I have this suspicion that I'll never understand much about my mind. The secret of our cognition is a mental tail chase that eluded centuries of thinkers. Yet critically, each successive generation builds upon their predecessors with fresh perspectives. Each generation will think they are closer than ever before.

We are closer than ever before.

The key to intelligence is Deep Learning (DL). It would fail by itself, but cognitive science provides it a foundation. Artificial general intelligence (AGI) should be explicitly modeled after our brains. Since DL approximates neural networks, it will support AGI. As a backbone it addresses the problems of cognitive philosophy and can be extended by fields like neuroscience and network science. Thus, cognitive science, by explaining our only proof of general intelligence, can bring DL to AGI.

True, AGI is distinct from human cognition. Yet, cognitive science grounds the otherwise intractable; we're hard pressed to invent an intuitive intelligence better than our own. Rather, it's likely that in our quest to understand our own brains, we'll stumble across insights that will solve AGI. Our higher cognition has only developed recently compared to the massive systems behind our primal senses, and the latter are already beginning to be faithfully replicated by DL systems. Of course, there's much work remaining to be done, but with fields like cognitive neuroscience, we can identify neural correlates for individual cognitive functions and approach intelligence systematically.

Current research constraints are practical, not technical. Labs lack funding and manpower, but not for long. "Narrow" DL research focuses on specific, usually commercial applications such as face recognition and simple translators, but scope is expanding. Computer vision has conquered classification, moving to object recognition, even semantic segmentation, in ways that mirror our understanding of human vision. Cross-modality efforts have produced embodied agents that answer questions in simulated environments with natural language. More biologically plausible models are developing in audio, and meanwhile performance benchmarks across the board creep upward. Tech giants, with all their monetary might, will inevitably consider AGI more seriously. Sizable efforts have already committed to the challenge.

Moreover, AGI is becoming popular again. For context, enthusiasm for AGI has come in cycles, once at the outset of modern AI systems in the 1960s, and again before the AI winter in the 1980s, when cognitive models such as SOAR and ACT-R were built. Interest is now booming, evidenced by the ubiquity of intelligent agents in popular culture (i.e. Jarvis in Iron Man) and the acclaim that followed

recent titles such as Superintelligence and The Singularity is Near. As digital assistants become pervasive on mobile phones and now home speakers, AGI will only be more prominent.

However, any attempt on intelligence should define its goals. Early philosophy establishes several broad question: mind's relation to body, the vast frontier of consciousness, and knowledge acquisition. We can consider DL in each of these contexts. I'm a physicalist and a reductionist. There is no reason for ideal forms or a disembodied mind (that mysticality reminds of God, which as from the NPR podcast is speculated to have evolutionary roots - I'm also an atheist). As for reduction, as Dennett phrased, consciousness is an emergent illusion. Functionalism is a useful interpretation for cognitive scientists, but everything reduces to a physical basis within us (probably not limited to our brains, but I digress). We needn't dance around the constraints of high level functionalism, as it's absurd that a sum shouldn't reduce into its constituents. The whole point of complexity theory is to disentangle the chaotic interactions, not dismiss it offhand as a black box.

Distributed representations are an irrefutable basis for physicalism. Consider the functionalist challenge to reconcile the infinite states of mind with the finite states of brain. However, while localized representations would bound our knowledge by some function on our number of neurons, distributed representations are continuous and have infinite space. Better yet, distributed representations are not static, and changes in frequency distribution comprise yet another factor of representational power, evidenced by the way locusts use frequency distributions to encode odor, as detailed in the Network approach. Though local representation is how we feel we perceive, it is highly unlikely. fMRIs visibly indicate how the mind is constantly globally activated.

Generally, we must restrain our instinctive protests and rely on neuroscientific evidence. Philosophy highlights how speculative reasoning and intuition alone makes little headway. DL therefore tables the debate over basis of mind by being behavior based — which is good. Not all researchers have to work on interpretability. It is both brave and reckless to relinquish the safety of complete understanding, but that step bottlenecked earlier thinkers. Around a decade ago, the computer vision community struggled to design complex visual features, as proposed in the pandemonium feature detector model, failing to find performant features that "made sense." In DL, many models converge on intermediate features that aren't intuitive to us. We have the luxury of simulating emergent phenomena, and our task is reduced to iteratively designing and understanding more complex hierarchies. Interpretability's role is for eventual reverse engineering and reduction of intelligence, but it's no longer a requirement. The ability to dissect

our systems is a unique opportunity to analyze what is actually essential for recognizable intelligence. The mind does emerge from body, and we're past philosophy.

Now, there's the harder problem of consciousness. The main philosophical contention with AGI is summarized by strong vs weak AI as posited by Searle in the Chinese Room argument. However, this explanatory gap is both insurmountable and contrived, much like idealism is a vacuous form of monism. Weak AI that acts entirely intelligent, protesting its own consciousness, would pass any fair test.

Weak AI consciousness is still difficult, since DL does not yet have good memory, which is likely essential to a continuous experience that we could call consciousness, to quote Clive Wearing. As more complex agents begin operating on DL-based sensory modules, we can experiment with memory, inspired by research on areas of the temporal lobe anomalous in Wearing's brain. Consider a simplified version of at least a sensory consciousness that could be the basis of a new DL model. We attend to certain aspects of our perception, constituting our qualia. By oversimplifying Hebb's rule, this representation ties together the actual senses along with any relevant abstract patterns, such as emotion. The hippocampus could replay these sensations, binding them to some (distributed) recall trigger, likely as we sleep. Such a hypothesis could eventually be tested with precise recording techniques. In retrieval, the trigger is activated, chaining to the associated sensory distribution and reintroducing our qualia. We relive these activation patterns every time we remember, as evidenced by how mental imagery activates visual sensory neurons. This is essentially what semantic networks, but we have to remember that these "memory nodes" are actually dense distribution patterns, perhaps 'locally' represented by large, percolating cell assemblies. The exact mechanics should be eventually explained by network science.

Thus network science research is critical for architecting higher order DL solutions. We will need to confront how functionally distinct regions exist naturally in the brain. Precise localization is unlikely, as we can see in revised models of our attentional mechanism in the Neuroscience approach. There, several components, such as the colliculus or thalamus, work together, but can loosely substitute for each other. This evidences extreme plasticity and domain generality. Unfortunately, such systems are incompatible with current stitched together DL agents. Yet these agents are in their infancy, and the more we impose communication pathways and introduce bias, the less they think for themselves. Such pathways would ideally be learned. Comparative neuroscience in primates or even rats ought to offer insight into how we should connect cognitive modules. We should start with known pathways, as the where/what pathways in audio or vision. Standardization of techniques like brain morphometries will identify how pathways

manifest themselves in different people. It is promising that with people all developing differently, only a small fraction developing visibly manifested disorders. This implies our cognitive modules must be large functional units and resistant to small design errors, connected in a stable small-world network with signal highways across the brain.

We're naturally pointed to research on cognitive development, and this leads to the problem of knowledge acquisition. Current systems converge to stable solutions, but much of the reinforcement learning community is focusing on lifelong or online learning, such that deployed systems can continuously evolve and improve themselves from constant input. This research could plausibly evolve into a system like the brain's, since the brain constantly receives sensory input. Neuroscience can investigate how the brain self-stimulates and persists activity, and determine how long signals persist in sensory deprivation. Simultaneously, the stability-plasticity tradeoff is driving research on memory representation, as discussed earlier. Hofstadter, a prominent thinker in AGI, provides a detailed argument for how we think by analogy, and argues we build up more complex ideas over time by compounding analogies. Perhaps we could create a model that forcibly compresses its own representations as it learns more. I'm optimistic about the time it will take, perhaps, but simply by considering our brain, some biological technique must solve these problems, which we can use to even outperform our limited "hardware."

Evolutionary perspectives encountered while researching development will raise a more fundamental, auxiliary question for researchers. How does the human genome encode the complexities of different cognitive modules such as memory and problem solving, forcing consistent development across individuals? Why is the left brain responsible for the sense of self, the concrete current, and the right brain responsible for tuning into the world, in almost everyone (according to Jill Bolte Taylor)? DNA determines an approximate connectome and must influence the LTP process, controlling plasticity itself. This is clear in the linguistic perspective. Lindbergh claimed a deadline for acquiring our innate language ability else it would be crippled. But even though language functions like grammar can be pruned away, regions involving concept representation remain plastic. Primates like Koko and Kanzi indicate they have such regions, since they mainly lack grammar control. Separation of knowledge from grammar also explains why Genie learned representations but not grammar. Similarly, convergently evolved language understanding of parrots and dogs are also evidence that our 'unique' communicative abilities must only be a short development off of what most animal brains have. Our genetic language abilities, including universal grammar, were driven by social pressures for communication. Meanwhile, concept representation must be developed since abstract concepts can only be culturally shaped. It's not apparent

if understanding the influence of DNA is necessary for AGI, but it certainly seems valuable. Somehow, our genome has specified our connectome to develop into these hypercomplex DL systems, such as language regions reflecting NLP architectures, or the occipital lobe reflecting VGG16 architecture.

What does this mean for the immediate future? Older, knowledge-based programming and designed cognitive architectures will be outpaced. For example, take Simon's computational model INFANT. In making formal problem solving agents we try to emulate the brain as a creative but still rigid logic system, which is biologically infeasible. The value of deep learning is that we needn't understand all we model. The largest debate will be deciding what cognitive functions are sufficient for recognizable intelligence. In a system like BabyX, the simulated baby is purportedly powered by DL architectures, which is good. However, BabyX also invests effort into simulating chemical responses, and is building completely bottom-up. Cognitive neuroscience can lead us astray. An endeavor of directly replicating the brain will likely take longer than the insights achievable by looser modeling and objective analysis of individual components. An approach like Sebastian Seung's connectome at least could be automated, at which point specific pathways could be selected for analysis. Given the quality of the KESM scans, it seems a much more feasible short-term goal to reproduce the neural pathways in his planned approach.

Though our brains are not computers, they can be modeled with one. There is no neuromythology here. Since it's probably complexity and not quantum shenanigans that powers our cognition, we only have to vaguely model neurons to capture our brain's 'behavior,' the mind. Now that we have DL to model flexibility of biological neural networks, we've arguably discovered an alternative implementation level from the Tri-Level hypothesis. Now we can go anywhere with research. You can pair headlines like "High-speed, 3D microscope captures stunning videos of fruit fly nerve cells in action," with projects like NeuroKernel, which aims to recreate the fruit fly brain in a connectome approach. Recent neuroscience headlines indicate breakthroughs we might expect, such as insight into language processing, or the role of genetics in human brain expansion, but range to the fantastical, such as deeper understanding of neuronal geometry, or the development of brain organoids from cultivated stem cells.

It's hard to be pessimistic. Our progress with cognitive science has gone far, and at last we're reaching a turning point where we can begin solving more mysteries than we reveal, using deep learning. We have ever more resources and interest to build models, and we're systematically revealing more about our own brains. We're not fated to be an incomplete system, we will understand ourselves. Nothing is inherently intractable about intelligence. It is a complexity just over the horizon.