Why Fake It When You Can Mimic It? (Intelligence)

Abhijay Metekar

Director, Ecosoft Global Pvt Ltd

abhijaymetekar@gmail.com

Bhushan Annokar

Savitribai Phule Pune University

bhushananokar72@gmail.com

Abstract

While significant progress has been made in AI's ability to replicate human behavior, these systems still lack an essential component: the internal **thought processes** that lead to decision-making and problem-solving. This paper proposes a novel approach to AI that focuses on **mimicking** human thinking rather than attempting to **fully replicate** the underlying brain activity. Though we do not fully understand how thoughts are formed neurologically, we can observe the **outcomes** of thinking—decisions, solutions, and actions. By feeding AI systems with **text, voice, and multimodal data**, alongside existing **EEG signals**, we suggest that these systems can **mimic human-like thinking**. This **thoughtful mimicking model** could revolutionize how AI solves problems by training it on the observed outcomes of human thinking, allowing it to replicate human decision-making strategies without the need for understanding the full neurological process.

1. Introduction and Preface

Traditional AI models, particularly in language processing and problem-solving, have made impressive strides. However, they still lack the ability to **mimic** human thought processes in a way that leads to **dynamic problem-solving**. Human thinking is often complex and nonlinear, where the internal cognitive process evolves dynamically in response to new data and contextual changes.

This paper presents an alternative architecture for training AI systems that emphasizes **mimicking human thought**, with the goal of enabling the AI to adapt and problem-solve like humans. By integrating EEG signals, which capture real-time brain activity, with text or image-based input, we aim to create a **"thoughtful mimicking model"** that learns to mimic human thinking and problem-solving behaviors across various domains.

---

2. Proposed Architecture: From Thoughts to Action

2.1. Initial Proposal: Encoder-Decoder Model for Thought-to-Text Conversion

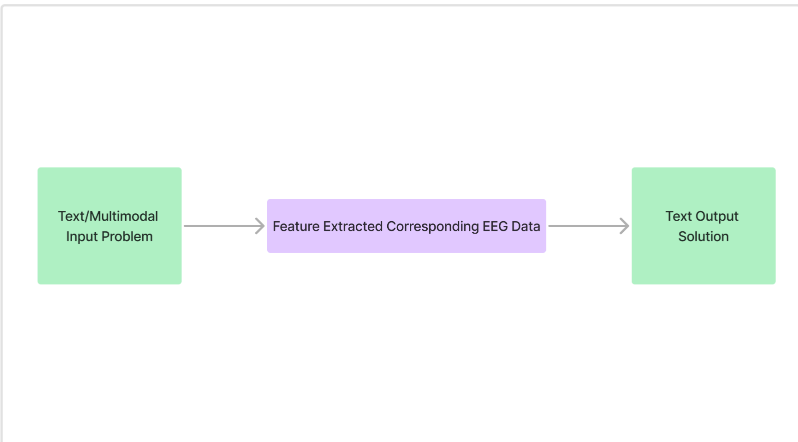
Initially, we proposed an architecture based on an **encoder-decoder model** aimed at converting EEG data into text. This approach would require encoding the EEG signals into an **intermediate representation** (such as tokens), which could then be decoded into actionable outputs, such as text or problem-solving decisions.

While this approach holds promise for understanding thought processes, it comes with the challenge of **semantic interpretation**—trying to map the EEG signals directly to the meaning behind a thought.

2.2. Simplified and Efficient Architecture

Rather than sticking to a complex encoder-decoder model that attempts to **directly translate EEG signals into semantic meaning**, we propose a **simpler, more efficient architecture**. This model is designed to be **agnostic** to the semantic content of the EEG data, focusing on mimicking the **reasoning and decision-making** process.

- **Model Input:** The input consists of a **problem statement** (e.g., a coding problem or any problem-solving task) provided to a group of people with similar **IQ levels**.
- **EEG Recording:** Throughout the problem-solving process, the **EEG signals** are recorded. These signals, representing real-time cognitive processing, are then added to the **model's input vocabulary**.
- **Data Tokenization:** The EEG data is tokenized similarly to textual input, allowing the model to incorporate **brainwave data** alongside textual or image-based input.



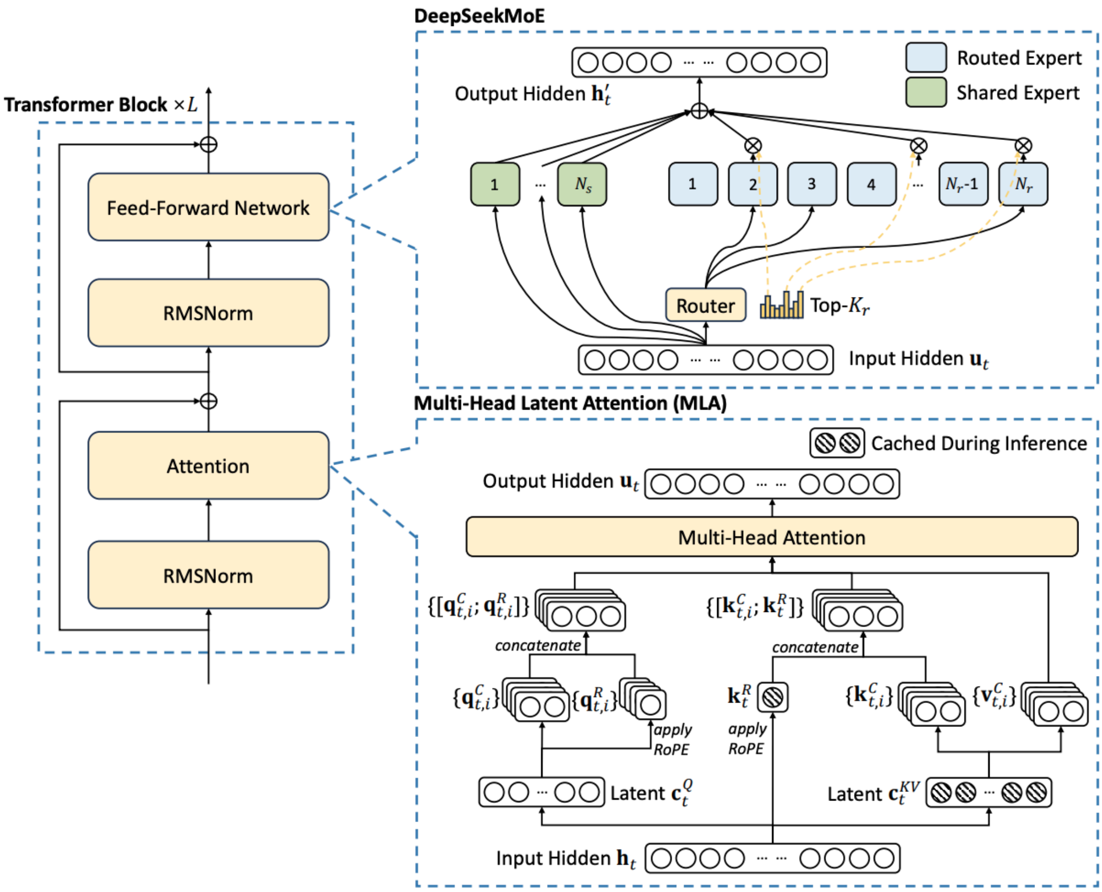
2.3. Key Modification: Context-Aware EEG Integration in Final Logit Calculation

In the final phase of the architecture, when calculating the logits (i.e., predicting outputs), we ensure that the model **effectively utilizes EEG tokens as intermediary representations** rather than as direct vocabulary tokens in the final output. Specifically, the model learns to allocate the **final 30% of its context length** exclusively for generating textual or multimodal data—ensuring that the EEG-based tokens contribute to the reasoning process but do not appear in the direct output.

This structured approach ensures that EEG data **guides** the problem-solving process but does not dominate the final representation, allowing the **true output data (problem solutions)** to remain the focal point. This prevents overfitting on EEG tokens while ensuring that the model leverages cognitive signal patterns to refine its decision-making and reasoning.

2.4. Actual Architecture

- Our proposed model artfully integrates the foundational principles of the **decoder-only Transformer architecture** from "Attention Is All You Need" (Vaswani et al., 2017) with the advanced innovations of **DeepSeek-V3** (DeepSeek, 2024). This synthesis results in a system adept at processing textual data while seamlessly incorporating **EEG-derived temporal features**, thereby capturing the nuances of human cognitive processes.
- Central to our design is the adoption of a **Mixture-of-Experts (MoE) architecture**, encompassing 671 billion parameters, with 37 billion activated per token. This structure ensures that only the most relevant parameters are engaged during each computation, optimizing both performance and resource efficiency (DeepSeek, 2024).



DeepSeek V3 Architecture (DeepSeek, 2024)

- To effectively integrate EEG data, we introduce a specialized **feature extraction neural network**. This intermediary network distills the continuous microvolt fluctuations over time into latent representations, emphasizing features crucial for problem-solving. These latent embeddings are then fed into the Transformer's attention mechanisms, allowing the model to assess the significance of each feature within context.
- A distinctive feature of our architecture is the strategic allocation of the model's context window. We designate the final 30% of this window exclusively for generating the actual output, whether textual or multimodal data. This deliberate partitioning ensures that while the model benefits from the rich, intermediary EEG-derived tokens during processing, the ultimate outputs remain coherent and directly pertinent to the task at hand.
- The training regimen is both rigorous and innovative. Groups of individuals with comparable IQ levels are presented with novel challenges within a specific domain, such as complex coding problems. As they navigate these tasks, their brain activity is meticulously recorded, capturing the nuanced electrical patterns associated with problem-solving. Upon reaching solutions, both the EEG data and the resultant outputs are collected, providing a

comprehensive dataset that encapsulates the journey from cognitive processing to problem resolution.

- This architecture not only mirrors the intricate dynamics of human thought but also exemplifies the potential of integrating neural signals into AI systems. By weaving together advanced Transformer designs with EEG-derived insights, we aim to craft a model that transcends mere simulation of intelligence, resonating instead with the authentic rhythms of human cognition.

2.5 Execution and Implementation:

The first step is to curate the data. To do this, choose a specific problem domain, or opt for a more generic approach. In this case, we focus on coding problems as an example.

A group of individuals with similar IQs within the same domain should be selected, as this ensures the EEG data remains consistent. As mentioned earlier, this approach helps in maintaining uniformity in the overall EEG graphs.

Next, the data is recorded for problem-solving analysis. The problem statements provided should be unforeseen to the participants, allowing critical thinking and genuine problem-solving EEG data to be generated.

Once the solution is found, the EEG data should be scaled to the same x-axis time scale for each individual. This enables the model to process all the data on a similar time frame, thus normalizing them in the time dimension.

It is important to curate a substantial amount of such data. If this is not possible initially, testing can be done with smaller datasets. The more diverse the domains and problems, the richer the critical thinking encoding and learning for the model.

The remainder of the previously outlined architecture is applied, and the training follows the same principles as an LLM, with the addition of the extra steps proposed by our architecture.

2.6. Benefits of This Approach

- **Mimicking Human Thought**: Rather than simply translating EEG signals into text, this architecture focuses on **utilizing cognitive signal patterns as an intermediary step** in problem-solving. The model learns to integrate these signals meaningfully, allowing it to capture **patterns of human reasoning** and adapt its responses accordingly.

- **Optimized for Problem Solving**: By ensuring that the final output space remains focused on problem-solving rather than raw EEG data, the model becomes fully optimized for **real-world applications**. It learns to **solve problems based on observed cognitive processing patterns**, leading to better **mimicry of human decision-making and reasoning skills** while maintaining clarity in output representation.

---

3. Human Cognition: The Importance of IQ and Commonality

3.1. Selecting Participants with Similar IQ Levels

To ensure the model effectively learns human cognitive behavior, we propose that **participants** (individuals who solve the problems) should have **similar IQ levels**. The reason for this is to **reduce the variance** in cognitive processing and maintain a **common cognitive framework**.

By choosing individuals with similar cognitive abilities, we ensure that the **EEG data** collected reflects a **consistent** and **shared problem-solving process**. This consistency is crucial for teaching the model to understand and mimic human thought patterns more accurately.

3.2. Commonality in Thought Process

The model must understand how individuals with similar IQ levels think and process information. If we were to mix data from individuals with widely different cognitive abilities, it would be harder for the model to discern consistent cognitive patterns.

This approach allows the model to **mimic not just one individual's thought process**, but the combined reasoning abilities of the selected group. This **superposition of thought patterns** enables the model to develop a more **generalized understanding** of human cognition.

---

4. Mimicking Superimposed Thought Processes

4.1. Merging Diverse Thought Patterns

One of the critical outcomes of this approach is that the model does not simply **inherit** the cognitive behavior of any single individual. Instead, it learns to **superimpose** the cognitive patterns of various individuals, creating a more robust, thoughtful **mimicking model**.

For example, if the model learns from both **my thinking patterns** and those of a renowned expert (e.g., Stephen Hawking), it would not just inherit my thought process or the expert's, but it would develop the ability to **mimic both thought patterns** simultaneously. This is akin to merging two distinct cognitive processes into a single unified problem-solving approach, which is a **more flexible and adaptive** way of thinking.

4.2. Theoretical Implications of Superimposed Thought

By training on a set of people with similar cognitive abilities and superimposing their thought processes, the model has the potential to become much more capable than traditional models. It can dynamically adapt to new information, solve problems more efficiently, and apply different reasoning strategies as needed.

Similar to how holograms appear to be real but are merely visual projections tricking the human brain into perceiving them as physical entities, the imitation of human thought processes in AI can function in a comparable way. While the AI may not actually experience thought in the human sense, its ability to mimic human cognitive behavior could lead to a perception of it 'thinking,' even though it exists as a sophisticated imitation rather than genuine human-like thought. The model's **superimposed cognition** enables it to mimic how humans would think and respond in various problem-solving scenarios, giving the impression of thought, even though it's based on imitation rather than original cognitive processes.

4.3. Continuous Learning with Reinforcement

To further **fine-tune** the model's ability to **mimic human thought processes**, we introduce the idea of **reinforcement learning** as a **continuous feedback loop**. This approach allows the model to **evolve** its problem-solving strategies over time, much like the way human brains adapt and improve with experience.

In reinforcement learning, the model learns by interacting with its environment (the problem-solving tasks) and receiving feedback based on the actions it takes (i.e., the **correctness** or **efficiency** of its solutions). By constantly adjusting its internal processes, the model improves its **decision-making and reasoning abilities**, enabling it to solve more complex tasks as it encounters them.

4.4. Metacognitive Mode: Problem-Solving Roulette for Continuous Evolution

To keep the model's **thinking behavior evolving**, we propose a **problem-solving roulette** approach. Here, the model doesn't just solve a fixed set of problems; rather, it is continuously provided with **new, unforeseen challenges** by another model that actively **surfs the internet** to discover novel problems.

This mechanism ensures that the model is not static or stuck in solving a predefined set of tasks. Instead, it operates in a **metacognitive mode**, where it is constantly made to **think about thinking**—reflecting on its own problem-solving strategies, adapting its reasoning processes, and refining its cognitive behavior over time. This mimics the **dynamic nature of human cognition**, where the brain is continuously presented with new stimuli, problems, and learning opportunities.

By integrating **metacognition** into the problem-solving roulette, the model develops the ability to **assess, monitor, and refine its own problem-solving approach**, rather than simply applying fixed strategies. This **self-reflective learning** ensures continuous improvement, much like how expert humans refine their problem-solving techniques through experience and self-evaluation.

While this dynamic learning process requires high computational resources, the cost is justified by the **emergence of an AI system that evolves and improves autonomously**—not merely imitating human thought but actively **thinking about how to think**. By combining **reinforcement learning** with metacognitive modeling, the system can **continuously enhance its ability to solve increasingly complex and diverse problems**, pushing the boundaries of AI cognition.

---

5. Conclusion, Limitations, and Future Directions

5.1. Conclusion

The proposed approach offers a novel method for training AI systems that can **mimic human thought** by focusing on **problem-solving behavior** rather than trying to replicate the precise neural mechanisms of

thought. By utilizing **EEG signals**, **reinforcement learning**, and **problem-solving roulette**, the system is capable of continuously adapting and improving its problem-solving skills. The outcome is a **dynamic model** that can handle an increasingly diverse range of tasks, making it more robust and flexible than traditional AI systems.

5.2. Limitations

While the proposed model demonstrates significant potential, several limitations must be considered:

- **High Computational Cost:** The reinforcement learning process and continuous problem-solving mechanism require substantial computational resources, making large-scale deployment expensive.

- **Data Availability:** The model's reliance on extensive, labeled EEG data poses a challenge for scalability, as acquiring high-quality, domain-specific cognitive datasets remains a constraint.

- **Ethical Implications:** The continuous evolution of AI models raises concerns about autonomy, control, and potential unintended consequences, necessitating careful oversight.

- **Context Retention Bottleneck:** Like existing AI systems, a major limitation is carrying forward contextual information across extended problem-solving sessions. To address this, we propose the continuous updating of a **Retrieval-Augmented Generation (RAG) database** dedicated to storing and retrieving "thoughts" separately. This mechanism enables the integration of relevant cognitive context into problem statements dynamically. However, this approach does not fully resolve the inherent context length limitations of Transformer architectures, marking it as an open area for further research.

5.3. Future Directions

Future work should focus on optimizing the computational efficiency of this model while expanding its ability to handle real-world problems. Additionally, further research into understanding how **superimposed cognitive processes** might lead to enhanced problem-solving across various domains would be critical to advancing this approach.

References:

- A. Vaswani et al., *Attention is All You Need*, 2017.
- DeepSeek AI, *DeepSeek-V3: Advancements in Reinforcement Learning and Latent Attention*, 2024.