Revista EIA

Cognitive Dual-Process Theories Applied to Artificial Intelligence

Revista EIA
ISSN 1794-1237
e-ISSN 2463-0950
Año XIX/ Volumen 22/ Edición N.44
Julio - diciembre 2025
Reia4432 pp. 1-11

Publicación científica semestral
Universidad EIA, Envigado, Colombia

Para citar este artículo / To reference this article /
Sarrazola-Alzate, A.
Cognitive Dual-Process Theories Applied to Artificial Intelligence

Revista EIA, 22(44), Reia4432 pp. 1-11
https://doi.org/10.24050/reia.v22i43.1854

Autor de correspondencia:
Sarrazola-Alzate, A.
Doctor en Matemáticas
Correo electrónico:
andres.sarrazola@eia.edu.co

Recibido: 17-02-2025
Aceptado: 20-06-2025
Disponible online: 01-07-2025

Andrés Sarrazola-Alzate[1]

1. Universidad EIA, Colombia

Abstract

In his 2011 book *Thinking, Fast and Slow*, Daniel Kahneman popularized the dual-process theory of cognition through the distinction between **System 1**—fast, automatic, and associative—and **System 2**—slow, deliberate, and rule-based. His research, which integrates economic theory with cognitive psychology, revealed the pervasiveness of cognitive biases and showed how the interaction between these systems can systematically lead to reasoning errors.

This article explores the applicability of Kahneman's dual-process framework to artificial learning systems, particularly in addressing phenomena such as hallucinations and inference failures in large language models. We examine the cognitive mechanisms involved in idea formation at the neural level and propose two postulates that outline the structural challenges in implementing a **System 2 analogue** in artificial intelligence.

Given that machine learning systems rely on mathematical formalisms, we introduce a simplified mathematical model of cognitive processing. This model suggests that an axiomatic understanding of synaptic behaviour may be crucial to identifying and mitigating systematic reasoning flaws in natural language processing systems.

Keywords: Biological Architectures, Artificial Learning, Artificial Intelligence, Large Language Models, Natural Language Models, decision-making.

Teorías Cognitivas de Procesamiento Dual aplicadas a la Inteligencia Artificial

Resumen

En su *Thinking, Fast and Slow* (2011), Daniel Kahneman popularizó la teoría del procesamiento dual de la cognición, distinguiendo entre el **Sistema 1** (rápido, automático y asociativo) y el **Sistema 2** (lento, deliberado y gobernado por reglas). Sus investigaciones, que integran teoría económica y psicología cognitiva, evidenciaron la presencia generalizada de sesgos cognitivos y mostraron cómo la interacción entre ambos sistemas puede conducir sistemáticamente a errores de razonamiento.

Este artículo examina la pertinencia de adaptar el marco teórico de Kahneman a los sistemas de aprendizaje artificial, con el objetivo de mitigar fenómenos como las alucinaciones o fallos de inferencia en modelos de lenguaje a gran escala. Analizamos los mecanismos cognitivos que intervienen en la formación de ideas a nivel sináptico y proponemos dos postulados que exponen los desafíos estructurales de implementar un **análogo del Sistema 2** en inteligencia artificial.

Dado que los modelos de aprendizaje automático se sustentan en fundamentos matemáticos, presentamos una modelización matemática simplificada del proceso cognitivo. Esta aproximación sugiere que una comprensión axiomática del comportamiento sináptico podría ser clave para identificar y corregir errores sistemáticos en el razonamiento de los sistemas de procesamiento del lenguaje natural.

Palabras clave: Arquitecturas biológicas, Aprendizaje artificial, Modelos de lenguaje de gran tamaño, modelos de lenguaje natural, toma de decisiones.

1. Introduction

In 1908, mathematician Henri Poincaré delivered a lecture on the origin of mathematical creativity. The French mathematician, a supporter of the intuitionist school, emphasized the role of intuition and the unconscious in mathematical discovery. He began by analyzing the intervention of the unconscious—or the subliminal self—rejecting its characterization as merely automatic and instead portraying it as a mechanism endowed with discernment. In other words, just as important as the conscious self.

Secondly, Poincaré revisited the automatic characterization of the unconscious self to explore a combinatorial hypothesis underlying problem-solving. He proposed that the subliminal self might evaluate all possible combinations, allowing only those that are "interesting" or harmonious to emerge into the domain of consciousness. Yet both perspectives face limitations—whether due to the implausibility of the proposed scenario or the philosophical ambiguity they entail.

The ideas developed by the French mathematician allow us to glimpse that invention is a process of selection. This selection takes place at an unconscious stage and is preceded by a period of intense theoretical reflection. Such a stage can only occur through extensive conscious intellectual work, which gradually shifts the search for a solution into a domain whose cognitive activity is imperceptible to our senses—or, if preferred, one of an automatic nature.

Based on this dual nature of mathematical invention, as presented in Henri Poincaré's narrative, we proceed to outline a scenario in which responses to stimuli are generated by two distinct engines, one of them represents fast, associative cognitive operations, while the other encompasses slower, rule-governed processes.

Table 1. *Two Cognitive Systems (Adapted from "Representativeness revisited: Attribute substitution in intuitive judgment". Kahneman D. & Frederick S. 2002)*

System 1 (Intuitive)	System 2 (Reflective)
Automatic	Controlled
Effortless	Effortful
Associative	Deductive
Rapi, Parallel	Slow, serial
Process opaque	Self-Aware
Skilled action	Rule application

The nature of these mechanisms has been the subject of study by numerous authors in the decades that followed. We will now explore the development of these theories alongside their relationship to artificial intelligence systems.

The philosophy of Artificial Intelligence (hereafter AI) lies in the mimicry of cognitive processes that originate in the synapses of the brain. In practice, brain activities are imitated depending on the situation to be addressed and based on diverse problem domains, various tools have been developed that have been widely studied in modernity (Wang, 2019).

To carry out the process described in the previous paragraph, research in the broader field of neuroscience has focused on a deep understanding of the cognitive architectures present in the biology of the brain (Lieto, 2018). As previously mentioned, although the dual-process model of brain functioning has been explored by various authors and labelled in different ways from as early as 1912 through the early 2000s (Chaiken & Trope, 1999), in this paper we will follow the terminology systematized by psychologist Daniel Kahneman. In his influential book "Thinking Fast and Slow" (Kahneman, 2011), D. Kahneman uses the terminology "Systems 1 and 2" originally proposed by Stanovich and West (Stanovich & West, 2000) to explain how our neural configuration processes information and, from it, makes a decision. In this article we will explain how such systems have been of considerable significance in the so-called General Artificial Intelligence and the way in which they are present. To do so, we will start by giving a brief explanation of each mentioned system.

***Table 2**. Content on which Processes Act (Adapted from Kahneman D. & Frederick S. 2002.)*

System 1	System 2
Affective	Neutral
Causal propensities	Statistic
Concrete, specific	Abstract
Prototypes	Sets

Revista EIA https://doi.org/10.24050/reia.v22i44.1890

System 1

Every Saturday Andrés gets up early and takes his car to do the week's shopping. The trip to the market is a fifteen-minute drive that he has made without fail for the past two years. Andrés, without being aware of it, knows where to turn and where the traffic lights, speed bumps and potholes are. Over the years, the Saturday activity has become an automatic process or, equivalently, a habit controlled by the so-called system 1.

System 1 then refers to the section of our brain that generates an automatic and rapid response to a situation or stimulus, which is based on information stored in our memory or governed by intuition. The most salient features of this system are operating automatically and effortlessly, making quick decisions based on experience, facial recognition, driving on a familiar route, and making judgments with a probability of error or heuristics. It is worth noting that system 1 is not free from bias-based errors, or from hallucinations and fallacies that affect unconscious decision-making (Kahneman, 2011).

System 2

Consider the number sequence 1, 2, 4, 5, 8, 1000. What is the next number? Most likely, if you have set yourself the task of answering this question, you will notice how your brain focuses on different arithmetic operations, trying to identify a pattern that allows you to find the general term of the sequence. With each attempt, your ideas will move from simple algebraic hypotheses to more complex calculations or even a lexicographical analysis of the sequence (and in fact the solution, which is 1001, results in identifying that all the numbers in the sequence share the characteristic of not having the letter e in their spelling). Your brain has identified that a fast, automatic process will not be able to solve the enigma and has gone into a slow, high-effort, rational mode, i.e. you have unconsciously called the "heavy cavalry" of system 2 in search of technical help for a problem that is claimed to be highly complex.

We will thus allude to system 2 by referring to the brain region in charge of complex processes, which involve analytical mechanisms that require time, patience and prior analysis before finding the expected response. Among the most notorious characteristics of this system are the use of critical thinking to solve highly complex problems, its functioning is executed in a conscious state and, in general, when it is executed, it follows an automatic response of system 1 which do not produce the expected result. Furthermore, its capacity is limited, and its prolonged use generates the sensation of tiredness, in other words, we can affirm that there is a relationship between the execution time and the performance of the system.

In the following sections, we will see how large language models (LLM) are artificial architectures that reflect the automatic behaviour of system 1, and how research has focused on reproducing an artificial analogue of system 2 that will act as a regulatory agent, avoiding the presence of hallucinations (Sarrazola-Alzate, 2023).

The System 1 and Large Language Models

Originally designed for text generation, large language models (LLMs) such as ChatGPT (Wu, et al., 2023), PalM (Anil, et al., 2023) and Perplexity (Perplexity.ai, 2022), have demonstrated remarkable versatility in addressing a wide variety of problems including logical reasoning, arithmetic and geometric problem-solving, and even the derivation of new theorems within Whitehead and Russell's propositional calculus (Whitehead & Russell, 1927). In light of these achievements, the pursuit of a hypothetical General Artificial Intelligence (Fjelland, 2020) has reignited debates about the actual limits of LLMs: which problems lie beyond their reach, and what alternative cognitive mechanisms might compensate for those limitations. Among the most promising of these alternatives are heuristics—non-algorithmic strategies grounded in prior experience, pattern recognition, and intuition.

According to Hungarian mathematician George Pólya (1945), heuristics are rooted in the experience accumulated during the direct or indirect resolution of problems (Pólya, 1945). This cognitive approach has been formalized using topological structures known as acyclic connected graphs—or simply, trees—which offer a geometrical representation of the mental pathways explored during problem-solving (Jungnickel, 2005). In their seminal works (Newell & Simon, Report on a General Problem-Solving Program, 1959) and

(Newell & Simon, Human Problem Solving, 1972) Newell and Simon proposed that the process of forming an idea can be mapped within such a space: nodes (or vertices) represent partial solutions, a set that we will denote by $\mathcal{PS}\sigma\ell$, and the branches are in one-to-one correspondence with operators between vertices of $\mathcal{PS}\sigma\ell$; in other words, a branch b is a function

$$b : \mathcal{PS}\sigma\ell \to \mathcal{PS}\sigma\ell$$

that satisfies combinatorial properties compatible with the acyclicity of the graph. Every single idea is represented by a sequence of branches $\mathcal{R}_{ik} = [b_1, \ldots, b_k]$ with $b_i \neq b_j$ for all $1 \leq i \neq j \leq k.$. The selection of branches and the mental shift from one to another are governed by heuristics, closely aligned with what Daniel Kahneman (2011) refers to as **System 1**-a fast, intuitive, and context-sensitive mode of cognition.1. This formalized structure exposes two key limitations of current LLMs when engaged in goal-oriented tasks (Yao, et al., 2024):

F.1- *Locally*, there is no axiomatic method by which LLMs infer valid relationships between branches in the problem space.

F.2- *Globally*, these models lack inherent mechanisms for planning, anticipation, or feedback—processes that emerge naturally and continuously within the heuristics of System 1, and that remain, for now, uniquely human.

Bearing in mind the fallacies described in the previous paragraphs, and in the interest of enabling systems to exert control over their own responses, this investigation suggests the implementation of **verification processes** within chains of thought, in order to preserve the robustness and reliability of the resulting assertions. In other words, it calls for the emulation of an artificial system capable of acting as a **control agent—**a mechanism akin to System 2, responsible for oversight, evaluation, and correction.

The System 2 and Control Models

To establish how the activities described in the previous paragraph are currently being replicated in LLM, let us consider $\mathscr{P}$ θ to be a pre-trained language model with set of parameters θ. We will use lower case letters x,y ,z,s… to denote language sequences x=(x[1],…,x[k]), where x[j] is a token. In other words, To establish how the activities described in the previous paragraph are currently being replicated in LLM, let us consider to be a pre-trained language model with set of parameters . We will use lower case letters to denote language sequences , where is a token. In other words,

$$\mathfrak{p}_\theta(x) = \prod_{i=1}^{k} \mathfrak{p}_\theta(x[i] \mid x[1], \ldots x[i-1]).$$

In (Wei, et al., 2022) the authors introduced a *Chain of Thoughts* $z_1,\ldots,z_n$ to send an input *x* to an output *y*, where each z_i symbolizes a coherent sequence of language, which serves as an intermediate step to solve the problem at hand. In practice

$$[x, z_1, \ldots, z_n, y] \sim p_\theta{}^{\mathrm{CoT}}\left(z_1, \ldots, z_n, y \mid x\right)$$

where $\mathscr{P}\theta^{CoT}$ refers to the probability determined by the chain of thoughts (CoT).

The previous technique, called “Chain-of-Thoughts Prompting”, is based on a chain of reasoning that decomposes the problem into a series of intermediate steps that are easier to evaluate, which, in turn, add additional computations to the interface. This sequence of reasoning allows a detailed look at the behaviour of the algorithm, suggesting how the model arrives at a solution and giving the opportunity to correct as appropriate (the underlying architecture that leads to specific outputs remains an open question(Hagendorff & Wezel, 2020)).

The weakness of “Chain-of-Thoughts Prompting” lies not only in the computational complexity, but also in the fact that sequences of reasoning in the form “let’s see it step by step”, prompt known as *Zero-Shot-CoT*, when confronted

Revista EIA https://doi.org/10.24050/reia.v22i44.1890

without examples make large language models tend to exhibit hallucinations that affect the effectiveness of the deductions. As mentioned above, this pathology arises due to the **F.1** and **F.2** deficiencies described in the previous section.

The first idea to remedy the presence of hallucinations consisted of a manual intervention in the arguments outlined by the programs, however, this human-machine interaction would imply a considerable effort on the part of both agents involved, which makes this initiative an unfeasible solution both on the human and computational side.

To optimize both computational and cognitive resources, modern research has focused on developing artificial analogues of system 2, which are intended to act as controlling agents to refine the chain of thoughts generated by $\mathscr{P}\theta^{CoT}$. To carry out this process, different techniques have been implemented, such as “Auto-Chain-of-Thought” (Auto-CoT) (Zhang, Zhang, Li, & Smola, 2022), “Pattern-Aware Chain-of-Thought” (PA-CoT) (Zhang, Wang, Wu, & Wang, 2024) and “Tree-of-Thought” (ToT) (Yao, et al., 2024). Each of the above methods allows LLMs to make decisions by considering multiple branches to decide on the next course of action, while maintaining a process of verification and feedback, enabling an overall view of each choice or branch . In other words, the above procedures are early prototypes of artificial systems 2.

2. Conclusions

Artificial intelligence has evolved from a paradigm rooted in the emulation of neural processes to a more ambitious attempt to reconstruct the epistemological structures of the self. This shift, though conceptually fertile, brings with it a host of challenges that remain largely unaddressed by current mathematical and computational frameworks.

As this paper has shown, AI systems continue to rely on formal mechanisms that reproduce certain cognitive behaviours, yet they lack the reflective control and evaluative capacity found in human reasoning. These limitations manifest most clearly in phenomena such as hallucinations, faulty deductions, and the absence of long-term coherence in decision-making.

To address this gap, future research must prioritize the development of artificial analogues of **System 2**—deliberative, rule-based architectures capable of overseeing and refining the fast, intuitive outputs of **System 1** processes. Such models would not merely add layers of computation but would introduce the capacity for **meta-cognition**: the ability to monitor, assess, and revise one's own inferences.

In short, the next frontier in artificial intelligence does not lie in scaling data or parameters alone, but in cultivating within machines the capacity to question themselves—a hallmark of human thought, and a prerequisite for true artificial reasoning.

3. Bibliography

Anil, R., Andrew, M. D., Orhan, F., Melvin, J., Dmitry, L., Alexandre, P., . . . al., e. (2023). Palm 2 technical report. *arXiv preprint arXiv:2305.10403*.

Chaiken, S., & Trope, Y. (1999). Dual-Process Theories in Social Psychology. Guilford Press.

Chowdhary, K. (2020). Fundamentals of artificial intelligence. *Springer*.

Fjelland, R. (2020). Why general artificial intelligence will not be realized. *Humanities and Social Sciences Communications*, 1-9.

Google AI. (n.d.). Retrieved from Google AI: https://ai.google/discover/palm2/

Hagendorff, T., & Wezel, K. (2020). 15 challenges for AI: or what AI (currently) can't do. *Ai & Society*, 355--365.

Jungnickel, D. (2005). Graphs, networks and algorithms. *Springer*.

Kahneman, D. (2011). *Thinking, fast and slow.* New York: Farrar, Straus and Giroux.

Kaplan, J. a., Gray, S., Radford, A., Wu, J., & Amodei, D. (2020). Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*.

Newell, A., & Simon, H. A. (1959). Report on a General Problem-Solving Program. *IFIP Congress*, 31.

Newell, A., & Simon, H. A. (1972). *Human Problem Solving.* Englewood Cliffs, NJ: Prentice-hall.

OpenAI. (2023). *GPT-4 Technical Report.* https://arxiv.org/abs/2303.08774.

Perplexity.ai. (2022). *Perplexity*. Retrieved from Perplexity: https://www.perplexity.ai/

Pólya, G. (1945). *How to solve it.* Princeton University Press.

Sarrazola-Alzate, A. (2023). Problemas estructurales en sistemas inteligentes. Una aproximación desde el programa ChatGPT. *Repositorio Institucional, Fondo Editorial EIA*.

Sel, B., Al-Tawaha, A., Khattar, V., Wang, L., Jia, R., & Jin, M. (2023). Algorithm of thoughts: Enhancing exploration of ideas in large language models. *arXiv preprint arXiv:2308.10379*.

Revista EIA | https://doi.org/10.24050/reia.v22i44.1890

Shunyu, Y., Yu, D., Zhao, J., Izhak, S., Griffiths, T. C., & Narasimhan, K. (2024). Tree of thoughts: Deliberate problem solving with large language models. *Advances in Neural Information Processing Systems*.

Stanovich, K., & West, R. F. (2000). Individual difference in reasoning: implications for the rationality debate? *Behavioral and Brain Sciences*, 645-726.

Wang, P. (2019). On definig artificial inteligence. *Journal of Artificial General Intelligence*, 1-37.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., . . . others. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 24824-24837.

Whitehead, A. N., & Russell, B. (1927). Principia mathematica. *Cambridge University Press*.

Wu, T., Shizhu, H., Jingping, L., Sun, S., Kang, L., Qing-Long, H., & Yang, T. (2023). A brief overview of ChatGPT: The history, status quo and potential future development. *Journal of Automatica Sinica*, 1122-1136.

Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T., Cao, Y., & Narasimhan, K. (2024). Tree of thoughts: Deliberate problem solving with large language models. *Advances in Neural Information Processing Systems*.

Zhang, Y., Wang, X., Wu, L., & Wang, J. (2024). Pattern-Aware Chain-of-Thought Prompting in Large Language Models. *arXiv preprint arXiv:2404.14812*.

Zhang, Z., Zhang, A., Li, M., & Smola, A. (2022). Automatic chain of thought prompting in large language models. *arXiv preprint arXiv:2210.03493*.

Zhao, Wayne, X., Kun, Z., Junyi, L., Tianyi, T., Xiaolei, W., . . . al., e. (2023). A survey of large language models. *arXiv preprint arXiv:2303.18223*.