Before AI colonizes our minds: Cognitive autonomy via System 1/2

Grzegorz Jamróz

November 9, 2026

Abstract

In this short note we argue that the threat of AI cognitive colonization highlighted by Ganapini et al. (2026) arXiv:2606.13658 can be mitigated by introducing an overlay, called System 1/2, which modulates the AI filtered information reaching humans in order to provide us with *cognitive autonomy*, which we define. We discuss examples of cognitive outcomes of AI interaction with humans with and without System 1/2 and call for the development of open-source tools to moderate the impact of AI systems at the deeply cognitive level.

Keywords: System 1/2, Cognitive autonomy, System 0, AI-human interaction

1 Introduction

Human cognition can be roughly decomposed into two systems, [11, 13]. System 1 is intuitive, fast and effortless. System 2, on the other hand, is deliberate, slow and requires conscious effort. In human cognition, System 1 engages first and produces quick judgments/decisions which however may be inaccurate or biased. System 2 is used to either override System 1 on certain tasks (e.g. withholding one's impulsive reaction or correcting optical illusions) or where System 1 fails or is inadequate (e.g. solving puzzles). An important caveat, recognized by the pioneers of the two-system theory is the assumption that all cognition happens inside a human brain. While this assumption is plausible in the traditional non-digital world, it seems to be incorrect in the contemporary AI-infused reality. Indeed, using automated navigation services, choosing travel itinararies with the help of chatbots and being fed user-profiled news from social media are but a few examples where human cognition is either enhanced or replaced by external systems. Depending on where this external component operates it has been called System 3 [10], if it interacts with Systems 1 and 2, System 0 [4], if it preprocesses information for Systems 1 and 2, still interacting with them down the line or Thinkframes ([1, 2]) if it enhances collective reasoning. The research mentioned, along with e.g. [5, 6, 7, 8, 9] has highlighted several concerning cognitive phenomena related to adoption of AI tools in our everyday lives. These include (at a higher level) cognitive surrender/cognitive offloading [10], homogenization of ideas [6] and cognitive colonization [8].

The authors of [8] argue that the AI working as System 0 is closest to how AI actually operates. Recognizing this contribution we propose a way out of the colonization threat by postulating the development of another system, which we call 'System 1/2', which directly engages

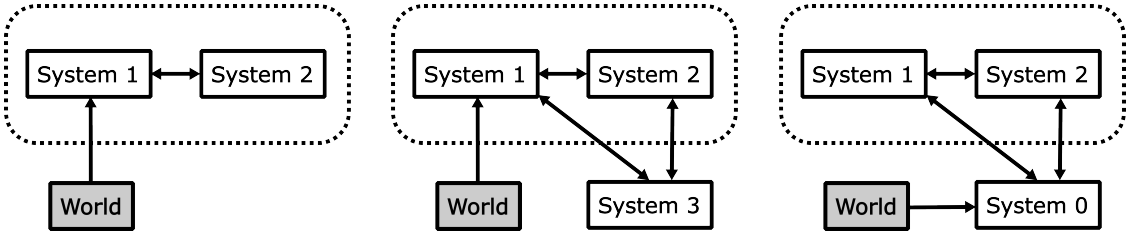


Figure 1: The classical two-system theory of cognition (left). The tri-system AI enhanced theory of cognition (middle). The tri-system AI pre-filtered theory of cognition(right). System 1 (fast, intuitive) and System 2 (slow, deliberate) reside in the human brain. System 3 is an AI system which can be consulted or work in autopilot mode. System 0 is an AI system which preprocesses cognitive information, reshaping human choice/judgment space.

with the users, giving them the opportunity to fully recognize and accept cognitive alterations and risks, which may be incurred by AI use and which are obscured by AI tool developers. Accordingly, we call for the development of open-source System 1/2 which would benefit from current and future advancements in behavioural research by fine-tuning the relevant cognitive cues System 1/2 would offer.

In Section 2 we introduce System 1/2 and show how to apply it to achieve cognitive autonomy. Section 3 discusses real world examples while Section 4 summarizes the proposed ideas.

2 System 1/2

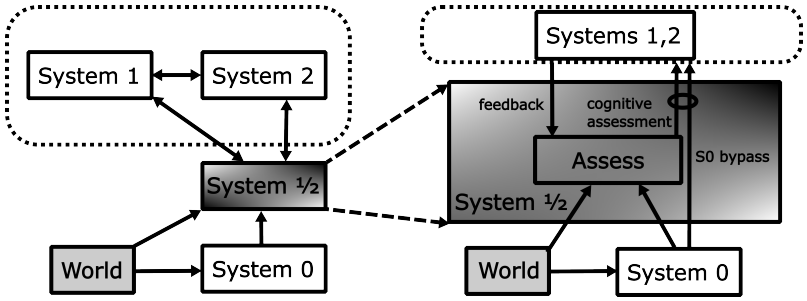


Figure 2: System 1/2 modulates the operation of System 0 (left panel), allowing the brain to deliberately choose the structure of cognition it desires, resulting in *cognitive autonomy*. It succeeds in the task by (right panel) comparing the world information with System 0 information and producing a cognitive assessment of the expected interaction of System 0 with Systems 1,2. Systems 1,2 receive bundled System 0 and cognitive assessment input which allows them to decide whether and how to process System 0 input. The feedback allows System 1/2 to adapt the cognitive assessment while Systems 1 and 2 learn how to obtain a desired cognitive structure.

In this section we introduce System 1/2, see Fig. 2, which moderates System 0's output and dynamical interaction with Systems 1 and 2 in order to provide the user with *cognitive autonomy*, which we define below. We characterize it computationally, psychologically and epistemically by directly relating it to characteristics of System 0 discussed in Table 1 in [8].

Definition 2.1. Cognitive autonomy *is a fully integrated individual awareness and control of cognitive processes that are used, in particular the amount and consequences of delegation of aspects of thinking to external entitites.*

2.1 Computational characterization

System 0 (reproduced from [8]) and System ½ are computationally characterized in Table 2.

Table 1: Computational characterization of System 0 (AI) and System ½ (AI modulator)

Core Claim (System 0)	System 0 consists of AI systems, data-driven processes that operate on user-specific behavioral data, pre-structure informational inputs before deliberation, and adapt continuously through feedback loops and personalization ([4]).
Related properties (System 0)	Anticipatory Personalization: the system continuously learns from the individual user's behavior and actively predicts and shapes future cognitive states rather than merely responding to explicit input ([3]).
Empirical marker (System 0)	Behavioral data ingestion; continuous model updating; individually calibrated outputs; pre-query content generation.
System ½ (this paper's contribution):	System ½ does not interfere directly with the operations of Anticipatory Personalization. Instead, it controls and anticipates AI systems' operations in order to extract the characteristics of anticipatory personalization that are being used. It relays the relevant information allowing the user to actively and consciously decide how to use a given AI system.

2.2 Psychological characterization

System 0 (reproduced from [8]) and System ½ are psychologically characterized in Table 2.

Table 2: Psychological characterization of System 0 (AI) and System ½ (AI modulator)

Core Claim (System 0)	AI meets Heersmink-style multidimensional integration (habitual reliance, trust, individualization, bidirectional information flow) and becomes part of the agent's cognitive architecture rather than an external tool consulted at arm's length. Integration includes a phenomenological dimension: AI outputs are experienced as the agent's own rather than as consulted external inputs.
Related Properties (System 0)	Adaptive Invisibility: the more sophisticated these systems become, the more seamlessly they blend into cognitive workflows, making the mediating role increasingly difficult to detect ([12]).
Empirical Marker (System 0)	Dependence producing deficits on removal ([5]) Phenomenological markers: difficulty distinguishing AI suggestions from own phrasing ([9]); Ghostwriter Effect ([7]); route treated as "the way" rather than as recommendation.
System ½ (this paper's contribution)	System ½ is integrated along the AI based System 0. This means that users experience the cognitive choice of its own architecture as the integral part of their own cognition. The user has a full control of the cognitive structure he/she uses. The Adaptive Invisibility is achieved by AI systems only if users wish it, otherwise AI component is noticed. There is no dependence deficit on removal as users are always aware that it would happen when AI system is removed. The ghostwriter effect is fully recognized at every stage by users. Users are fully aware and reminded (if they wish) of the fact that recommendations and not ground truths are provided.

2.3 Epistemic characterization

In Table 3 we present the epistemic characterization of System 0 (reproduced from [8]) and the corresponding characterization of System 1/2.

Table 3: Epistemic characterization of System 0 (AI) and System 1/2 (AI modulator)

Core Claim (System 0)	The cognitive consequence of the computational pre-structuring: AI has already filtered, ranked, and framed the informational environment before Systems 1 and 2 engage. The epistemic space is shaped prior to deliberation rather than in response to queries the agent has already formulated.
Related Properties (System 0)	Automation of Relevance Judgment: decisions about what deserves attention are made algorithmically before the user formulates a question
Empirical Marker (System 0)	Ranked/curated/framed content encountered before any query; algorithmic feeds; AI overviews; recommendation systems that operate in the absence of explicit user requests.
System 1/2 (this paper's contribution)	The epistemic space is continuously enhanced by the choice which systems to use and awareness of consequences. The users are constantly reminded that only a fraction of reality and potential choices are presented. This fact can be fully internalized as the users learn potential consequences of various AI operating schemata. Users are reminded that AI systems offer pre-judged data and have the opportunity to use alternative sources of information. Automated AI feeds are described as typically biased to the tastes of a user and various alternative views exploration is invariably offered.

3 System 1/2 in practice

In this section we describe examples of how System 1/2 could be implemented in practice, see Table 4. Crucially, the presented examples are relatively straightforward or of moderate complication. Nevertheless, we expect the cognitive impact of AI technologies to become more intricate and the full scope of necessary future research is not clear as of now.

Conjecture 3.1. *We conjecture that it is possible to develop a System 1/2 resulting in full cognitive autonomy, allowing the user to consciously choose their preferred cognitive structure (which will be user-dependent).*

The tasks discussed in Table 4 seem to have relatively simple solutions which however require fine-tuning via experiments. More involved cognitive phenomena emerging as AI-based System 0 takes up larger and larger portions of our cognitive lives require further investigation.

4 Discussion

We introduced the idea of System 1/2 which can actively moderate the AI systems' interaction with brain-based cognitive Systems 1 and 2. It enhances the user AI experience by making it fully conscious (*cognitively transparent*), and in particular, allowing the user to make a fully-informed choice which parts of cognition can be be 'surrendered' to AI and which the user

Table 4: Examples of cognitive challenges facing humans resulting from AI use and proposed solutions via System 1/2

Cognitive Challenge	**System 0 role**	**Solution via System 1/2**
News feeds in social networks are biased in order to maximize engagement or include sponsored content.	Based on previous passive and active usage, System 0 makes a profile of the user and recommends news in order to satisfy its internal objective (e.g. income from ads).	News feeds are augmented via the System 1/2 overlay with a warning that they are potentially biased which could result in artificial augmentation of importance of one's information circle (bubble) and disregarding the role and importance of alternative views. The user being fed this information every time internalizes the fact that only a fraction of reality is presented.
Itinerary choice via chatbot use leading tourists into dangerous places and recommeding reckless actions (see [14])	System 0 is deficient and its recommendations do not follow the safety rules and or the technical level of users.	System 1/2 identifies the potential vulnerabilities of System 0 and reminds the user constantly that the safety is their own responsibility and recommendations may be inadequate. It also reminds the users that their planning skills may decline.
AI assistance when writing resulting in difficulty distinguishing AI sugestions from own phrasing [9]	Recommendations are accurate and the user has little incentive to take conscious effort to phrase their own sentences. Using recommendations is more efficient.	The user is constantly reminded that long-term use of phrase recommendations is likely to deteriorate their ability to write. The user may modify the intensity of recommendation use (e.g. only in specific situtatoins) in order to preserve their cognitive abilities while streamlining their work.
Ghostwriter effect where users consider themselves authors while not owning the text [7]	The contribution of System 0 is cognitively unclear and hard to grasp for humans.	Reminding the user that their contribution to text creation is perhaps only the AI prompt and they are not authors of a given text. Letting the users consciously decide whether they are happy with this situation.
Usage of chatbots in learning	The presented information may be inaccurate. The chosen learning strategy may be inefficient.	Reminding the users that the presented information can be biased or hallucinated. Reminding the users about the efficiency of different ways of learning (e.g. superficial/deep understanding).
Health recommendations based on pulse, temperature, skin conductance, pupil dilation and other physiological readings	System 0 recommends actions which the user is supposed to undertake on their own (e.g. resting, moving to a warmer place, calming down when angry etc.)	Reminding the user of the potential threat of deteriorated auto-regulation ability and allowing the user to adjust which physio-readings and recommendations are truly advantageous long-term

would prefer to preserve. The implementation of such an open-source system is a challenge which will rely not only on technological prowess but also advancements in cognitive science, which will have to address the emerging cognitive phenomena related to AI use, not only as cognitive structure modifications facing adults but also perhaps AI-from-birth cognitively integrated structures.

References

[1] Branda, F. (2026). *When artificial intelligence shapes the way we think.* Philosophy & Technology, 39(1), 42.

[2] Cassinadri, G. (2026) *Thinkframes as AI-mediated cognitive ecologies.* Philosophy & Technology, 39(38), 1-5.

[3] Chen, M. Y., Lughofer, E., Rubio, J. D. J., and Wu, Y. J. (2021). Anticipatory Systems: Humans Meet Artificial Intelligence. Frontiers in psychology, 12, 721879.

[4] Chiriatti, M., Ganapini, M., Panai, E., Ubiali, M., and Riva, G. (2024). *The case for human-AI interaction as system 0 thinking.* Nature Human Behaviour, 8(10), 1829-1830.

[5] Dahmani, L., and Bohbot, V. D. (2020). *Habitual use of GPS negatively impacts spatial memory during self-guided navigation.* Scientific reports, 10(1), 6310.

[6] Doshi, A. R., and Hauser, O. P. (2024). *Generative AI enhances individual creativity but reduces collective diversity.* Science Advances, 10(28), eadn5290.

[7] Draxler, F., Werner, A., Lehmann, F., Hoppe, M., Schmidt, A., Buschek, D., and Welsch, R. (2024). *The AI ghostwriter effect: When users do not perceive ownership of AI-generated text but self-declare as authors.* ACM Transactions on Computer-Human Interaction, 31(2), 1-40.

[8] Ganapini, M. B., Chiriatti, M., Panai, E., and Riva, G. (2026). *Before You Think: System 0, AI-Mediated Cognition and Cognitive Colonization.* arXiv preprint arXiv:2606.13658.

[9] Hohenstein, J., Kizilcec, R. F., DiFranzo, D., Aghajari, Z., Mieczkowski, H., Levy, K., ... and Jung, M. F. (2023). *Artificial intelligence in communication impacts language and social relationships.* Scientific reports, 13(1), 5487.

[10] Shaw, S. D., and Nave, G. (2026). *Thinking fast, slow, and artificial: How AI is reshaping human reasoning and the rise of cognitive surrender.* Available at SSRN 6097646.

[11] Stanovich, K. E., and West, R. F. (2000). 24. *Individual differences in reasoning: implications for the rationality debate.* Behav. Brain Sci, 23, 645-665.

[12] Susser, D. (2019, January). *Invisible influence: Artificial intelligence and the ethics of adaptive choice architectures.* In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (pp. 403-408).

[13] Kahneman, D. (2011). *Thinking, fast and slow.* Farrar, Straus and Giroux.

[14] `https://www.wired.com/story/chatbots-are-leading-hikers-astray-with-dangerous-consequences/` Accessed: 2 Nov, 2026.