual domains (Turing, 1950: 441).

The concept of the Turing machine thus laid the foundations for understanding symbolic representation and algorithmicinformationprocessing. Asanabstract

model of computation, it showed that any process expressible as a sequence of discrete steps can be formalized, which encouraged researchers to conceptualize human thought as a kind of symbol manipulation. Whatever reservations one may have regarding this analogy, Turing argued that we cannot know „what it is like to be a computer", and that this fact is sufficient to support the idea of functional similarity between human cognition and computational processes (Turing, 1950: 443). Consequently, the question „Does a machine really think?" is, on his operationalist view, misconceived, only behavior observable under appropriate conditions is relevant.

The practical beginnings of AI as a discipline were marked by the work of Allen Newell and Herbert Simon. In 1956, they constructed the first computer program capable of proving theorems from *Principia Mathematica*. The program, *Logic Theorist*, demonstrated that computers can solve symbolically represented problems using procedures that are not merely numerical but structurally analogous to human reasoning. It relied on explicit representations, rules of inference, primarily modus ponens (referred to in their text as the „rule of detachment") (Newell & Simon, 1956: 26), and search mechanisms. This provided the first concrete demonstration that at least some cognitive processes could be formalized and simulated computationally. It also showed that symbolic models of cognition, in which knowledge is represented as sequences of symbols governed by rules, can be empirically testable. This success shifted the field toward more precise inquiries into how humans solve logical problems, while also raising new challenges, most notably, how to account for creativity, intuition, and other complex forms of thought not easily reducible to symbolic operations.

A philosophical articulation of this framework was provided by Hilary Putnam's functionalism in the early 1960s. Putnam developed a model of the human mind inspired by computational systems, in which mental states are defined in terms of their relations to other mental states and their causal– functional roles. His justification for this view relied directly on the Turing machine, which he interpreted as a device with a finite number of internal configurations, each belonging to one of a finite number of states (Putnam, 1960: 364). This gave rise to the thesis of *multiple realizability*, the idea that a single mental state can be realized in different physical substrates, provided it fulfills the same functional role. Functionalism therefore asks what a system does, rather than what material substrate it is made of. For the cognitive revolution, already well underway, this represented a further consolidation of its central commitments. It created a conceptual space in which computational systems could be treated as legitimate candidates for possessing cognitive properties, insofar as they instantiate the appropriate functional relations. Putnam's theory thus liberated the concept of intelligence from biological constraints and allowed artificially constructed systems to be regarded as genuine candidates for intelligence.

The work of Turing, Newell, and Simon, integrated into a functionalist framework, established the foundations of early AI. For the first time, it became methodologically possible to formally describe cognitive functions, which in turn made it possible to conceptualize intelligence as a systematic and algorithmically realizable process. Within this triad, empirical method, symbolic representation, and functional analysis, the first stable paradigm of AI emerged.

5. The Importance of Linguistics and Representational Models

One of Chomsky's contributions, as we have already seen, lay in his critique of behaviorism. This critique redirected psychology toward internal structures. However, his main contribution to the cognitive revolution and its later offshoots, among which AI is included, lies in the formalization of knowledge about language as an abstract, representational system. Naturally, today our first association is his generative grammar, which offered one of the first precisely specified models of a mental capacity. It is a system of rules that, in a clearly defined manner, assigns structural descriptions to sentences. According to Chomsky, every speaker of a language has internalized generative grammar, which expresses their knowledge of that language, even though they are not consciously aware of that knowledge (Chomsky, 1965: 8).

This model fits seamlessly within the functionalist framework because linguistic competence, as the internal knowledge of the linguistic system, is something that is measured exclusively through the causal-functional relations between the constituent elements of linguistic competence. Linguistic performance, in the sense of the practical, actual use of language, is a completely separate matter from linguistic competence. Chomsky insists on this distinction because once it is accepted, language is no longer viewed as a set of habits but as an innate cognitive capacity of the human being that is realized only through use. The very fact that humans possess

innate mechanisms enabling them to understand an almost infinite number of sentences from a limited set of elements shows that the structures Chomsky describes can be viewed as a way of gaining insight into the structure of human cognition.

The most important contribution of this approach to the development of AI lies in the fact that generative grammar treated language as a system of formal representations. These formal representations can be subjected to algorithmic analysis, and therefore syntax and grammar became suitable candidates to serve as the first carriers of internal mental representations. Such a paradigm made it possible for natural language to be trained, modeled, and processed within early computational systems, which directly led to the first programs for parsing sentence structure and to symbolic systems for natural language processing (NLP).

Thus, the importance of linguistics is not merely historical. The ideas of generative grammar have continued, and still continue, to shape the development of language-processing systems. Its influence is visible even in statistical approaches, in which programs use information about grammar as a basis for sentence analysis, as well as in contemporary models based on neural networks. These models, although they do not use explicitly defined grammatical rules, nevertheless organize linguistic patterns and relations among words during training. These patterns and relations align with what Chomsky referred to as generative grammar.

6. Limitations of Classical AI

Thus, the development of early, symbolically based AI demonstrated that it was possible to formalize certain cognitive functions and simulate them through procedures structurally similar to human reasoning. However, this success simultaneously marked the boundary of limitations whose critical re-evaluation gave rise to new approaches. The most significant critiques were not directed at the technical aspects of symbolic models, but at a major, overarching conceptual difficulty. A system that possesses only symbols, along with rules for manipulating them, has no way to „understand“ the framework within which it already operates.

This issue was forcefully articulated by John Searle’s Chinese room argument, which provided the first precise articulation of the symbol-grounding problem – structures alone are insufficient to provide themselves with meaning. It is possible to achieve formal success in manipulating symbols without any grounding that would explain the relation between the symbols and what we call the „world“. This means that a language model that successfully manipulates symbols, transformations, and rules is neither a guarantee nor an explanation of its actual competence, unless it is clear how those representations are anchored to perceptual or action streams (Searle, 1980: 418–419). Searle’s thought experiment shows that it is possible for a system, human or machine, inside a room, by manipulating symbols according to rules, to provide correct answers in Chinese while having no understanding of the meaning of those

symbols. The symbols are only formally connected, but are not „grounded“ in real meaning or experience.

As a supplement to the formulation of this problem, connectionist models emerge, emphasizing autonomous learning from data and the dynamic character of cognitive processes. Discrete symbols are no longer central, instead, patterns of activation within a network have become the central focus. Predefined rules no longer exist, rather, learning and adaptation through experience have been introduced. This dynamic character offers elements that the symbolic approach lacked – a concrete connection to data, autonomous learning, and the gradual formation of meaning. Thus, it became clear that representations must be not only formally specifiable but also adaptive, context-sensitive, and formed through interaction with data. Symbolic models simply could not offer this.

Therefore, the limitations of classical AI do not represent the end of a paradigm, but rather a transitional point at which it becomes clear that intelligence requires more than symbol manipulation. It is also a point at which all previous achievements of the cognitive revolution culminate – the initial critique of behaviorism, continued through formal models and functionalism, ultimately led to the insight that the demands of learning and dynamism cannot be ignored when discussing intelligence. This reveals the first major limitation of the information metaphor. Its absolutization leads to the metaphysical thesis that intelligence is exclusively symbol manipulation. However, its reinforcement by connectionist insights lays the foundations for further progress, both in scientific research in general and in AI research. This insight marks the beginning of the development of contemporary deep learning models,

which do not have a static symbolic structure but are systems capable of adaptation and autonomous formation of meaning through data. These models constitute the first step beyond the limitations of classical AI.

7. Neural Networks as an Extension of the Classical Cognitive Paradigm

The development of deep neural networks is the most significant shift in the advancement of AI since the emergence of the symbolic paradigm. Contemporary deep learning architectures have further developed the principles of distributed representation toward more complex and dynamic forms of representation, completely abandoning predefined symbolic structures. Parallel distributed processing (PDP) emerges as well. These models use networks of simple units (artificial neurons) that are interconnected and learn through adaptation. The task of a single unit is to receive input data from neighboring units and, as a function of those inputs, calculate an output value that it sends back. All of this operates in parallel because many units within the network can perform computations simultaneously (Rumelhart, Hinton & McClelland, 1986: 47). This illustrates the decentralized nature of processing within connectionist models, and this decentralized nature is precisely the specific difference in relation to symbolic models of AI, where each representation is explicitly specified and processing proceeds sequentially. Thus, knowledge within a model of parallel distributed processing is not stored in the form of rules or instructions but is distributed and scattered across the

weights of the connections between units. This enables the network to adapt to new tasks because the representation is dynamic and constantly distributed (Rumelhart, Hinton & McClelland, 1986: 48). Hence, the very structure of the network and the configuration of weights within it are what determine the system's knowledge, rather than any predefined and fixed set of rules. Because of all this, the system can evolve – what a single unit represents can change with experience, and the system can begin to operate in very different ways (Rumelhart, Hinton & McClelland, 1986: 46).

At that point it becomes clear that connectionist approaches do not emerge as an alternative but as a consistent extension of what the cognitive revolution had led to. The emphasis on internal structures, representations, and information processing remains consistent with the earlier cognitive paradigm. However, these approaches introduce dynamism and learning, elements that early symbolic models had to predetermine and fix in advance.

The development of connectionist approaches also led to the emergence of deep learning, which today forms the basis of most advanced AI systems. Deep learning involves the use of neural networks with multiple layers (deep networks), where each layer processes data at an increasingly higher level of abstraction. This architecture enables models to automatically extract and recognize complex patterns in data, which is crucial for tasks such as image recognition, speech processing, language translation, and many other cognitive functions. One of the most important types of deep neural networks are convolutional neural networks (CNNs), which are particularly efficient in processing images and other

spatial data. The basic principle of CNNs relies on convolution and pooling operations. Convolution enables the network to detect local patterns in data (e.g., edges, shapes), while pooling reduces data dimensionality and extracts the most important features, making the network more robust to variations and noise in the input data. This organization enables the model to gradually construct increasingly complex representations, from simple to highly abstract (Buckner, 2024: 205–210).

This gradual hierarchy of abstraction parallels classical models of linguistic structure. While Chomsky formally described the structures of language, deep networks discover these structures on their own during learning. In doing so, they functionally achieve something that classical linguistics regarded as a necessary property of mental representations, but achieve it as a consequence of processing rather than through a predefined grammar. This further reinforces Chomsky's position and once again reveals his role in the overall development of the cognitive revolution and, as we see, AI as a separate discipline.

In addition to CNNs, other architectures also play a significant role, such as recurrent neural networks (RNNs), which are suited for processing sequential data (e.g., text, speech, time series), as well as transformer models, which are now dominant in natural language processing. Transformer models are perhaps the most important example of how the ideas of the cognitive revolution are transformed into algorithmic systems. The attention mechanism enables the model to evaluate the relevance of input data at each step and to organize representations depending on context. Such models reconstruct the idea of mental structures in the form of vector relationships within latent spaces.

Within deep learning, generative models have also been developed, such as generative adversarial networks (GANs). GANs consist of two networks, a generator, which attempts to produce new data similar to real data, and a discriminator, which attempts to distinguish real from generated data. This competitive process drives the progressive refinement of both networks, resulting in increasingly realistic and sophisticated generated content. The success of the discriminator prompts the generator to further refine itself in order to „fool“ it, while each improvement of the generator requires the discriminator to become even more sensitive to subtle differences between real and generated data. The generative and discriminative networks may or may not share the same latent space, depending on the model's architecture (Buckner, 2024: 204–211).

To understand the above paragraphs, it is necessary to explain the term „latent space“. It represents a multidimensional mathematical space in which the network stores its internal representations of data. During learning, the network learns to map input data into this space such that similar data points cluster together, while different ones are placed farther apart. In this way, the latent space enables the model to recognize, classify, and even generate new data based on learned patterns. In deep neural networks, knowledge is not explicitly encoded in the form of rules or symbols but is distributed across the weights of the connections between units (neurons). This distributed representation allows the model to be flexible, adapt to new tasks, and process complex, irregular data in a manner reminiscent of certain aspects of human cognition.

Therefore, deep learning, viewed in the broader context of the development of cognitive science,

represents the endpoint of a trajectory that begins with the abandonment of behaviorism, progresses through formal models of representation and functionalism, and is then transformed through connectionism. Deep learning is the concretization of the fundamental ideas of the cognitive revolution, ideas that claimed that it is possible to formalize cognitive processes, that the mind is a system of internal representations, and that understanding must be explained through mechanisms of information processing. Contemporary AI systems are therefore theoretical models that continue the tradition initiated by the cognitive revolution.

8. Conclusion

The cognitive revolution demonstrated that understanding the mind is impossible without accepting internal structures, representations, and processes that manipulate information. The fall of behaviorism, Chomsky's critique, the development of computational models of thought, and the formalization of cognitive functions created the theoretical space in which it became possible to formulate the first concepts of machine intelligence. In this sense, AI did not emerge independently of cognitive science but as its intrinsic experimental continuation. The first wave of AI, symbolic and functionalist models, was a direct product of this new conception of intelligence. They presupposed that cognitive functions can be described as the manipulation of representations, that knowledge can be treated as a structural relation among symbols, and that „thinking procedures" are algorithmically executable. This paradigm allowed computers for the first time to

simulate logical and linguistic processes, to establish the notion of an intelligent machine, and to treat the formalization of thought as a fully legitimate goal of scientific inquiry.

At the same time, it was precisely symbolic models that revealed their own limitations. Searle's experiment, the grounding problem, and the inability of symbolic systems to learn autonomously pointed to the fact that formal structures, no matter how precise, are insufficient to explain understanding. This moment also revealed the limitation of the information metaphor itself. If intelligence is merely a sequence of symbol transformations, then there is no explanation of how symbolic representations acquire meaning or how they become connected to the real world.

Connectionism, and later deep learning, emerge precisely as responses to these limitations. In them, representations are no longer predefined but are formed through learning, experience, and relational structures within data. Distributed representations, latent spaces, and dynamic networks introduce elements that symbolic models could never provide. The gradual formation of meaning and the ability of a system to organize structures that are not predetermined meant that these new systems were capable of learning. Therefore, deep learning is the extension of the cognitive revolution in which its concepts are preserved but given a new practical and technical realization. For this reason, it becomes clear that the cognitive revolution is not merely the historical context in which AI emerged but also the intellectual framework that defined what AI can be. It provided the concept of representation, the model of the mind as an information- processing system, and the methodology

that allows complex cognitive processes to be formalized. Yet at the same time, it revealed its own limitations in the absence of dynamism and learning. It left a gap that only connectionism could address and fill.

Contemporary AI models, deep neural networks and transformer models, stand at the intersection of these two legacies. They retain the structural and algorithmic discipline of the classical cognitive paradigm but introduce adaptation, learning, and contextual representations. For this reason, it can be said that today's AI is the furthest extension of the cognitive revolution thus far. It is its most consistent and technically successful realization, but also a correction of its initial limitations. Indeed, the relationship between symbolic structure and dynamic learning determines the theoretical significance of the cognitive revolution for the contemporary notion of AI. It shows that intelligence is neither purely symbolic nor purely statistical, but a hybrid system in which representation, computation, and learning co-determine one another. The cognitive revolution did not provide a final answer to what intelligence is, but it created the conditions for the question to be posed in a scientifically fruitful way. This is likely its greatest and most enduring contribution.

References

1. Buckner, C. J. (2024). *From deep learning to rational machines*. Oxford University Press.
2. Chomsky, N. (1959). A review of B. F. Skinner's *Verbal Behavior*. *Language*, 35(1), 26–58. https://doi.org/10.2307/411334
3. Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.

4. Gardner, H. (1985). *The Mind's New Science: A History of the Cognitive Revolution*. Basic Books.
5. Marr, D. (2010). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* MIT Press. (Original work published 1982)
6. Miller, G. A. (2003). The cognitive revolution: A historical perspective. *Trends in Cognitive Sciences*, 7(3), 141–144. https://doi.org/10.1016/S1364-6613(03)00029-9
7. Newell, A., & Simon, H. A. (1956). The logic theory machine: A complex information processing system. *IRE Transactions on Information Theory*, 2(3), 61–79. https://doi.org/10.1109/TIT.1956.1056810
8. Putnam, H. (1960). Minds and machines. In S. Hook (Ed.), *Dimensions of Mind* (pp. 138–164). New York University Press.
9. Rumelhart, D. E., Hinton, G. E., & McClelland, J. L. (1986). A general framework for parallel distributed processing. In D. E. Rumelhart, J. L. McClelland, & PDP Research Group (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. 1, pp. 45–76). MIT Press.
10. Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417–457. https://doi.org/10.1017/S0140525X00005756
11. Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 2(42), 230–265.
https://doi.org/10.1112/plms/s2-42.1.230

12. Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433
13. Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, 20(2), 158–177. https://doi.org/10.1037/h0074428