AI Mindscape: The Intersection of Cognitive Neuroscience and Artificial Intelligence

Neuroscience as Model for AI ITAI-4374

Prof: Patricia McManus

Sumesh Surendran

April 8, 2025

Section 1: Neuroscience Concept Definition

Predictive coding is a theory in neuroscience proposing that the brain is fundamentally a prediction machine. It constantly generates expectations about incoming sensory data based on an internal model of the world and compares these predictions to the actual sensory input (Huang & Rao, 2011). In essence, perception is viewed as a process of hypothesis testing—where higher-level areas send down predictions of what lower-level sensory areas should perceive, and only the mismatch, or *prediction error*, is propagated upward. This ongoing feedback loop allows the brain to update its mental model of the environment considering new evidence (Rauss & Pourtois, 2013).

Each level of the cortical hierarchy attempts to predict the level below it, with only unpredicted residuals (prediction errors) transmitted forward (Huang & Rao, 2011). This aligns with the Bayesian brain hypothesis, where prior expectations are combined with sensory evidence to construct perception (Rauss & Pourtois, 2013). For instance, phenomena such as the brain "filling in" the blind spot or being tricked by optical illusions can be explained through predictive coding—the brain’s top-down expectations strongly influence what we consciously perceive.

Section 2: Connection to Artificial Intelligence

The concept of predictive coding has influenced the design of artificial intelligence (AI), especially in deep learning and sequence modeling. AI systems inspired by this concept are built to predict inputs and adjust their internal representations based on prediction error, mirroring how the human brain processes sensory data (Lotter et al., 2017). This is particularly evident in unsupervised learning models and recurrent neural networks used for predicting video frames, text, or future states in an environment.

A clear example is the **PredNet** architecture developed by Lotter, Kreiman, and Cox (2017), a deep learning model inspired directly by predictive coding. Each layer in PredNet tries to predict the activity of the next input, and prediction errors are passed upward to adjust the model. This mirrors how the brain transmits only mismatches between prediction and reality. PredNet showed strong performance in predicting future frames in video data and understanding driving scenes, supporting the efficiency of prediction-based learning.

Additionally, predictive coding has inspired alternatives to backpropagation, such as **predictive coding networks (PCNs)**. These use local learning rules based on minimizing prediction errors at each layer, offering a biologically plausible alternative to gradient descent (Lee et al., 2022). Such models have demonstrated robustness in incremental and few-shot learning tasks and reduce issues like catastrophic forgetting, showing practical benefits over traditional training approaches.

Section 3: Exhibit Proposal (Conceptual Explanation)

Exhibit Title: *Predictive Minds: How Our Brains and AI Anticipate the World*
Format: Augmented Reality (AR) Experience

Content Summary

This AR exhibit immerses visitors in an experience where they can see how both human brains and AI systems make predictions and adjust them based on real-time feedback. Visitors wear AR glasses or use handheld devices to interact with real-world objects overlaid with visualizations of predictions. For example, the system might show a partially hidden object and overlay what the brain likely predicts it to be, contrasted with an AI model's best guess.

When the actual object is revealed, the AR interface highlights discrepancies between prediction and reality, representing **prediction error**. This demonstrates the active role of the brain in constructing perception and mirrors how AI systems are designed to adjust internal parameters in response to error, just as in **PredNet** or PCNs (Lotter et al., 2017; Lee et al., 2022).

Interactive Component

An activity within the exhibit allows visitors to tap out a rhythm or draw a pattern on a touchscreen. The AR system then displays predictions from both a human “brain” avatar and an AI model in real-time. When the pattern is disrupted, visitors see both systems make incorrect predictions and watch how quickly each adjusts. This interactive setup reinforces the core idea of learning from prediction errors.

A secondary experience involves optical illusions—such as the **hollow mask illusion**—where the brain's predictive model overrides actual sensory input. Visitors can view these illusions side-by-side with how AI systems interpret the same visuals (typically correctly, as they lack prior assumptions). This highlights both the strengths and limitations of human and artificial perception (Huang & Rao, 2011).

Visual Elements

Visualizations include:

- Diagrams showing hierarchical predictive flows in the brain (top-down and bottom-up).
- Neural network graphics show prediction and error layers.
- AR overlays showing side-by-side predictions and outcomes.

These make complex mechanisms more intuitive by using accessible symbols (colored arrows, animated errors, etc.) to represent the flow of predictions and adjustments.

Section 4: Reflection and Justification

I selected predictive coding as the focus of this exhibit because it captures a powerful, central idea shared by both neuroscience and AI: perception is not passive, but active prediction. This concept elegantly connects how our brains construct reality and how machines are being trained to understand the world. Predictive coding is conceptually rich yet intuitive enough to be demonstrated visually and interactively, making it ideal for a public exhibit.

By using augmented reality, the exhibit can turn invisible brain processes into something visible, interactive, and memorable. AR also allows for real-time user engagement and side-by-side comparisons of brain and AI behaviors. The inclusion of illusions and error-driven learning experiences supports better public understanding of how perception and learning work biologically and computationally.

From a practical standpoint, predictive coding provides a framework for developing more adaptive, efficient, and human-like AI systems. Public understanding of this paradigm fosters awareness of AI capabilities and limitations, as well as empathy for how the brain sometimes "fails" (e.g., hallucinations or illusions). This exhibit empowers visitors to think critically about their own perception—and that of machines—and makes abstract concepts feel personal and relevant.

References

- Huang, Y., & Rao, R. P. N. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(5), 580–593. https://doi.org/10.1002/wcs.142
- Rauss, K., & Pourtois, G. (2013). What is bottom-up and what is top-down in predictive coding? *Frontiers in Psychology, 4*, 276. https://doi.org/10.3389/fpsyg.2013.00276
- Lotter, W., Kreiman, G., & Cox, D. (2017). Deep predictive coding networks for video prediction and unsupervised learning. *arXiv preprint* arXiv:1605.08104. https://arxiv.org/abs/1605.08104

- Lee, J., Jo, J., Lee, B., Lee, J.-H., & Yoon, S. (2022). Brain-inspired predictive coding improves the performance of machine-challenging tasks. *Frontiers in Computational Neuroscience, 16*, 1062678. https://doi.org/10.3389/fncom.2022.1062678
- Wikipedia contributors. (2025, January 10). *Predictive coding*. In Wikipedia. https://en.wikipedia.org/wiki/Predictive_coding