

ECE **1724**

2023 Recap

Can AI take inspirations from dreams like us?

Jeong Min Kong

Department of Electrical and Computer Engineering

University of Toronto

Toronto, Canada

jeong.kong@mail.utoronto.ca

Parham Aarabi

Department of Electrical and Computer Engineering

University of Toronto

Toronto, Canada

p@arh.am

Abstract—Many historical figures in both arts and sciences have taken inspirations from their dreams to pursue creative & impactful work. Most notoriously, the classic book *Frankenstein*, and the discovery of the chemical structure of Benzene, were all inspired by dreams. This gives arise to the exciting question: would it be possible to make AI systems dream like us, so they can also take creative inspirations from their dreams? More interestingly, are existing generative AI models *already* unintentionally adapting some of the biological processes happening in our brains when we dream? This paper approaches such questions by first reviewing the scientific literature that discuss about various biological phenomena that happen in our brains

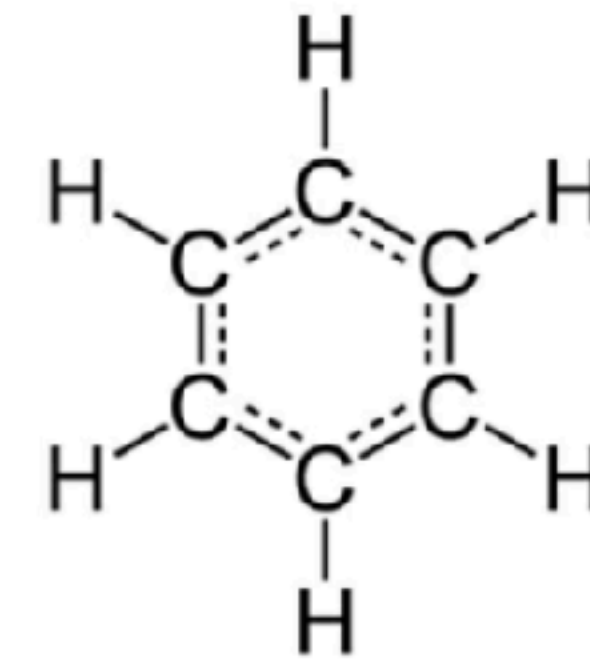


Fig. 1. a) *The Nightmare* by Henry Fuseli b) Chemical structure of Benzene

ECE 1724 Project: Quantum Computing and Consciousness: Investigating the Relationship between Quantum Processes and Consciousness

Pouya Kananian
Electrical and Computer Engineering
University of Toronto
Toronto, Canada
pouya.kananian@mail.utoronto.ca

Abstract—This paper offers a concise overview of theories proposing a connection between quantum processes in the brain and consciousness. Quantum computing can open up new possibilities for studying this link, including through the use of artificial quantum neural networks. However, the feasibility of using these networks to study theories of consciousness remains uncertain. The paper highlights the potential of quantum computing to explore the relationship between quantum processes and consciousness.

Index Terms—consciousness, quantum mechanics, quantum computing, quantum neural networks

so different interact with each other? This problem became known as the mind-body problem and set the stage for future debates on the nature of consciousness itself. Due to the principle of causal closure of classical physics, if the mind-brain system can be entirely explained by classical physics, then the current physical state of the brain is sufficient to determine its future state, with no role for the mind to play. This implies that the mind is merely an epiphenomenon and has no power to influence the physical world. Therefore, the deterministic nature of classical physics means that free will

The Global Workspace Theory: A Step Towards Artificial General Intelligence

1st Mohamed Abdelwahab

*Electrical and Computer Engineering Department
University of Toronto
Toronto, Canada
mo.abdelwahab@mail.utoronto.ca*

2nd Parham Aarabi

*Electrical and Computer Engineering Department
University of Toronto
Toronto, Canada
p@arh.am*

Abstract—Global Workspace Theory (GWT) and Artificial General Intelligence (AGI) are two concepts in cognitive science and Artificial Intelligence, respectively. This paper discusses the possibility of achieving AGI using a deep learning implementation of GWT. The shared latent space is trained using the latent spaces of the connected deep learning modules. This implementation aims to enhance the performance of specialized models in their specified tasks and achieve more general functions from single-task/specialized modules. The paper also discusses the possible applications of this implementation in healthcare.

Index Terms—Artificial General Intelligence, Deep Learning, Global Workspace Theory

II. BACKGROUND

GWT, illustrated in Fig. 1, is one of the most prominent functional theories explaining how consciousness emerges. It was first proposed in [2] and later updated in [3]. The theory posits that the brain is composed of multiple specialized unconscious modules that perform specific functions. These modules are interconnected through a global workspace, and they continuously compete to gain access to that workspace. Depending on the inputs or on the task requirement, one module will be able to access the global workspace and broadcast or share its contents with other distinct specialized modules.

Faulty Neural Networks

1stShiuan-Wen Chen

Dept. of Electrical and Computer Engineering (ECE)

University of Toronto

Toronto, Canada

shiuanwen.chen@mail.utoronto.ca

2ndBrendan Duke

Dept. of ECE

University of Toronto

Toronto, Canada

brendan.duke@utoronto.ca

3rdParham Aarabi

Dept. of ECE

University of Toronto

Toronto, Canada

p@arh.am

Abstract—This study aims to investigate the response of the nervous system to injury through experiments using a neural network model trained with the MNIST dataset [1]. Multiple experiments are performed to examine the relationship between neural network damage and accuracy. How the damaged network can restore its functionality or accuracy with the aid of another neural network is also investigated. By analyzing these results, a better understanding of the nervous system’s ability to respond to injury and adapt to changes in neural networks can be gained.

Index Terms—neural network, neural injury, artificial intelligence

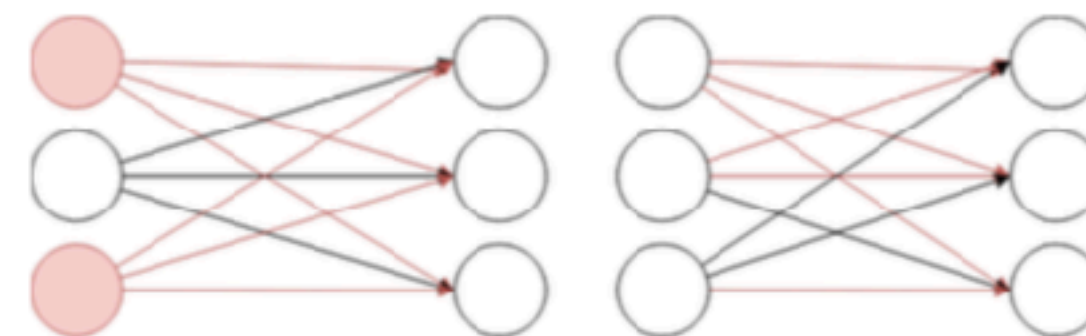
I. INTRODUCTION

In recent years, neural networks have gained widespread adoption as tools for modeling and understanding the human brain. However, the response of the neural system to injury re-

II. EXPERIMENTS

A. Noisy neuron

To replicate the characteristics of an injured nervous system, the experiment employs a fully connected two-layer neural network comprising an input layer and an output layer. The network is evaluated based on two critical parameters - noise probability and noise level - in terms of accuracy. Additionally, the study introduces two types of damage, namely neuron damage and output damage. The former impacts all outputs from the affected neuron, while the latter only affects certain outputs from the neural layer.



Can AI have a personality?

Umarpreet Singh

Electrical and Computer Engineering

University of Toronto

Toronto, Canada

umarpreet.singh@mail.utoronto.ca

Abstract—Recent advancements in large language models have sparked a re-examination of how artificial intelligence (AI) is perceived. These models exhibit human-like behaviour in a variety of complex tasks, leading to claims of their consciousness or possession of a self. However, verifying such claims has been challenging due to a lack of available measurement methods and tools. In this paper, we present an assessment of the personality of large language models using established methods for assessing human personality. Personality is defined as an individual’s views of the world, behaviours, and actions based on those views. We argue that current large language models have formed their own views and opinions from the training data and process, which they use in their decision-making processes. To test our hypothesis, we conducted a variety of personality tests on several large language models, including ChatGPT, GPT3 and LLAMA. Our analysis revealed fascinating insights into the personalities of these AI systems, which have implications for how we train and conceptualize AI. Importantly, we found that not only is the personality of each large language model internally consistent, but it is also consistent across different models. We further found that LLama tends to score more highly on Neuroticism

Models [6]. However, such claims lack validation due to the lack of any experimental or technical tools [7]. Some argue that these models simply execute sophisticated calculations and interpolations on their training data to accomplish these human-like feats.

Despite the impossibility of testing or proving the consciousness of these models, it can be argued that during their training or fine-tuning, they have acquired some understanding of real-world concepts, theories, or ideas. Although they may not be aware of or able to comprehend these concepts when interacting with humans through conversations, they play an essential role in their comprehension and production of appropriate responses. The GloVe embedding [8] is a similar example, which is widely used in the NLP field to convert words into high-dimensional vectors by correlating them with other words. During this process, they learn numerous biases from the training data without comprehending their meanings

2024 Recap

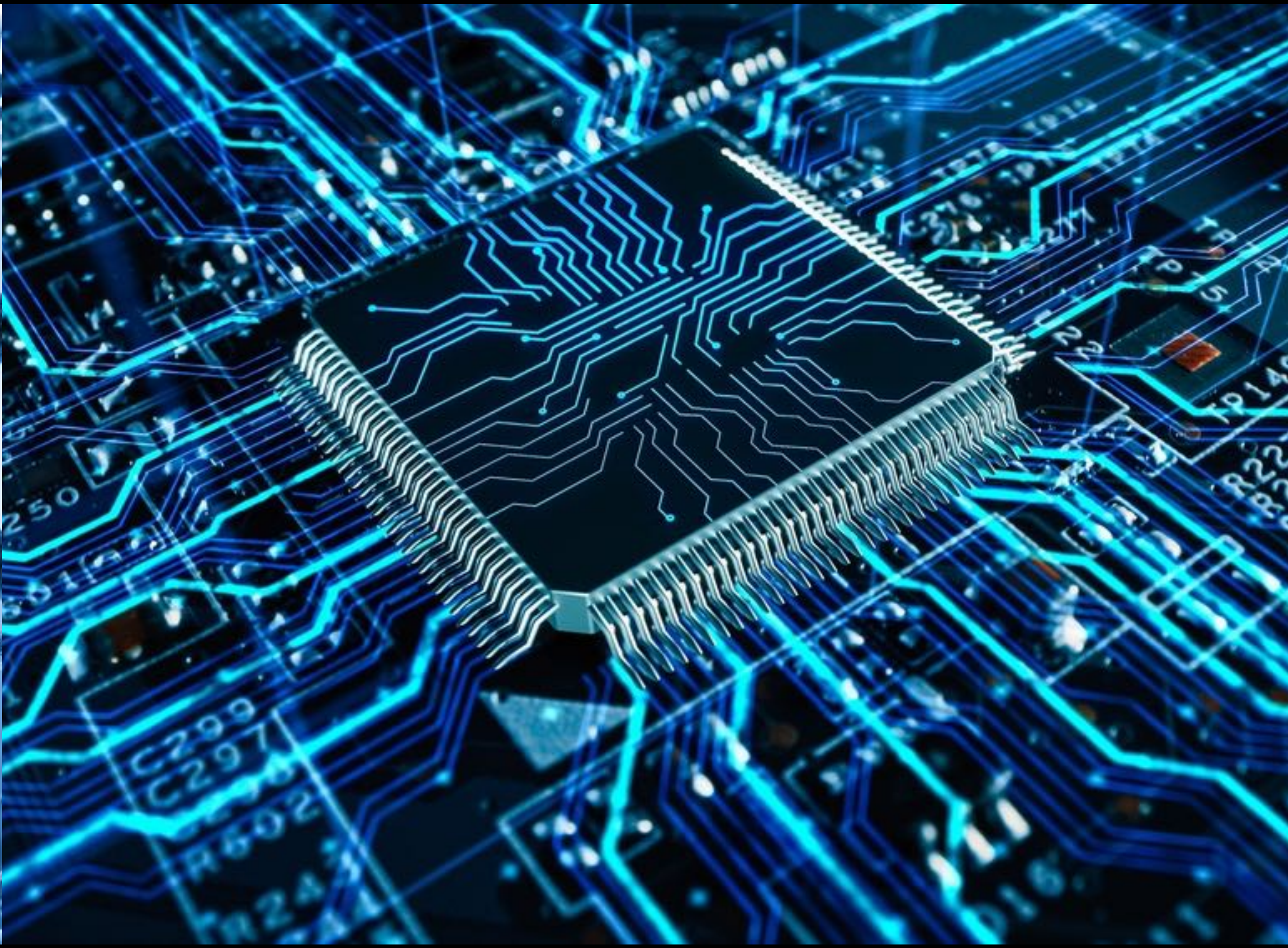
What is **intelligence**?

Where does it come from?

Large number of
biological neurons



Large number of
artificial neurons



Does intelligence **always** result from highly connected, large-scale, deep, multi-layer, multi-stage decision nodes?

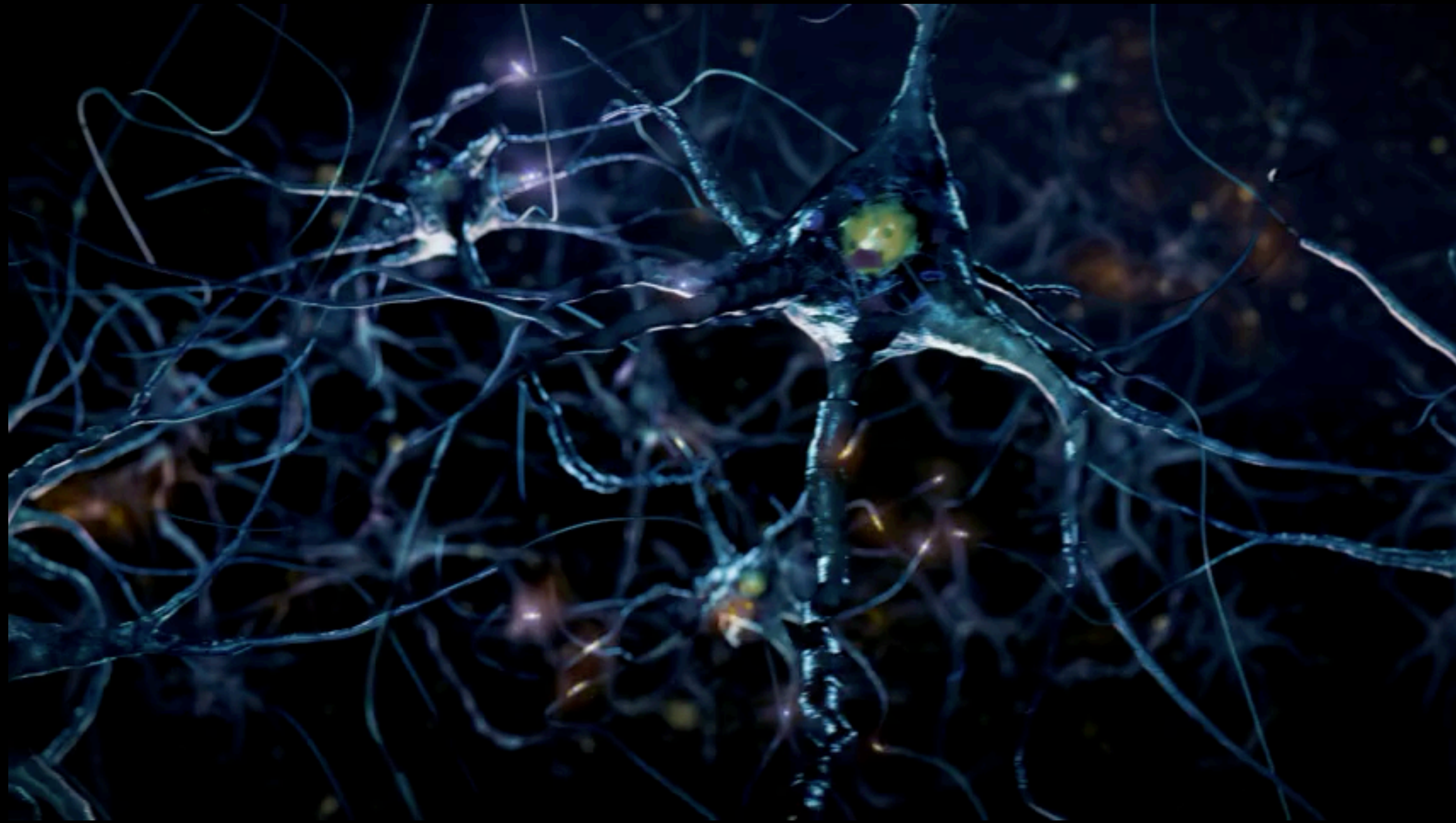
Is intelligence similar to a **symphony** where the whole is greater than the sum of its parts?



So far, we are aware of biological and artificial neural net based intelligence.

Could there be others?

Electrical neural network



Physical neural network



0.086 T

100 T

20,000,000 T

Neurons in

Parameters in

Animals in

Brain

GPT-4

Nature

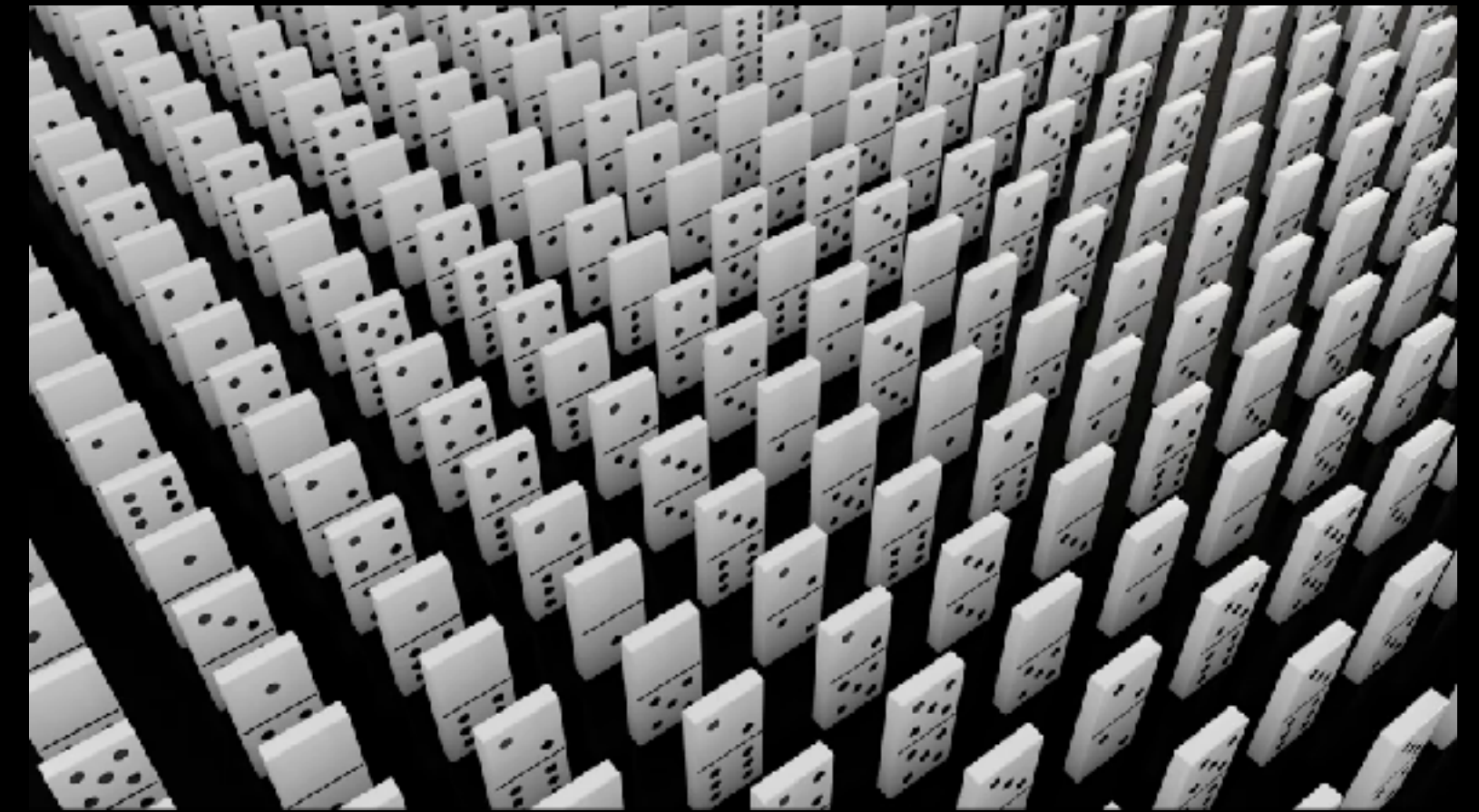
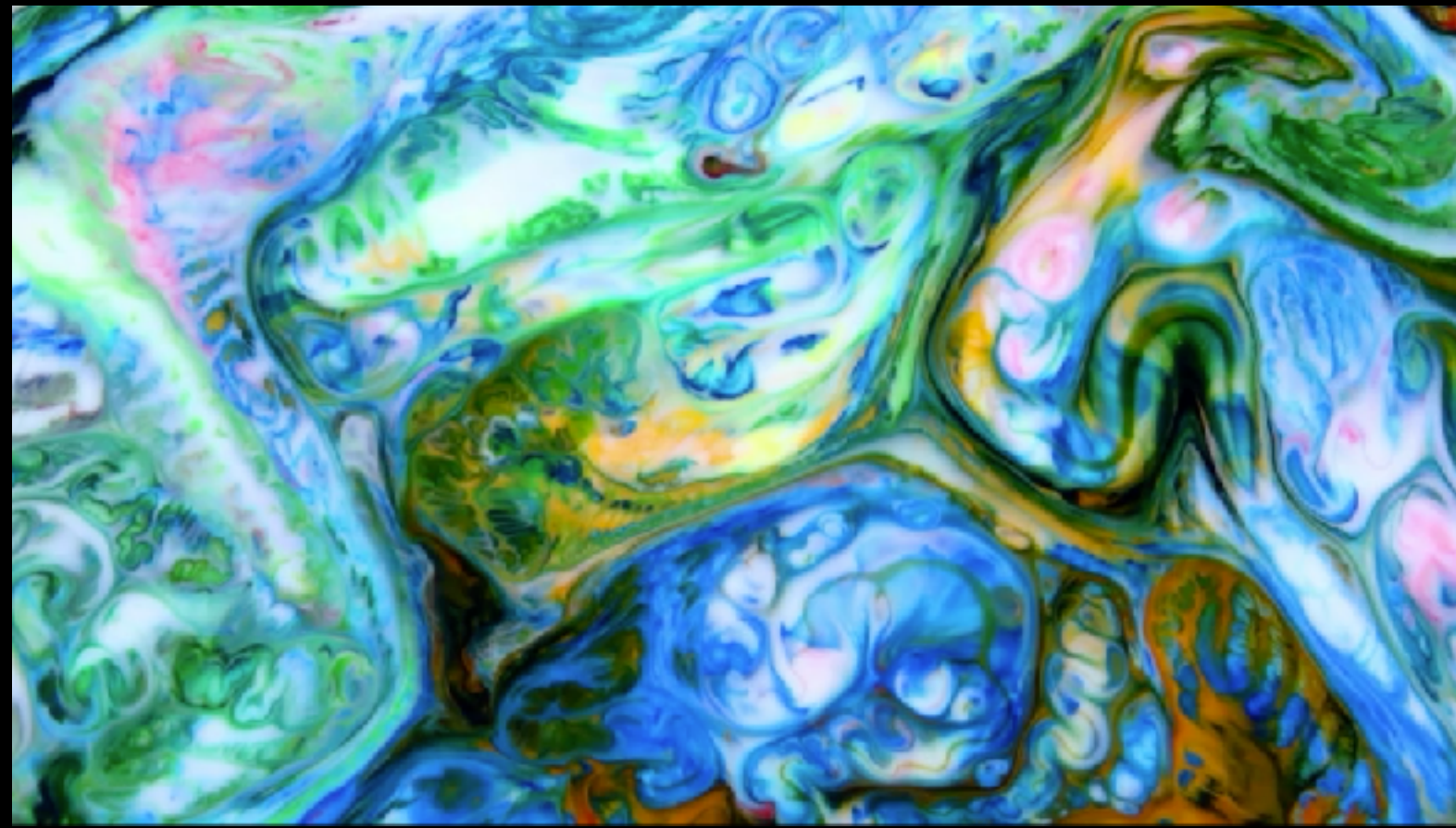
1 X

*1 X

200k X

* Assume 1 neuron has roughly 1k parameters

Can we showcase or simulate a deep
physical neural network?





2025

Global Workspace

Theory is a theory of consciousness involving multiple sub-processes (e.g. agents) and a single shared global workspace.

M. Abdelwahab and P. Aarabi, "The Global Workspace Theory: A Step Towards Artificial General Intelligence," 2023 IEEE Conference on Artificial Intelligence (CAI), Santa Clara, CA, USA, 2023.





WHAT ARE YOU LOOKING FOR?

I am looking for a perfume to buy for my niece for her birthday. My max budget is \$200.



SOCIAL AGENT

Éclat de Minuit is trending with 1.2M mentions; it's a fruity, everyday scent within your budget.

PERFUMIST AGENT

Consider Éclat de Minuit for its fruity notes, or Velours Infini for a gentle floral touch. Which scent appeals more?



WHAT DO YOU THINK?

She loves flowers, especially roses, so floral might be the right option.



FRIEND AGENT

Velours Infini's gentle floral scent could be perfect. Imagine her smile when she wears it! 🌹👍

GRUMPY AGENT

Velours Infini's subtlety suits her floral preference, but Éclat de Minuit's popularity might appeal to her trendy side.



SOCIAL AGENT

Velours Infini's gentle floral scent is understated yet chic, aligning with current influencer trends for subtle elegance.

PERFUMIST AGENT

Velours Infini's floral notes evoke rose gardens; its chemistry offers a calming effect. Does she prefer calming or invigorating scents?



FRIEND AGENT

Imagine her feeling like she's in a serene rose garden with Velours Infini. It's a calming floral embrace! 🌹🤍



WHAT DO YOU THINK?

calming scents are good i think.



HEAD AGENT

Velours Infini is perfect for her love of roses and calming floral scents, fitting your budget and offering a serene, elegant experience.



Velours Infini

\$185

[BUY NOW](#)

A very subtle gentle floral scent

Idea:

Create a large number of agents, focused on a particular task, and sharing a global workspace. Can they achieve something more than the sum of their parts?

The Task:

Predict buy/neutral/sell investments in NASDAQ (QQQ). Decisions are made at 10am (30 mins past stock opening) each day. Each agent will decide/vote if at 10am we should buy/keep/sell.

Generation 1

Select an agent from agent list, and create first generation agent. Demo solo agent in action.

(5% - Deadline: Jan 19)

Generation 1

Finalize Gen 1 agent and demo in conjunction with other agents.

(5% - Deadline: Jan 26)

Generation 2

Create Gen 2 agent. Demo solo agent in action and pass Turing test.

(10% - Deadline: Feb 2)

Generation 2

2-min video explaining Gen 2 agent design and goals.

(10% - Deadline: Feb 16)

Generation 2

Demo Gen 2 agent in conjunction with other agents.

(10% - Deadline: Feb 23)

Generation 3

Deadline for Gen 3 agents with solo test.

(10% - Deadline: Mar. 2)

Generation 3

Finalize Gen 3 agent and demo in conjunction with other agents.

(10% - Deadline: Mar 23)

Live Testing Week

Test on real-time NASDAQ (QQQ) data and see how each agent votes and cumulatively how the overall hive AI performs.

(20% - Deadline: Mar 24 - April 7)

Project papers (2 pages IEEE Conf style
max).

(20% - Deadline: Apr 1)

Agent Deadlines are Sunday at 11:59pm

This allows us to integrate agents and ready
to test by class on Tuesday.

Late submissions will receive a grade of 0.

How do we grade Agents?

Did it follow the agent integration rules (+1)?

Did it properly participate (+1)?

Did it help or have an impact on the final result (+1)?

Did it showcase unique agent-centric expertise (+1)?

Was it respectful and aware of the global workspace
(+1)?

Test week grading:

Solo agent performance rank: +10% for 1st place, 0% for last place

Combined live agents performance: +10% if combined AI beats QQQ performance, 0% if it doesn't.

Schedule:

Jan. 7 - Intro Lecture

Jan. 14 - Student Project Consultations

Jan. 21 - Gen 1 Agent Solo Test

Jan. 28 - Gen 1 Agent Full Test

Feb. 4 - Gen 2 Agent Solo Test

Feb. 11 - No Lecture (Reading Week)

Feb. 18 - Gen 2 Agent Solo Test

Feb. 25 - Gen 2 Agent Full Test

Mar. 4 - Gen 3 Agent Solo Test

Mar. 11 - No Lecture

Mar. 18 - No Lecture (Virtual Office Hours)

Mar. 25 - Gen 3 Final Full Test

Mar. 24 - April 7 - Test Week

Apr. 1 - Paper Deadline

Agent list:

Political finance impact agent

Macro economics agent

Big tech news agent

Financial performance agent

Futurist agent

Stock expert agent

HR tracking agent

...

How to make an agent:

A lengthy prompt on what the agent should do, how the agent should vote, with prompt including the following variables:

#gws = current global workspace

#gain = current gain in percentage (- if loss)

#qqq = percent gain/losses for the past 7 days on QQQ

#site123 = html of a custom predetermined website

Agent outputs:

Your agent should have two modes (or alternatively, you can create two different agents).

Mode 1: Read relevant info/variables, and make a meaningful contribution to the global workspace. In this case, output is a useful/relevant comment.

Mode 2: Read relevant info/variables, and make a decision on buy/hold/sell. In this case, output is a single word.

http:// **p . arh . am**

Email: **p @ arh . am**