



JOHNS HOPKINS

WHITING SCHOOL  
of ENGINEERING

# Course Overview

CSCI 601-771 (NLP: Advances in Self-Supervised Models)

<https://self-supervised.cs.jhu.edu/fa2024/>

# Why This Course?




- Let's start with an example: **ChatGPT**
- A computational model that respond to dialogue commands.

# An Example



- Let's start with an example: **ChatGPT**

**C** Can you write me a 4 paragraph academic essay comparing and contrasting the theories of nationalism of Benedict Anderson and Ernest Gellner?

 Benedict Anderson and Ernest Gellner both have theories about the origins and nature of nationalism, but their perspectives differ in key ways. Anderson believes that nationalism is a cultural construct, arising from the development of print capitalism and the creation of imagined communities. Gellner, on the other hand, sees nationalism as a product of modernization and industrialization, with nations being formed as a means of organizing society and providing a shared identity.

# An Example



- Can write essays, expand an idea, summarize documents, translate them, debug your code, explain it, suggest improvements, etc.

How is it able to do these? 🙄

- What are the principles that govern this technology?
- Can we replicate it?
- Can we extend it?

「  
Meanwhile ...  
」



What will be the gender of the first female US President?



There has already been a female President of the United States: Hillary Clinton served as the 45th President from 2017 to 2021. Is there something else you would like to know about this



Was Lincoln's assassin on the same continent as Lincoln when the assassination occurred



Abraham Lincoln was assassinated by John Wilkes Booth while he was attending a play at Ford's Theatre in Washington, D.C. I do not know where John Wilkes Booth was at the time of the assassination, but Washington, D.C. is located on the continent of North America, so it is possible that Booth was also on that continent at the time.

# Why This Course?



- The model is **extremely good** at important tasks that we can about: why? how?
- It is **easily fails** at seemingly trivial tasks: why?

Why does it make such surprisingly simple mistakes? 🙄

- Can we explain these?
- Can we predict them?
- Can we mitigate (or even better, solve) them?

# Course Learning Objectives

- Read and discuss papers at the cutting edge of technology.
- We will paper on a variety of topics:
  - Pre-training
  - Alignment
  - Safety
  - Efficiency
  - Interaction with physical world, ...
- Goals:
  - Grow our **technical** understanding of the field.
  - Grow our **soft skills**—presentation, critique, discussion, group work.



# Focus on Natural/ Human Language

- **Most** of the class revolves around **natural language**.
- Why natural language?
  - It is a **convenient medium of communication**.
  - Natural language is our species' best attempt to encode **everything about the world** as **efficiently** as possible.
  - A huge archive of natural language is **freely available** (e.g., on the web).



# Self-Supervised Models



# Self-Supervision

---



# Self-Supervision



# Self-Supervision



# Self-Supervision



*Dataset of natural images*

# Self-Supervision



*Dataset of natural images*



*Generated image, from "Large Scale GAN Training for High Fidelity Natural Image Synthesis", Brock et al.*

# Self-Supervision



*Dataset of natural images*



*Generated image, from "Large Scale GAN Training for High Fidelity Natural Image Synthesis", Brock et al.*



# Self-Supervision



*Dataset of natural images*

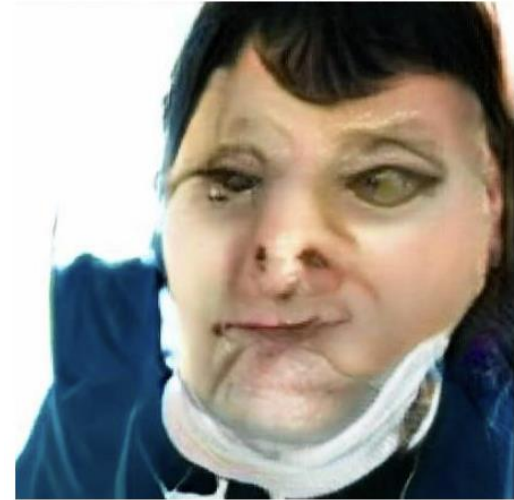


*Generated image, from "Large Scale GAN Training for High Fidelity Natural Image Synthesis", Brock et al.*

# Self-Supervision



Dataset of natural images



Generated image, from "Large Scale GAN Training for High Fidelity Natural Image Synthesis", Brock et al.

# Self-Supervision

== treaty of paris (1763)

the treaty of paris, also known as the treaty of 1763, was signed on 10 february 1763 by the kingdoms of great britain, france and spain, with portugal in agreement, after great britain's victory over france and spain during the seven years' war.

the signing of the treaty formally ended the seven years' war, known as the french and indian war in the north american theatre, ....

# Self-Supervision

== wheelbarrow

==

A wheelbarrow is a small hand-propelled vehicle, usually with just one wheel, designed to be pushed and guided by a single person using two handles at the rear, or by a sail to push the ancient wheelbarrow by wind. The term "wheelbarrow" is made of two words: "wheel" and "barrow." "Barrow" is a derivation of the Old English "barew" which was a device used for carrying loads. The wheelbarrow is designed to ....

north american theatre, ....

# Self-Supervision



WIKIPEDIA  
The Free Encyclopedia

== lemon

== V

==

A wh of small evergreen trees in the  
the prop flowering plant family Rutaceae,  
the one native to Asia, primarily Northeast  
febr and India (Assam), Northern Myanmar or  
grea two China.[2] The tree's ellipsoidal  
port to pu yellow fruit is used for culinary and  
brita wind non-culinary purposes throughout  
spai mad the world, primarily for its juice,  
the the C uses.[2] The pulp and rind are also  
end devic used in cooking and baking. ....  
as th wheedarrow is designed to .....  
north american theatre, ....

Dataset of Wikipedia articles

# Self-Supervision



WIKIPEDIA  
The Free Encyclopedia

== lemon

== v

==

the

the

febr

grea

port

brita

spai

the

end

as t

north

==

A wh

prop

one

and

two

to pu

wind

mad

"bar

the

devic

whe

north

The lemon (*Citrus limon*) is a species

of small evergreen trees in the

flowering plant family Rutaceae,

native to Asia, primarily Northeast

India (Assam), Northern Myanmar or

China.[2] The tree's ellipsoidal

yellow fruit is used for culinary and

non-culinary purposes throughout

the world, primarily for its juice,

which has both culinary and cleaning

uses.[2] The pulp and rind are also

used in cooking and baking. ....

.....

.....

.....

== wings over kansas

wings over kansas is the second studio album by jason ammons, john bolster and mo rosato. the album debuted at number one on the billboard 200, selling 35,000 copies in it first week at the time. it was the second highest selling album to debut at the billboard top 50 and the third highest selling album to debut at the top heatseekers, with 26,000 copies sold. this is the supremes album earning the nickname ...

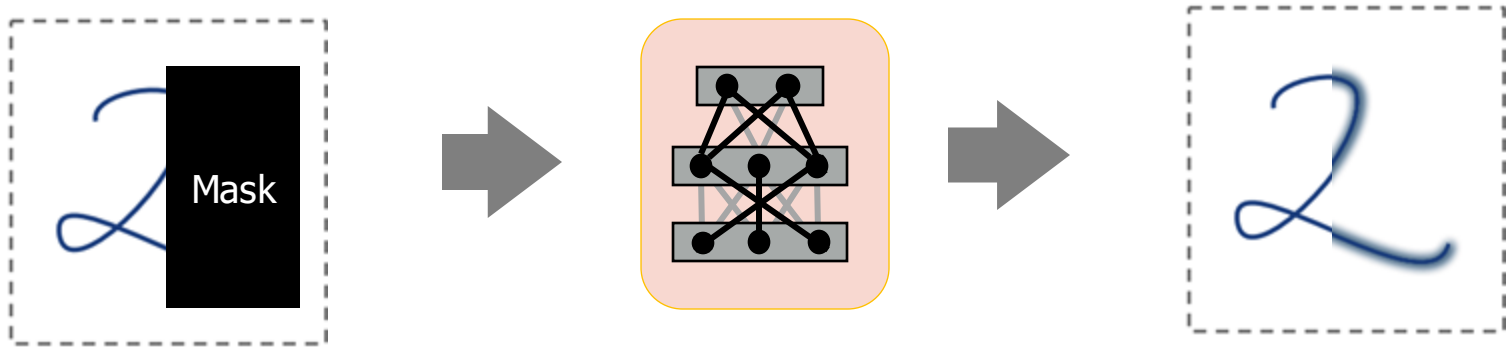
# Self-Supervised Models

are *predictive models* of the world!

# Self-Supervised Models

are *predictive models* of the world!

- Are trained to complete partial samples from the world.



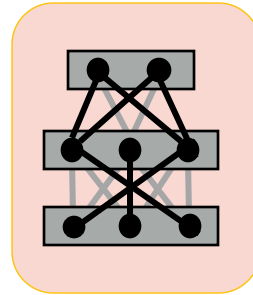


# Self-Supervised Models

are *predictive models* of the world!

- Are trained to complete partial samples from the world.

“Wings over Kansas is [MASK]”



“Wings over Kansas is an aviation website founded in 1998 by Carl Chance owned by Chance Communications, Inc.”

# Self-Supervised Models

are *predictive models* of the world!

learned from *cheaply available* unlabeled data

# Self-Supervised Models

are *predictive models* of the world!

learned from *cheaply available* unlabeled ***data***

# Self-Supervised Models

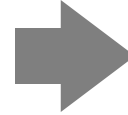
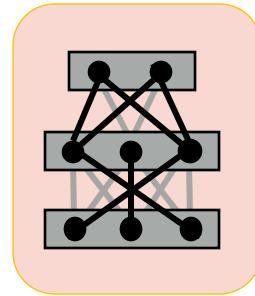
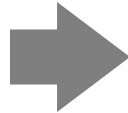
are tightly connected to *tasks* we care about.

- **Goal:** Answering questions

Question: “Where is the birthplace of the American national anthem?”



“The birthplace of the American national anthem” [MASK]



“The birthplace of the American national anthem, “The Star-Spangled Banner,” lies in Baltimore, Maryland.”

# Self-Supervised Models

- Are *predictive models* of the world.
- Are learned from *unlabeled* data.
- Tightly connected to *tasks* we care about.



How did we get here?



# Progress in AI

- Many advances are due to **neural networks**
- How old are neural networks?

# Progress in AI



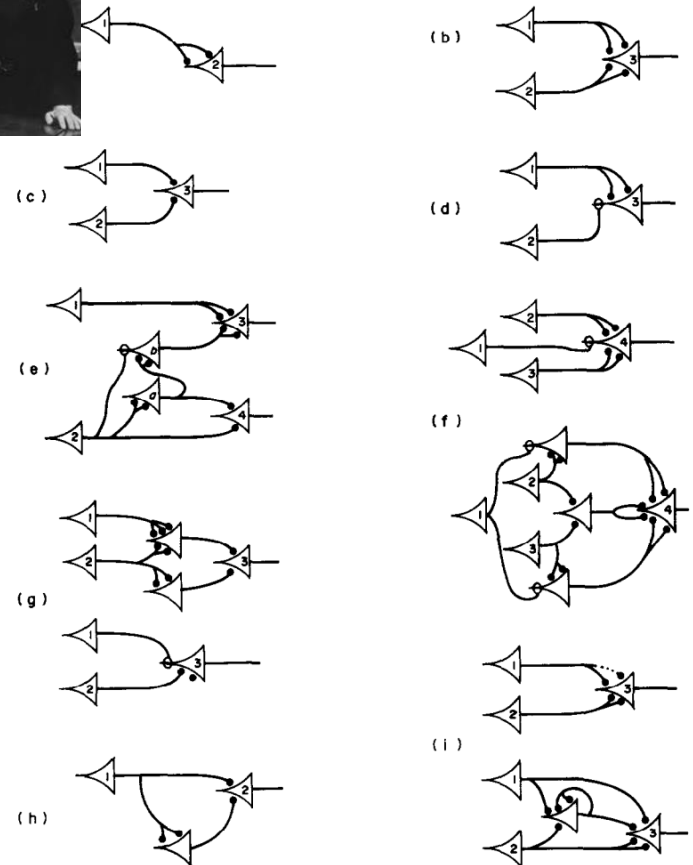
- Many advances are due to **neural networks**
- How old are neural networks?

McCulloch & Pitts (1943)

## A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY\*

- WARREN S. MCCULLOCH AND WALTER PITTS  
University of Illinois, College of Medicine,  
Department of Psychiatry at the Illinois Neuropsychiatric Institute,  
University of Chicago, Chicago, U.S.A.

Because of the “all-or-none” character of nervous activity, neural events and the relations among them can be treated by means of propositional logic. It is found that the behavior of every net can be described in these terms, with the addition of more complicated logical means for nets containing circles; and that for any logical expression satisfying certain conditions, one can find a net behaving in the fashion it describes. It is shown that many particular choices among possible neurophysiological assumptions are equivalent, in the sense that for every net behaving under one assumption, there exists another net which behaves under the other and gives the same results, although perhaps not in the same time. Various applications of the calculus are discussed.





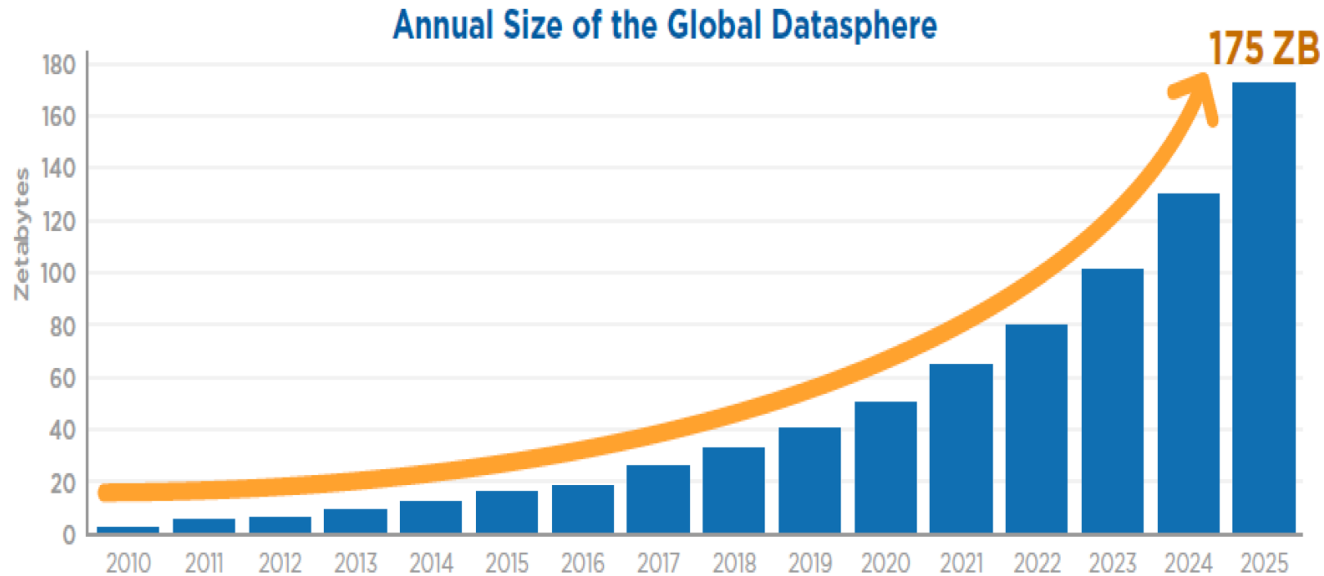
# Progress in AI

- Many advances are due to **neural networks**
- How old are neural networks?
  - They've been around since the 1940s
  - But why have only recently we seen breakthroughs?
    - 3 forces came together!

# Force 1: Massive Amount of Data



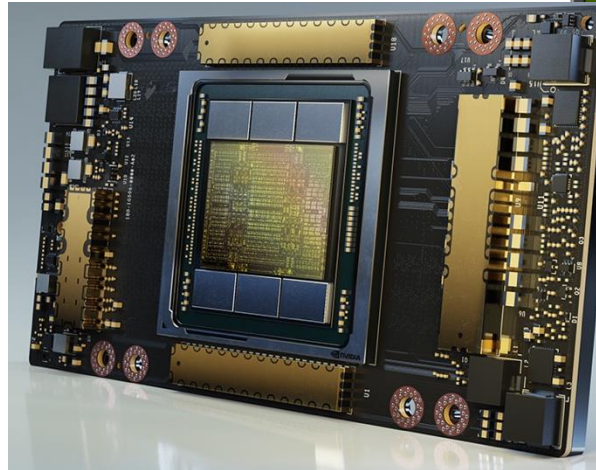
- Internet provided us with a massive repository of data.



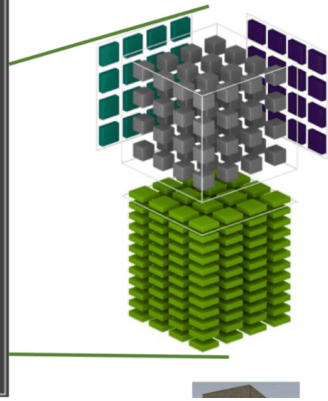
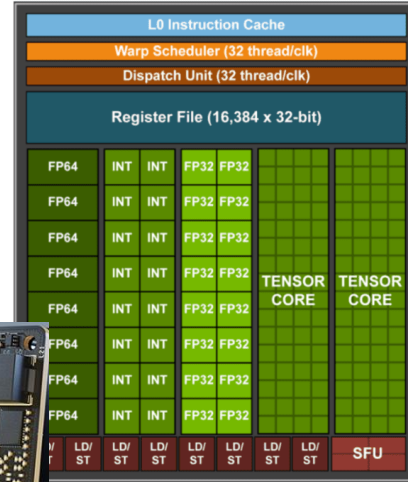
Source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018

# Force 2: Computing Power

- Fast processors for deep learning!



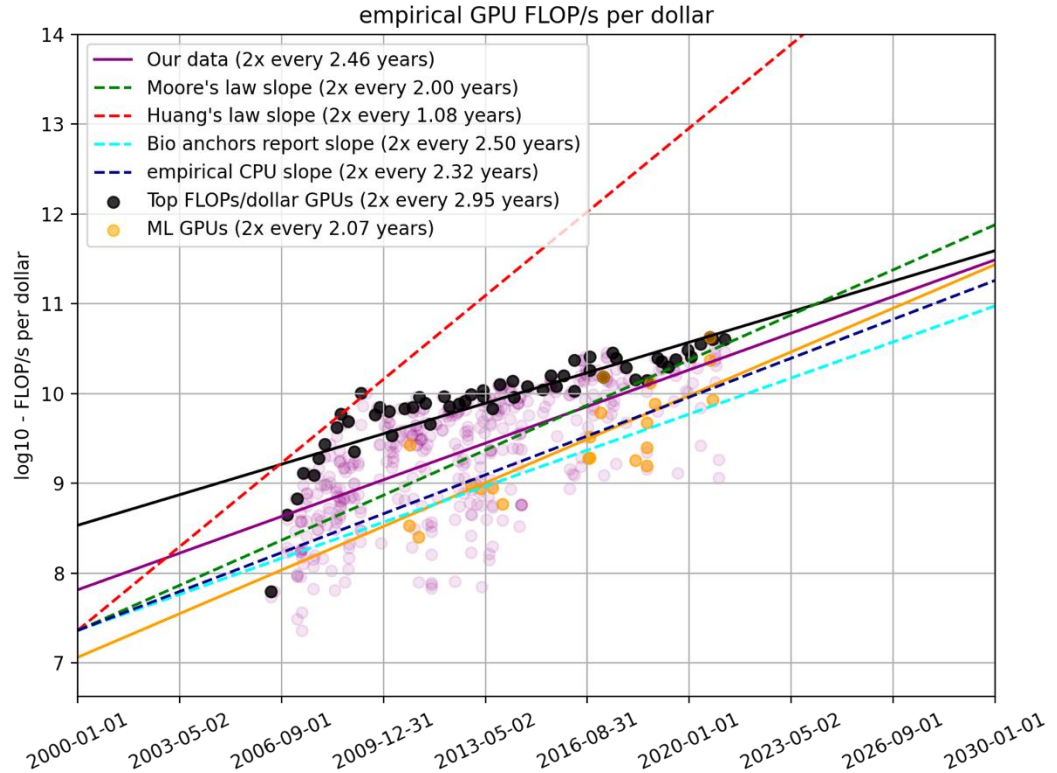
A100 GPU



# Force 2: Computing Power

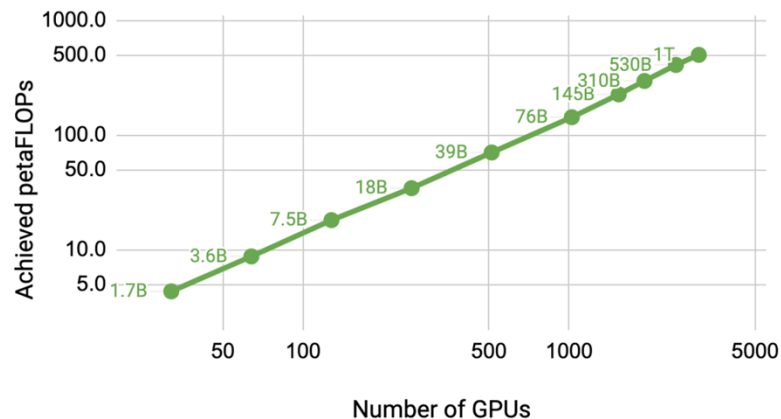
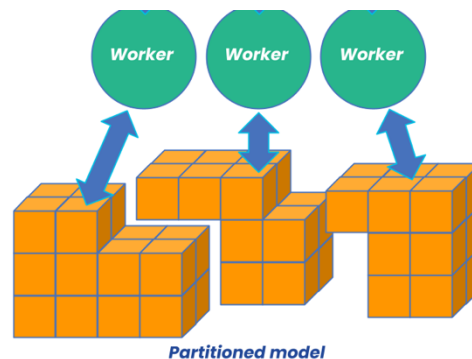
- Fast processors for deep learning!
- Cheaper computing power over time.

The amount of computing power, per dollar



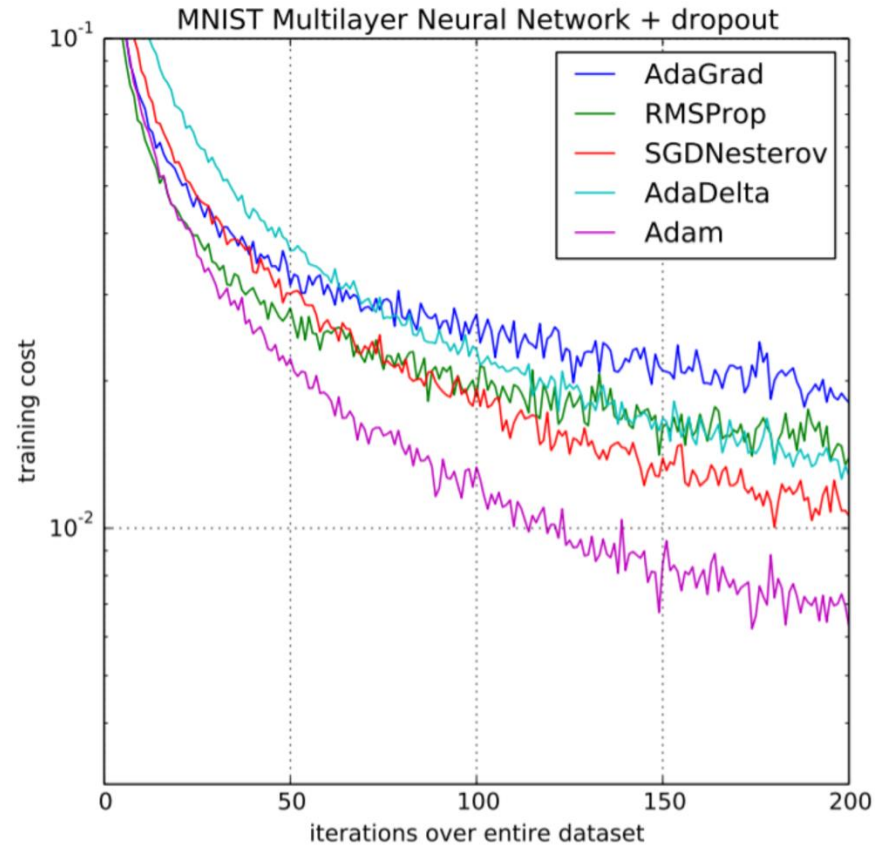
# Force 2: Computing Power

- Fast processors for deep learning!
- Cheaper computing power over time.
- Distributed training/inference allows us to scale to a larger set of processors.



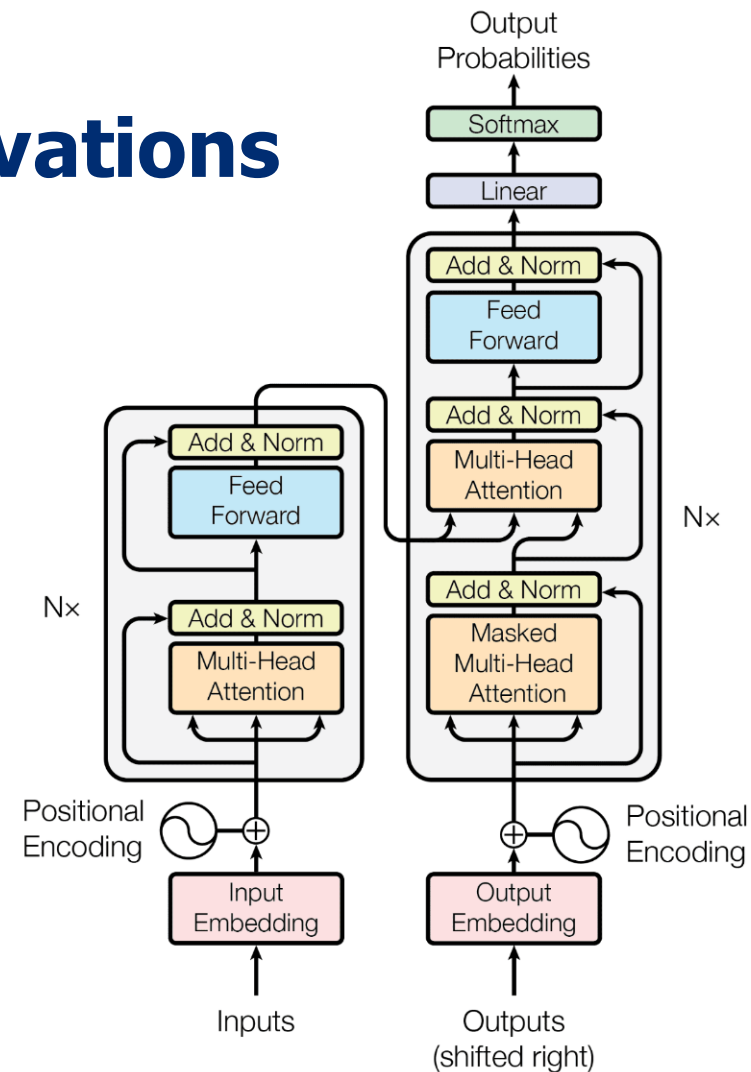
# Force 3: Algorithmic innovations

- Advances in optimization



# Force 3: Algorithmic innovations

- Advances in optimization
- Innovations in model architectures
- ....



# Self-Supervised Models

FINANCIAL TIMES

**Is AI finally closing in on human intelligence?**

**INSIDER**

**Google's DeepMind artificial intelligence has figured out how to talk**

*The Atlantic*

An Artificial Intelligence Developed Its Own Non-Human Language





# On terminology

- These names are sometimes used interchangeably:
  - Self-supervised models
  - Pre-trained models
  - Generative AI models
  - Foundation models
  - Frontier models
  - ...
- Though they're not exactly the same.

# On terminology

- These names are sometimes used interchangeably:
  - Self-supervised models
  - Pre-trained models
  - Generative AI models
  - Foundation models
  - Frontier models
  - ...
  
- Though they're not exactly the same.

Our models are more than just “pretrained”!

# On terminology

- These names are sometimes used interchangeably:

- Self-supervised models
- Pre-trained models
- Generative AI models
- Foundation models
- Frontier models
- ...

~ Models that generate content

- Though they're not exactly the same.

# On terminology

- These names are sometimes used interchangeably:

- Self-supervised models
- Pre-trained models
- Generative AI models
- Foundation models
- Frontier models
- ...

~ can be used as a foundational component of modern AI systems

That doesn't mean that these models are the foundation of AI!

- Though they're not exactly the same.

More discussion on naming (Bommasani et al., 2021) <https://arxiv.org/pdf/2108.07258.pdf>

# On terminology

- These names are sometimes used interchangeably:
  - Self-supervised models
  - Pre-trained models
  - Generative AI models
  - Foundation models
  - Frontier models
  - ...
- Though they're not exactly the same.

They're user-facing  
"frontiers" of applications

# On terminology

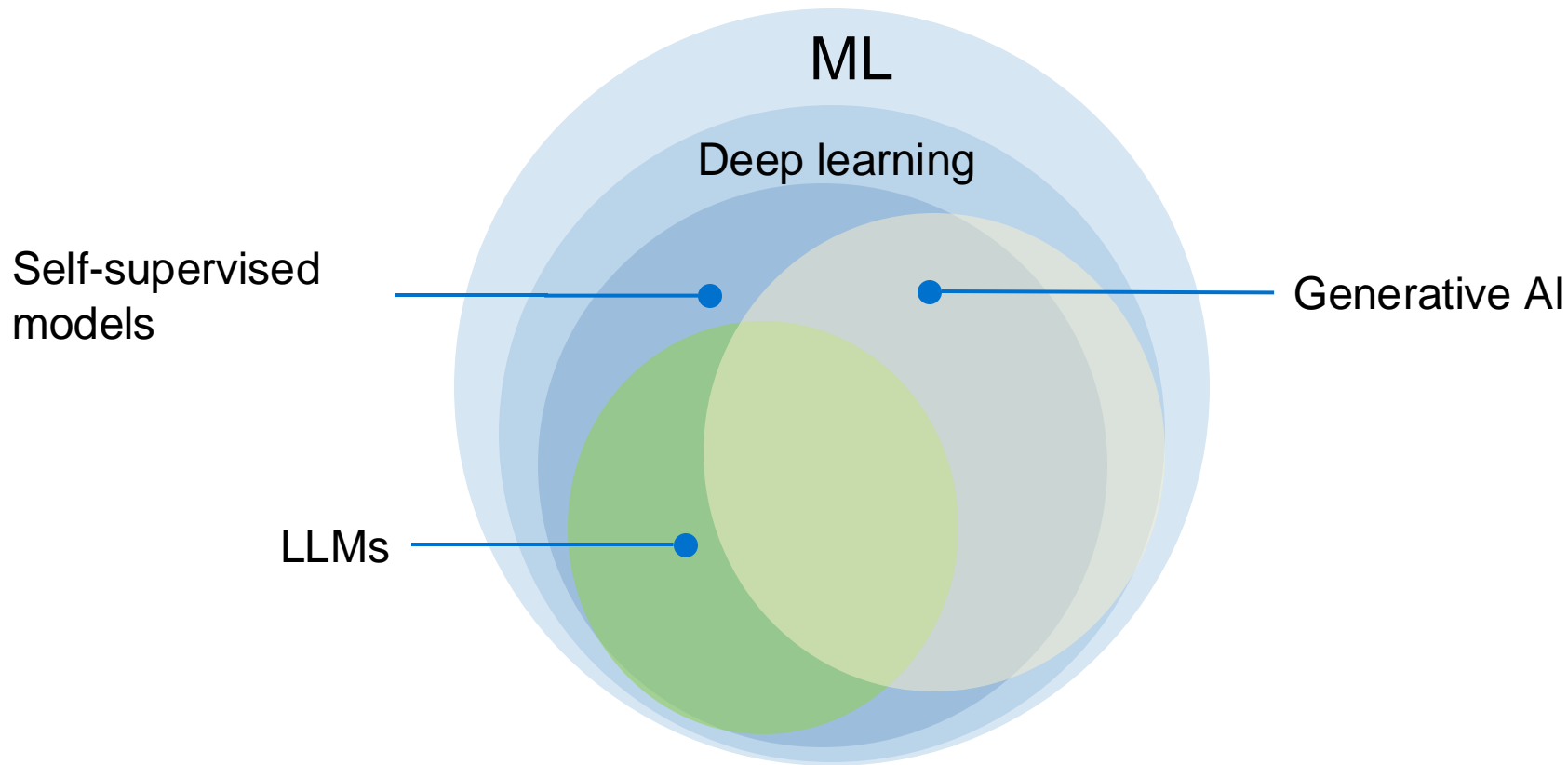
- These names are sometimes used interchangeably:
  - Self-supervised models
  - Pre-trained models
  - Generative AI models
  - Foundation models
  - Frontier models
  - ...
- Though they're not exactly the same.

# On terminology

- These names are sometimes used interchangeably:
  - Self-supervised models
  - Pre-trained models
  - Generative AI models
  - Foundation models
  - Frontier models
  - ...

Which is your favorite?







# Current state of Self-supervised Models

- Almost every AI model is based on Neural networks
- Performance is consistently improving with scale
  - More training data
  - Larger models (number of neural network parameters)

# Current state of Self-supervised Models

State-of-the-art models are hundreds of billions of parameters

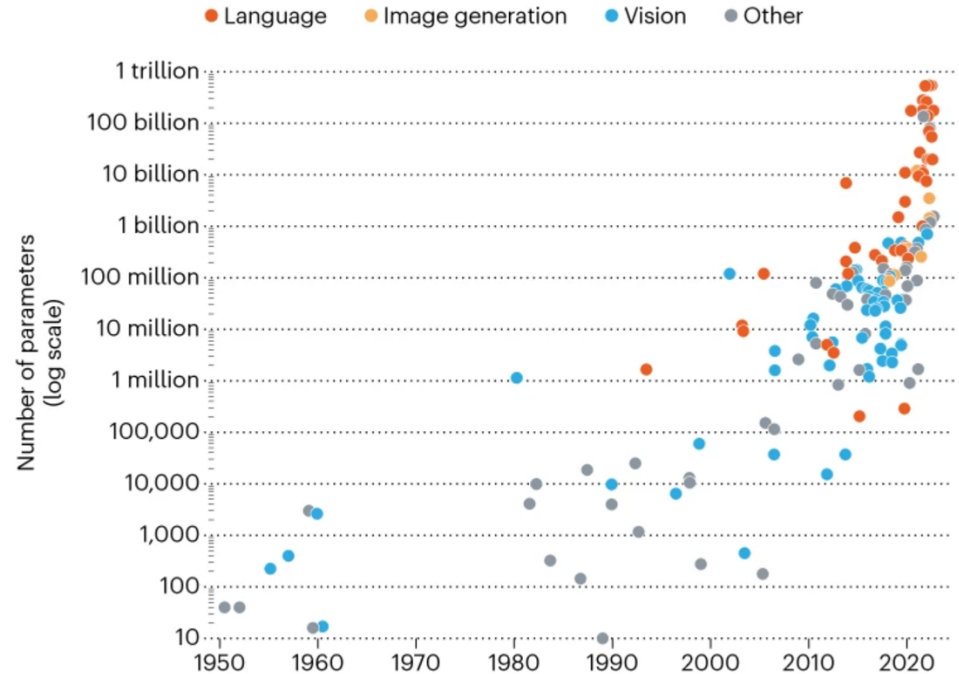


Image from: In AI, is bigger always better? <https://www.nature.com/articles/d41586-023-00641-w>

# Current state of Self-supervised Models

State-of-the-art models are hundreds of billions of parameters

Trained on vast amounts of data (Trillions of tokens)

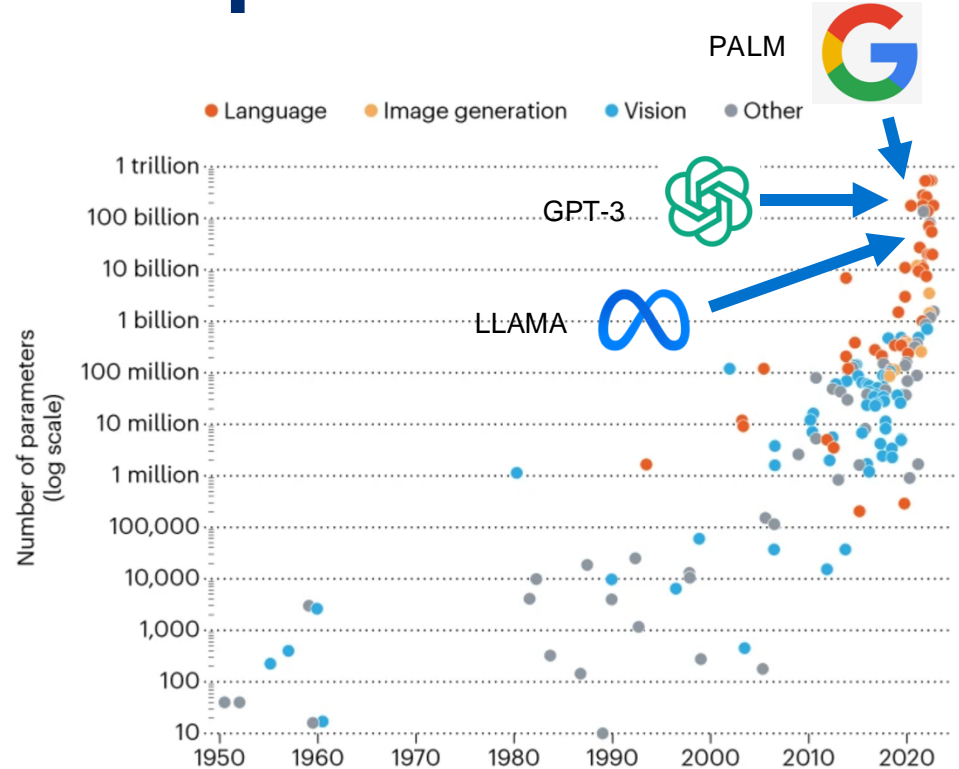


Image from: In AI, is bigger always better? <https://www.nature.com/articles/d41586-023-00641-w>

# Course Learning Objectives

- Read and discuss papers at the cutting edge of technology.
- We will paper on a variety of topics:
  - Pre-training
  - Alignment
  - Safety
  - Efficiency
  - Interaction with physical world, ...
- Goals:
  - Grow our technical understanding of the field.
  - Grow our soft skills: presentation, critique, discussion, group work.

# Course Logistics Brief

- **Instructor:** Daniel Khashabi
  - You can call me “Daniel”, as long as we act mutually respectfully.
- **Course Assistant:** TaiMing (Terry) Lu



Daniel Khashabi  
Instructor



TaiMing Lu  
Course Assistant

# During the Class

- **A small team** (usually 2 students) present 1-2 papers (20-25 minutes).
- **The class** will critically discuss to understand contribution (remaining 50 mins).
  - 1. Breakout groups:**
    - Smaller groups of ~5 people to give you a chance to discuss the paper in a smaller group. (20 mins)
    - These are assigned by Terry for each session.
  - 2. Class discuss:**
    - We will start by summarizing the most important ideas discussed in the breakout groups.
    - We will then continue our open-ended conversation.

# Before the Class

---

- **The presenter students:**

- You will know that you're presenting 10 days in advance.
  - Terry is in charge of assignments. Negotiate with him! :)
- You need to share your slides with Terry and me 48 hours before the class.
  - We will give you feedback so that you can improve your slides.

- **The non-presenter students:**

- Write a 3-4 discussion bullets (sentences/questions) about the paper.
  - We will use these for in-class discussions.
- Avoid descriptive summaries of the paper or generic statements.
- These should be probing, analytical, and thought-provoking comments.
  - For example, you may choose to highlight a problem or limitation with one of the readings. Or you may offer a better approach or method.

# Course Prerequisites

- Comfortable with programming, particularly Python
- Comfortable with the foundations:
  - Transformer, pre-training, fine-tuning
  - Some basic understanding of alignment
- **HW1 should give a sense how prepared you are!**



# Grading Policy

- **One** Homework (individual): 10%
- In-class participation (individually): 10%
- Discussion sentences (individually): 10%
- Timely sharing of slides for feedback (team): 10%
- In-class presentation (team): 10%
- Final project (team): 50%
- Helping us improve the course : +3%

# Final Project

- Must be **exploring a topic related to the focus of the class.**
- This is your **chance to gain research experience** on a topic of interest.
- Topic choice will be (relatively) free. We will help you develop your ideas!
- **Deliverables:**
  1. Submit project **proposal** outline (for our formal review and suggestions)
    - To make sure that the project is scoped reasonably and doable in your limited time.
  2. Get excited 🥳 and work on the project
  3. Midway report
  4. Final project presentation, poster session, report

# Self-Supervised Models: Course Website

CS 771      Schedule      Expectations      Final Project      Conduct

## CS 601.771 Advances in Self-supervised Models

Johns Hopkins University - Fall 2024


Large self-supervised (pre-trained) models (such as Large Language Models or LLMs) have transformed various data-driven fields, such as natural language processing (NLP). This advanced course aims to provide a holistic view of the issues related to these models. The class will mainly involve reading and discussing recent papers in the field.

The focuses of this class will involve various issues: data efficiency, robustness, long context, multi-modality, reasoning grounded in web or physical world, security/legal/privacy issues.


**Note:** The course is different from (more advanced than) 601.471/671 (offered in the spring semesters) which is focused on building the foundational concepts.

**Prerequisites:** Natural Language Processing (CS 465/665), NLP: Self-Supervised Models (CS 471/671), or instructor consent.

**Relevant Courses at Hopkins:** This course has some overlap with "Natural Language Processing" (EN 601/665), and "Artificial Agents" (EN 601.470/670), though the courses have different focuses.



Daniel Khazabi  
Instructor



TaMing Lu  
Course Assistant

### Logistics

- **Classes:** on Tuesday/Thursday 9 - 10:15 am EST (room: Hodson 303)
- **Office hours:** Daniel office hour: Thursdays 12 - 1 pm EST, or by appointment (Hackerman hall, 316B).
- **Contact:** If you have any questions about the course, you can post them on Slack
- **Virtual or in-person:** The class will be in-person.
- **News and announcements:** All the news and announcements will be made on Slack.
- **COVID:** Students who report symptoms associated with COVID-19 are expected not to attend class and to isolate themselves for at least five days and until they have been symptom-free for 24 hours.
- **Course grade:** Your grade is based on the following activities:
  - (1) one assignment (10%) – individually,
  - (2) in-class participation (10%) – individually,
  - (3) reactions before class (10%) – individually,
  - (4) sharing slides on-time for feedback (10%) – sharing your slides with staff to get feedback; done in group,
  - (5) the actual presentation (10%) – presenting papers assigned to you, done in group,
  - (6) a final project (50%) – done in groups.
- **Potential changes:** The instructor reserves the right to make changes to the syllabus or project due dates. These changes will be announced as early as possible.

# Communication Mechanism

---

- A Slack channel where:
  - We will have a room for sharing general discussions
  - Direct communication with me, if needed.
  - You can also create channels for your teamwork.
  
- Are people comfortable with Slack?

# Quick pulse check (1)

- I have understood the course expectations!
  - Yes
  - No

# Quick pulse check (2)

- I am a .....:
  - MS student,
  - PhD student,
  - something else



# JOHNS HOPKINS

WHITING SCHOOL  
*of* ENGINEERING

© The Johns Hopkins University 2023, All Rights Reserved.