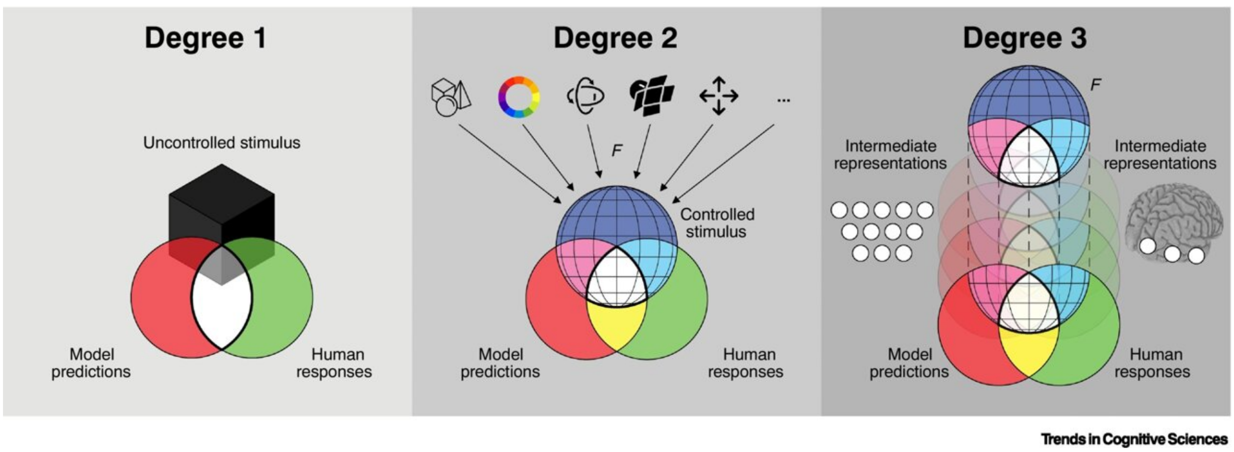
Deep neural networks have become increasingly powerful in everyday real-world applications

October 10 2022



Three degrees of increasingly constrained algorithmic equivalence. First degree: typical equivalence (white intersection) tested between human responses (green) and 2D images of real-world 3D faces, bodies, objects, and scenes, and deep neural network (DNN) predictions (red). Such images typically form an experimental 'black box' because we cannot control and therefore explicitly test how real-world 3D features (projected on these 2D images) cause similar responses between humans and DNN models (white intersection). Second degree: a stimulus model generates 2D images from variations of its generative features F (blue), and provides a coordinate system to chart and systematically explore the match between human and models. Such generative psychophysics controls the images, thereby enabling decomposition of human and DNN responses into: (A) the white triple intersection of similar human and DNN responses to the same controlled 3D F, (B) the cyan remainder of human responses to stimulus features F that the DNN model cannot predict, (C) the magenta remainder of DNN predictions from F that are dissimilar to human responses, and (D) the crucial yellow remainder flagging that the F of this particular stimulus model fails to account for the total similarity of human and DNN behavior (suggesting testing of other F from other generative stimulus models). Third degree: equivalence of the brain and DNN algorithms (across schematized brain and DNN layers) whose similar computations process the same stimulus features F to produce the same responses. Credit: *Trends in Cognitive Sciences* (2022). DOI: 10.1016/j.tics.2022.09.003

Researchers use deep neural networks, or DNNs, to model the processing of information, and to investigate how this information processing matches that of humans.

While DNNs have become an increasingly popular tool to model the computations that the brain does, particularly to visually recognize real-world "things," the ways in which DNNs do this can be very different.

New research, published in the journal *Trends in Cognitive Sciences* and led by the University of Glasgow's School of Psychology and Neuroscience, presents a new approach to understanding whether the human brain and its DNN models recognize things in the same way, using similar steps of computations.

Currently, deep neural network technology is used in applications such as face recognition, and while it is successful in these areas, scientists still do not fully understand how these networks process information.

This opinion article outlines a new approach to better this understanding of how the process works: first, that researchers must show that both the brain and the DNNs recognize the same things—such as a face—using the same face features; and, secondly, that the brain and the DNN must process these features in the same way, with the same steps of computations.

As a current challenge of accurate AI development is understanding whether the process of machine learning matches how humans process information, it is hoped this new work is another step forward in the creation of more accurate and reliable AI technology that will process information more like our brains do.

Prof Philippe Schyns, dean of research technology at the University of Glasgow, said, "Having a better understanding of whether the human brain and its DNN models recognize things the same way would allow for more accurate real-world applications using DNNs.

"If we have a greater understanding of the mechanisms of recognition in human brains, we can then transfer that knowledge to DNNs, which in turn will help improve the way DNNs are used in applications such as facial recognition, where there are currently not always accurate.

"Creating human-like AI is about more than mimicking human behavior—technology must also be able to process information, or 'think,' like or better than humans if it is to be fully relied upon. We want to make sure AI models are using the same process to recognize things as a human would, so we don't just have the illusion that the system is working."

The study, "Degrees of Algorithmic Equivalence between the Brain and its DNN Models" is published in *Trends in Cognitive Sciences*.

More information: Philippe G. Schyns et al, Degrees of algorithmic equivalence between the brain and its DNN models, *Trends in Cognitive Sciences* (2022). DOI: 10.1016/j.tics.2022.09.003

Provided by University of Glasgow