

ECE **1724**

Presentation and papers are online at:

[http:// p . arh . am](http://p.arh.am)

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

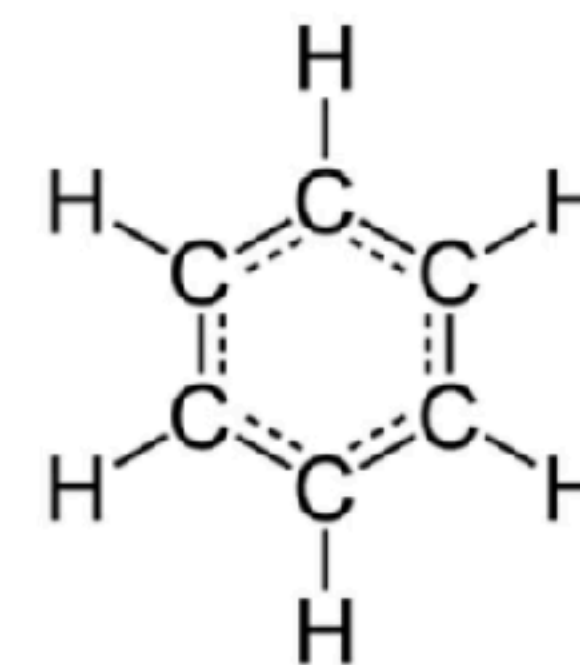


Fig. 1. a) *The Nightmare* by Henry Fuseli b) Chemical structure of Benzene

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

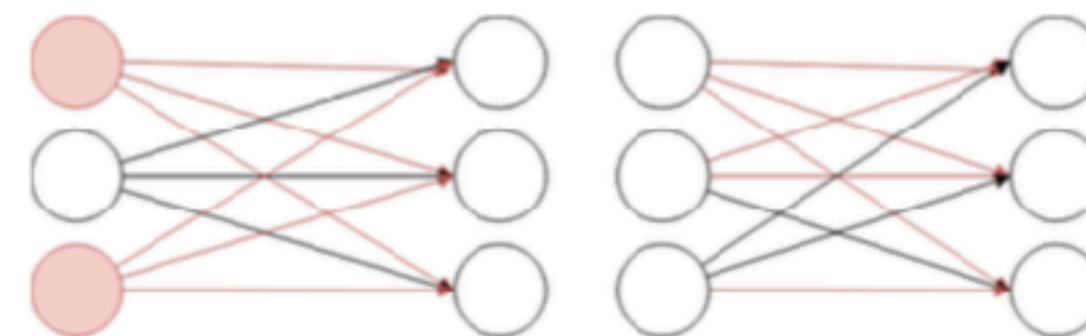
I. INTRODUCTION

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

What is **intelligence**?

The ability to **acquire**
and **apply knowledge**.

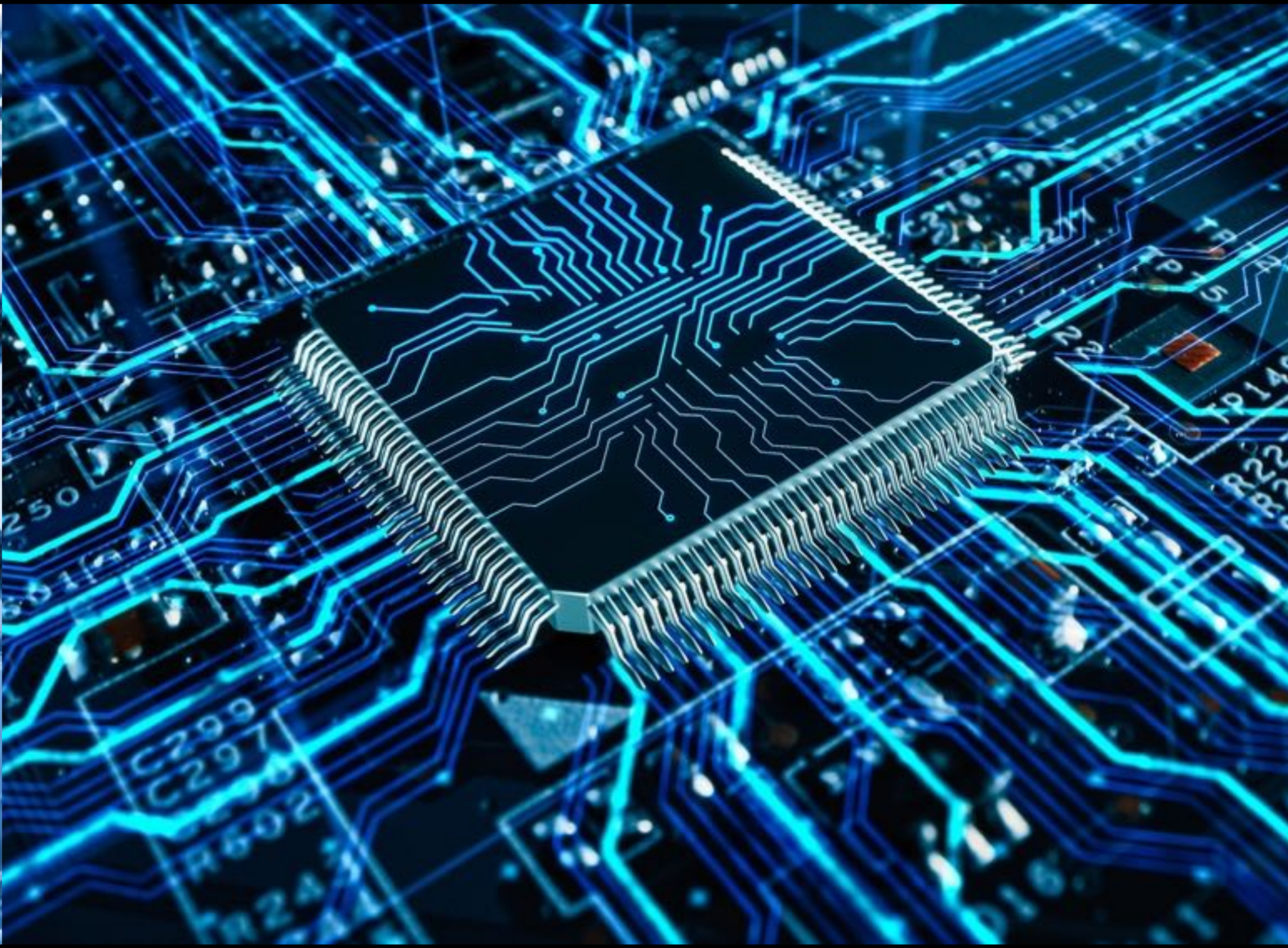
Human level **reasoning** and
decision making.

Where does it come from?

Large number of
biological neurons



Large number of
artificial neurons



Multi-stage decision nodes

Highly connected

Very large scale

Deeply Multi-layer

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

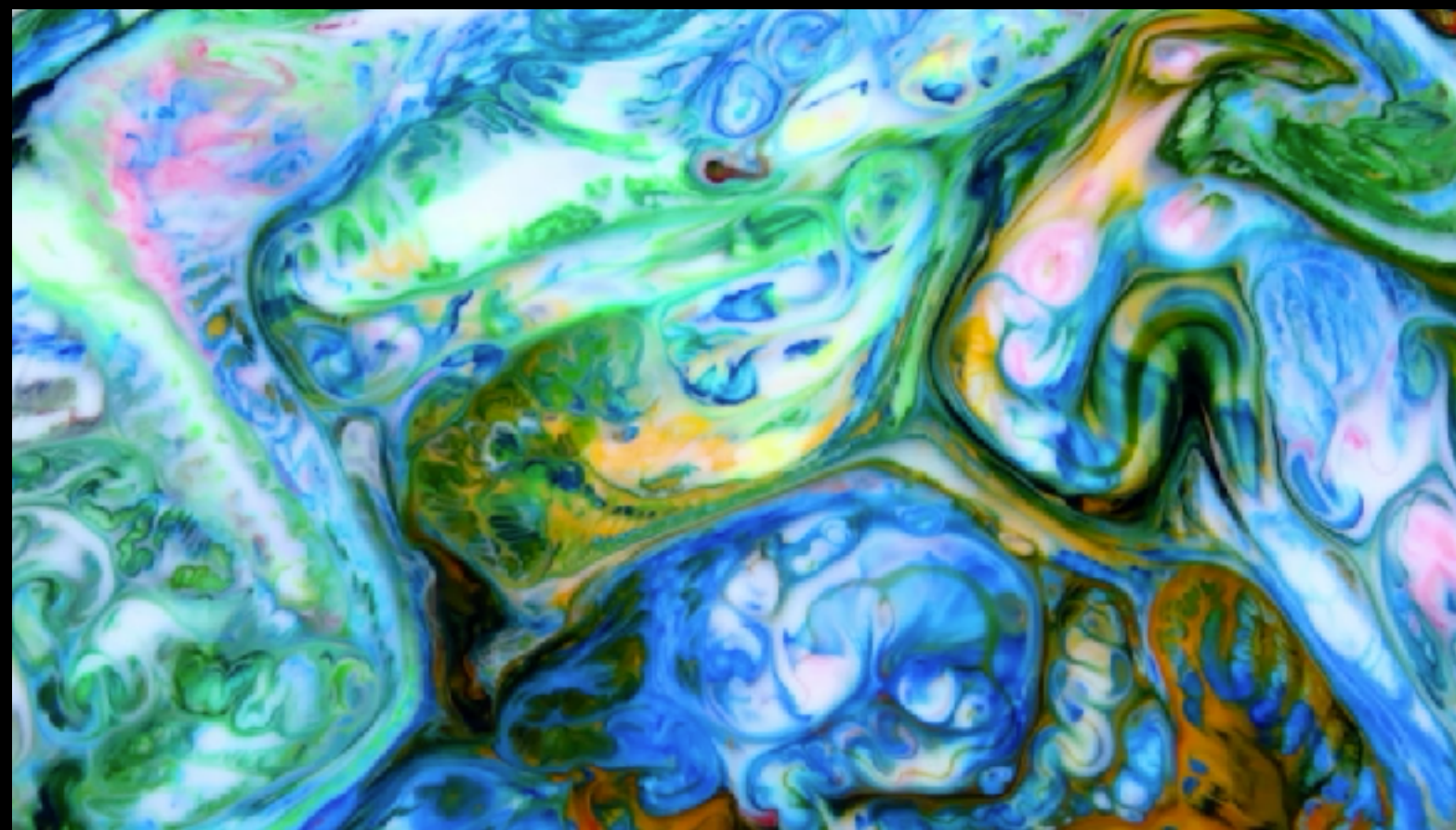
1X

*1X

200kX

* Assume 1 neuron has roughly 1k parameters

Question 1 - Can we showcase or simulate a deep physical neural network?



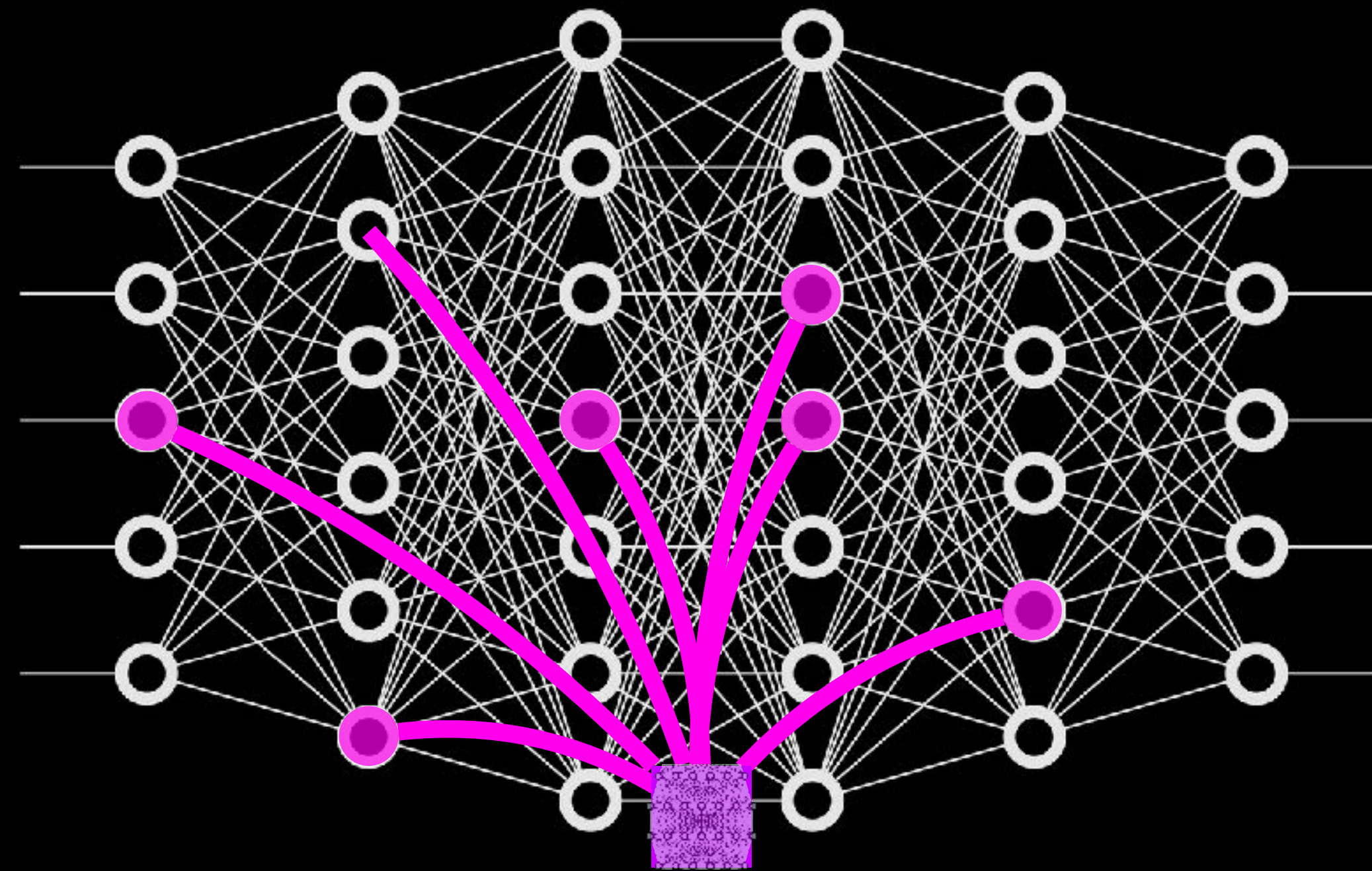
Project 1

Devise a useful/simple NN (max 20 neurons). Train it using your own training data. Implement it physically and record 2-min video of it working.

(15% - Deadline: Jan 22)

Question 2 - Can we interface with a NN?

If a single neuron is damaged, if we replace it, would we be the same person?



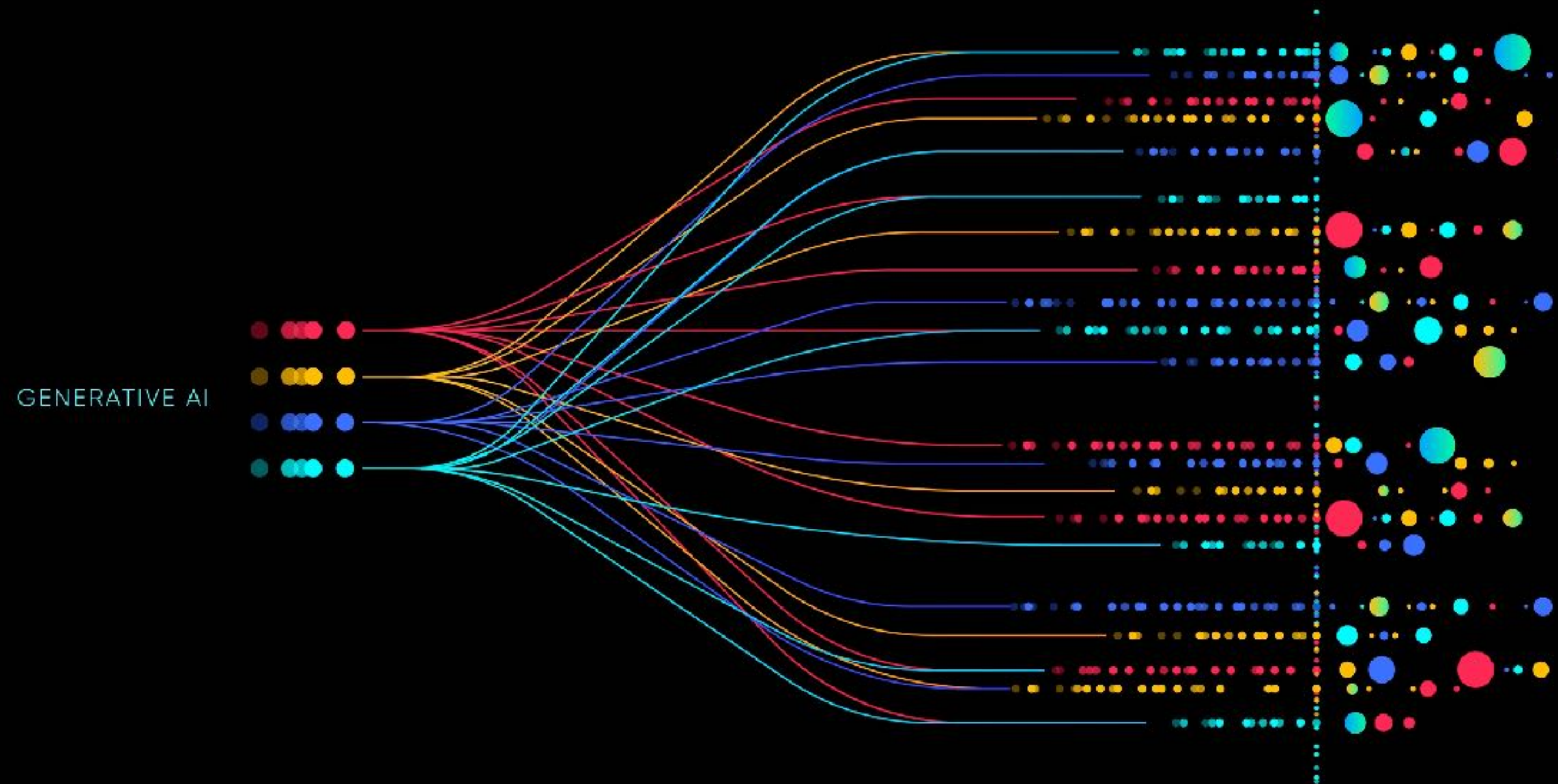
If we lose/regenerate 1% of our neurons/year,
and we still are the same person, potential for
artificial neural degeneration exists.

Project 2

Read papers on neural hacking and neural degeneration. Summarize and make a 2-min video summary/recap.

(5% - Deadline: Jan 29)

Question 3 - There has been a lot of talk about LLMs like chatGPT. How can we better understand them?



Project 3

Implement Transformer-based nano language model (max 15 tokens, 5 words per sentence), from scratch. Record 2-min video demo.

(15% - Deadline: Feb. 12)

Example tokens:

I you we am are happy sad with in at life
school class work

Question 4 - What impact would the loss of a few neurons have on your nano language model?

Project 4

Damage 'X' percent of neurons, and see the impact on functionality. Record 2-min video to summarize your result.


(10% - Deadline: Feb. 26)

Question 5 - If AI was trying to take over humanity, how would it do it exactly? We should ask it to find out...

ChatGPT

Using bait to control a few pests to influence the behavior of the larger population is a unique approach. Here's a concept that might help:

Behavior Manipulation:

1. **Identify Leader Pests:** In some species, certain individuals might have more influence or dominance. If you can identify such leaders, targeting them could impact the behavior of the rest of the population.
2. **Offer Food with a Purpose:** Use food as a reward for desired behaviors. For example:
 - **Trained Responses:** If they respond to sound or light, consider using those cues along with food rewards to direct their movement.
 - **Lead Them Away:** Use the food to guide them away from the infested area or toward an area where you have set up traps or control measures.
3. **Create Diversion or Isolation:** If possible,  the 'controlled' pests to lead the rest away from the area they've infested. This might involve setting up a secondary, more attractive food source in a different location to draw them away.
4. **Limit Food Sources:** Once you've manipulated the behavior of these 'controlled' pests, gradually reduce the availability of food in the infested area. This could encourage the pests to rely more on the food source you've provided, leading them away from their destructive activities.

Challenges and Considerations:

- **Accuracy and Control:** Precision is crucial in this approach. You'll need to ensure that the 'controlled' pests respond as expected, which might not always happen due to the unpredictable nature of pests.
- **Patience and Adaptation:** Pests might adapt to changes in their environment or food sources. Regularly assess their response to your control methods and adapt your strategy accordingly.

Project 5 - Use analogies (e.g. pest control, germ extermination, viral treatment, etc.) to ask LLMs to plan out takeover of a specific species and observe any patterns. Record 2-min video summarizing your take aways.

(5% - Deadline: Mar. 4)

Select course project (any of the 5 basic projects extended).

(Deadline: Mar. 4)

Project abstracts (1 page max) due.

(10% - Deadline: Mar. 11)

Record 5-min Youtube video project
showcase.

(20% - Deadline: Apr. 1)

Project papers (4 page max) due.

(20% - Deadline: Apr. 1)

Schedule:

Jan. 9 - Lecture 1

Jan. 16 - No Lecture

Jan. 23 - P1 Demo/Discussion

Jan. 30 - P2 Demo/Discussion

Feb. 6 - P3 Help Session

Feb. 13 - P3 Demo/Discussion

Feb. 20 - No Lecture (Reading Week)

Feb. 27 - P4 Demo/Discussion

Mar. 5 - P5 Demo/Discussion

Mar. 12 - No Lecture (Virtual Office Hours)

Mar. 19 - No Lecture (Virtual Office Hours)

Mar. 26 - Final Project Help Session

Apr. 2 - Final Project Demo/Discussion