



Course Overview

CSCI 601-471/671 (NLP: Self-Supervised Models)

<https://self-supervised.cs.jhu.edu/sp2024/>

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

- Learn about **architectures** that serve as foundation for tools like ChatGPT
- Understand a wide range of self-supervised **algorithms**
- Fearlessly build and train self-supervised **models** 😅
- Gain intuitions about the **capabilities** and **limitations** of our models
- Reflect on **societal impacts**: their misuse, biases, etc.

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

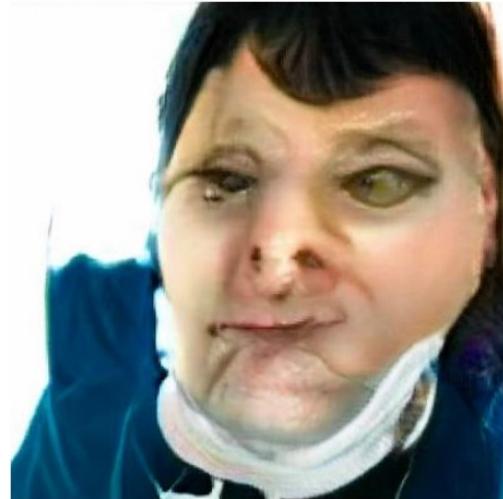


Generated image, from "Large Scale GAN Training for High Fidelity Natural Image Synthesis", Brock et al.

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



`== lemon`

WIKIPEDIA
The Free Encyclopedia

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.

Dataset of Wikipedia articles

Self-Supervision



== lemon

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 ...

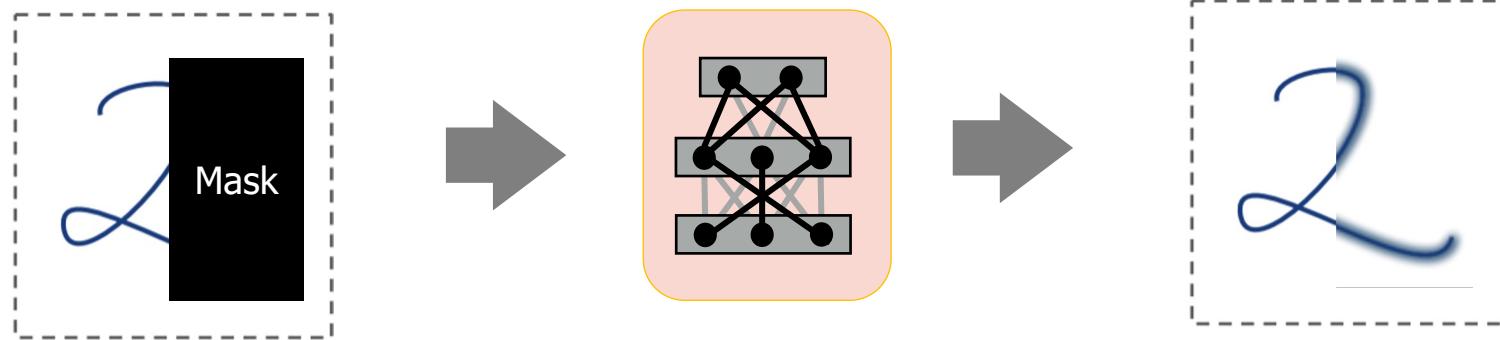
Self-Supervised Models

*are ***predictive models*** of the world!*

Self-Supervised Models

are predictive models of the world!

- Are trained to complete partial samples from the world.

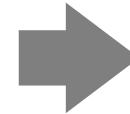
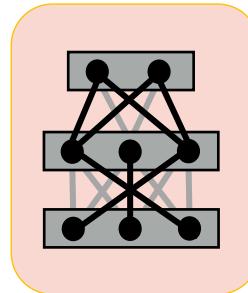
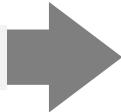


Self-Supervised Models

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

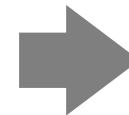
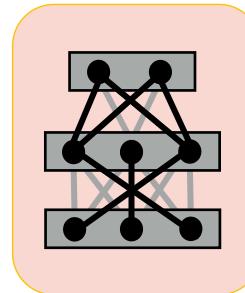
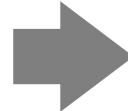
*are tightly connected to **tasks** we care about.*

- **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.



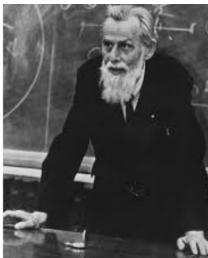
How did we get here?



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*

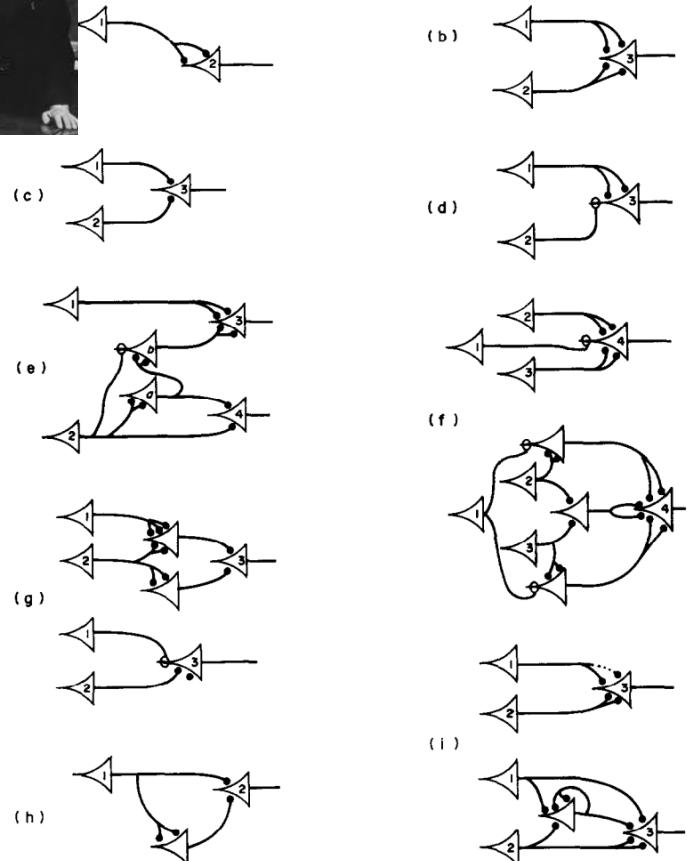
■ WARREN S. McCULLOCH AND WALTER PITTS

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.



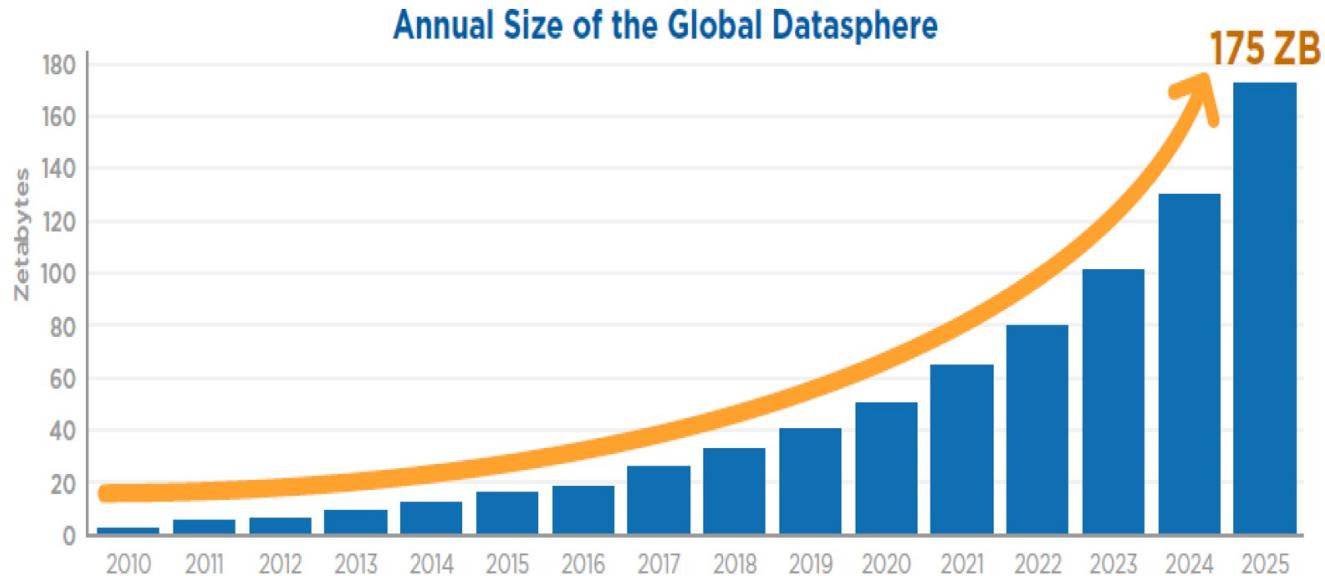
Progress in AI

- 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



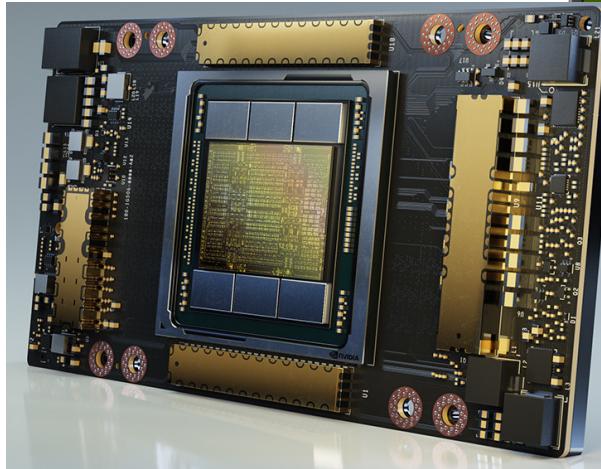
- Internet provided us with a massive repository of data.



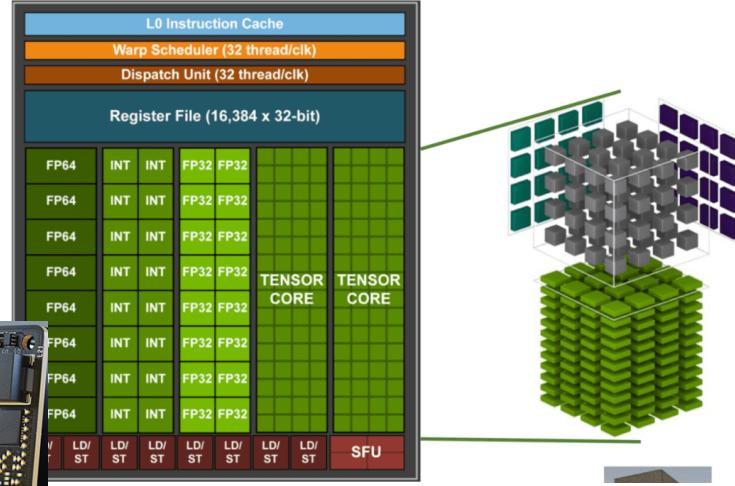
Source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018

Force 2: Computing Power

- Fast processors for deep learning!



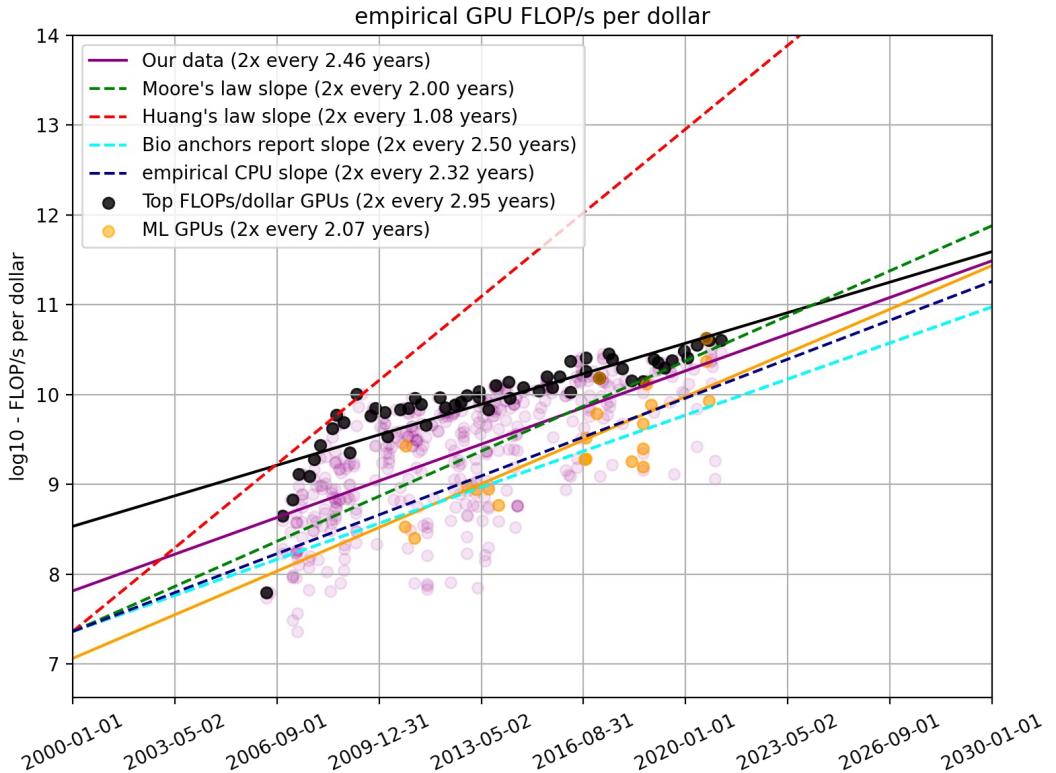
A100 GPU



Force 2: Computing Power

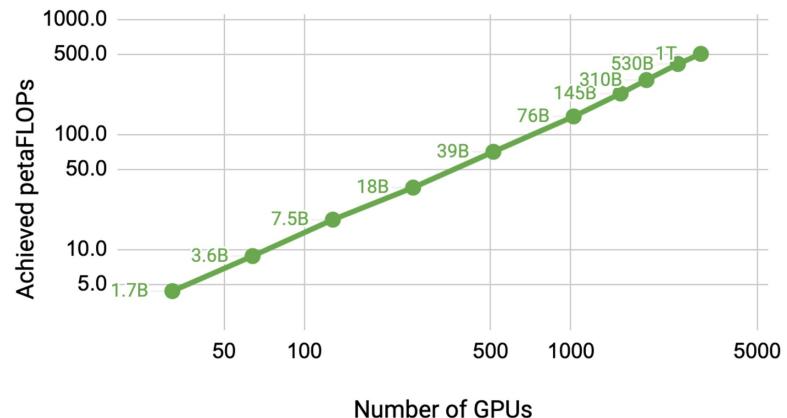
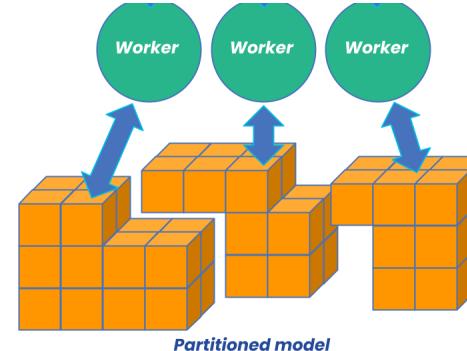
- Fast processors for deep learning!
- Cheaper computing power over time.

The amount of computing power, per dollar



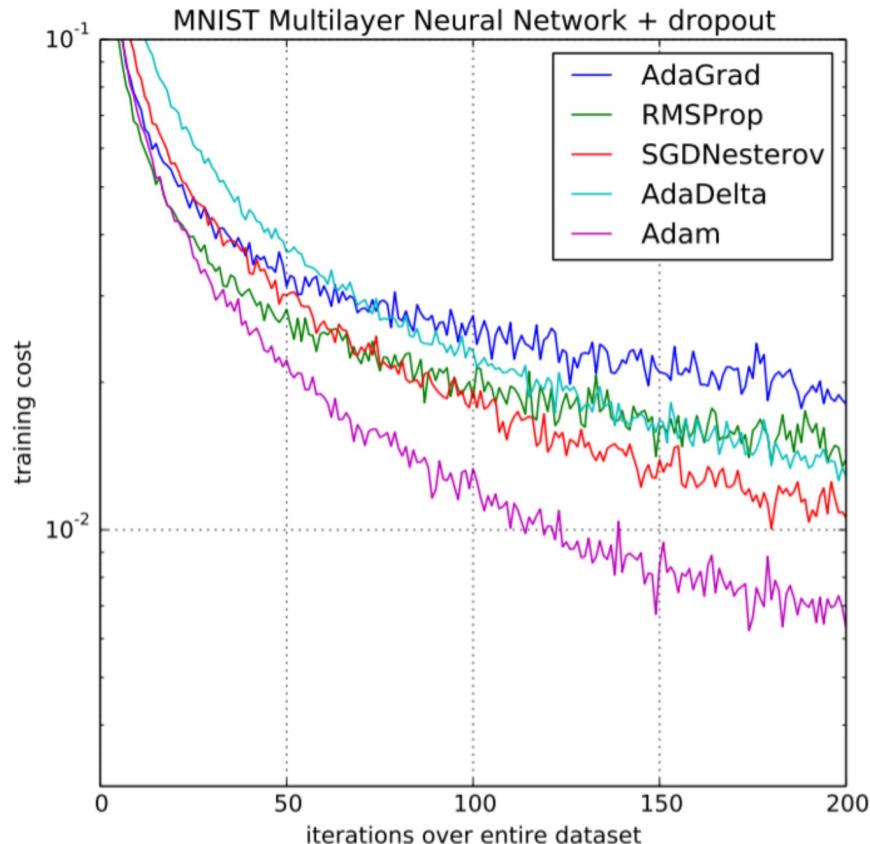
Force 2: Computing Power

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



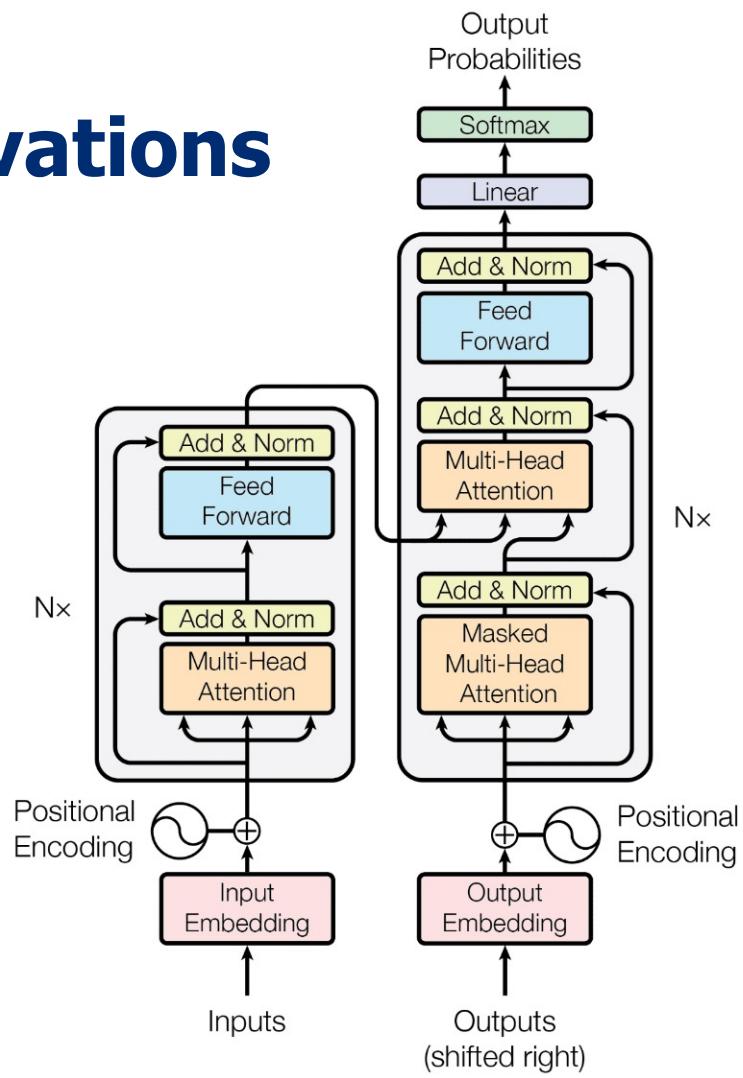
Force 3: Algorithmic innovations

- 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.
- ~ Models that generate content

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

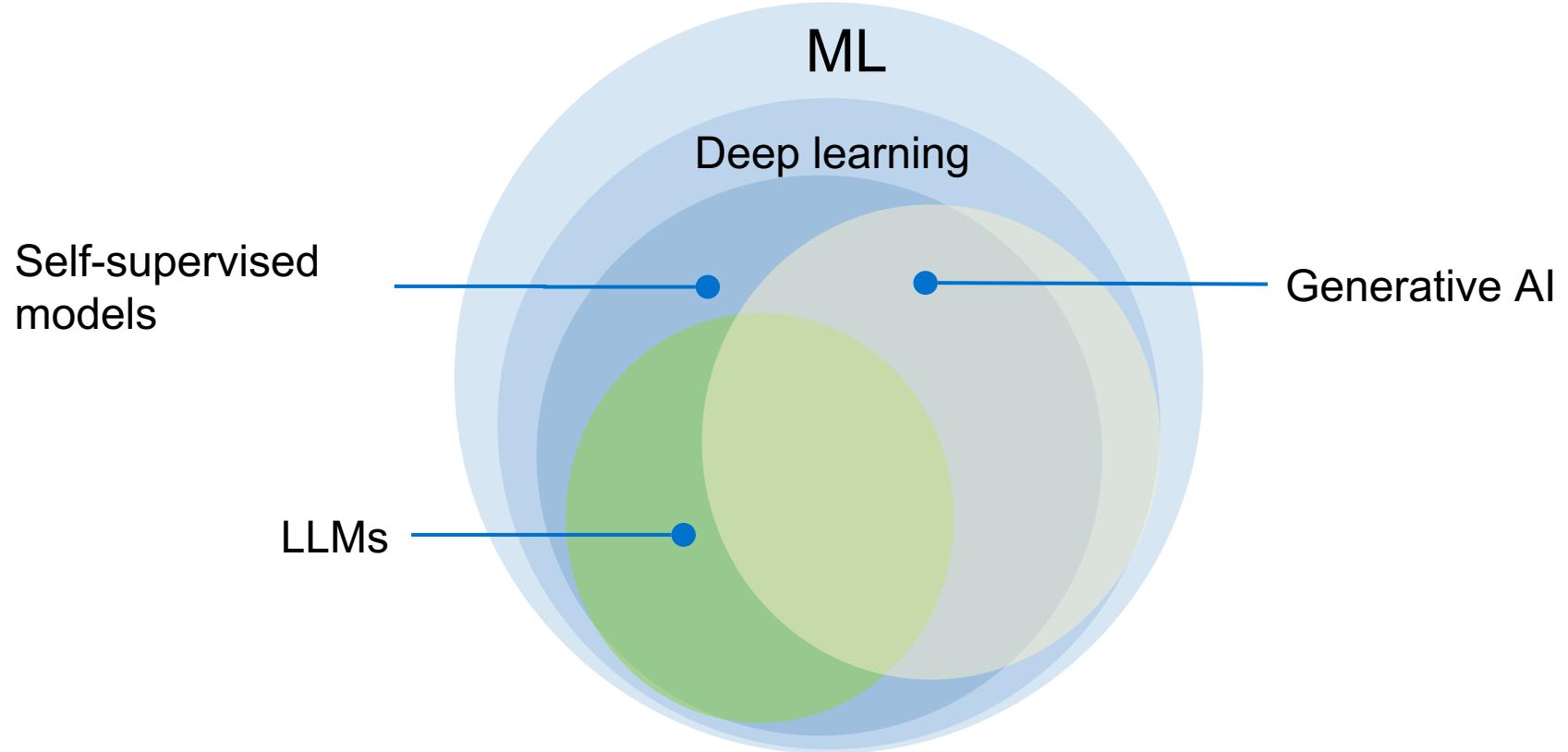
- These names are sometimes used interchangeably:
 - Self-supervised models
 - Pre-trained models
 - Generative AI models
 - Foundation models
 - Frontier models
 - ...
- Though they're not exactly the same.

We will see it!

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

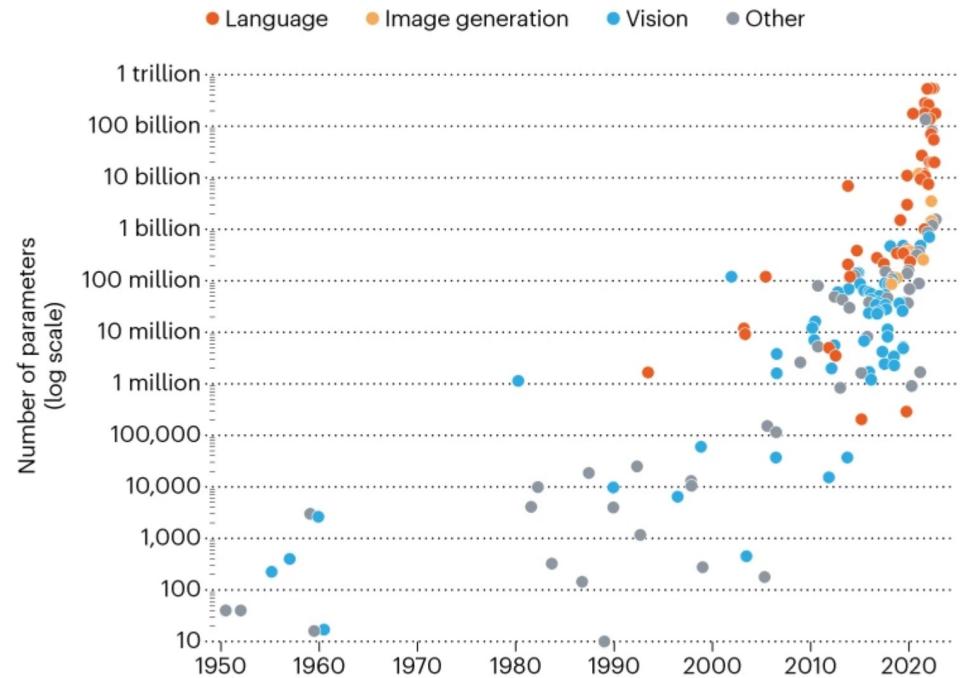


Image from: In AI, is bigger always better? <https://www.nature.com/articles/d41586-023-00641-w>

[Slide credit: Arman Cohan]

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)

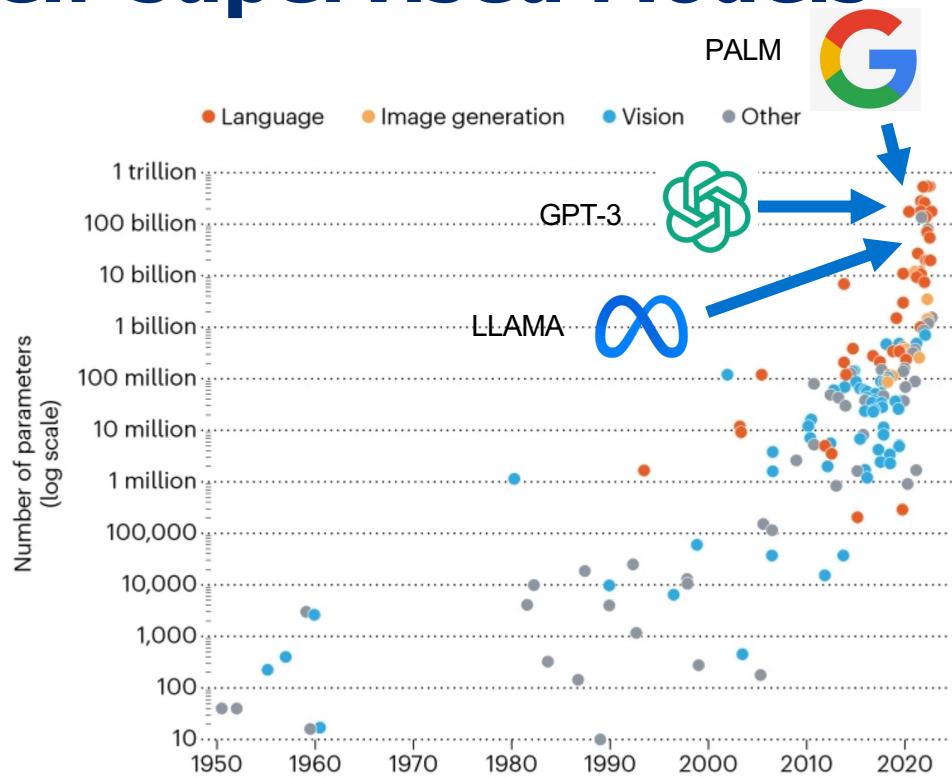


Image from: In AI, is bigger always better? <https://www.nature.com/articles/d41586-023-00641-w>

Taco Tuesday



Jacqueline Bruzek ×

Taco Tuesday

Hey Jacqueline,

Haven't seen you in a while and I hope you're doing well.

A screenshot of the Google Translate mobile application. At the top, the status bar shows the time as 6:56, signal strength, Wi-Fi, and battery level. Below the status bar is a blue header with the "Google Translate" logo and a user profile icon. The main interface shows the source language "English" on the left and the target language "Spanish" on the right, separated by a bidirectional arrow icon. Underneath this, there is a speaker icon followed by "ENGLISH" and a close "X" button. The text "I would like a beer please." is displayed in English. Below the text input area are three icons: "Camera" (with a camera icon), "Conversation" (with a people icon), and "Transcribe" (with a microphone icon). A horizontal line separates these from the bottom section. The bottom section features a blue bar with the word "SPANISH" on the left and a star icon on the right. The translated text "Me gustaría una cerveza por favor." is displayed in Spanish.

6:56

Google Translate

English ⇌ Spanish

ENGLISH X

I would like a beer please.

Camera Conversation Transcribe

SPANISH ☆

Me gustaría una cerveza por favor.



Hi, how can I help?

what time will my Hermes parcel arrive



Your parcel from Hermes
will arrive today



DALL-E Generation

“Photo of hip hop cow in a denim jacket recording a hit single in the studio”



Not Morgan Freeman!



<https://www.youtube.com/watch?v=oxXpB9pSET0>

Course Learning Objectives

- Learn about **architectures** that serve as foundation for tools like ChatGPT
- Understand a wide range of self-supervised **algorithms**
- Fearlessly build and train self-supervised **models** 😅
- Gain intuitions about the **capabilities** and **limitations** of our models
- Reflect on **societal impacts**: their misuse, biases, etc.

Disclaimer

- We will not be exhaustive.
- The goal is to give you a bird's-eye view of the field.





Course Logistics

CSCI 601-471/671 (NLP: Self-Supervised Models)

<https://self-supervised.cs.jhu.edu/sp2024/>

Course Logistics Brief

- **Instructor:** Daniel Khashabi
 - You can call me “Daniel”, as long as we act mutually respectfully.
- **TA:** Jiefu Ou
- **CAs:** Kevin, Sungwon, Tianjian, Yining, Camden



Daniel Khashabi
Instructor



Jiefu Ou
Teaching Assistant



Kevin Kim
Course Assistant



Sungwon Kim
Course Assistant



Tianjian Li
Course Assistant



Yining Lu
Course Assistant



Camden Shultz
Course Assistant

Self-Supervised Models: Course Website

CS 471/671 Schedule Assignments Project Suggest an edit

CS 601.471/671 NLP: Self-supervised Models

Johns Hopkins University - Spring 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). In this course, students will gain a thorough introduction to self-supervised learning techniques for NLP applications. Through lectures, assignments, and a final project, students will learn the necessary skills to design, implement, and understand their own self-supervised neural network models using the Pytorch framework.

Note: The course is different from [601.771](#) (offered in the fall semesters) which is focused on advanced topics in recent papers and is geared toward grad students that want to specialize in the latest developments in self-supervised models.

Prerequisites: (1) Data Structures (601.226), (2) Background in Natural Language Processing & Machine Learning or having finished one of the relevant courses such as Machine Learning (CS 475/675), Machine Learning: Deep Learning (CS 482/682), Natural Language Processing (CS 465/665), Machine Translation (CS 468/668). (3) All the class assignments will be in Python/PyTorch. If you don't know Python or PyTorch but have experience in other programming languages (Java, C++, etc.) you can probably pick Python/PyTorch pretty quickly. (4) Calculus and linear algebra: you should be comfortable with matrix operations (matrix multiplication, transpose, inverse, dot product, gradients). (5) Probability: basic probability properties (conditionals, marginals, mean, standard deviation), distributions (normal, categorical, etc.).

Relevant Courses at Hopkins: This course has some overlap with "Natural Language Processing" (EN.601/665), "Introduction to Human Language Technology" (601.467/667), and "Artificial Agents" (EN.601.470/670), though the courses have different focuses.

Self-Supervised Models: Other websites



piazza

Course Prerequisites

- Comfortable with programming, particularly Python
- Calculus, linear algebra, probability
- Familiarity with Natural Language Processing
- HW1 should give a sense how prepared you are!

Course Logistics Brief

- Lots of important information on the **website**:
 - <https://self-supervised.cs.jhu.edu/sp2024>
 - Syllabus, lectures, videos, tutorials, etc.
- **Pre-recorded videos**: will be available on the course website and/or Piazza.
- **Office hours**: regular office hours; will be announced on Piazza.
- **Questions or discussions**: Piazza

Course Work and Grading Policy

- **Homework (team optional): 40%**
 - 7 x 1-week assignments
- **Midterm quizzes (individually): 30%**
- **Final course project (team optional): 30%**
- **Bonus points:**
 - **Extra credit questions (HW and quizzes): +10%**
 - **Helping us improve the course : +3%**

Homework

- **Timeline:**
 - Released **every Tuesday** and due in a week.
 - Each have both **conceptual (theory)** and **empirical (coding)** parts.
 - All assignments will be submitted to Gradescope.
- **Late day policy**
 - 10 free late days across all assignments; afterwards, 5% off course grade per day late
 - Assignments not accepted more than 7 days late (unless given permission in advance)
- **Collaboration policy:** Allowed within certain limits:
 - Needs to be well-documented; acknowledge working with other students.
 - Avoid: copy code off the web or ChatGPT! — write your own solutions.
- **HW1 is released today! Due next week. Submitted via Gradescope.**

Final Project

- No HW after spring break to let you focus on your 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

Attendance Policy

- Not mandatory!

Attendance Policy

- Not mandatory!
- No additional grade if you attend.
- However, it is highly encouraged!!
 - Participation in class is our chance to learn more effectively.
 - Come to the class and participate in the discussions!

Quick pulse check (1)

- I am a:
 - CS major,
 - CS minor,
 - MS student,
 - PhD student,
 - studying something else

Quick pulse check (2)

- Which one best describes you?
 1. NLP and DL are rather new to me. Looking forward to learning more.
 2. I am a DL, NLP or self-supervised learning enthusiast – I read casually on these topics.
 3. I am practitioner of NLP or ML and have already worked with a variety of self-supervised models.
 4. I am a pro. I could be part of the teaching staff of this class.

Quick pulse check (3)

- I have read (and understood) the syllabus!
 - Yes
 - No

Wrapping it up!

- HW1 is released!
- If you're not going to take this, drop the course!
 - Many are on the waitlist.



JOHNS HOPKINS

WHITING SCHOOL *of* ENGINEERING