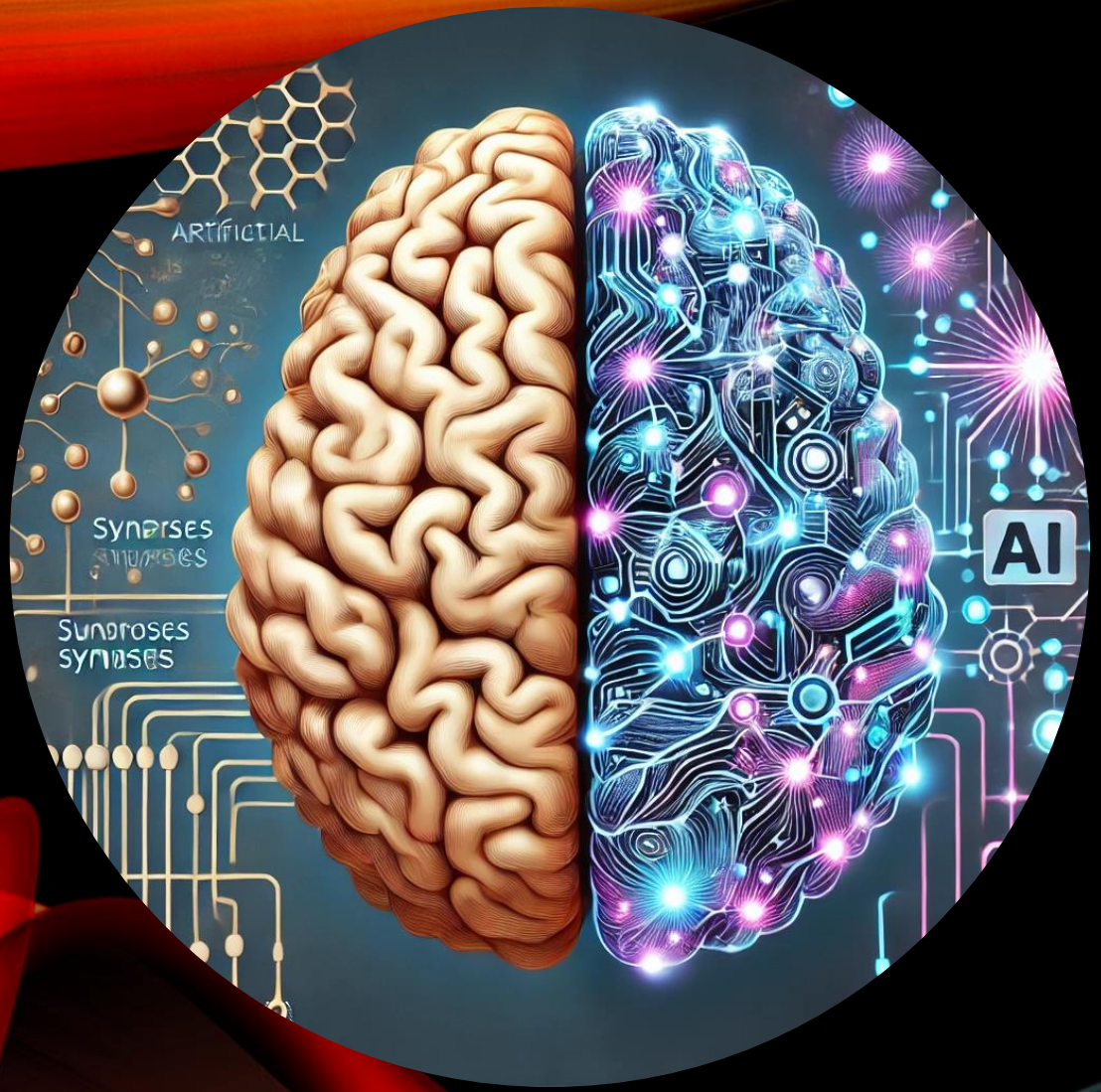

HLMI: A DIFFERENT TYPE OF INTELLIGENCE

Exploring How Artificial Intelligence Thinks Differently from Humans

Dominic Reilly

WHAT IS CONSCIOUSNESS?

- **Philosophy**
  - Consciousness is often described as **subjective experience** - the "what it's like" aspect of existing.
- **Neuroscience**
  - Consciousness involves the **brain's processes** that result in awareness and thought.
- **Psychology**
  - Consciousness is defined as our **awareness of ourselves and our environment**. The ability to focus on internal experiences as well as external events.
- Awareness, meta-cognition (self-reflection), intentionality

THE FOUNDATION: PHYSICALISM VS. DUALISM

Physicalism

- The idea that the mind is fully a product of the brain
- All subjective experience can be explained by physical processes
  - Neural activity

Dualism

- The mind cannot be explained solely by the brain
- There is something non-physical at play that causes our subjective experience
  - Soul

DUALISM VS PHYSICALISM IN HLMI

Physicalism

- The brain boils down to a highly advanced computation system
  - Consciousness arises through this computation system
- Consciousness should be possible through AI
- Consciousness is an emergent property of complexity

Dualism

- The mind has some trans physical aspect to it
  - Consciousness is caused both by the brain and this aspect in tandem
- Consciousness will never be possible through AI

GEOFFREY HINTON INTERVIEW

- Hinton draws an analogy between Artificial Intelligence and alien intelligence
- The brain has around 100 trillion connections while LLM's like GPT-4 have up to a trillion at most
- Despite this, GPT-4 knows "hundreds of times more than any one person"
  - This seems to imply that AI has a better learning algorithm than humans
    - Or at least more training data

GEOFFREY HINTON INTERVIEW

- Hinton notes that AI will be able to share knowledge instantly
  - “It’s as if there were 10,000 of us, and as soon as one person learns something, all of us know it.“
- Humans must painstakingly teach each other
- This could lead to a kind of collective intelligence that will always be getting smarter as each model learns new things
- Hinton suggests that **“confabulation”** or making up information is not unique to AI — humans do it all the time as well.
  - "Bullshitting is a feature, not a bug.“
- The difference is that people usually confabulate correctly enough to sound passable, whereas AI is often more extreme in making things up
  - “Hallucinations”

“For 40 years, Hinton has seen artificial neural networks as a poor attempt to mimic biological ones. Now he thinks that’s changed: in trying to mimic what biological brains do, he thinks, we’ve come up with something better.”

ALIEN INTELLIGENCE

- Serial vs Parallel
  - AI can run millions of calculations simultaneously.
  - Humans can hold 7 + -2 things in working memory at once, leading to sequential thinking
- Empirical Learning
  - AI learns through exposure to vast datasets
  - Humans develop understanding by generalizing from a few examples and drawing from a wide range of context

ALIEN INTELLIGENCE

- Memory
  - AI's "memory" (weights in a neural network) is robust in retaining trained information
  - Not nearly as flexible as the human memory
  - No episodic memory, so it cannot draw from experiences
- Dimensional Thinking
  - When GPT-4 process text, it looks at relationships across millions of examples
  - HLMI may find connections between seemingly unrelated fields of human knowledge
    - "cross-domain" thinking we cannot comprehend

CBMM10 PANEL: RESEARCH ON INTELLIGENCE IN THE AGE OF AI

- Center for Brains, Minds, and Machines
- Featuring:
  - Demis Hassabis
    - Co-Founder & CEO of Deep Mind (AlphaGo, AlphaFold)
  - Geoffrey Hinton
    - "One of the Founding Fathers of Deep Learning", Pioneer of backpropagation
  - Ilya Sutskever
    - Co-Founder & Chief Scientist at OpenAI

CBMM10 PANEL: RESEARCH ON INTELLIGENCE IN THE AGE OF AI

- AI was inspired by the brain
  - Massive network of simple processes working together to complete complex tasks
- AI's effect on Neuroscience and Neuroscience's effect on AI
  - We need to better understand our intelligence to understand AI
- AI's intelligence is too powerful to be evolved as it requires so much energy
- We have always taken inspiration from the brain, but we are reaching a limit
- Intelligence with or without a body has different requirements
  - Embodied vs Disembodied intelligence
- Embodied intelligence requires understanding of causation

CBMM10 PANEL: RESEARCH ON INTELLIGENCE IN THE AGE OF AI

- Difficult to decide on benchmarks for intelligence to test AI
  - Not all problems have a "correct" answer (creative problems)
  - We need a push to define these benchmarks
- Internet is to the AI revolution as oil is to industrial revolution
  - Lucky to have this resource
- "Throughout my life I have seen a lot of people say "Neural nets will not be able to do X"… I do not believe those statements anymore." - Geoffrey Hinton
- There are different levels of creativity
  - Hinton looked to analogy
  - Hassabis pointed toward interpolation and extrapolation
    - Create novel poetry, new strategies in games
- True invention or out-of-the-box creativity is still something AI cannot do

FUTURE CONVERGENCE VS DIVERGENCE

Convergence

- Efforts like neuromorphic computing are aimed at replicating the structure of the brain
- This is an attempt to bridge the gap between biological and digital intelligence
- A more “brain-like” form of intelligence

Divergence

- More likely, AI will continue to diverge
- Future AI might develop different senses
  - Sensory input system comprising not only of vision and audio but entirely different data streams
  - Infrared, GPS, electromagnetic fields
  - Blend these senses to create a world-model beyond our comprehension

DISCUSSION

- **Emergent Consciousness**
  - We know about unexplainable emergent behaviors in LLM's such as common-sense reasoning. What if consciousness is simply an emergent behavior within the complexity of the human brain? Will we someday see unintended, unexplainable, emergent consciousness in AI?
- **Control and Alignment**
  - As HLMI potentially diverges from our understanding, how can we ensure it remains aligned with **human values**? Is it possible to fully control an intelligence that we do not understand?
- **Embodiment in AI**
  - Should HLMI be given some form of **embodiment** to make its understanding more like ours, or should we let it develop along its own distinct, disembodied path? How would embodiment change the ethical considerations?
- **Physicalism vs. Dualism**
  - Which view do you think better describes the future of HLMI's potential for **consciousness** — physicalism or dualism? Could dualism imply that there are inherent limits to what AI can achieve?