Proceedings of CONF-MLA 2025 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2025.LD24677

A Study on Artificial Intelligence Models Inspired by the Mechanisms of Brain Attention

Fangyao Liu

Faculty of Science and Technology, Beijing Normal University - Hong Kong Baptist University United International College, Zhuhai, China
winniezaiye@163.com

Abstract. The integration of brain-inspired attention mechanisms into artificial intelligence (AI) has revolutionized fields like natural language processing and computer vision. This paper investigates the biological foundations of attention in the human brain, focusing on neural networks such as the dorsal and ventral attention systems, and examines how these principles have been adapted into AI architectures, particularly Transformer-based models. By comparing biological and artificial attention mechanisms, this study aims to evaluate the capability of AI models to handle key tasks like text summarization, image classification, and multi-modal integration. The key findings demonstrate that while AI models perform exceptionally well in specific tasks, they fall short in terms of the dynamic adaptability and energy efficiency exhibited by biological systems. Furthermore, challenges such as high computational costs and limited interpretability are highlighted. To address these limitations, future directions, including neuromorphic computing and hybrid models, are proposed. The paper highlights the potential of interdisciplinary approaches to connect neuroscience and AI, enabling the development of more robust, human-like cognitive systems.

Keywords: Attention Mechanism, Artificial Intelligence, Neural Networks, Neuromorphic Computing, Transformer Models

1. Introduction

The ability to concentrate on important information while eliminating distractions is a key function of attention, a vital cognitive process. In neuroscience, attention is regulated by intricate neural networks like the dorsal attention network (DAN) and the ventral attention network (VAN), which work together to allocate cognitive resources effectively. Inspired by these biological principles, artificial intelligence (AI) researchers have developed attention-based models to enhance machine learning capabilities [1]. The introduction of the Transformer architecture in 2017 marked a pivotal moment, thus showcasing the power of attention mechanisms in tasks like machine translation and image recognition. Despite these advancements, a gap persists between biological and artificial attention systems. While AI models excel in specific tasks, they often fall short in areas such as dynamic adaptability, energy efficiency, and contextual awareness, which are inherent to the human brain. This paper explores the fundamental differences between biological and artificial attention mechanisms, examines the performance of AI attention models in real-world applications like

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.LD24677

natural language processing and computer vision, and investigates the challenges that limit their scalability and biological plausibility. By synthesizing insights from neuroscience and AI, this study seeks to provide a comprehensive understanding of attention mechanisms and their implications for future AI development. The findings are expected to advance AI research while also contributing to the broader objective of creating more interpretable, efficient, and human-like cognitive systems.

2. Foundations of brain attention mechanisms

2.1. The definition and mechanism of attention

Efficient information processing and decision-making rely on attention, a fundamental cognitive process that allows the brain to focus on relevant stimuli and filter out distractions. In neuroscience, attention is typically classified into two types: bottom-up (stimulus-driven) attention and top-down (goal-directed) attention.

Bottom-up attention is involuntary and triggered by salient external stimuli, such as a bright flash of light or sudden noise, primarily regulated by the sensory cortices and superior colliculus, which quickly respond to novel inputs. In contrast, top-down attention is guided by internal goals or expectations, such as focusing on a conversation in a noisy environment, and involves brain regions like the prefrontal and posterior parietal cortices, which regulate sensory processing based on task relevance [2]. These two forms of attention are interrelated and interact dynamically. For example, the prefrontal cortex can modify the saliency map generated by bottom-up processes, balancing the response to new stimuli with focus on goal-directed tasks [2]. At the neural level, attention is governed by the coordinated action of neuromodulators, especially acetylcholine and dopamine [3]. Acetylcholine, released from the basal forebrain, boosts the signal-to-noise ratio in sensory cortices by increasing neuronal responsiveness to relevant stimuli and dampening responses to distractions. This mechanism has been demonstrated in animal studies, where cholinergic activation improved performance in visual attention tasks. Dopamine, released from the ventral tegmental area, helps prioritize stimuli linked to rewards and motivation. For example, fMRI studies indicate dopamine levels rise in the striatum when focusing on reward-predicting cues, thus highlighting a neural basis for selective attention driven by expected outcomes [4]. And these neuromodulatory systems allow the brain to dynamically allocate resources based on external stimuli and internal goals, a flexibility that current AI systems struggle to replicate.

2.2. The neural network underlying attention

The brain's attention system relies on the DAN and the VAN. The DAN is primarily responsible for goal-directed attention and spatial awareness, encompassing the frontal eye fields (FEF) and the intraparietal sulcus (IPS), which help individuals focus attention on specific locations or objects based on task requirements [5]. For instance, in visual search tasks, the DAN guides eye movements, manages visual focus, and enhances sensitivity to target stimuli. Research has shown that when the DAN is activated, the IPS and FEF synchronize with visual cortical areas, thus enhancing spatial selectivity and perceptual accuracy. In contrast, the VAN primarily detects unexpected or prominent stimuli and is crucial for reorienting attention when new or significant information emerges. The key regions of this network include the temporoparietal junction (TPJ) and the ventral frontal cortex (VFC). For example, when a bright spot appears at the edge of the screen, the VAN quickly shifts attention to the new stimulus, interrupting the DAN's focus. This dynamic interaction allows the brain to maintain focus on the task at hand while staying responsive

DOI: 10.54254/2755-2721/2025.LD24677

to external changes, with the DAN sustaining goal-directed attention and the VAN redirecting focus when necessary.

In addition ,the default mode network (DMN), which is active during rest, regulates attention both internally and externally. The DMN supports self-reflection and internal thought processes, while the DAN and VAN are focused on responding to external stimuli [6]. These networks allow the brain to stay focused on targets in visual tasks while also reacting to unexpected stimuli. In multi-step tasks, the DAN helps sustain focus on the current goal, but when the task requires switching from one goal to another like from color categorization to shape categorization, the DAN's activity temporarily weakens. At this point, the VAN detects task-switching cues and shifts attention, prompting the DAN to refocus for a smooth transition between goals. In rule-switching tasks, increased TPJ activity highlights its role in detecting changes and redirecting attention.

3. The function of attention mechanisms in artificial intelligence

3.1. The inception and evolution of attention mechanisms

The development of attention mechanisms in AI has undergone a transformative journey, marked by key milestones. Early models like recurrent neural networks (RNNs) and long short-term memory (LSTM) networks were constrained by their sequential processing nature, which hindered their ability to capture long-range dependencies in data. These models relied on hidden states to propagate information over time, but their performance deteriorated as sequence lengths increased. The introduction of the Transformer architecture in 2017 marked a turning point, pioneering the use of self-attention mechanisms [7]. In contrast to RNNs, Transformers process input data in parallel, enabling them to assess the importance of different elements simultaneously. This innovation gave rise to advanced models such as BERT (Bidirectional Encoder Representations from Transformers) [8] and GPT (Generative Pre-trained Transformer), which have achieved significant success in tasks, such as machine translation, text summarization, and image recognition. The Transformer's multi-head attention mechanism enables it to focus on different aspects of the input data, effectively mimicking the brain's ability to prioritize relevant information.

3.2. Key differences between biological and artificial attention systems

Despite the impressive capabilities of Transformer-based models in fields such as natural language processing and computer vision, fundamental differences remain between biological and artificial attention systems, particularly in adaptability, energy efficiency, and interpretability [2]. In the human brain, attention functions as a flexible, context-aware system that can adapt in real time to new stimuli. This adaptability is enabled by dynamic neural circuits, including the DAN and VAN, which integrate top-down goals with bottom-up sensory input [7]. Moreover, neuromodulators like acetylcholine and dopamine help optimize cognitive resource allocation by modulating attentional focus [7].

In contrast, artificial attention relies on fixed weights, limiting dynamic adaptation to new tasks [9]. The brain operates complex, multimodal tasks on just 20 watts, while models such as GPT and BERT consume hundreds of megawatt-hours yet lack real-time adaptability [10]. Enhancing its energy efficiency, the brain leverages mechanisms such as sparse activation, localized processing, and feedback loops. Interpretability also distinguishes the two systems. In neuroscience, techniques such as functional MRI (fMRI) allow researchers to study the interplay between the DMN and task-positive networks, offering insights into cognitive mechanisms. AI attention weights, although

DOI: 10.54254/2755-2721/2025.LD24677

mathematically transparent, often fluctuate across tasks and runs, making them unreliable indicators of causal reasoning. This limits the safe deployment of AI in high-stakes fields like healthcare and autonomous systems [10]. These differences highlight the need for interdisciplinary collaboration to build more brain-inspired and adaptable artificial attention systems.

3.3. Specific comparisons of attention mechanisms in the brain and AI

Both systems use attention to prioritize relevant information, but their underlying processes differ profoundly. In the human brain, attention is not governed by fixed equations but emerges from dynamic neural interactions influenced by neurotransmitters such as acetylcholine and dopamine. This biological flexibility allows real-time adjustment of attention to stimuli, tasks, and internal states. Thus, human attention generalizes seamlessly across domains such as vision, language, and decision-making without the need for retraining. In contrast, Transformer-based AI uses dot-product attention to assess relevance via vector similarity. For each token in the input, the model computes three vectors: a query (Q), a key (K), and a value (V). Attention scores are derived from the dot product of Q and K, scaled by the square root of the dimensionality to avoid gradient explosion, and passed through a softmax function to weight the value vectors, producing a weighted sum that highlights relevant information [8]. Despite its efficiency, the process is inherently static. The Q, K, and V vectors are generated using weight matrices fixed after training, meaning the model cannot dynamically adapt its attention without retraining. This contrasts sharply with the brain's plastic neural circuits, where attention is continuously modulated via feedback loops, neuromodulation, and asynchronous firing. Networks such as the DAN and VAN reconfigure their activity in response to goals and stimuli, a level of adaptability not yet achievable in AI.

Moreover, energy efficiency marks another major divide. The human brain accomplishes flexible attention tasks with just around 20 watts of power, using sparse activation and localized processing to minimize energy cost. In contrast, Transformer-based models like BERT and GPT require megawatt-hours of energy to train and still lack real-time adaptability [9]. Interpretability further separates the two. Brain attention can be studied using neuroimaging tools like fMRI, which reveal coordinated activity across networks such as the DMN and task-positive systems. While attention in AI is mathematically visible, the resulting weights often vary widely across tasks and runs, making them unreliable for explaining the model's reasoning, thus raising concerns in applications where transparency is vital.

4. The application and performance evaluation of attention mechanisms

4.1. Revolutionizing computer vision with attention mechanisms

Vision Transformers (ViTs) represent a shift from traditional convolutional neural networks (CNNs) by replacing convolutions with self-attention. By dividing images into patches and modeling their relationships globally, ViTs capture both local and long-range dependencies, thus leading to strong performance on benchmarks like ImageNet, where ViT-Large achieved 88.55% top-1 accuracy, surpassing many CNNs. While ViTs offer better global context modeling, they require large datasets and significant computational resources. They also struggle with fine-grained details, making tasks like segmentation more challenging. In comparison, CNNs excel at extracting local features but often miss global context. ViTs address this by modeling relationships across all image patches, though CNNs still outperform ViTs in tasks like object detection and segmentation due to their stronger local feature extraction. Recent models like DETR and Swin Transformer have adapted

DOI: 10.54254/2755-2721/2025.LD24677

ViTs for these tasks: DETR uses end-to-end attention-based detection without anchor boxes, while Swin Transformer introduces hierarchical attention for improved segmentation and detection. ViTs excel in image classification when trained on large datasets, show competitive results in object detection through models like DETR, and hybrid models combining ViTs and CNNs, such as Swin Transformer, outperform pure CNNs in image segmentation benchmarks.

4.2. Enhancing natural language processing with attention models

BERT and GPT have advanced natural language processing by using attention mechanisms to improve machine understanding and generation of human language. Both models use self-attention to capture word relationships, but they differ in their architecture and use cases [7]. BERT employs bidirectional attention, considering context from both the left and right of a word, making it highly effective for tasks like sentiment analysis, question answering, and named entity recognition (NER), where understanding full context is essential. This approach allows BERT to generate accurate, contextually rich responses [10]. In contrast, GPT utilizes unidirectional attention, processing text from left to right, which makes it particularly effective for text generation tasks such as dialogue and text completion, where fluency is essential. However, GPT's unidirectional design limits its ability to capture full context, thus affecting tasks like sentiment analysis [7]. In question answering tasks like SQuAD, BERT's bidirectional approach captures the full context, giving it an advantage over GPT, which may miss key details and reduce accuracy. Similarly, BERT outperforms GPT in sentiment analysis, as it can capture both preceding and following words to understand sentiment nuances, while GPT's one-way processing makes it less effective in this regard.

4.3. Leveraging attention mechanisms for multimodal tasks

Multimodal tasks, which combine data from sources like text and images, have been significantly improved by attention mechanisms. Models such as Contrastive Language-Image Pretraining and A Large-scale ImaGe and Noisy-text leverage attention mechanisms to align visual and textual data, powering applications like image captioning, visual question answering, and cross-modal retrieval [7]. Both models leverage attention to connect images with text, facilitating zero-shot learning and task generalization without additional training, a major advantage of attention-based multimodal models. Multimodal transformers like VisualBERT and ViLT integrate visual and textual inputs to address complex tasks. These models achieve state-of-the-art performance in visual reasoning and multimodal sentiment analysis, where understanding both image content and textual context is crucial. Their attention mechanisms enable the model to focus on the most relevant features in each modality, improving integration and performance [8]. Despite their strengths, multimodal attention models face challenges such as data alignment, which is critical for pairing text and image features accurately. Misalignments can degrade performance, particularly in tasks like cross-modal retrieval. Moreover, processing multiple modalities simultaneously introduces high computational complexity, making real-time applications difficult to scale. Ongoing research focuses on improving alignment techniques and optimizing computational efficiency without compromising accuracy.

5. Challenges and future outlook

5.1. Existing challenges

Despite their success, AI attention models like GPT and BERT face scalability issues, requiring vast computational resources and raising environmental concerns due to high energy consumption [10].

DOI: 10.54254/2755-2721/2025.LD24677

The underlying causes of these challenges stem from the scale and complexity of these models. As transformer-based architectures, especially those with self-attention mechanisms, scale up, their computational demands grow exponentially, creating bottlenecks in both training and inference and hindering efficient data processing. Though GPUs and TPUs are built to accelerate training, they are still limited in handling large models, which intensifies computational load and energy consumption, ultimately driving up operational costs and environmental impact. Also, the lack of interpretability in AI attention models is a key barrier, especially in critical areas like healthcare and autonomous driving, where transparency is vital. Though attention mechanisms highlight key input features, the decision-making process remains opaque, hindering the use of these models in high-accountability fields [11]. Moreover, AI attention models face challenges in generalizing across different tasks and domains. Unlike the human brain, which adapts to new tasks with minimal retraining, AI models often need extensive retraining for unfamiliar contexts [12]. This limitation underscores the need for more flexible AI architectures that generalize across domains with fewer resources [13].

5.2. Future developments

Future research should explore neuromorphic computing to create more efficient AI models by mimicking the brain's architecture and using spiking neural networks [10]. Another important area is the exploration of hybrid models that combine symbolic AI with neural networks. Symbolic AI, which relies on explicit rules and logic, can enhance the interpretability and reasoning capabilities of attention-based models. This hybrid approach could enable AI systems to perform complex tasks with greater transparency and reliability. Besides, progress in transfer learning and few-shot learning can help minimize the data and computational needs of AI models. By enabling models to leverage knowledge from one domain to another, these techniques can improve generalization and reduce the need for extensive retraining [10]. In addition, interdisciplinary collaboration between neuroscientists and AI researchers is crucial for developing more biologically plausible attention mechanisms. By mimicking the brain’s dynamic and adaptive processes, researchers can develop AI systems that are both more efficient and better aligned with human cognition.

6. Conclusion

This paper explores the integration of brain-inspired attention mechanisms into AI models, highlighting both the achievements and limitations of this approach. Despite significant progress in natural language processing and computer vision, AI attention models lack the dynamic adaptability and energy efficiency of biological systems. Key challenges include the high computational cost of training large models, lack of interpretability, and difficulties in generalizing across tasks. However, AI attention models still hold substantial potential, particularly when combined with advancements in neuromorphic computing, hybrid models, and transfer learning. Future research should bridge AI and biological systems through hardware innovations such as spiking neural networks to improve adaptability and energy efficiency. By integrating neuroscience insights with technical innovations, AI systems that are more efficient, interpretable, and biologically plausible can be developed.

References

[1] Bahdanau, D., Cho, K. and Bengio, Y. (2014) Neural Machine Translation by Jointly Learning to Align and Translate. arXiv preprint arXiv: 1409.0473.
[2] Posner, M.I. and Petersen, S.E. (1990) The Attention System of the Human Brain. Annual Review of Neuroscience, 13(1), 25-42.

DOI: 10.54254/2755-2721/2025.LD24677

[3] Sarter, M., et al. (2005) Unraveling the Attentional Functions of Cortical Cholinergic Inputs: Interactions Between Signal-Driven and Cognitive Modulation of Signal Detection. Brain Research Reviews, 48(1), 98-111.
[4] Schultz, W. (2007) Multiple Dopamine Functions at Different Time Courses. Annual Review of Neuroscience, 30, 259-288.
[5] Kastner, S. and Ungerleider, L.G. (2000) Mechanisms of Visual Attention in the Human Cortex. Annual Review of Neuroscience, 23(1), 315-341.
[6] Raichle, M.E. (2015) The Brain's Default Mode Network. Annual Review of Neuroscience, 38, 433-447.
[7] Vaswani, A., et al. (2017) Attention is All You Need. Advances in Neural Information Processing Systems, 30.
[8] Devlin, J., et al. (2018) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.
[9] Dosovitskiy, A., et al. (2020) An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale.
[10] Lake, B.M., et al. (2017) Building Machines That Learn and Think Like People. Behavioral and Brain Sciences, 40, e253.
[11] Lindsay, G.W. (2020) Attention in Psychology, Neuroscience, and Machine Learning. Frontiers in Computational Neuroscience, 14, 29.
[12] Corbetta, M. and Shulman, G.L. (2002) Control of Goal-Directed and Stimulus-Driven Attention in the Brain. Nature Reviews Neuroscience, 3(3), 201-215.
[13] Hassabis, D., et al. (2017) Neuroscience-Inspired Artificial Intelligence. Neuron, 95(2), 245-258.