

# Beyond Deterministic AI: A Probabilistic Multi-Path Model for Subconscious-Like Thinking

Saeed Sabzeh

B.Sc. in Computer Science

[Saeed.Sabzeh.80@gmail.com](mailto:Saeed.Sabzeh.80@gmail.com)

March 2025

## Abstract

Current AI models, particularly large language models (LLMs), operate on deterministic or probabilistic token selection, where the highest-probability next token is chosen while lower-probability alternatives are discarded. This paper proposes an alternative multi-path probabilistic branching approach, where instead of selecting a single highest-probability token, the AI assigns attention weights to multiple possible branches and lets them evolve in parallel. This structure simulates subconscious thought processes, where lower-probability ideas persist and can influence final decision-making. We explore how this mechanism aligns with human intuition, emotional responses, and the "shotgun effect" described in *Thinking, Fast and Slow* by Daniel Kahneman. The proposed approach could lead to AI with emergent intuition, creative thinking, and deeper decision-making processes beyond linear next-token prediction.

## 1. Introduction

Most AI models today, including GPT-based LLMs (Brown et al., 2020), follow a single-path probabilistic model—choosing the next token based on the highest probability. However, human thought does not function this way. Our minds frequently explore multiple possibilities at once, many of which are subconscious, before a final conscious thought emerges (Baars, 1997; Kahneman, 2011). This leads to:

- Unconscious intuition influencing decisions before logical reasoning catches up.
- Creative insights appearing unexpectedly from background thought processes.
- Emotional biases affecting how certain ideas gain prominence in the mind.

This paper proposes a multi-branch probabilistic LLM architecture, where low-probability token paths are given weighted attention instead of being ignored. These paths persist as subconscious-like thoughts, competing for dominance over time. This approach can:

- Simulate intuition and gut feelings by allowing weakly probable ideas to surface later (Gladwell, 2005).
- Generate emotion-driven responses by reinforcing past-weighted experiences (Damasio, 1994).
- Model the "shotgun effect," where multiple ideas fire simultaneously before conscious thought forms (Kahneman, 2011).

This work suggests that by allowing AI to explore parallel thought paths, we can move closer to a model of subconscious cognition, bridging the gap between AI reasoning and human-like intuition.

## 2. The Proposed Multi-Path Probabilistic Model

## 2.1 Current LLM Limitations

- Next-token selection is linear, meaning once a token is chosen, other possible paths are discarded (Vaswani et al., 2017).
- Lower-probability tokens are not explored further, leading to rigid and overly rational responses.
- LLMs lack a mechanism for background thinking—everything is forward-driven and explicit.

## 2.2 Multi-Path Token Selection

Instead of eliminating low-probability tokens, this model assigns:

- **Primary Focus (PF):** Highest-probability token (conscious thought).
- **Subconscious Paths (SP):** Lower-probability tokens that continue with reduced weight.
- **Delayed Activation (DA):** Mechanism for reintroducing subconscious paths into conscious output.

Each SP token path continues to generate potential completions in parallel, creating branching thought streams that persist beyond immediate selection.

## 3. Cognitive Science Perspective: Alignment with Human Thought

### 3.1 Subconscious Thought and Intuition

- The human brain does not think linearly; multiple thoughts emerge before one is consciously chosen (Baars, 1997).
- Weakly probable thoughts remain active in the background, influencing later decisions (Damasio, 1994).
- This explains gut feelings, where subconscious cues surface before logical reasoning is applied (Gladwell, 2005).

By allowing lower-probability paths to persist, AI can mimic this subconscious intuition, generating responses that feel less mechanical and more naturally insightful.

### 3.2 The Shotgun Effect (*Thinking, Fast and Slow*)

- Kahneman (2011) describes how the brain fires multiple related ideas at once, only some of which reach awareness.
- This happens before logical analysis begins, allowing spontaneous insights to occur.
- In our AI model, this happens through multi-path token exploration, where subconscious thoughts emerge dynamically.

### 3.3 Emotional Reinforcement in Thought Weighting

- In humans, emotions shape how certain thoughts gain strength over time (Damasio, 1994).

- If a past experience was emotionally intense, similar thoughts get higher subconscious weighting.
- In AI, this could be achieved by reinforcing paths associated with past high-impact outputs, leading to more emotionally aware AI.

## 4. Applications and Implications

### 4.1 AI That Thinks in Layers

- Traditional AI is linear, but this model allows thought layers to persist, influencing future outputs.
- AI could generate unfinished ideas that develop over time, leading to more nuanced responses.

### 4.2 Enhanced Creativity and Divergent Thinking

- Instead of a single answer, AI can generate competing possibilities, simulating human brainstorming (Runco, 2014).
- Could be used in creative writing, strategic decision-making, and abstract reasoning.

### 4.3 Intuitive AI Decision-Making

- AI could develop a sense of intuition, favoring certain paths based on experience rather than immediate logic.
- This could improve human-AI interactions, where AI responds with a more natural "gut feeling."

## 5. Future Research and Challenges

- **Computational Cost:** Maintaining multiple active token paths increases complexity.
- **Managing Noise:** Not all subconscious paths are useful—filtering mechanisms must be optimized.
- **Ethical Considerations:** AI that mimics subconscious thinking could influence decision-making in unpredictable ways.

## 6. Conclusion

This paper presents a multi-path probabilistic branching model for AI that allows subconscious-like thought emergence. Unlike traditional LLMs, this model maintains multiple parallel thought paths, weighted by probability, which persist and influence later decisions. This approach aligns with:

- Human subconscious processing, where ideas form before reaching awareness.
- Kahneman's "shotgun effect," explaining how multiple ideas fire simultaneously.
- Emotional reinforcement mechanisms, allowing AI to mimic intuitive decision-making.

By embracing this model, AI research moves beyond deterministic token selection and into the realm of emergent, self-evolving thought processes, marking a significant step toward true human-like cognition.

## References

- Baars, B. J. (1997). *In the Theater of Consciousness: The Workspace of the Mind*. Oxford University Press.
- Brown, T., et al. (2020). *Language Models are Few-Shot Learners*. NeurIPS.
- Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. Penguin.
- Gladwell, M. (2005). *Blink: The Power of Thinking Without Thinking*. Little, Brown.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.
- Runco, M. A. (2014). *Creativity: Theories and Themes: Research, Development, and Practice*. Academic Press.
- Vaswani, A., et al. (2017). *Attention is All You Need*. NeurIPS.