

Cognition and Artificial Neural Nets: A Propitious Encounter

Karthik Desingu*

Department of Computer Science and Engineering, Anna University, India

***Corresponding Author:** Karthik Desingu, Department of Computer Science and Engineering, Anna University, India.

Received: July 12, 2023; **Published:** August 30, 2023

Neural network models have been exploited since the late '80s and '90s to solve real-world problems, based on the then contemporary developments such as the back propagation learning algorithm. Over the last decade, the availability of massive data sets, enhanced computational resources, and further developments in algorithms have led to an explosive growth in the use of neural networks in machine learning and artificial intelligence. They showcase the state-of-the-art performance in artificial intelligence applications, all the way from speech comprehension and image recognition to cross-modality translation. In parallel, however, an analogously related perspective, elegantly relating the spheres of cognitive neurobiology and machine learning, has spurred interest. All the more, both sides of this stance seem to entail far-reaching implications.

Given the excellent performance of contemporary research in neural networks on real-world AI objectives, neural network models seem like highly potent tools for modeling human cognition and revamping our understanding of complex brain functions. For instance, scientists have long struggled to comprehend the reasons behind the specializations within the brain for various tasks. They have wondered, not just why different parts of the brain do different things, but also why the differences can be so specific: say, why does the brain have an area for recognizing objects in general, but also one for faces in particular? Deep neural networks now are showing that such specializations may be the most efficient way to solve problems in perception. Computational neuroscientists are finding that deep neural networks can be good explanatory models for the functional organization of living brains. But the implications of such models are not limited to enhanced understanding of brain function. Neural models can be pivotal to stimulate further cognitive computational research using the tools, methods, and insights drawn from AI. Recently, advances have been significant in developing biologically grounded cognitive theories and in mechanistically explaining, on the basis of these brain-constrained neural models, hitherto unaddressed issues regarding the nature, localization and ontogenetic and phylogenetic development of higher brain functions. A range of neural models including localist, auto-associative, hetero-associative, deep and whole-brain networks have been developed to this effect.

To address these ambitious goals, and to forage ahead for a representative model in cognitive neuroscience, however, the underlying AI techniques and architectures need an extent of neurobiological realism. Although neural networks have advanced dramatically in recent years, and have even managed to achieve human-like performance on complex perceptual and cognitive tasks, their similarity to aspects of brain anatomy and physiology is largely imperfect. There is a need to improve their biological plausibility. Moreover, neural network models of cognition today are largely limited to small synthetic domains, and are heavily based on architectures from the late 20th century. A comprehensive approach to progress in this direction can advance brain-constrained modeling, and can consequently even have broad-ranging clinical and theoretical culminations in neuroscience.

On the flip side of this, rather natural analogy, lies an exigency to understand and explain the behavior of deep neural networks. It remains one of the most important challenges of modern deep learning solutions — an entire field of research that concerns *explainability and interpretability* of neural networks. This quest for improving the interpretability of deep learning models has moved one strain of researchers to draw inspiration from biology, neuroscience and psychology. The case for using cognitive psychology to explain neural networks is particularly inspiring and intuitive, given the deep-rooted organic origins of the very core of artificial neural networks. Conceptually, cognitive psychology attempts to bolster the understanding of different processes of the mind, ranging from

attention, language use and memory to perception, problem solving, creativity, and thinking in general. Although some of the earliest ideas behind cognitive psychology can be traced back to the 1600s, it wasn't until the 1950s when American psychologists sparked heterodox ideas that confronted the bases of the dominant school of thought at the time, namely behaviorism, and adopted a different model of the mind based on advancements in neuroscience. The term cognitive psychology is attributed to the German-American psychologist, Ulric Neisser. According to him: "... it is apparent that cognition is involved in everything a human being might possibly do; that every psychological phenomenon is a cognitive phenomenon ...". This relation forms the central subject of a remarkable line of research.

At DeepMind, a subsidiary of Alphabet Inc, for instance, a research team borrows the realism behind this perception and posits an eventual line of explainability techniques based on biological cognition to understand decisions of neural networks. Specifically, the team focuses on a cognitive model called *one-shot learning* that explains the remarkable ability of humans to guess the meaning of a word after seeing just one specimen. This ability, in humans, is believed to be based on the discriminative application of some very specific inductive biases: whole object bias that a word refers to a whole object rather than a part of it; *taxonomic* bias that a name refers to the shallowest level of classification of an object; and shape bias that a noun is influenced most by its shape, as opposed to other characteristics like texture, size or color. A closely related concept in deep learning is (also) called one-shot learning, and it draws on concepts from a class of computational learning techniques called *few-shot learning and meta-learning*.

The experiments at DeepMind tested a hypothesis that one-shot learning architectures might also be based on similar inductive biases. The team designed a curated dataset to investigate the physical feature -based biases of two state-of-the-art family of neural networks — *Inception and Matching networks*. Not only did these experiments validate the hypothesis, but they also exhibited other glaring similarities with human cognition! Shape biases in neural networks emerged gradually over the course of early training. This is densely reminiscent of the emergence of shape bias in humans: young children show smaller shape bias than older children, and adults show the largest bias. All the more, akin to how psychologists draw inferences based on multiple trials, and don't rather rely on a single subject, the levels of bias in neural networks positively correlates with the sample size used to train them: larger and multiple training experiments are imperative to drawing sustainable conclusions about these architectures. Such striking similitude has encouraged the exploration of psychological techniques to reveal and represent a better understanding of neural networks, and has formed a key element of AI explainability research.

In essence, a *symbiotic relationship* embodies at the intersection of the understanding of artificial neural networks, and well, the biological neural networks! Progressive understanding and deciphering of one, benefits the other in achieving far-reaching and wide-ranging applications. Understanding and unshathing from cognitive learning techniques has proven focal to *interpret* the black-box that deep neural networks are, whilst the very close-resembling complexity of these neural networks with the biological brain, and their successful utility in real-world applications make them perfect candidates to uncover cognitive mysteries. A propitious encounter indeed.

Volume 5 Issue 3 September 2023

© All rights are reserved by Karthik Desingu.