Human Interaction and Emerging Technologies (IHIET-FS 2025), Vol. 196, 2025, 198–205
https://doi.org/10.54941/ahfe1005968

SmarAI: Enhancing Scene Understanding by Combining Different AI Technologies

Adnan Agbaria and Yael Dubinsky

Department of Software Engineering, Kinneret Academic College, Israel

ABSTRACT

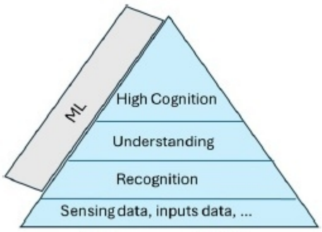
This paper introduces SmartAI, a novel framework that integrates Machine Learning (ML) and Knowledge Representation and Reasoning (KRR) to enhance AI capabilities in reasoning and adaptability. Inspired by Daniel Kahneman's Thinking, Fast and Slow theory, SmartAI leverages ML for rapid, intuitive processing (System 1) and KRR for deliberate, analytical reasoning (System 2). The framework emphasizes modularity, enabling seamless orchestration of these technologies without altering their core components. A case study on scene understanding demonstrates SmartAI's effectiveness in combining fast pattern recognition with in-depth contextual reasoning, achieving superior interpretive outcomes. Beyond scene understanding, SmartAI lays the foundation for context-aware AI applications in diverse fields such as healthcare, education, and autonomous systems. This work sets a precedent for integrating specialized AI technologies to achieve human-like cognitive flexibility. However, it introduces new challenges in effectively managing and orchestrating interactions between these complementary technologies, opening avenues for future research.

Keywords: Machine learning, Knowledge representation and reasoning, Scene understanding, Semantic networks

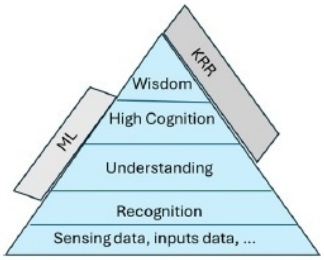
INTRODUCTION

Artificial Intelligence (AI) has experienced unprecedented advancements in recent years, primarily driven by innovations in machine learning (ML) and knowledge-based technologies (Booch et al., 2021, Chowdhury et al., 2023). These breakthroughs have enabled AI systems to achieve remarkable outcomes in narrowly defined tasks, excelling in specific domains with superhuman precision. However, the journey toward creating flexible, human-like AI remains incomplete. Current AI technologies often struggle with adaptability and contextual reasoning, limiting their ability to tackle more complex, dynamic problems. To address these challenges, integrating diverse AI technologies has become an essential avenue for advancing the field (Fahlman 2012).

Generative AI models, such as GPT, have pushed the boundaries of ML by showcasing exceptional capabilities in natural language generation, rapid decision-making, and pattern recognition (Goyal et al., 2020). Figure 1(a) highlights the features achievable by ML models, organized by complexity, effort, and sequence. Despite substantial progress, achieving higher cognitive functions like “wisdom” necessitates incorporating Knowledge Representation and Reasoning (KRR) (Brachman et al., 2004, Donadello et al., 2017). Figure 1(b) illustrates wisdom at the apex of AI capabilities, underscoring its reliance on structured reasoning, context-awareness, and logical decision-making.



(a) Features with ML



(b) For wisdom, KRR mostly required

Figure 1: Listed features that can be achieved using AI.

This paper introduces SmartAI, a dual-system framework inspired by Daniel Kahneman’s theory of Thinking, Fast and Slow (Kahneman 2011). By integrating the rapid, intuitive processing of ML (System 1) with the deliberate, analytical reasoning provided by KRR (System 2), SmartAI aims to bridge the existing gap between AI capabilities and human cognitive flexibility.

In related work, Chain-of-Thought (CoT) prompting has emerged as a promising strategy for enhancing the reasoning capabilities of large language models (LLMs) (Zihan et al., 2023). CoT prompting aligns closely with System 2 thinking, emphasizing step-by-step, structured reasoning to enable logical progression and traceability. By incorporating intermediate reasoning steps, this method mirrors the deliberate and analytical nature of human cognition. Research frameworks also suggest that embedding systematic search processes, process supervision, and reinforcement learning within LLMs can improve their ability to handle complex reasoning tasks (Violet et al., 2025). These advancements provide foundational insights into how SmartAI combines CoT-inspired reasoning approaches with the structured logic of KRR systems (Akkaladevi et al., 2021), achieving a balanced and complementary integration.

SmartAI is particularly well-suited for scene understanding - a field of computer vision focused on enabling machines to interpret and comprehend visual scenes in a human-like manner (Zhou et al., 2018). Scenes often involve intricate relationships between objects, spatial contexts, and activities, requiring AI systems to extract meaningful insights from images or video sequences. While traditional ML models excel at recognizing individual objects, they struggle with the deeper reasoning needed to interpret context, relationships, and underlying intentions. SmartAI addresses these limitations by leveraging ML for fast recognition and KRR for in-depth, context-aware reasoning, ensuring both efficiency and accuracy in scene interpretation.

Beyond scene understanding, SmartAI represents a paradigm shift in AI development, advocating for the integration of specialized technologies rather than isolated solutions. Its architecture reflects a novel approach where ML and KRR components are treated as modular, independent systems, orchestrated seamlessly without internal modifications. This collaborative framework not only enhances adaptability but also sets a precedent for applying SmartAI to a range of domains, including healthcare, education, autonomous systems, and beyond. By building on existing research and integrating state-of-the-art methodologies (Adomavicius et al., 2011; Shuai et al., 2011), SmartAI paves the way for the next generation of adaptable, context-aware AI systems capable of addressing the complexities of real-world scenarios.

The remainder of this paper is organized as follows. The next section describes the SmartAI architecture, describing each component and the relationship between the components. Then, we give a scene understanding use case example and show how SmartAI addresses this use case. In the last section we conclude our work.

SMARTAI ARCHITECTURE

In this section, we introduce the SmartAI architecture, our integrated approach designed to achieve wisdom-like capabilities by combining Machine Learning (ML) models with Knowledge Representation and Reasoning (KRR) (Brachman et al., 2004). As illustrated in Figure 2, SmartAI comprises three main components: a large language model (LLM) for intuitive, rapid inference (System 1), a semantic-network-based KRR system for analytical, deliberate reasoning (System 2), and an Orchestrator component that manages interactions between them. In our current implementation, ChatGPT 4o is utilized for System 1, and the Scone system (Fahlman 2011) for System 2.

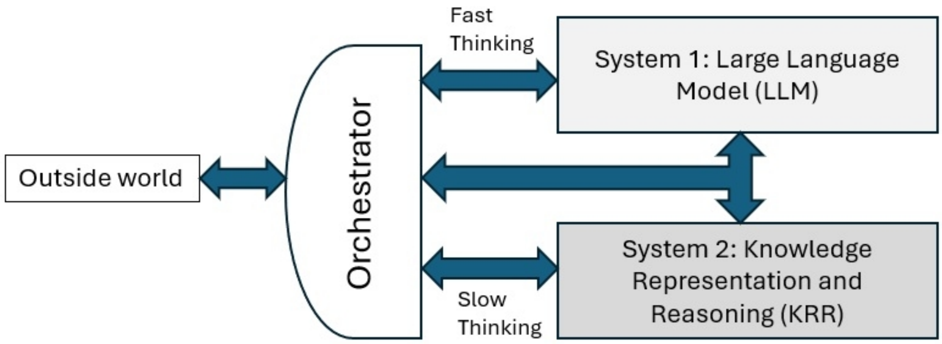


Figure 2: SmartAI architecture - ML models for fast thinking and KRR for slow thinking.

SmartAI's uniqueness lies in its modular architecture, enabling seamless integration of existing LLM and KRR technologies without internal modifications. However, this modularity introduces a primary challenge-effective orchestration of the two complementary AI technologies. To address this, the Orchestrator component plays a critical role in managing input queries and coordinating responses, closely resembling human cognitive processes. Although our current version supports only images as input, the architecture is designed to handle diverse input modalities in future implementations.

The Orchestrator serves several essential functions. It receives and analyses incoming queries, decides how queries should be processed, and coordinates responses from System 1 and System 2. Upon receiving a query, the Orchestrator forwards it directly to the ML-based System 1 for immediate intuitive inference. Based on System 1's response, the Orchestrator evaluates the quality, detail, and context-awareness of the generated description. If the response sufficiently addresses the query, it is immediately returned to the user. If not, the Orchestrator triggers additional processing by System 2 for deeper, contextually aware reasoning. Once responses from both systems are available, the Orchestrator merges them into a single coherent output.

For the ML component (System 1), SmartAI employs ChatGPT 4o due to its proven capability in rapid inference and pattern recognition. While this provides immediate effectiveness, future SmartAI iterations plan to incorporate customized, domain-specific LLMs to enhance accuracy and adaptability.

For the KRR component (System 2), we have selected semantic networks with marker-passing reasoning (Sowa 1987), supported by the Scone system (Fahlman 2011). Semantic networks provide robust representation for context-aware knowledge, and marker passing offers efficient reasoning capabilities. Previous research demonstrates their effectiveness in dynamic, real-world scenarios (Adomavicius et al., 2011; Shuai et al., 2011). The main challenge with the Scone-based approach is building comprehensive semantic networks containing sufficient contextual knowledge for accurate reasoning.

The integration and cooperation of these components are clearly exemplified through the scene understanding scenario, a fundamental task in computer vision aimed at enabling AI to interpret visual scenes similarly to humans (Zhou et al., 2018). Scene understanding demands recognizing not only individual objects but also relationships, context, and underlying intentions within scenes. SmartAI addresses these complexities by harmonizing System 1's rapid object recognition with System 2's deeper contextual analysis.

Specifically, when SmartAI receives an image (as illustrated in Figure 3), the following processing steps occur:

1. The Orchestrator forwards the image directly to both System 1 and System 2.
2. System 1 generates a description of what it sees (text description). The Orchestrator and System 2 handle this description.
3. The Orchestrator assesses System 1's description for completeness and accuracy:

a. If the description fully addresses the query, the Orchestrator directly returns this result to the user.
b. If further context or detail is required, the Orchestrator formulates and sends a query to System 2 for additional analysis. Once System 2 returns a detailed response, the Orchestrator synthesizes both systems' outputs into a comprehensive, contextually enriched description.

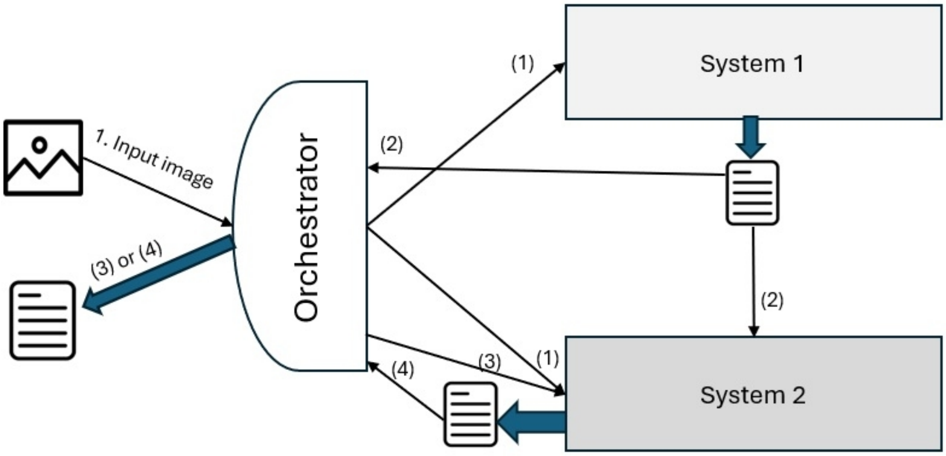


Figure 3: The operation process for scene understanding.

This structured and collaborative workflow allows SmartAI to effectively balance intuitive and analytical processing, optimizing both performance and cognitive effort. Such architecture lays the groundwork for broadening SmartAI's application across various domains, including healthcare, education, autonomous systems, and beyond.

USE CASE: SCENE-UNDERSTANDING

In this section, we demonstrate SmartAI's effectiveness through a practical example of scene understanding. We submitted the image shown in Figure 4 first to ChatGPT 4o and ChatGPT o1, requesting descriptions and interpretations of the scene.

ChatGPT 4o responded as follows: "This image shows a curved road passing through a forested area with trees on both sides. In the lower-left corner, there is a sign that appears to have been edited or superimposed onto the image. The sign resembles a triangular warning sign, but its design includes a red border and additional black and white elements, which may not correspond to a standard traffic sign."

ChatGPT o1, which has enhanced reasoning capabilities, provided a more detailed response: "It appears to be a winding, two-lane asphalt road running through a forest. Tall conifers and other trees line both sides, and the ground is covered with fallen leaves or needles. There are no people or vehicles visible. In the lower-left portion of the image, a warning road sign (red and white with a black symbol) has been overlaid or edited in, suggesting caution for drivers-possibly indicating a narrow bridge, a humpback bridge, or another road hazard ahead. The overall setting is a quiet, wooded area with a gently curving road."

Figure 4: The provided image for scene-understanding query.

Despite detailed descriptions, both ChatGPT models failed to precisely identify the road sign's meaning due to a lack of contextual reasoning. To address this limitation, we utilized SmartAI. Within SmartAI, ChatGPT functions as System 1 (fast thinking), providing initial intuitive responses. Using ChatGPT's initial descriptions, additional queries were then generated for System 2, our KRR component (Scone). Leveraging context-aware knowledge embedded within its semantic network, Scone correctly interpreted the road sign, identifying it as indicating a "slight right" curve ahead.

In this initial phase, human operators acted as the Orchestrator by interpreting System 1's output and formulating targeted queries for System 2. Future implementations of SmartAI will automate this process, training a dedicated AI model to analyze System 1's intuitive outputs systematically and generate precise, contextually relevant queries for System 2. This automation will enhance SmartAI's responsiveness, scalability, and cognitive efficiency in practical scenarios.

CONCLUSION

SmartAI demonstrates significant potential to transform AI capabilities by seamlessly integrating Machine Learning (ML) and Knowledge Representation and Reasoning (KRR). Utilizing the Orchestrator to effectively balance intuitive and analytical reasoning, SmartAI overcomes key limitations inherent in traditional AI models, providing enhanced adaptability and context-awareness in complex scenarios. Its innovative architecture, particularly illustrated through the effectiveness in scene-understanding tasks, establishes a solid foundation for broader applications requiring nuanced reasoning and rapid decision-making.

Although initially focused on visual scene understanding, SmartAI's modular design allows for versatile expansion. Extending input capabilities to text, audio, or multimodal forms could empower SmartAI to address challenges in natural language processing, robotics, and human-machine interactions. Further enriching the semantic networks in the KRR component and developing tailored LLMs optimized for specific domains could enhance both accuracy and scalability. Addressing these technical enhancements will position SmartAI to operate robustly within dynamic environments, delivering effective solutions to real-world complexities.

Future research should also emphasize practical impacts in critical sectors such as healthcare, education, and autonomous systems. By integrating deeper cognitive capabilities, SmartAI has the potential to revolutionize these domains, providing personalized insights and significantly enhancing decision-making quality. Continued development and refinement of SmartAI promise to bridge existing gaps between artificial and human cognitive flexibility, opening pathways toward more advanced and human-like AI technologies.

REFERENCES

Adomavicius, G. Mobasher, B. Ricci, F. and Tuzhilin, A. (2011). Context-Aware Recommender Systems. AI Magazine, Volume 32, No. 3.

Akkaladevi, S. Plasch, M. Hofmann, M. and Pichler, A. (2021). Semantic knowledge based reasoning framework for human robot collaboration. In Proceedings of 8th CIRP Conference of Assembly Technology and Systems, volume 97, pp. 373–378.

Booch, G. Fabiano, F. Horesh, L. Kate, K. Lenchner, J. Linck, N. Loreggia, A. Murugesan, K. Mattei, N. Rossi, F. and Srivastava, B. (2021) Thinking Fast and Slow in AI. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 35, pp. 15042–15046.

Brachman, R. and Levesque, H. (2004). Knowledge Representation and Reasoning. Elsevier.

Chowdhury, T. Ling, C. Zhang, X. Zhao, X. Bai, G. Pei, J. Chen, H. and Zhao, L. (2023). Knowledge-enhanced Neural Machine Reasoning: A Review.

Donadello, I. Serafini, L. and Garcez. A. S. (2017). Logic Tensor Networks for Semantic Image Interpretation. CoRR, abs/1705.08968.

Fahlman, S. E. (2011). Using Scone's Multiple-context Mechanism to Emulate Human-like Reasoning. In Proceedings of the AAAI Fall Symposium on Advances in Cognitive Systems.

Fahlman, S. E. (2012) Beyond Idiot-Savant AI. Advances in Cognitive Systems, 1:15–22.

Goyal, A. and Bengio, Y. (2020). Inductive Biases for Deep Learning of Higher-Level Cognition. CoRR, abs/2011.15091.

Kahneman, D. (2011) Thinking, Fast and Slow. Macmillan.

Shuai, L. Xueyan, Z. Xiaodong, S. Xiaohan, Y. Ruichun, T. and Qingyun. J. (2019). Survey on Context-aware Systems and Their Applications. In Proceedings of the 9th IEEE International Conference on Electronics Information and Emergency Communication (ICEIEC), pp. 234–237.

Sowa, J. F. (1987) Semantic Networks. In Encyclopedia of Artificial Intelligence. Wiley.

Violet, X. Snell, C. Gandhi, K. Albalak, A. Singh, A. Blagden, C. and Phung D. (2025). Towards System 2 Reasoning in LLMs: Learning How to Think with Meta Chain-of-Though. arXiv preprint arXiv:2501.04682.

Zhou, B. Zhao, H. Puig, X. Xiao, T. Fidler, S. Barriuso, A. and Torralba, A. (2018). Semantic Understanding of Scenes Through the ADE20K Dataset. International Journal of Computer Vision, 127:302–321.

Zihan, Y. He, L. Wu, Z. Dai, X. and Chen, J. (2023). Towards better chain-of-thought prompting strategies: A survey. arXiv preprint arXiv:2310.04959.