Can Mental Imagery Improve the Thinking Capabilities of AI Systems?

Slimane LARABI[1[0000-0001-8994-5980]]

USTHB University, BP 32 EL Alia, 16111, Algiers, Algeria slarabi@usthb.dz
http://perso.usthb.dz/ slarabi/accueil.html

Abstract. Although existing models can interact with humans and provide satisfactory responses, they lack the ability to act autonomously or engage in independent reasoning. Furthermore, input data in these models is typically provided as explicit queries, even when some sensory data is already acquired. In addition, AI agents, which are computational entities designed to perform tasks and make decisions autonomously based on their programming, data inputs, and learned knowledge, have shown significant progress. However, they struggle with integrating knowledge across multiple domains, unlike humans. Mental imagery plays a fundamental role in the brain's thought processes, which involves performing tasks based on internal multisensory data, planned actions, needs, and reasoning capabilities. In this paper, we investigate how to integrate mental imagery into a machine thinking framework and how this could be beneficial in initiating the thinking process. Our proposed machine thinking framework integrates a Cognitive thinking unit supported by three auxiliary units: The Input Data Unit, the Needs Unit, and the Mental Imagery Unit. Within this framework, data is represented as natural language sentences or drawn sketches, serving both informative and decision-making purposes. We conducted validation tests for this framework, and the results are presented and discussed.

Keywords: Machine thinking · Mental image · Reasoning.

Creating machines capable of autonomous thinking remains a significant challenge in artificial intelligence (AI). Machine thinking holds great promise for revolutionizing human–machine interaction by enabling machines to perform tasks based on internal data, planned actions, needs, and reasoning processes. Unlike existing AI models that primarily respond to human queries or follow predefined instructions, machine thinking aims to empower systems to simulate autonomous thought, adapt to dynamic environments, and make independent decisions based on contextual demands.

Despite advances in deep learning, computer vision, and natural language processing, current systems fall short of achieving human-level machine thinking. Prior research has explored key components of machine cognition, including reasoning models [6], sensory data integration [19], and imagination-driven tasks [18]. However, these efforts often lack a unified framework that bridges internal motivations, sensory perception, and reasoning.

Our aim is not to replicate human thought processes but to take inspiration from lifelong experience with how we think and act. Accordingly, in this paper, we present a framework for machine thinking composed of several key elements: the Needs Unit, which captures the machine's internal needs (scheduled actions, problems to solve, etc.); the Input Data Unit, which captures sensory data such as perceived scenes, sounds, and tactile information; and the Mental Imagery Unit, which enables the simulation and prediction of potential scenarios and provides insight into how AI systems might utilize generative capabilities for reasoning. The Cognitive Thinking Unit (CTU) integrates these auxiliary units to simulate reasoning, plan actions, and generate contextual responses.

To validate the proposed framework, we present experiments across diverse applications: image captioning, which evaluates the system's ability to describe visual inputs; matching needs and context, which tests the system's ability to align internal motivations with external data; and sentence generation by the Cognitive Thinking Unit (CTU), which demonstrates the framework's reasoning and decision-making capabilities.

The remainder of this paper is organized as follows: Sections 1 and 2 discuss related work in the domains of machine thinking and mental imagery. Sections 3 and 4 describe the proposed framework, detailing its components, functionalities, and how it performs reasoning through the generation of mental images. Section 5 presents the experimental results, and finally, Section 6 concludes the paper.

1 Related Works to Machine Thinking

Machine thinking which refers to the capacity of systems to infer actions or statements from context, is a multifaceted challenge that intersects with several domains of artificial intelligence, including commonsense reasoning, natural language inference (NLI), multi-modal learning, and contextual reasoning.

Commonsense reasoning is a critical yet underdeveloped area in machine intelligence due to the complexity and implicit nature of everyday knowledge [8]. Approaches include large pretrained models fine-tuned for commonsense tasks (e.g., GPT, T5 [17]), and structured knowledge bases such as ConceptNet [11], COMET [10], and ATOMIC [9], which support inferential reasoning using semantic relationships and textual inferences.

Natural language inference (NLI) aims to identify the relationship between a premise and a hypothesis (entailment, contradiction, or neutrality). Datasets like SNLI and MultiNLI [12] have enabled the training of powerful inference models such as BERT and RoBERTa, improving sentence-level comprehension.

Multi-modal learning combines heterogeneous inputs such as text and images to perform tasks like Visual Question Answering and cross-modal retrieval. Notable models include CLIP [13], Flamingo [15], and VilBERT [7], which align visual and linguistic data to facilitate integrated reasoning.

Contextual and goal-oriented reasoning systems leverage models like BERT and GPT to perform attention-based reasoning over full input sequences [14].

They are central to applications such as autonomous agents and natural language understanding. Additional approaches include rule-based inference systems and multi-modal architectures for action prediction [6].

Recently, Del Ser et al. [29] proposed theoretical key research areas for constructing world models based on Piaget's theory of cognitive development [30] , emphasizing learning through interaction, active knowledge construction, and dynamic reorganization of mental structures. To operationalize this vision, they identify six interconnected research areas essential for building reasoning capable World Models:
- Embodied AI, physics-informed ML, grounding knowledge in real-world interaction.
- Neuro-symbolic integration, combining statistical learning with symbolic reasoning.
- Causal inference understanding action effect relationships.
- Open-world learning, adapting to novel environments.
- Human-in-the-loop, incorporating human oversight and common sense.
- Trustworthy, responsible AI, ensuring safety and accountability.

Crucially, there is still no framework that enables machines to engage in autonomous thinking (reasoning initiated without an external query). Existing models remain reactive, and although approaches like reinforcement learning agents show promise, they are often task-specific and fail to generalize to broader notions of thinking. These systems also exhibit significant cognitive gaps that limit deeper reasoning, such as difficulties with counterfactual thinking and logically anticipating events in unfamiliar contexts. Moreover, they lack the capacity to internally simulate scenarios before acting, a skill that, in humans, supports planning, foresight, and creative problem solving. Mental imagery offers a potential path forward by enabling AI to rehearse possible outcomes, assess their plausibility through internal simulation, and reason about hypothetical futures. Embedding such an imagination module into AI architectures could help shift systems from reactive pattern-matching toward proactive, context-aware decision-making.

2 Mental Imagery

Mental imagery has long been recognized as a foundational component of human cognition, playing a crucial role in memory, planning, decision-making, spatial navigation, and emotional processing [23]. It supports core cognitive functions by enabling individuals to simulate future scenarios and reconstruct past experiences [24] [25]. These capabilities are essential not only for adaptive behavior but also for understanding and treating mental health conditions.

Recent neurocognitive studies have revealed that mental imagery and perception share overlapping neural substrates. For instance, Stecher et al. [22] demonstrated that both imagined and perceived natural scenes activate similar cortical patterns, particularly in the alpha frequency band, suggesting that imagery and late-stage perception engage shared neural codes. Key brain regions involved in these processes include the ventromedial prefrontal cortex, anterior hippocampus, posterior parahippocampal cortex, and retrosplenial cortex [1], reinforcing the biological grounding of visual imagination.

The functional utility of mental imagery extends beyond human cognition into artificial intelligence. Inspired by human intuitive planning, Li et al. [20] proposed Simulated Mental Imagery for Planning (SiMIP), a subsymbolic framework that enables robots to "imagine" sequences of actions through internal simulation. SiMIP integrates CNNs and GANs to generate and evaluate visual outcomes of planned actions, thus emulating the role of mental imagery in adaptive behavior and re-planning.

In educational psychology, Commodari et al. [26] emphasized the role of mental imagery and visuospatial processing in academic performance and creative problem-solving. Their findings advocate for instructional strategies that leverage visual thinking and mental simulation, especially in abstract domains like mathematics and science, to foster deeper cognitive engagement and innovation.

From an AI reasoning perspective, mental imagery offers a significant computational advantage. As shown by Kunda [21], imagery-based representations allow for direct inspection of relationships within a scene, bypassing the resource-intensive chaining required by propositional logic systems. This efficiency becomes critical as the complexity and volume of data increase, highlighting the potential of visual reasoning frameworks.

Despite these insights, current AI systems lack mechanisms for simulating internal mental imagery as part of their reasoning processes. While generative models can synthesize images from latent representations, they do not yet operate within a unified framework that links internal imagery to cognitive states or decision-making. Moreover, the source data for these generated images, i.e., what memory, knowledge, or perceptual cues they are based on, remains unspecified and disconnected from dynamic reasoning contexts.

In [31], Zeyuan Yang et al. present a novel approach to integrating mental imagery into machine reasoning by using latent visual tokens for internal simulation, avoiding the need for explicit image generation. While this method successfully demonstrates enhanced multimodal reasoning within a model's latent space, it is fundamentally constrained by its reliance on a task-specific supervised signal. For each new problem domain, the method requires an a priori manual creation of helper images that serve as a ground-truth for training the latent reasoning trajectory. This reliance on human-in-the-loop supervision for each task presents a significant challenge to the framework's scalability and autonomy, limiting its practical application to a small, predefined set of problems rather than enabling true generalization in open-ended environments.

In essence, the ability of the human brain to reconstruct experiences into mental images is key to flexible thinking and imagination. However, this mechanism is largely absent in state-of-the-art AI systems. Bridging this gap requires the development of new models that explicitly incorporate internally generated visual simulations into reasoning architectures, making mental imagery a first-class component of artificial cognition.

3 Machine Thinking: The Proposed Framework

Our goal is to propose a framework for machine thinking, inspired by the way the human brain operates. Given the limited understanding of how the brain executes cognitive processes, we draw inspiration from recent advances in neuroscience.

We define machine thinking as the process of inferring decisions and generating informative content from various data sources, with the results presented as natural language sentences and sketches. These sketches represent a series of machine-generated mental images, mimicking a common aspect of human cognition.

The synoptic scheme of the proposed framework, illustrated in Figure 1, comprises a **Cognitive Thinking Unit** which processes information received from the three auxiliary units: The **Needs Unit**, the **Input Data Unit** and the **Mental Imagery Unit**. The Cognitive Thinking Unit **(CTU)** synthesizes information from multiple sources and produce an output including both informative and decision-making content.

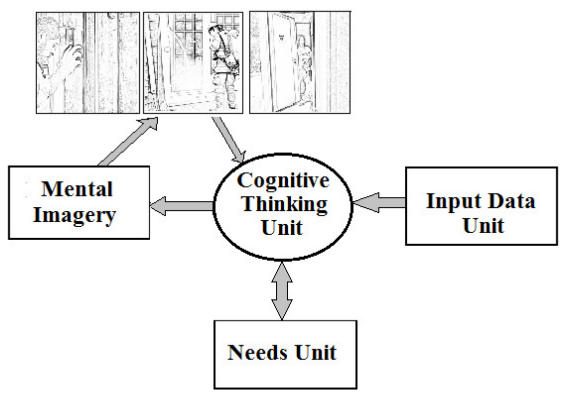

Fig. 1. Diagram of Machine Thinking Framework.

3.1 The "Needs" Unit

This unit contains data that may include:(1) A set of actions to be performed, typically scheduled as a result of prior reasoning. These actions are generated by the Cognitive Thinking Unit (CTU) and transferred to the Needs Unit for execution planning. (2) Knowledge acquired by the machine, representing internal goals or obligations.

For a human analogy, examples of knowledge include:
"At 8:00 AM, I must be at my desk."
"According to my agenda, I have a meeting at 10:00 AM with my PhD student in the next building."

Examples of actions to be performed, derived through reasoning, include:
"To get to the next building, take the stairs."
"Go to the Queens Building using the bike."

3.2 The "Input data" Unit

Is a unit which processes the acquired images, the heard audio signals and produces them as a set of natural-language sentences describing the received data. The recent advances in computer vision using neural approaches may be used such as YOLO for object detection and recognition [2] and DenseCap for image captioning [16]. In the same way, the audio signal is converted into useful sentences. This involves several key steps in the pipeline of Automatic Speech Recognition (ASR) using for example DeepSpeech and Wav2Vec 2.0 and Natural Language Understanding (NLU) using Transformers-based models (BERT, GPT). The sens of touch is also integrated for input data. Indeed, by using the tactile sensors, machine learning, and natural language generation, the sense of touch can be effectively translated into meaningful sentences that describe the physical properties of objects and interactions.

3.3 The "Mental Imagery" Unit

The Mental Imagery Unit receives stimulus from the Cognitive Thinking Unit (CTU), draws imagined images and communicates the convenient ones to the (CTU) which refine the result of reasoning.

3.4 The Cognitive Thinking Unit (CTU)

Thinking which is the ability to reach logical conclusions on the basis of prior information is central to human cognition [5]. Therefore, the advent of neuroimaging techniques has increased the number of studies related the study of examined brain function [3] and to the neural basis of deductive reasoning [4].

The Cognitive Thinking Unit refers to a central processing module or system that is responsible for Interpreting sensory or internal stimuli, Initiating reasoning processes, Making inferences based on knowledge, Integrating information from memory, perception, and imagination and Triggering mental imagery when needed for simulation or problem-solving.

In this paper, in addition to inferring action from a set of data (sentences) and a specific need, we study how mental images may generate new information for the machine.

4 Generating Mental Images and Thinking

After receiving a stimulus (A sentence or semantic description representing a situation, Hypothetical query, Goal-oriented query) from the CTU (Cognitive Thinking Unit), the Mental Imagery Unit (MIU) generates one or many sequences of mental images corresponding to possible scenario. These sequences of mental images are processed and the best one is selected, transmitted to the Needs Unit for further performing. In the next how a mental images sequence is generated and how this can be implemented.

4.1 Mental Image Generation

We assume that each piece of knowledge is stored in memory as a sentence and can be represented as a sketch derived from previously seen images. In computational terms, this implies the existence of a neural model with appropriate parameters capable of producing such visual representations.

When a new hypothesis (resulting from reasoning) is integrated into existing knowledge, the corresponding image is modified to incorporate the additional information. This enables the generation of continuously updated images as reasoning progresses and new knowledge is formed. The Mental Imagery Unit then performs further processing based on these visualized images.

4.2 Thinking: A Computational Perspective

The seminal work of Kahneman and Tversky demonstrated that human judgment is not always rational but is guided by mental shortcuts, or heuristics [32]. Mental imagery functions as a particularly powerful heuristic by allowing individuals to quickly and intuitively simulate outcomes. When a text conveys a potential gain (e.g., "a 2% increase in value"), the brain's analytical and effortful system must decode this abstract number. In contrast, a vivid, emotionally resonant mental image of that same gain (e.g., picturing oneself receiving a bonus) engages the intuitive system almost instantly. Such imagery evokes a stronger affective response, often overshadowing the slower, more deliberate processing of numerical or textual information [33] [34].

Neuroscience supports this primacy with biological evidence. Brain imaging studies reveal that emotionally charged or visually rich stimuli activate subcortical regions, including the amygdala, that govern emotion and reward processing more intensely than abstract text. These activations trigger rapid, unconscious "gut feelings." By contrast, the evaluation of textual or statistical information relies more on the prefrontal cortex, associated with deliberate reasoning and logical analysis. In situations where both systems compete, the fast emotional response linked to mental imagery often dominates decision-making, effectively biasing judgments before the analytical system can intervene [34]. This finding justify generating mental images from text, and their evaluation.

We consider thinking as the process of exploring possible ways to achieve an event by generating a time series of mental images. Once a selected need has been communicated to the Cognitive Thinking Unit (CTU) (see Fig. 2), the system first processes the received data and performs reasoning to identify potential scenarios. The selected scenarios are then represented by the Mental Imagery Unit as sequences of mental images, depicting the progression from the initial state to the final state. Mental images are generated based on contextual knowledge, inference, and the ability to introduce modifications to their content, either to add details, partially correct the event, or completely change it.

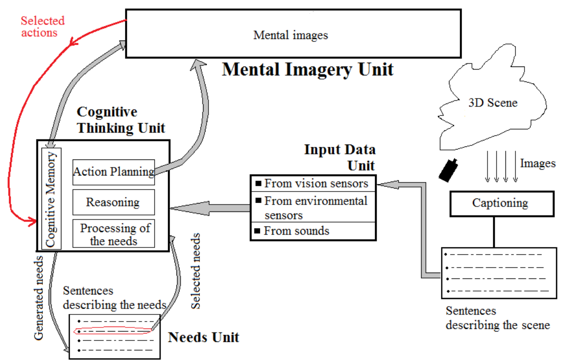

Fig. 2. Producing mental images and the selection of the optimal image sequence.

While the Cognitive Thinking Unit (CTU) possesses an inherent understanding of the semantic content of each image, the explicit generation and structuring of these visual representations are not merely for display; they serve as a rigorous simulation space. This simulated environment allows the Mental Imagery Unit to perform constraint verification against real-world rules that are often difficult to model abstractly, such as physical laws, spatial feasibility, execution time, and potential losses. To ensure that the chosen scenario is not only plausible but also optimal, the system evaluates all candidate mental image sequences ($s \in S$) using a formalized objective function, $J(s)$. This function integrates various constraints into a single metric:

$$J(s) = w_T\, T(s) + w_C\, C(s) + w_L\, L(s) + \dots$$

In this function, $T(s)$, $C(s)$, and $L(s)$ represent the estimated time, cost, and losses, respectively, for a given scenario s. The weighting coefficients (w_T, w_C, w_L) allow the system's priorities to be adjusted based on the specific goal, whether it is to minimize time, cost, or risk. This quantitative approach enables a direct comparison between different potential scenarios, ensuring the system can make a rational, data-driven decision.

The process of generating these sequences is dynamic and draws upon both creative generation and memory recall. Once, the Mental Imagery Unit proposes scenarios in response to a need communicated by the Needs Unit, the system can efficiently leverage past experiences by searching its stored mental images (in Cognitive Memory subunit) for image that corresponds to the beginning or a central piece of a new scenario. This memory-based approach is particularly useful for common tasks or when a crucial piece of information is missing. For instance, if the system is tasked with finding a misplaced object, it can retrieve a mental image of the last known location of that object, using it as an anchor point to explore a more constrained and probable set of solutions. This combination of generating new possibilities and retrieving past ones allows the Mental Imagery Unit to select the optimal solution ($s^\star = \arg\min_{s \in S} J(s)$) from a diverse set of options, ultimately ensuring that the chosen scenario is both effective and physically plausible before it is communicated to the Needs Unit.

How mental images can improve thinking

Words alone can be abstract and lack real-world context. A mental image acts as a visual grounding, providing a concrete, simulated environment for the AI to "see" and evaluate a scenario. This helps the AI perform constraint verification against physical laws, spatial feasibility, and other real-world rules that are difficult to model with just text. The recall of mental images from a stored memory (e.g., in a "Cognitive Memory" unit) is a crucial step that moves beyond stateless, short-term reasoning. It provides the system with a form of long-term episodic memory that can be accessed and leveraged.

5 Experiments

5.1 Input Data Unit Implementation

The Input Data Unit converts raw sensory data into a set of informative, natural-language sentences. While the framework can accommodate various sensors, our experiments focus on camera input, which delivers sequences of images from the environment. The processing pipeline transforms these images into a semantic description of the scene.

This pipeline employs a two-stage deep learning approach. First, an object detection model, Faster R-CNN, analyzes each image to identify and localize objects, providing bounding boxes for each. The model, pretrained on the extensive COCO dataset, detects a wide range of common objects. To ensure accuracy, detections with low confidence scores are filtered out.

Next, a BLIP (Bootstrapped Language Image Pretraining) model, a transformer-based image captioning model, takes the cropped region of each detected object and generates a corresponding natural-language description. This process allows the pipeline to provide detailed captions for specific regions of an image, such as "a red car is parked in a parking" as shown by figure 3(a). These sentences serve as the structured input data for the Cognitive Thinking Unit (CTU), enabling it to reason about the scene. An example of the sentences generated for a single image (see Fig. 3(b)) is:

sentences = [a laptop computer sitting on top of a wooden desk, a computer mouse and a mouse pad, a box with a pair of gloves in it, a laptop with a red arrow pointing up, a pile of papers on a table, a black leather seat with a metal handle, a laptop computer with a bunch of keys on it.] For more details, see the source code in [35]

(a) (b)

Fig. 3. (a) Some of captions are: "a car is seen in this image", "a bus is driving down the street", "a red car is parked in a parking", "a man in a white shirt and black pants is seen through a red square". (b) A sample of used image for captioning. Detected objects are illustrated with their associated bounding box.

5.2 Cognitive Thinking Unit Implementation: Reasoning and Action Inference

The Cognitive Thinking Unit (CTU) serves as the core of the framework, processing needs and contextual information to infer logical conclusions and actionable plans. It first receives a specific need, such as "I need the keys to open the door and go out," from the Needs Unit, along with a list of descriptive sentences representing the current environment from the Input Unit.

To connect the need with the available context, the CTU performs semantic matching. We use the *sentence-transformers* library and a pre-trained model like *all-MiniLM-L6-v2* to convert both the need and each context sentence into a high-dimensional vector representation. By calculating the cosine similarity between the need's vector and each context vector, the system identifies the most semantically relevant information. For instance, given the need to find keys, the system correctly identifies "a laptop computer with a bunch of keys on it" as the most relevant sentence with a high similarity score, while other sentences have significantly lower scores.

Once the most relevant context is identified, the CTU uses a large language model from the *Hugging Face Transformers* library to infer an appropriate action. This process involves a structured prompt that combines the contextual observations and the specific need into a single query. The model, such as a *Falcon* model, then generates a natural language response that proposes a set of actions.

For example (see Fig. 4), given the context "a laptop computer with a bunch of keys on it" and the need "need the keys to open the door and go out," the model generates actionable steps like: "Pick up the keys and open the door - Take the keys and unlock the door - Use the keys to open the door and go out." Similarly, for the need "I need to drink water to stay hydrated," the model correctly infers the action: "Take a sip of water from the bottle on the table."

This two-step process, combining semantic matching with a large language model, allows the CTU to bridge the gap between abstract needs and concrete environmental data, producing intelligent, context-aware actions.

Fig. 4. A sample of used image sequence for captioning.

5.3 Cognitive Thinking and Mental Imagery Units Implementation: Thinking and Mental Images Generating

This subsection presents a proof-of-concept demonstration of integrating a Cognitive Thinking Unit (CTU) and a Mental Imagery Unit to simulate human-like mental visualization from sequential actions. The goal is to assess the system's ability to generate meaningful, sketch-style mental images from natural language descriptions of actions, mimicking how humans imagine future scenarios during reasoning or planning.

For this demonstration, we implemented two key components:

-The Action Planning Module: A subunit of the Cognitive Thinking Unit (CTU) that simulates a basic decision-making process. For this experiment, the module was pre-programmed to generate a predefined sequence of natural language action descriptions: "a man takes the keys which are on the desk," "the man goes towards the door," and "the man opens the door."

-The Mental Imagery Unit: This module uses the pre-trained *Stable Diffusion* model *(runwayml/stable-diffusion-*$v1-5$*)* to generate visual representations of the actions. To abstract these outputs and mimic a stylized, internal visualization process, the photorealistic images are processed through a sketch transformation pipeline. This process, inspired by the way human mental imagery focuses on structural content over photorealistic detail, converts the images into grayscale sketches using Gaussian blurring and image division techniques.

Static Action Plan Demonstration

For each predefined action, the system successfully produced a corresponding sequence of mental sketches. As shown in Figure 5(First row), the sketches effectively capture the essential spatial and semantic elements of each action, including object interaction, spatial transition, and environmental interaction.

These results demonstrate that combining text-to-image diffusion models with a sketch-based abstraction is a viable computational approach for mental simulation. The sketch style supports cognitive tasks by highlighting structural content without unnecessary visual noise, more closely aligning the system's output with how humans imagine events in simplified visual terms. While this initial demonstration used a predefined action plan, it validates the core premise that mental imagery generation from action sequences can effectively support machine reasoning.

Dynamic Problem-Solving Demonstration

Beyond the static demonstration, this framework allows for a form of visual reasoning. The CTU can pose logical questions, such as, "What if the key doesn't open the door?" In response, the system can dynamically propose a new action, like "the person is nervous," and generate a corresponding new mental image. This iterative process can be repeated to explore alternative scenarios and their outcomes. For instance, the system can explore the options "the person breaks down the door" or "the person calls the firefighters for they open the door," with each step visualized through a new mental image as shown in Figure 5(second row). This ability to generate and analyze dynamic visual sequences is a crucial step towards incorporating more sophisticated, goal-oriented scene evolution into future work.

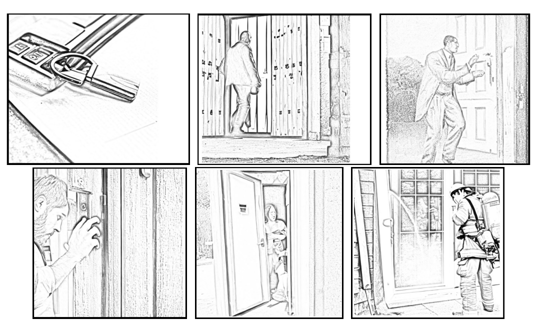

Fig. 5. (First row) Mental images generated from the three prompts. (Second row) The new mental images generated receiving from CTU unit new prompts.

6 Conclusion

In this paper, we addressed the ambitious challenge of enabling machine thinking by proposing a comprehensive framework inspired by recent advancements in neuroscience, deep learning, and artificial intelligence. While significant progress has been made in developing systems capable of processing sensory data, reasoning, and interacting with humans, our approach goes beyond existing models by integrating the concept of a Mental Image to simulate machine cognition.

We presented the limitations of current models, including their reliance on explicit queries, domain specificity, and the absence of intuitive reasoning mechanisms are discussed. These gaps motivated the design of our framework which consist of a Cognitive Thinking Unit (CTU) and three auxiliary units: The Needs, Input Data, and Mental Image which work in synergy to mimic human cognitive processes. This novel architecture enables machines to reason, plan, and infer decisions autonomously, utilizing both sensory inputs and internally generated representations.

The proposed framework have been implemented and the results demonstrated the framework's potential to bridge perception, reasoning, and imagination, thereby addressing key limitations of existing systems.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Anna M. Monk, Gareth R. Barnes and Eleanor A. Maguire, The Effect of Object Type on Building Scene Imagery—an MEG Study, Front. Hum. Neurosci., 2020.
2. J. Redmon, S. Divvala, R. Girshick and A. Farhadi, You Only Look Once: Unified, Real-Time Object Detection, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.779-788, 2016.
3. Bartley, J.E., Riedel, M.C., Salo, T. et al., Brain activity links performance in science reasoning with conceptual approach, npj Sci. Learn. 4, 20, 2019
4. L. Wang, M. Zhang, F. Zou, X. Wu, Y. Wang, Deductive-reasoning brain networks: A coordinate-based meta-analysis of the neural signatures in deductive reasoning, Brain and Behavior, Volume10, Issue12, 2020.
5. V. Goel, G. Navarrete, I. A. Noveck and J. Prado, Editorial: The Reasoning Brain: The Interplay between Cognitive Neuroscience and Theories of Reasoning, Front. Hum. Neurosci., 2017.
6. Sridharan, M., Mota, T. Towards combining commonsense reasoning and knowledge acquisition to guide deep learning. Auton Agent Multi-Agent Syst 37, 4 (2023). https://doi.org/10.1007/s10458-022-09584-4
7. H. Tan, M. Bansal LXMERT: Learning Cross-Modality Encoder Representations from Transformers. arXiv:1908.07490. 2019
8. C. Zhu. Machine Reading Comprehension. Algorithms and Practice. 2021, Pages 3-26. Elsevier 2021.

9. Sap, M., et al. ATOMIC: An Atlas of Machine Commonsense. AAAI. Pages 3027-3035. 2019. https://doi.org/10.1609/aaai.v33i01.33013027
10. A. Bosselut, H. Rashkin, M. Sap, C. Malaviya, A. Celikyilmaz, Y. Choi COMET: Commonsense Transformers for Automatic Knowledge Graph Construction. 2019. https://doi.org/10.48550/arXiv.1906.05317
11. R. Speer and C. Havasi. 2012. Representing General Relational Knowledge in ConceptNet 5. In Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12), pages 3679–3686, Istanbul, Turkey.
12. A. Williams, N. Nangia, S. R. Bowman. (2018). A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference. NAACL. https://doi.org/10.48550/arXiv.1704.05426
13. A. Radford et al. (2021). Learning Transferable Visual Models from Natural Language Supervision. https://doi.org/10.48550/arXiv.2103.00020
14. Clark, C., et al. What Does BERT Look at? An Analysis of BERT's Attention. (2019). https://doi.org/10.48550/arXiv.1906.04341
15. Alayrac, J.-B., et al. (2022). Flamingo: A Visual Language Model for Few-Shot Learning. DeepMind. 36th Conference on Neural Information Processing Systems (NeurIPS 2022)
16. J. Johnson, A. Karpathy, Li Fei-Fei. DenseCap: Fully Convolutional Localization Networks for Dense Captioning. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, Pages: 4565-4574 DOI: 10.1109/CVPR.2016.494
17. C. Raffel et al., Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer, Journal of Machine Learning Research, 2020, 21(140), pages 1–67.
18. F. Qi, W. Wu. Human-like machine thinking: Language guided imagination. 2019. https://doi.org/10.48550/arXiv.1905.07562
19. F. Zhao, C. Zhang, B. Geng. Deep Multimodal Data Fusion. ACM Comput. Surv., Vol. 56, No. 9, Article 216. Publication date: April 2024
20. S. Li, T. Kulvicius, M. Tamosiunaite, and F. Worgotter. Simulated Mental Imagery for Robotic Task Planning. arXiv:2211.08321v3 [cs.RO] 27 Jul 2023
21. M. Kunda. Visual mental imagery: A view from artificial intelligence. Cortex, Volume 105, August 2018, Pages 155-172.
22. Stecher, R., Kaiser, D. Representations of imaginary scenes and their properties in cortical alpha activity. Nature, Sci Rep 14, 12796 (2024). https://doi.org/10.1038/s41598-024-63320-4
23. J. Pearson, T. Naselaris, E. A. Holmes, S. M. Kosslyn. Mental Imagery: Functional Mechanisms and Clinical Applications. Volume 19, Issue 10, pp. 590-602, October 2015.
24. Moulton, S.T., Kosslyn, S.M. Imagining predictions: mental imagery as mental emulation Philos. Trans. R. Soc. Lond. B: Biol. Sci. 2009; 364:1273-1280
25. Gilbert, D.T., Wilson, T.D. Prospection: experiencing the future Science. 2007; 317:1351-1354
26. E. Commodari, J. Sole, M. Guarnera, V. Lucia La Rosa. Mental imagery in education: What impact on the relationships with visuospatial processing and school performance in junior high school students? Thinking Skills and Creativity, Volume 54, December 2024, 101667.
27. N. Koide-Majima, S. Nishimoto, K. Majima. Mental image reconstruction from human brain activity: Neural decoding of mental imagery via deep neural network-based Bayesian estimation. Neural Networks 170 (2024) 349–363.

28. H. Caselles-Dupre, C. Mellerio, P. Herent, A. Lopez-Persem, B. Beranger, M. Soularue, P. Fautrel, G. Vernier, M. Cord. Mind-to-Image: Projecting Visual Mental Imagination of the Brain from fMRI. arXiv:2404.05468v4 [q-bio.NC] 4 May 2024.
29. J. Del Ser, et al. Towards Cognitive Artificial Intelligence: A Comprehensive Perspective on Enabling Cognitive Mechanisms. arXiv preprint* arXiv:2503.15168, 2025.
30. Jean Piaget. The construction of reality in the child. Routledge, 1954. London.
31. Z. Yang, X. Yu, D. Chen, M. Shen, C. Gan. Machine Mental Imagery: Empower Multimodal Reasoning with Latent Visual Tokens.2025. https://www.arxiv.org/pdf/2506.17218
32. Kahneman, D., Tversky, A. Judgment under uncertainty: Heuristics and biases. Science, 185(4157), 1124–1131, 1974.
33. Tversky, A., Kahneman, D. Prospect theory: An analysis of decision under risk. Econometrica, 47(2), 263–291, 1979.
34. Slovic, P., Finucane, M. L., Peters, E., MacGregor, D. G. Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. Risk Analysis, 24(2), 311–322, 2004.
35. LNCS Homepage, https://github.com/slarabi/Machine_Thinking, last accessed 2025/11/14