



JOHNS HOPKINS

WHITING SCHOOL
of ENGINEERING

Course Overview

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

<https://self-supervised.cs.jhu.edu/sp2025/>

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**



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?



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



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



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

- We will cover a variety of inter-related topics:
 - Architectures
 - Pre-training
 - Alignment and safety
 - Efficiency
 - Interaction with the world (code, physical world, etc.)
 - Impacts on humans -- their misuse, biases, etc.
- Skills:
 - **Technical**—understanding of the algorithms and implementing them.
 - Gaining intuitions about **capabilities** and **limitations** of models.
 - **Soft skills**—algorithms, PyTorch, SLURM, intuition about capabilities, teamwork.

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 "bar which has both culinary and cleaning
ende the C uses.[2] The pulp and rind are also
as th devic used in cooking and baking.
north wheebarrow is designed to
north american theatre,

Self-Supervision



WIKIPEDIA
The Free Encyclopedia

== 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.

the lemon is designed to

north american theatre,

== 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 ...

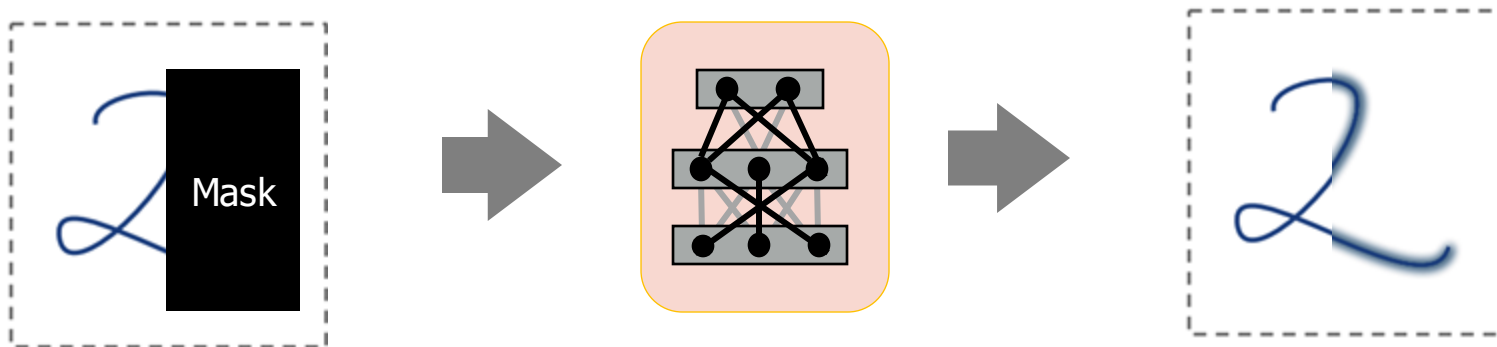
Self-Supervised Models

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

Self-Supervised Models

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

- Are trained to complete partial samples from the world.

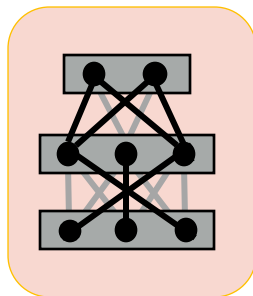


Self-Supervised Models

*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.*

Here is an example ...

Self-Supervised Models

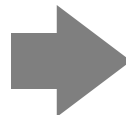
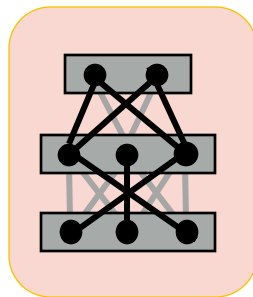
*are tightly connected to **tasks** we care about.*

- **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

1. Are *predictive models* of the world.
2. Are learned from *unlabeled* data.
3. 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



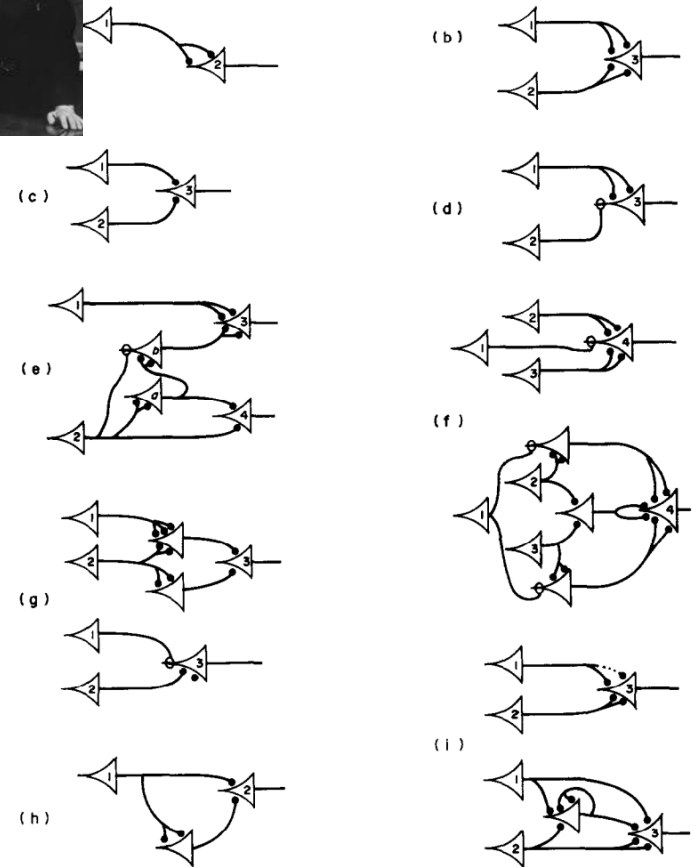
- Many advances are due to **neural networks**
- How old are neural networks?

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.



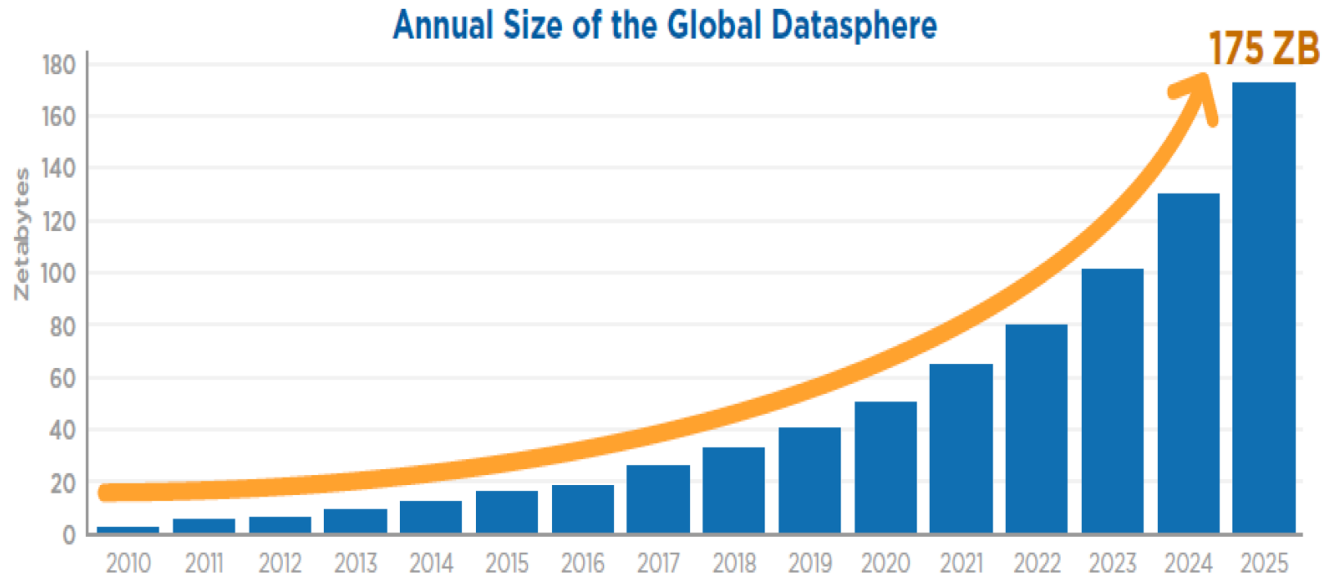
Progress in AI

- 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 necessary forces had to come together!**

Force 1: Massive Amount of Data



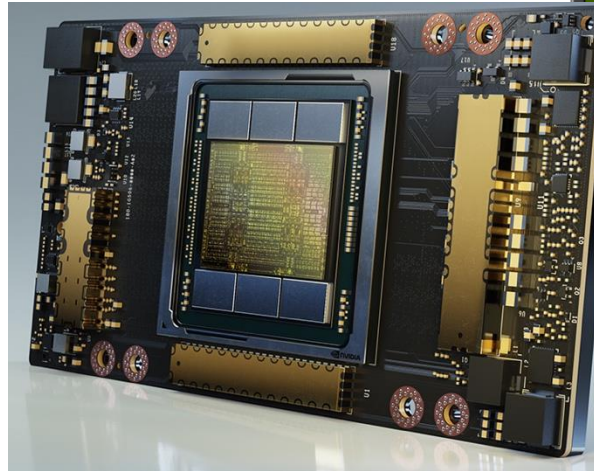
- Internet provided us with a massive repository of data.



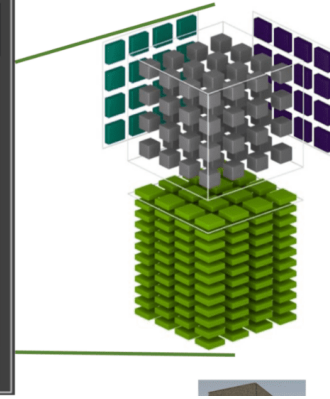
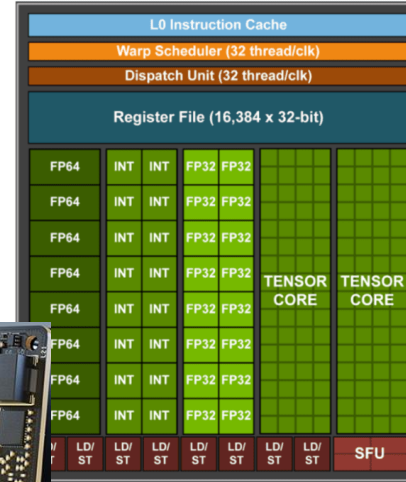
Source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018

Force 2: Computing Power

- Fast processors for deep learning!



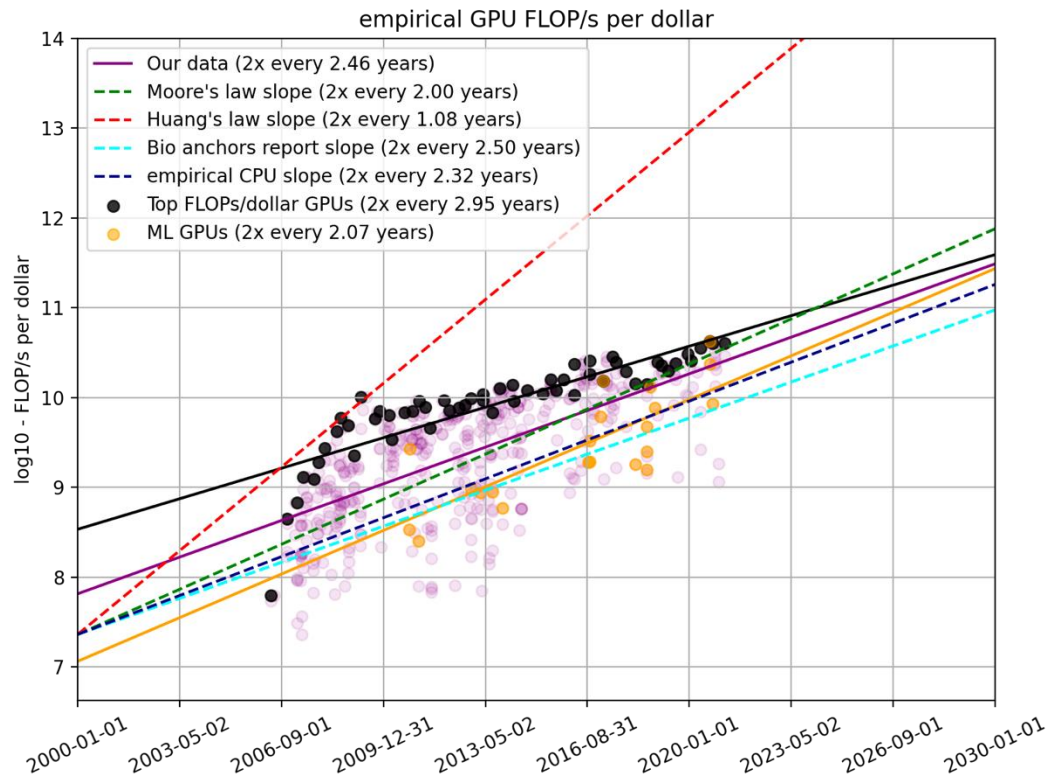
A100 GPU



Force 2: Computing Power

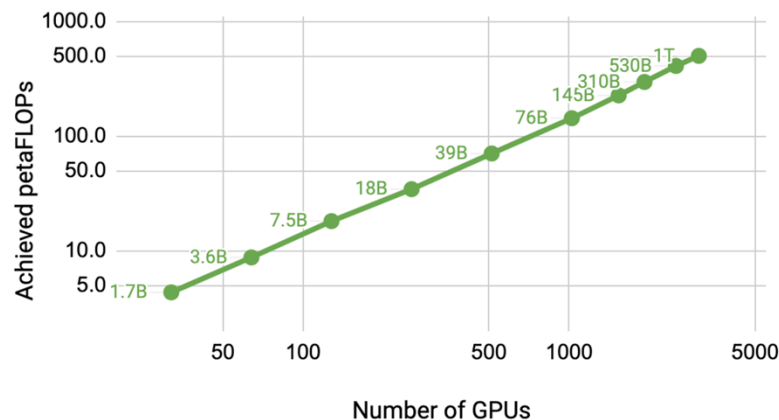
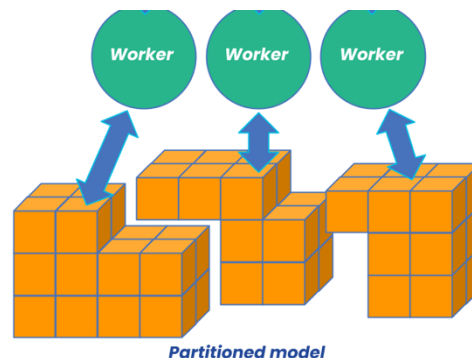
- Fast processors for deep learning!
- Cheaper computing power over time.

The amount of
computing power,
per dollar



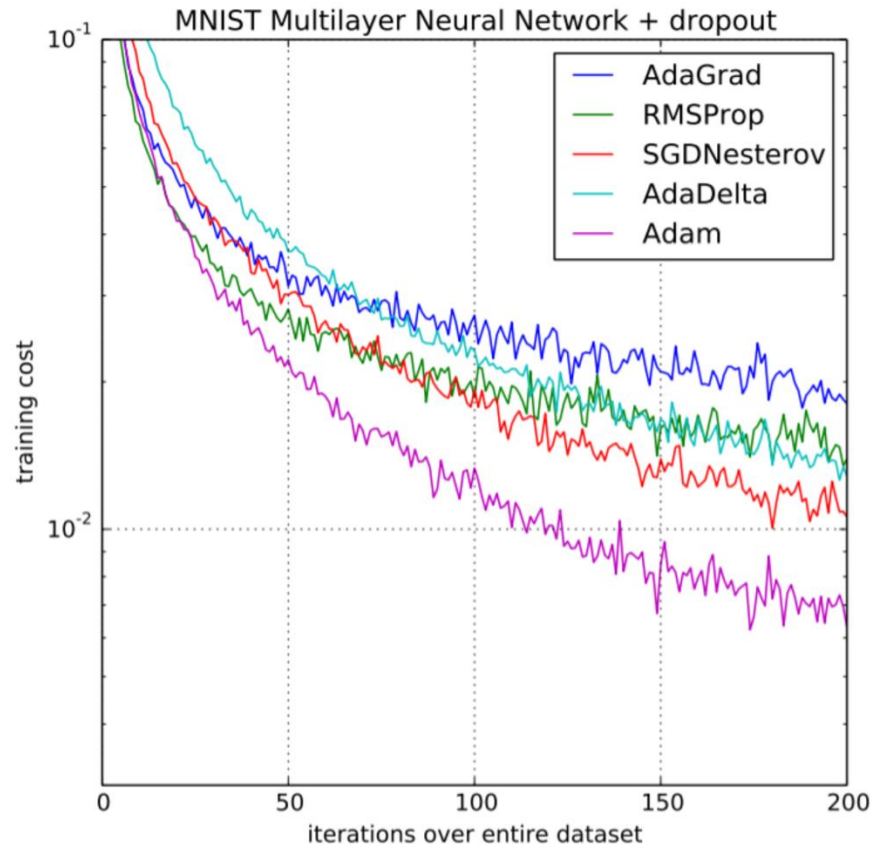
Force 2: Computing Power

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



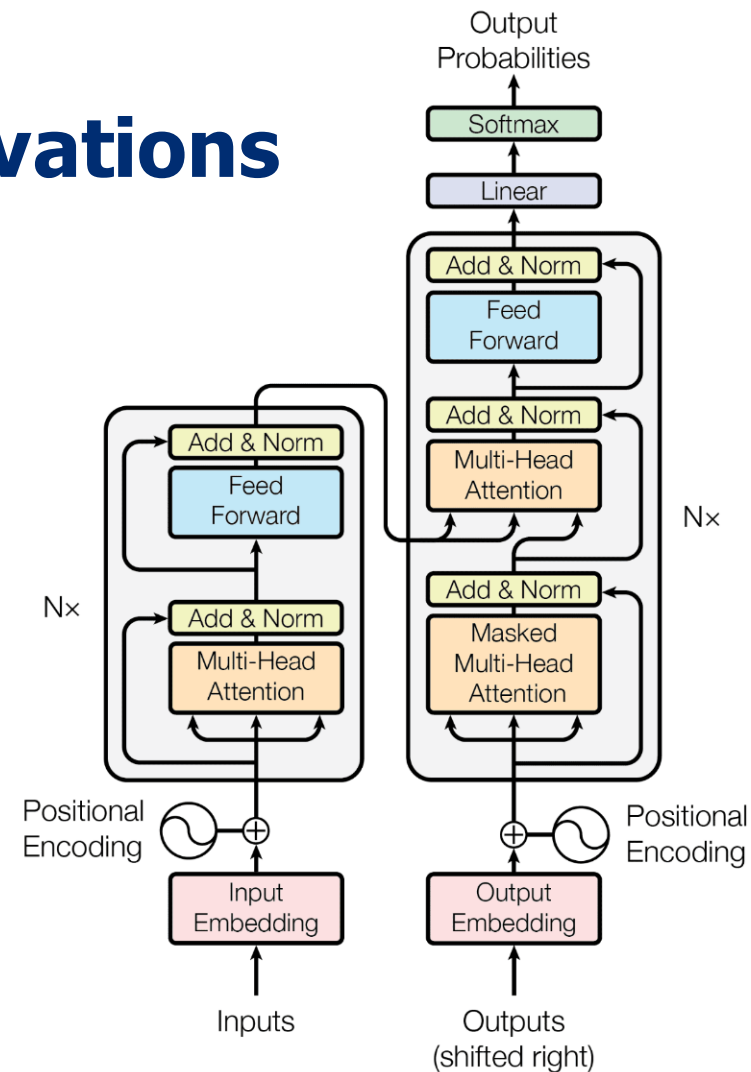
Force 3: Algorithmic innovations

- Advances in optimization

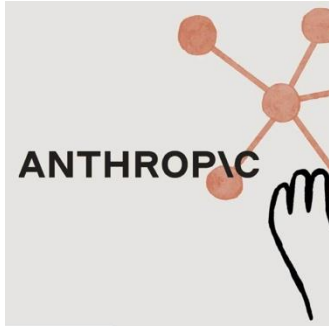


Force 3: Algorithmic innovations

- Advances in optimization
- Innovations in model architectures
-



The success we dreamed of



Language models that are remarkably capable at solving many important NLP benchmarks.

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

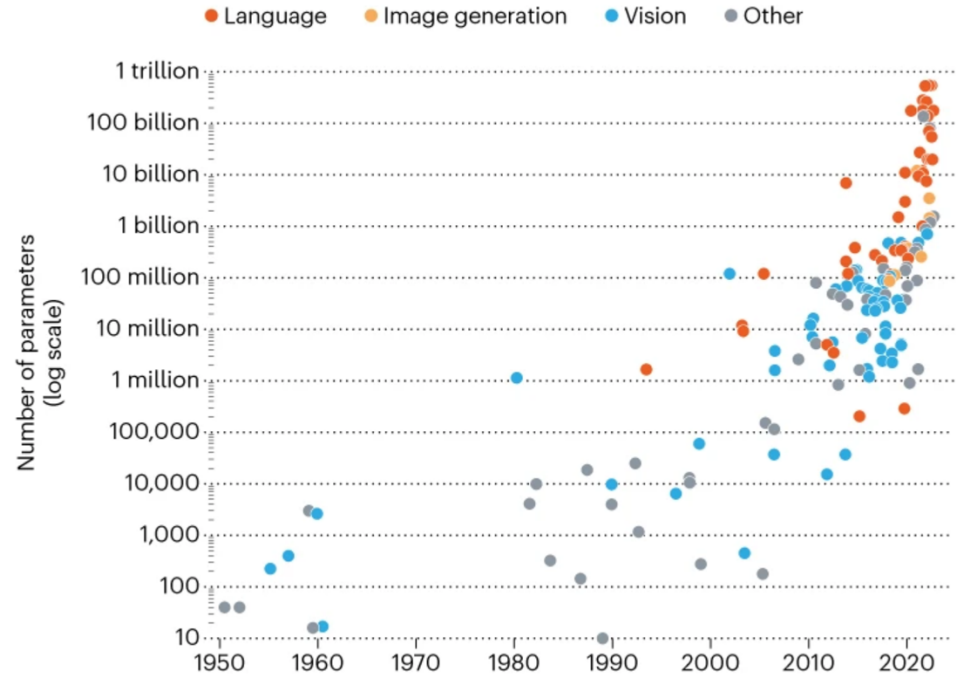


Image from: In AI, is bigger always better? <https://www.nature.com/articles/d41586-023-00641-w>

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)

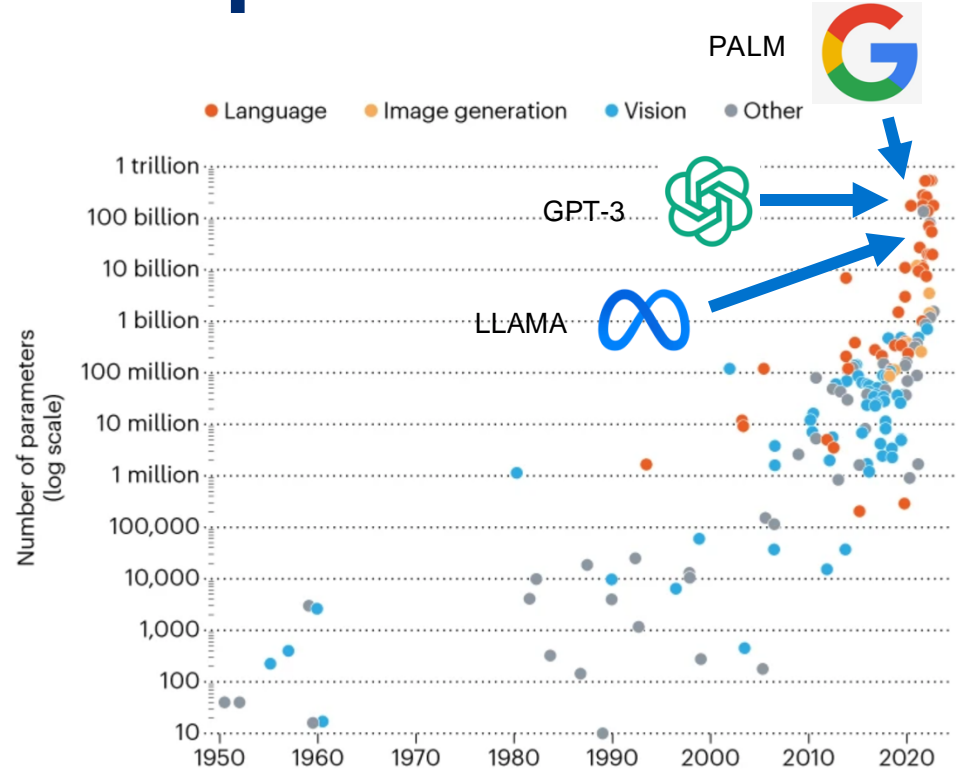


Image from: In AI, is bigger always better? <https://www.nature.com/articles/d41586-023-00641-w>

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
- ...

Our models are
more than just “pretrained”!

- 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
- ...

~ 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

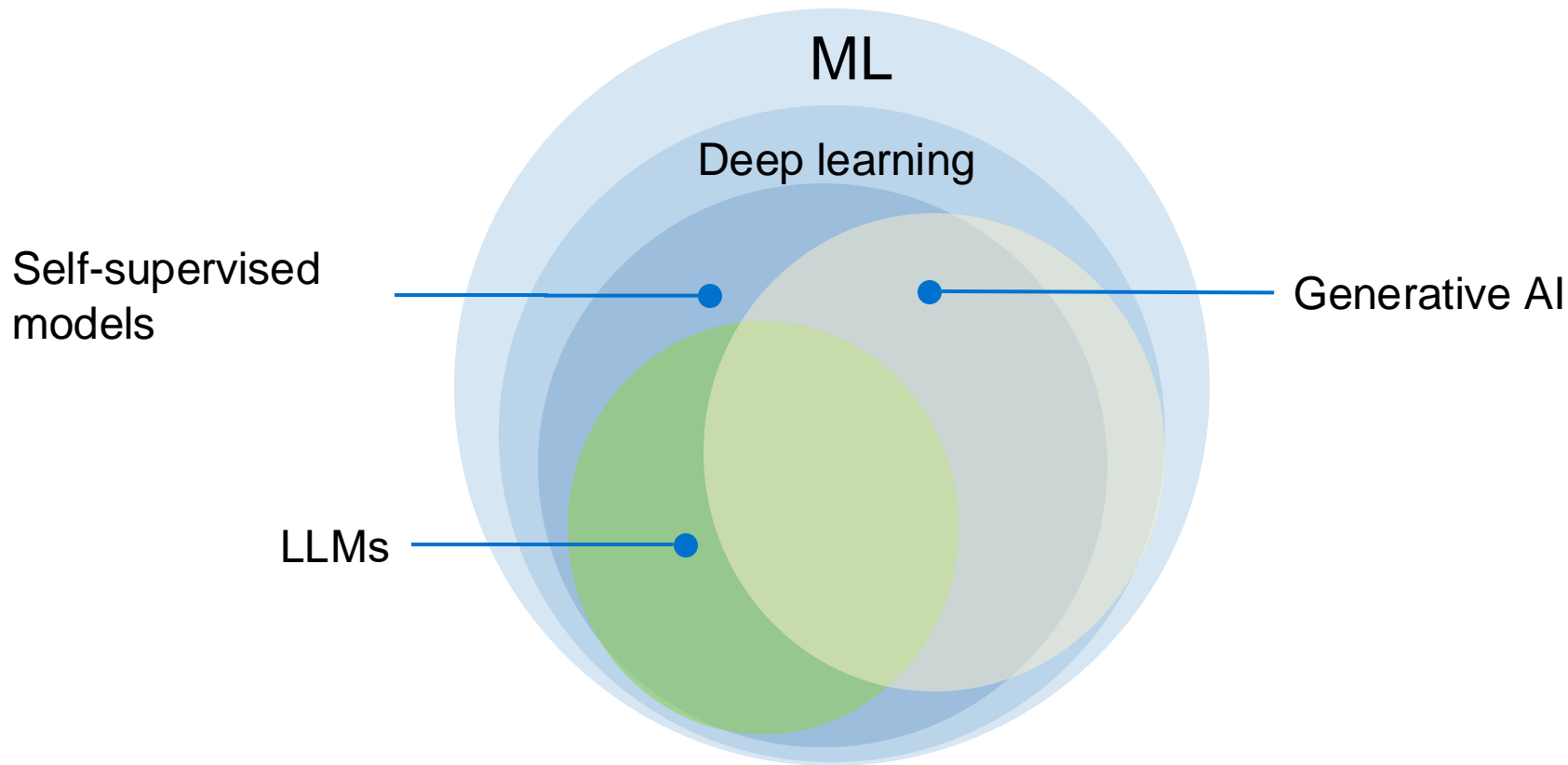
- These names are sometimes used interchangeably:
 - Self-supervised models
 - Pre-trained models
 - Generative AI models
 - Foundation models
 - Frontier models
 - ...
- Though they're not exactly the same.

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?





Course Learning Objectives

- We will cover a variety of inter-related topics:
 - Architectures
 - Pre-training
 - Alignment and safety
 - Efficiency
 - Interaction with the world (code, physical world, etc.)
 - Impacts on humans -- their misuse, biases, etc.
- Skills:
 - **Technical**—understanding of the algorithms and implementing them.
 - Gaining intuitions about **capabilities** and **limitations** of models.
 - **Soft skills**—algorithms, PyTorch, SLURM, intuition about capabilities, teamwork.

Disclaimer

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





JOHNS HOPKINS

WHITING SCHOOL
of ENGINEERING

Course Logistics

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

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

Course Logistics Brief

- **Instructor:** Daniel Khashabi
 - I prefer just “Daniel”, as long as we act mutually respectfully.
 - No "prof Daniel"!! 🙅
- **TA:** Tianjian Li
- **CAs:** Sungwon, Tanay, Adi, TaiMing, Prabhav, Rishitha



Daniel Khashabi
Instructor



Tianjian Li
Teaching Assistant



Adi Asija
Course Assistant



Prabhav Singh
Course Assistant



Rishitha Kalicheti
Course Assistant



Sungwon Kim
Course Assistant



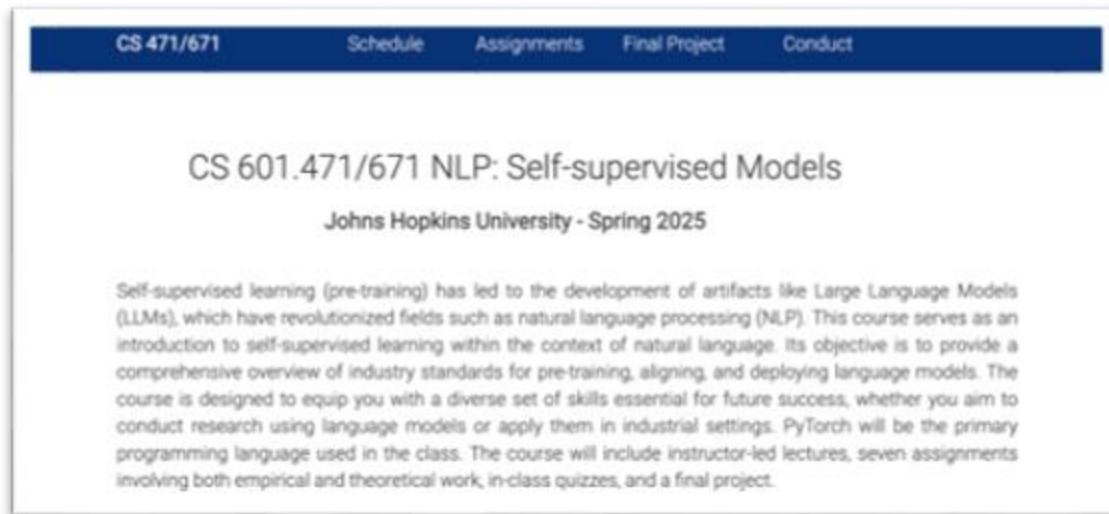
TaiMing (Terry) Lu
Course Assistant



Tanay Nayak
Course Assistant

Course Website

- Lots of important information on the **website**:
 - <https://self-supervised.cs.jhu.edu/sp2025>



Other Relevant Websites



plazza

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!

Quick pulse check

- Do I understand how “gradient descent” works?
 - Yes
 - No

Course Logistics Brief

- **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
- **3xMidterm quizzes (individually): 40%**
- **Final course project (team optional): 20%**
- **Bonus points:**
 - **Extra credit questions (HW and quizzes): +10%**

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**

- 7 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: copying code off the web or ChatGPT! — write your own solutions.

HW1 is released!

- Due Tuesday.
 - Has both theory (background on algebra, etc.) and programming (building a classifier).
- A **baseline** for self-assessment.
 - Future homework will be more advanced.

“Is Typesetting Mandatory?”

- No, but ...
- But **10x strongly** recommended.
- It is a must-have skill if you're considering going to any research field.

The screenshot shows a LaTeX editor interface with a dark theme. The top bar includes a 'Menu' button, a home icon, an 'Upgrade' button, and the document title 'CS-601.471-671-spring2023-homework1'. On the left, a file explorer shows 'header.tex', 'hw1.tex' (selected), and 'macros.sty'. Below it, a 'File outline' panel lists sections like 'Linear Algebra Review', 'Basic Operations', 'Matrix Algebra Rules', 'Probability Review', 'Calculus Review', 'One-variable derivati...', 'Multi-variable deriva...', 'Algorithms and Data Str...', and 'Programming'. The main editor area shows LaTeX source code with line numbers 79 to 92. The code includes `\item` and `\solution` commands for a list of problems, and a `\subsection` command for 'Matrix Algebra Rules'. The right pane shows the rendered output, which includes the title 'CS 601.471/671 NLP: Self-supervised Models', the subtitle 'Homework 1: Background Review + Word Representations', and instructions for the assignment. The rendered text includes a deadline check, a collaborator name field, and a list of prerequisites.

CS 601.471/671 NLP: Self-supervised Models

Homework 1: Background Review + Word Representations

For homework deadline, check the calendar on the course website'

Name: _____

Collaborators: _____

Sources used for your homework: _____

This assignment it combines knowledge and skills across several disciplines. The purpose of this is to make sure you are prepared for this course. We anticipate that each of you will have different st weaknesses, so don't be worried if you struggle with some aspects of the assignment. But if you find this to be very difficult overall, that is an early warning sign that you may not be prepared to take this course

To succeed in the course, you will need to know or very quickly get up to speed on:

- Math to the level of the course prerequisites: linear algebra, multivariable calculus, some probabi
- Statistics, algorithms, and data structures to the level of the course prerequisites.
- Python programming, and the ability to translate from math or algorithms to programming and l
- Some basic LaTeX skills so that you can typeset equations and submit your assignments.

How to hand in your written work: via Gradescope.

Collaboration: Make certain that you understand the course collaboration policy, described on the coo. You may discuss the homework to understand the problems and the mathematics behind the various le

How do I edit the overleaf file?

- Open the link; click on “Menu”; click on “Copy Project” to clone it.

The screenshot displays the Overleaf web interface. On the left, a sidebar contains options like 'Download', 'Source', 'PDF', 'Actions' (with 'Copy Project' and 'Word Count'), 'Sync' (with 'Dropbox' and 'Git'), and 'Settings' (with various configuration options like 'TeX Live version', 'Spell check', 'Dictionary', 'Auto-complete', 'Auto-close Brackets', 'Code check', and 'Editor theme'). The central pane is the 'Visual Editor' showing LaTeX code for a document titled 'CS-601.471-671-spring2024-homework1'. The code includes package declarations, color settings, and a definition for a custom style. The right pane shows the rendered document content, which includes a problem statement about evaluating expressions and a list of seven questions involving vectors, matrices, and norms. The questions are:

- $\sum_{i=1}^n x_i y_i$ (inner product). [Answer:](#)
- $\sum_{i=1}^n x_i z_i$ (inner product between orthogonal vectors).
- $\alpha(\mathbf{x} + \mathbf{y})$ (vector addition and scalar multiplication).
- $\|\mathbf{x}\|$ (Euclidean norm of \mathbf{x}). [Answer:](#)
- \mathbf{x}^T (vector transpose). [Answer:](#)
- $A\mathbf{x}$ (matrix-vector multiplication). [Answer:](#)
- $\mathbf{x}^T A \mathbf{x}$ (quadratic form). [Answer:](#)

Note, you do not need to show your work.

1.2 Matrix Algebra Rules

Assume that $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ are $n \times 1$ column vectors and A, B are $n \times n$ real-valued matrices, and I is the identity matrix of size n . State whether each of the below is true in general (you need to show your work).

- $\mathbf{x}^T \mathbf{y} = \sum_{i=1}^n x_i y_i$. [Answer:](#)

Final Project

- **No HW after spring break** to let you focus on your final project.
- Topic:
 1. Default projects: we will give you well-defined, scoped projects.
 2. Custom projects: can bring your own idea, as long as it is related to the class.
- This is your **chance to hone your technical skills** on a topic of interest.
- **Deliverables:**
 1. Submit project **proposal** outline (for our formal review and suggestions)
 - To make sure the proposals are well scoped and doable in limited time.
 2. Get excited 🥳 and work on the project
 3. Midway report
 4. Final project poster session and 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,
 - None of the above??

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 for the 671 session (not 471).



JOHNS HOPKINS

WHITING SCHOOL
of ENGINEERING

© The Johns Hopkins University 2023, All Rights Reserved.