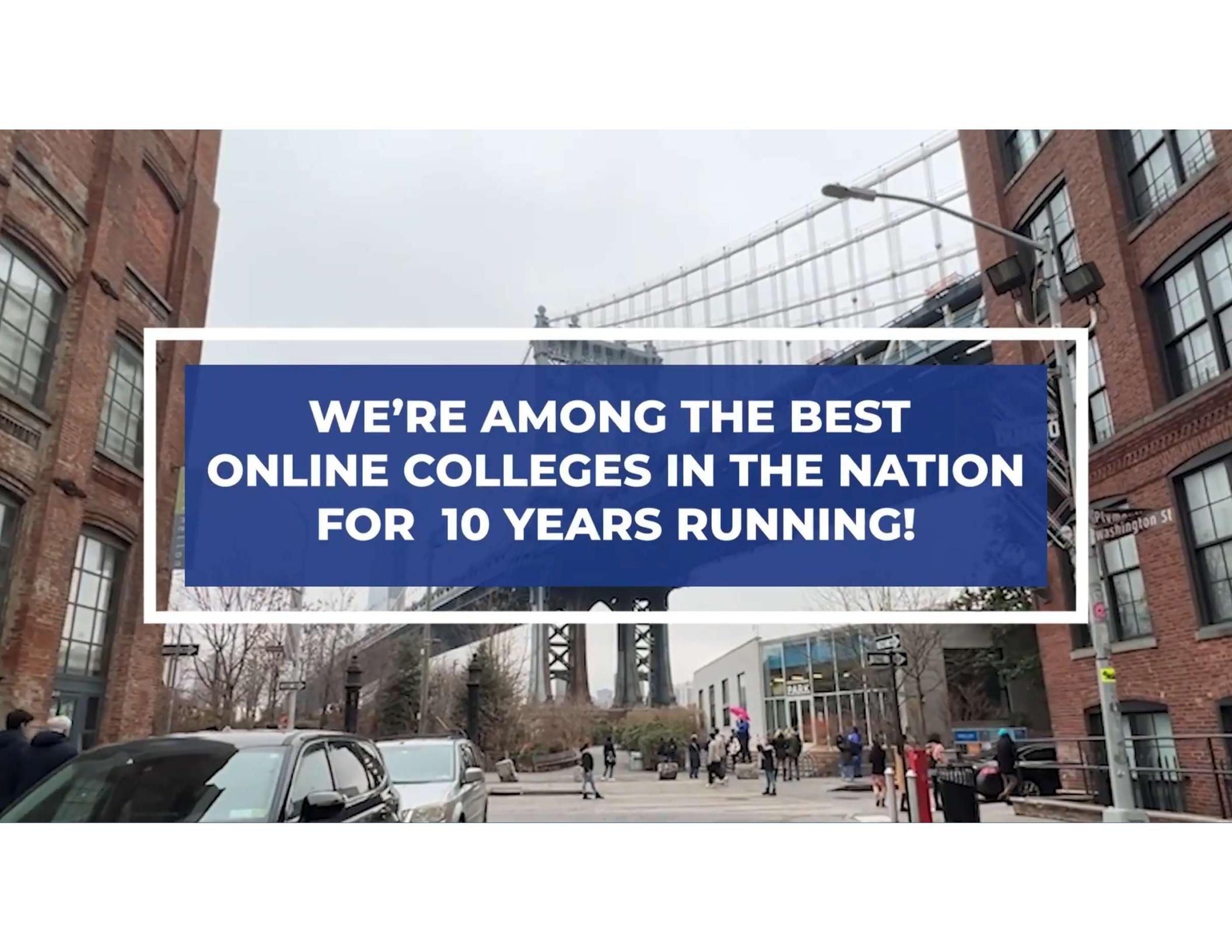


MASTER OF SCIENCE IN GENERATIVE AI

WEEK 1: INTRODUCTION



**HIGHEST RANKED
ONLINE COLLEGE IN
NEW YORK CITY!**

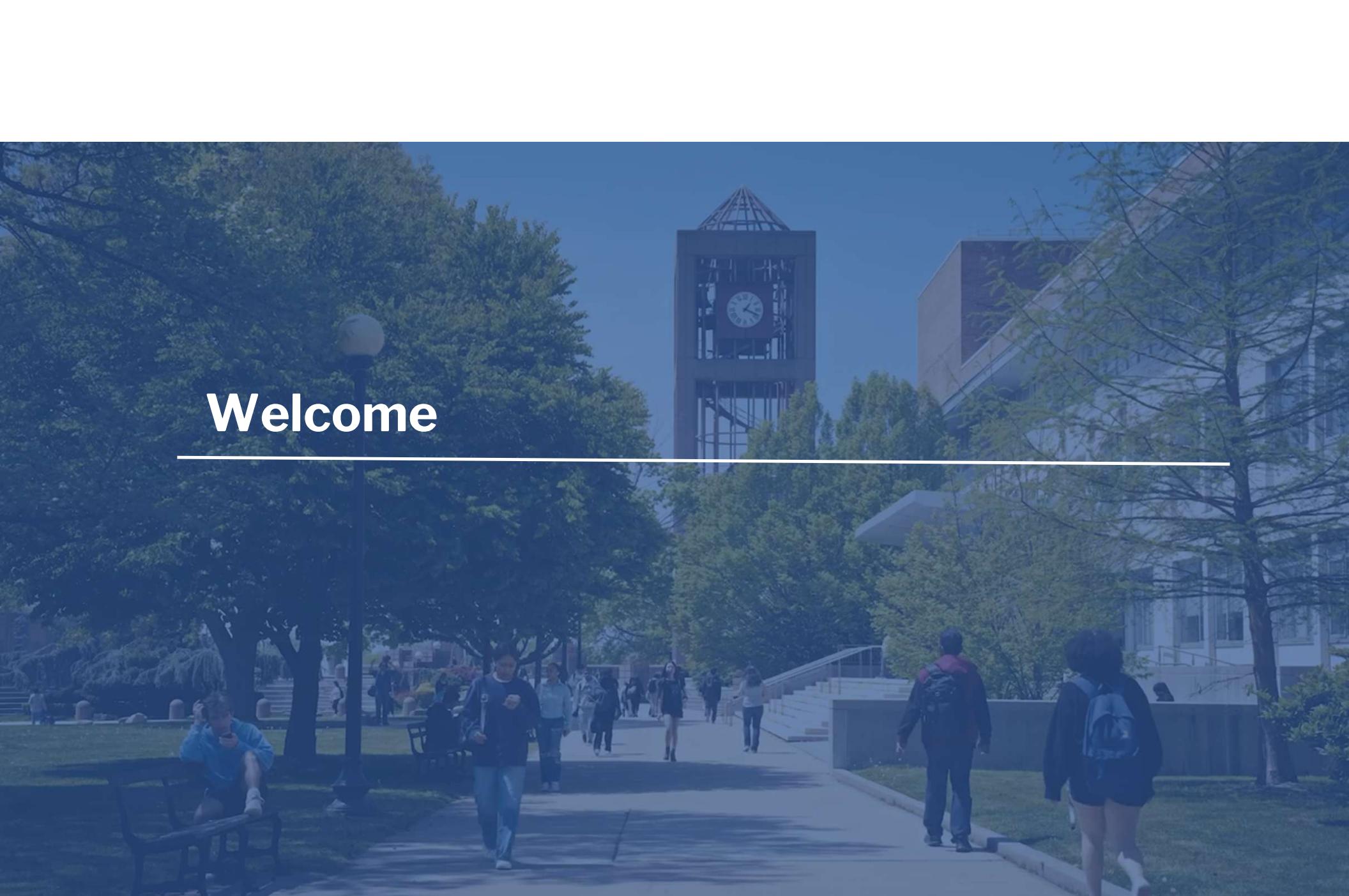


**WE'RE AMONG THE BEST
ONLINE COLLEGES IN THE NATION
FOR 10 YEARS RUNNING!**



Introduction to Machine Learning (GAI601)

Joe Sabelja



Welcome

Syllabus

1. Introduction to Machine Learning
2. Linear and Logistic Regression Models
3. Classification: Discriminant Analysis & k-Nearest Neighbors (kNN)
4. Classification: Naive Bayes & Integrated Methods
5. Decision Trees
6. Ensemble Methods: Bagging, Random Forests, and Boosting (Part 1)
7. Ensemble Methods: Boosting (Part 2) and Additive Trees
8. Support Vector Machines (SVMs)
9. Unsupervised Learning: Clustering
10. Unsupervised Learning: PCA and Distance Metrics
11. Resampling, Model Selection, and Additional Topics
12. Bias-Variance Tradeoff
13. Neural Networks and Foundation Models
14. Ethics and Responsible AI/ML

Schedule

| January | | | | | | | |
|---------------------------|---------|-----------|----------|--------|----------|--------|--|
| MONDAY | TUESDAY | WEDNESDAY | THURSDAY | FRIDAY | SATURDAY | SUNDAY | |
| 29 | 30 | 31 | 01 | 02 | 03 | 04 | |
| 05 | 06 | 07 | 08 | 09 | 10 | 11 | |
| 12 | 13 | 14 | 15 | 16 | 17 | 18 | |
| 19 | 20 | 21 | 22 | 23 | 24 | 25 | |
| 26 | 27 | 28 | 29 | 30 | 31 | 01 | |
| First Day of Class Week 1 | | Class 7PM | | | | | |

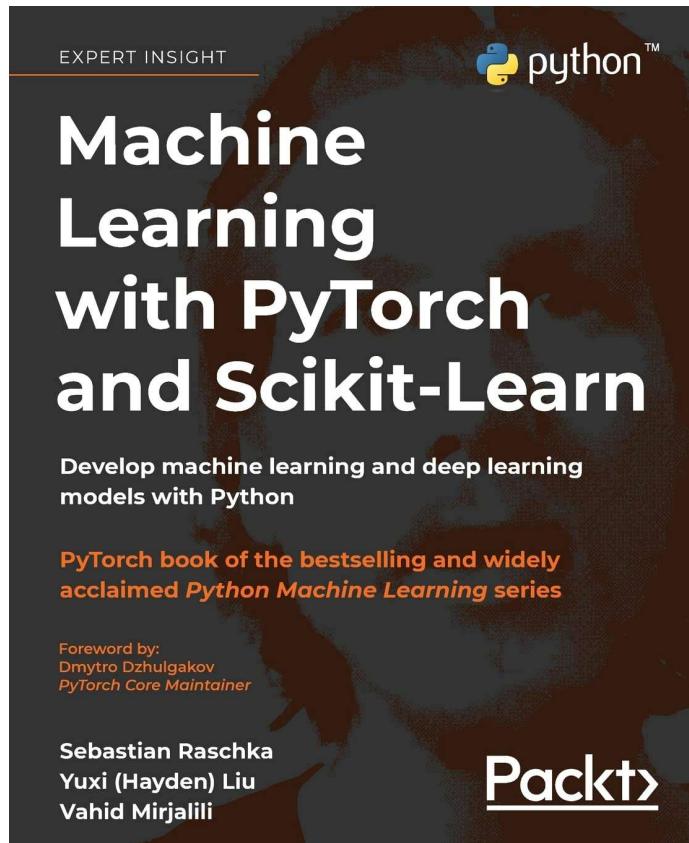
| February | | | | | | | |
|---|---------|-----------|--|--------|----------|--------|--|
| MONDAY | TUESDAY | WEDNESDAY | THURSDAY | FRIDAY | SATURDAY | SUNDAY | |
| 02 | 03 | 04 | 05 | 06 | 07 | 08 | |
| Week 2 | | | Class 7PM | | | | |
| 09 | 10 | 11 | 12 | 13 | 14 | 15 | |
| Week 3 | | Class 7PM | Lincoln's Birthday – CUNY SPS is closed, no classes scheduled* | | | Quiz 1 | |
| 16 | 17 | 18 | 19 | 20 | 21 | 22 | |
| Presidents' Day - no classes scheduled* | Week 4 | Class 7PM | | | | Quiz 2 | |
| 23 | 24 | 25 | 26 | 27 | 28 | 01 | |
| Week 5 | | Class 7PM | | | | Quiz 3 | |
| 02 | 03 | Notes: | | | | | |

| March | | | | | | | |
|--------|---------|-----------|-----------|--------|----------|--------|---------------------|
| MONDAY | TUESDAY | WEDNESDAY | THURSDAY | FRIDAY | SATURDAY | SUNDAY | |
| 02 | 03 | 04 | 05 | 06 | 07 | 08 | |
| Week 6 | | | Class 7PM | | | | Assignment 1 |
| 09 | 10 | 11 | 12 | 13 | 14 | 15 | |
| Week 7 | | | Class 7PM | | | | Online Midterm Exam |
| 16 | 17 | 18 | 19 | 20 | 21 | 22 | |
| Week 8 | | | Class 7PM | | | | Mid-semester survey |
| 23 | 24 | 25 | 26 | 27 | 28 | 29 | |
| Week 9 | | | Class 7PM | | | | Assignment 2 |
| 30 | 31 | Notes: | | | | | |

| April | | | | | | | |
|---------|---------|--------------|----------|--------|--------------|--------|--|
| MONDAY | TUESDAY | WEDNESDAY | THURSDAY | FRIDAY | SATURDAY | SUNDAY | |
| 30 | 31 | 01 | 02 | 03 | 04 | 05 | |
| | | Spring Break | | | | | |
| 06 | 07 | 08 | 09 | 10 | 11 | 12 | |
| | | Spring Break | | | | | |
| 13 | 14 | 15 | 16 | 17 | 18 | 19 | |
| Week 10 | | Class 7PM | | | Quiz 4 | | |
| 20 | 21 | 22 | 23 | 24 | 25 | 26 | |
| Week 11 | | Class 7PM | | | Assignment 3 | | |
| 27 | 28 | 29 | 30 | | | | |
| Week 12 | | Class 7PM | | | | | |

| May | | | | | | | |
|-----------------------------------|------------------|--------------------|----------|-------------------|----------|--------|-------------------|
| MONDAY | TUESDAY | WEDNESDAY | THURSDAY | FRIDAY | SATURDAY | SUNDAY | |
| | | | | 01 | 02 | 03 | |
| | | | | | | Quiz 5 | |
| 04 | 05 | 06 | 07 | 08 | 09 | 10 | |
| Week 13 | | Class 7PM | | | | | Assignment 4 |
| 11 | 12 | 13 | 14 | 15 | 16 | 17 | |
| Week 14 | | Class 7PM | | Last day of Class | | | Online Final Exam |
| 18 | 19 | 20 | 21 | 22 | 23 | 24 | |
| | | Final Examinations | | | | | |
| 25 | 26 | 27 | 28 | 29 | 30 | 31 | |
| Memorial Day - CUNY SPS is closed | Spring Term Ends | | | | | | |

Textbook



Machine Learning with PyTorch and Scikit-Learn

by Sebastian Raschka, Yuxi (Hayden) Liu, Vahid Mirjalili

February 2022

ISBN-10: 1801819319

ISBN-13: 978-1801819312

This is not a programming course

To help you understand the technical concepts and what data scientists do we will do vbe coding.

If you prefer R let me know and I will provide alternatives

Class Policies

1. Please leave your camera on

- See: https://www.cuny.edu/wp-content/uploads/sites/4/page-assets/academics/faculty-affairs/Camera-Use-Guidance-for-Online-and-Hybrid-Courses_FINAL-JUNE-20-2024.pdf

2. Review the Course Policies on Brightspace

- Respect, Online Etiquette and Anti-Harassment
- Lateness Policy

3. Read the guidelines on Generative AI & Academic Integrity on Brightspace

- Provide Snapshots with answers
- Do not use Generative AI for quizzes and exams
- You (the human) manage AI, not the other way around
- Don't cheat yourself

Grading

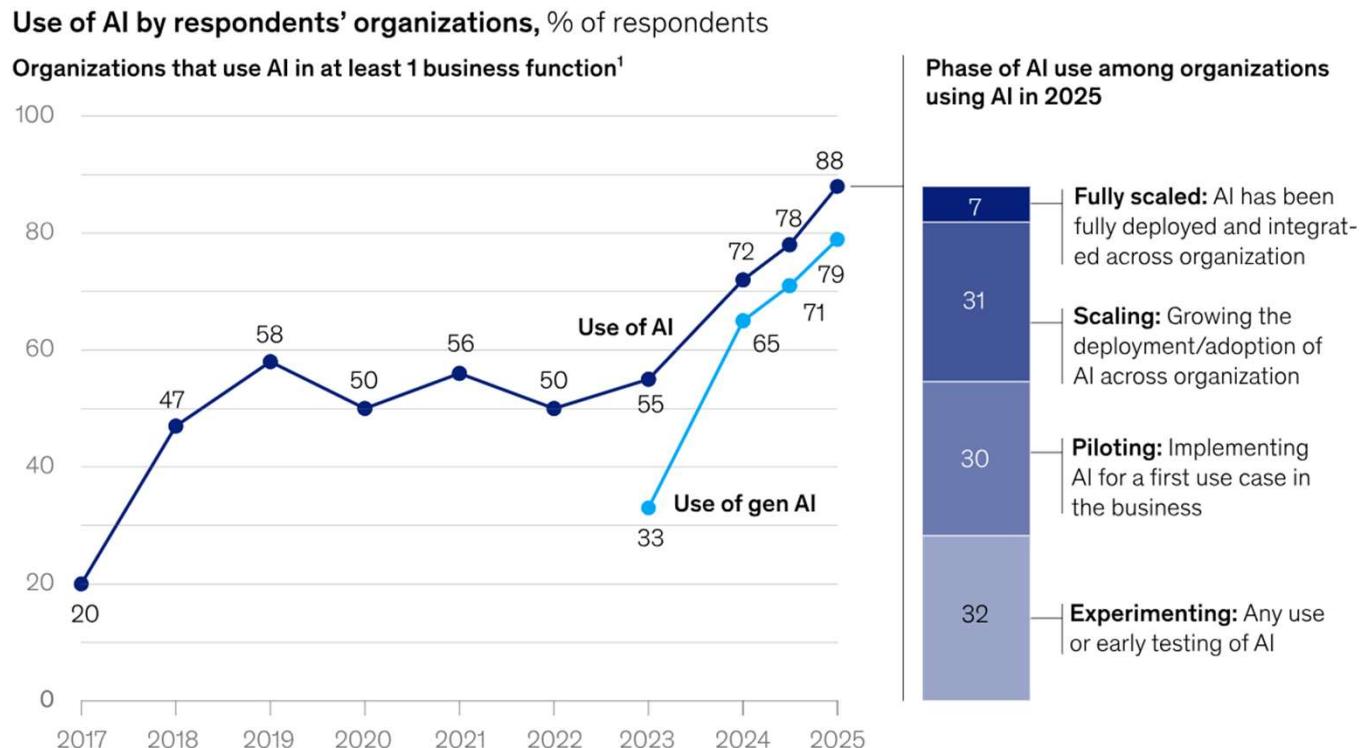
| Component | Number | Total Value |
|-----------------------|--------|-------------|
| Quizzes | 5 | 20% |
| Assignments | 4 | 20% |
| Exam: Mid-term | 1 | 30% |
| Exam: End of semester | 1 | 30% |
| | | 100% |



Introduction

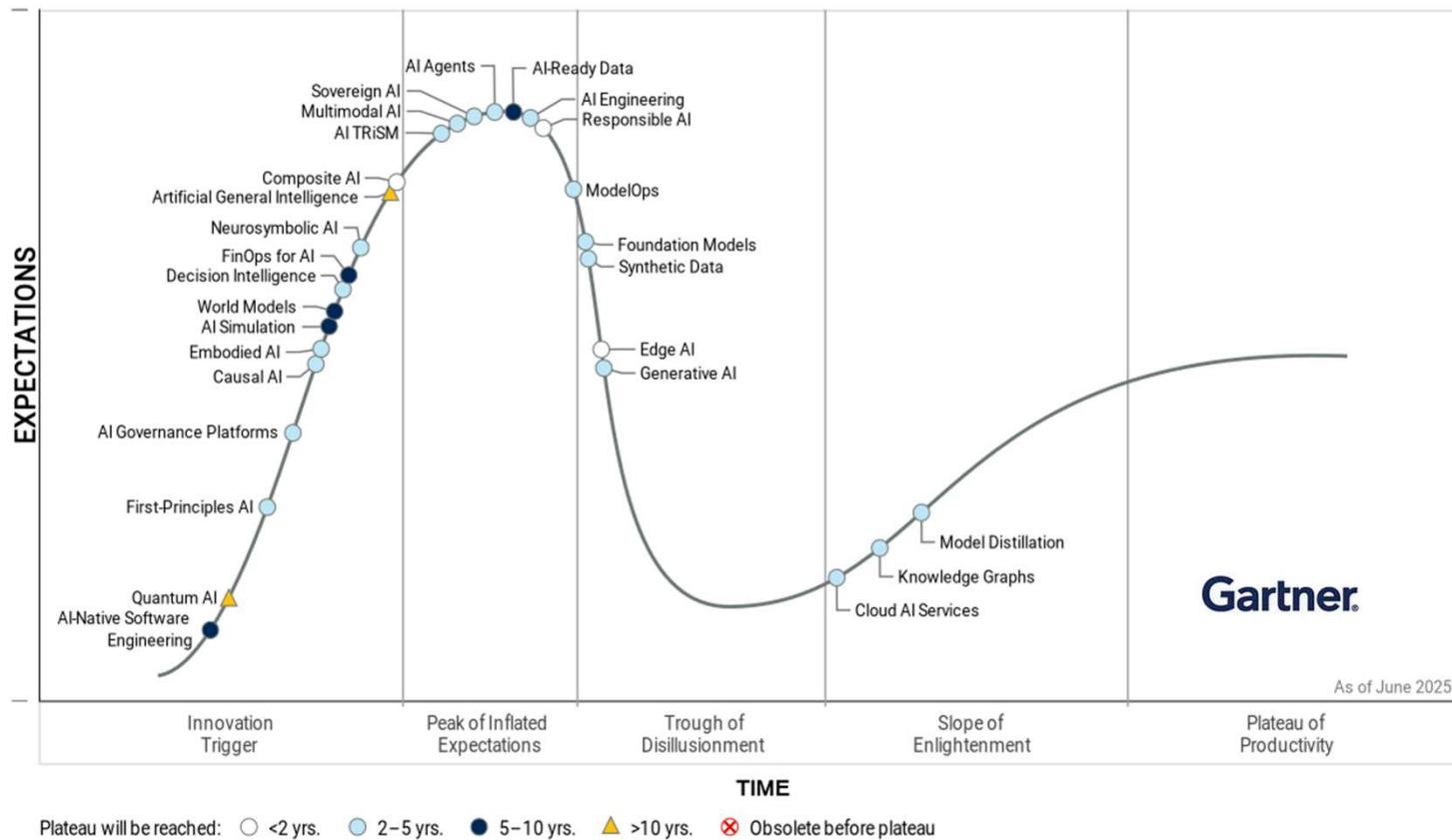
AI has become a critical need for firms

Reported use of AI in at least one business function continues to increase.



Source: McKinsey

There is a lot of excitement



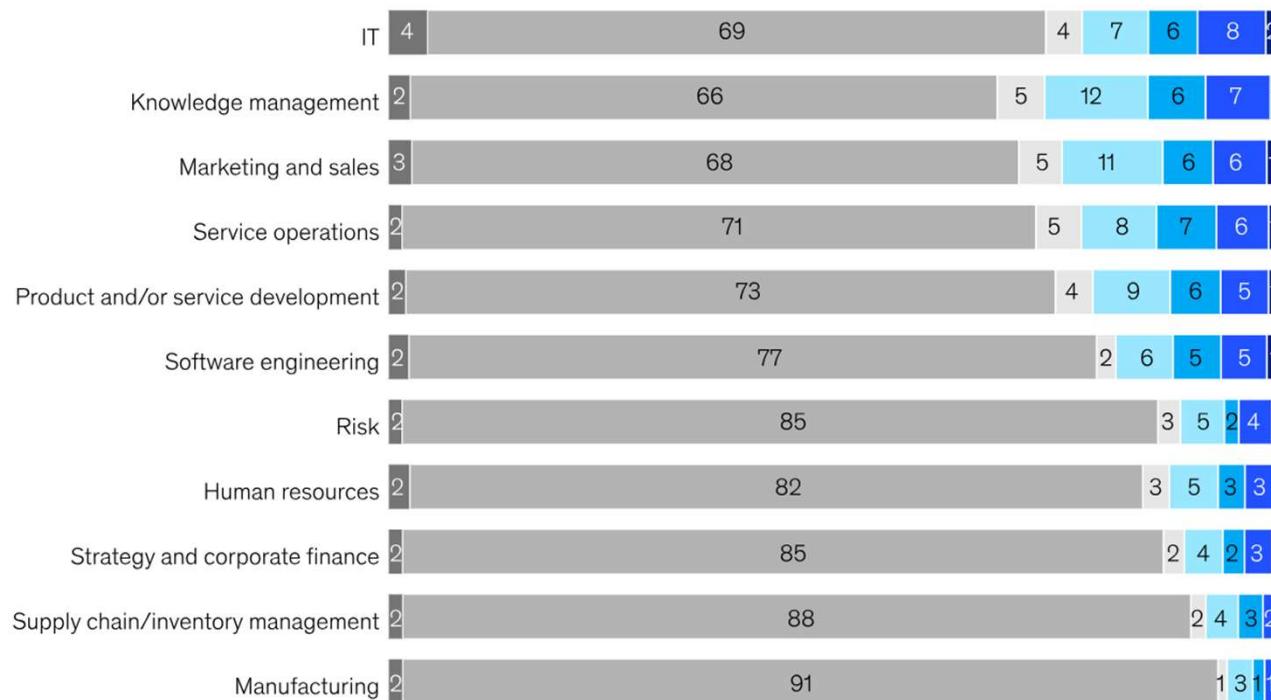
Source: Gartner

Real adoption of Generative AI is slow

No more than 10 percent of respondents report scaling AI agents in any individual function.

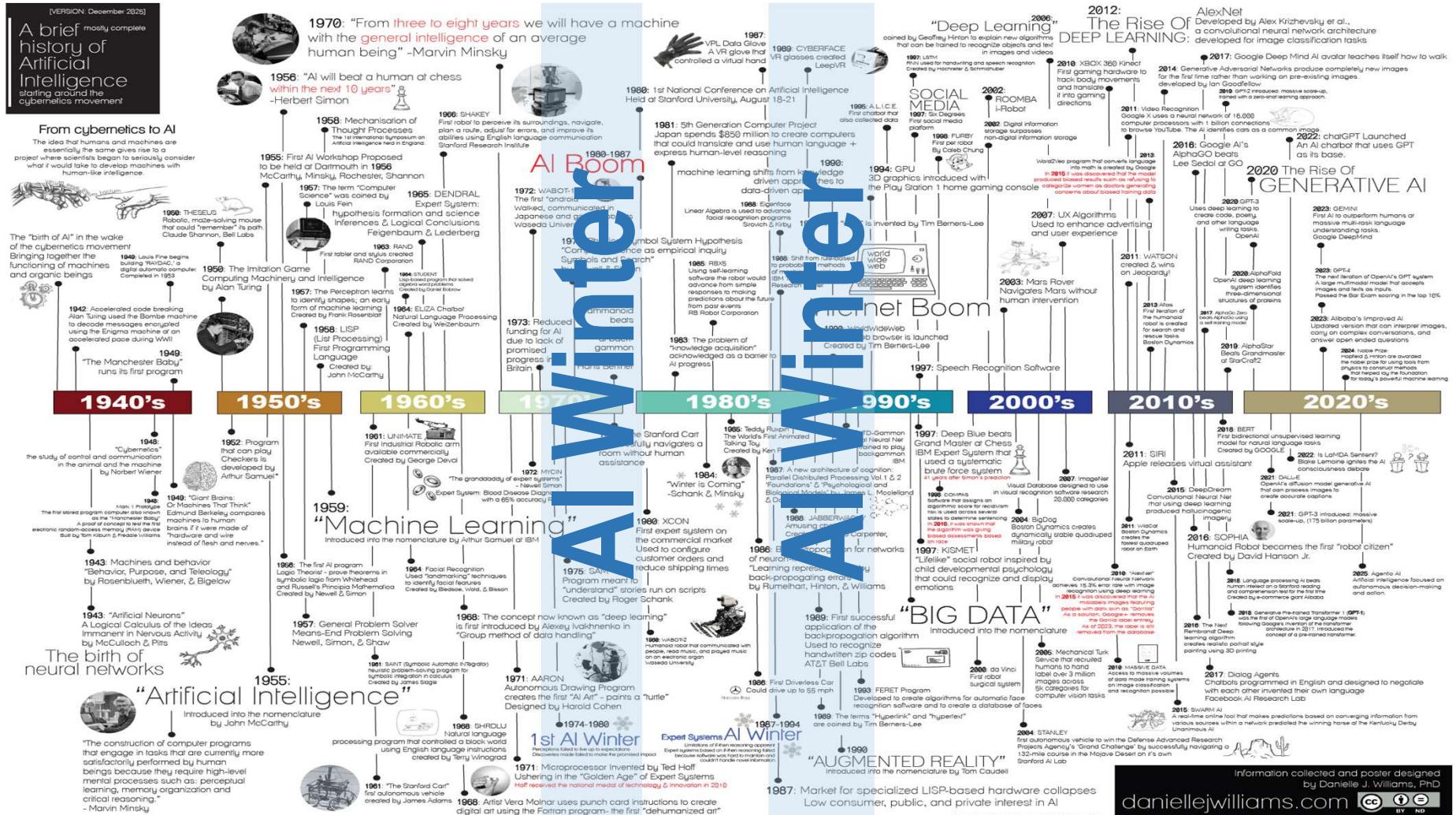
Phase of AI agent use at respondents' organizations, by business function,¹ % of respondents (n = 1,933)

■ Don't know ■ Not at all ■ Planning to use within year ■ Experimenting ■ Piloting ■ Scaling ■ Fully scaled

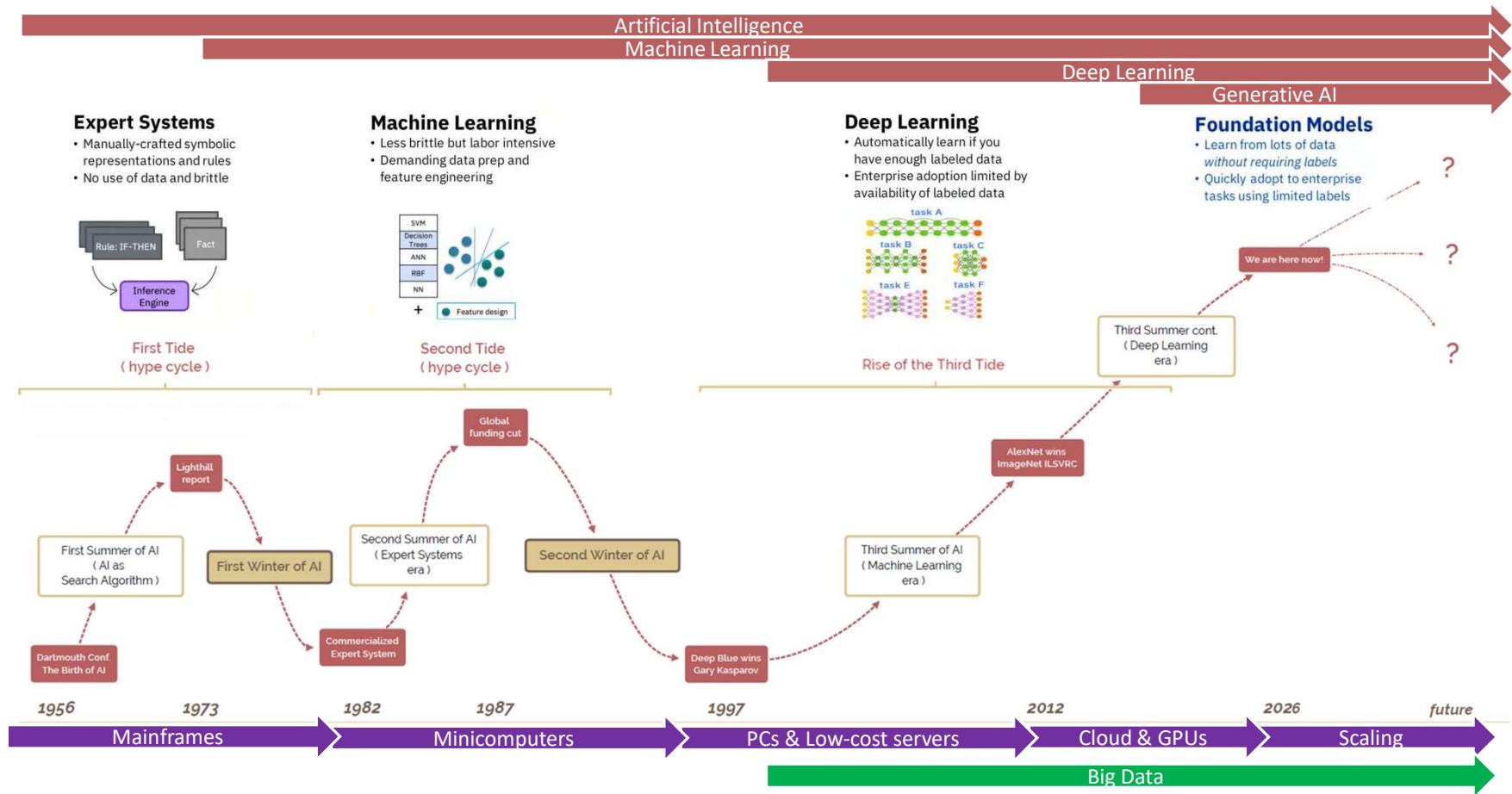


Source: McKinsey

Ai is not a recent phenomenon



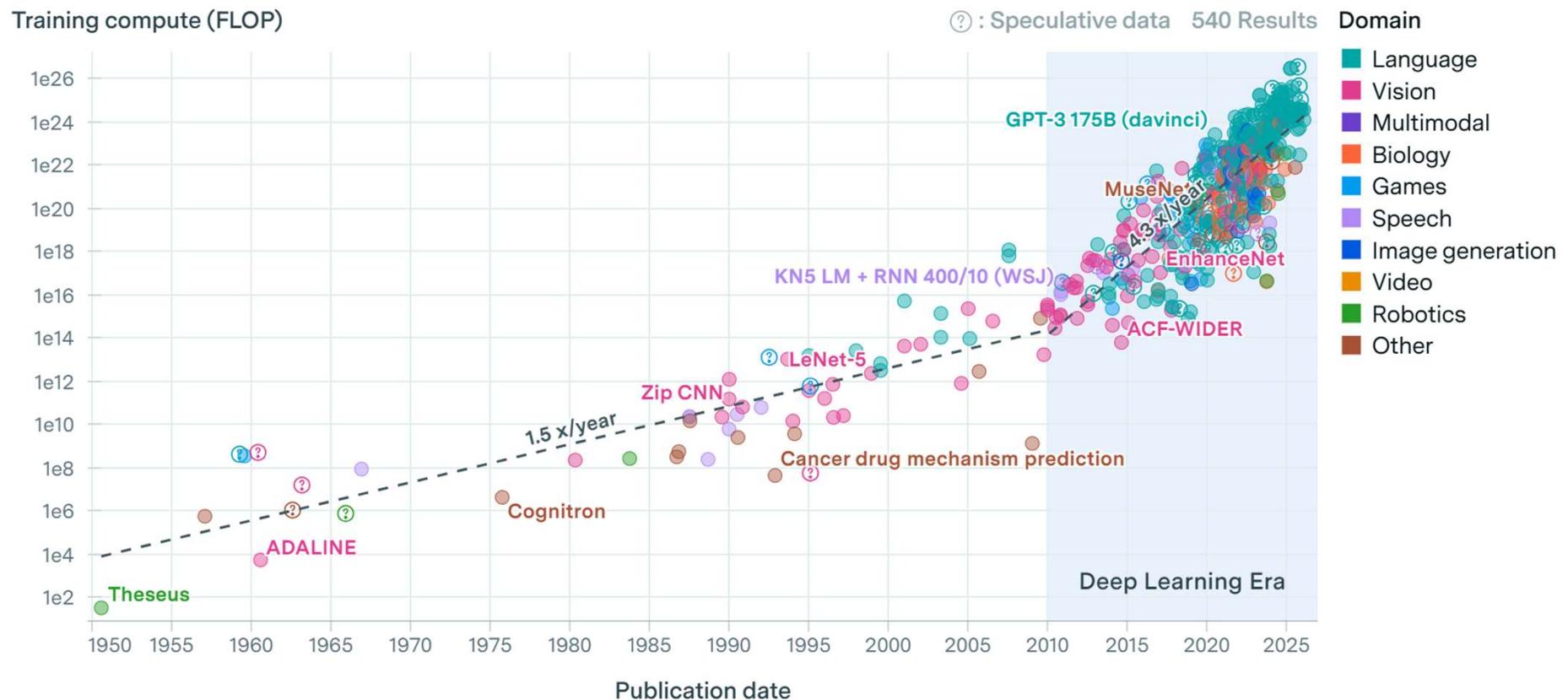
AI Winters



Growth is accelerating – due to GPUs

Notable AI models

EPOCH AI

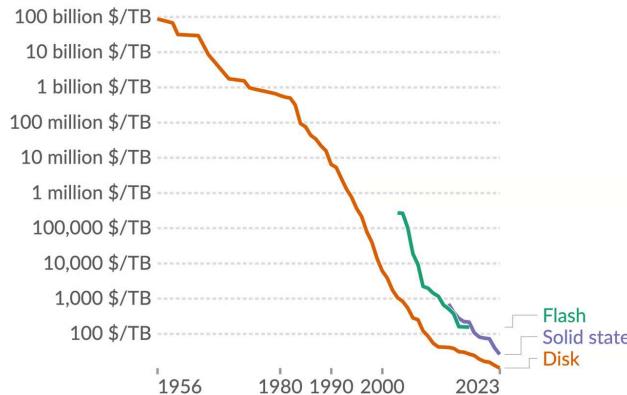


AI is becoming cheaper

Historical price of computer storage

Our World in Data

Expressed in US dollars per terabyte (TB), adjusted for inflation.
"Disk" refers to magnetic storage, "flash" to memory used for rapid data access and rewriting, and "solid state" to solid-state drives (SSDs).



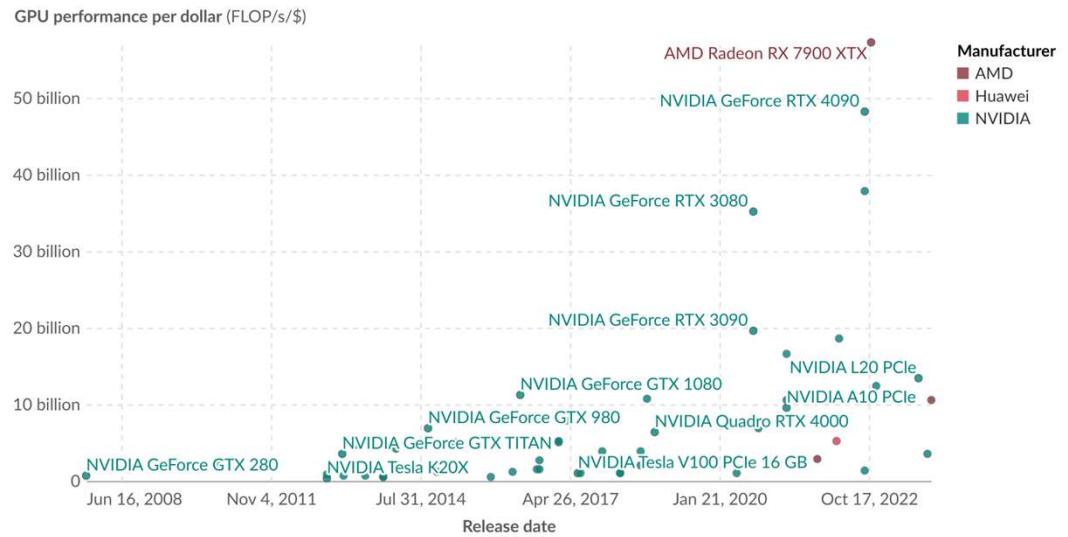
Data source: John C. McCallum (2023)

CC BY

GPU computational performance per dollar

Our World in Data

Hardware computational performance shown in floating-point operations¹ per second (FLOP/s) per US dollar, adjusted for inflation.



Source: Our world in data

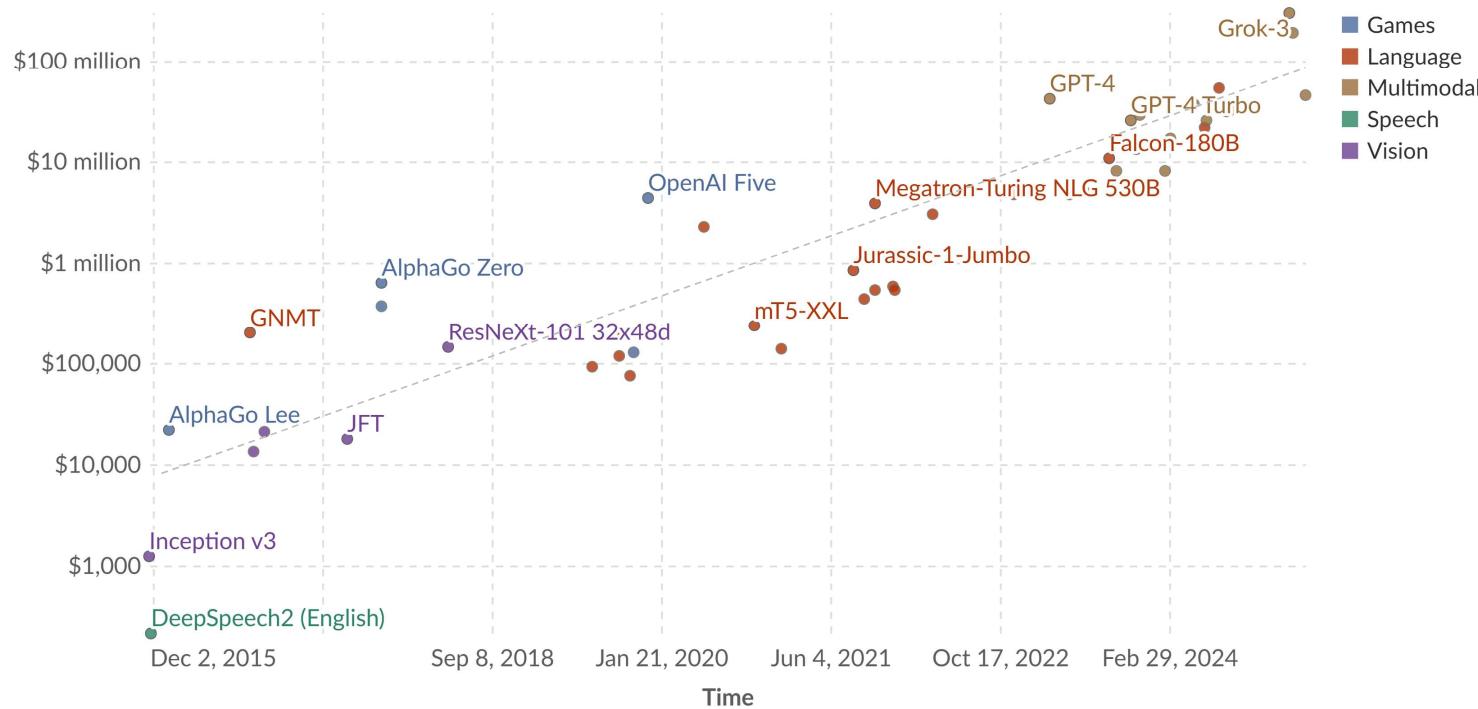
Generative AI models are getting bigger

Hardware and energy cost to train notable AI systems

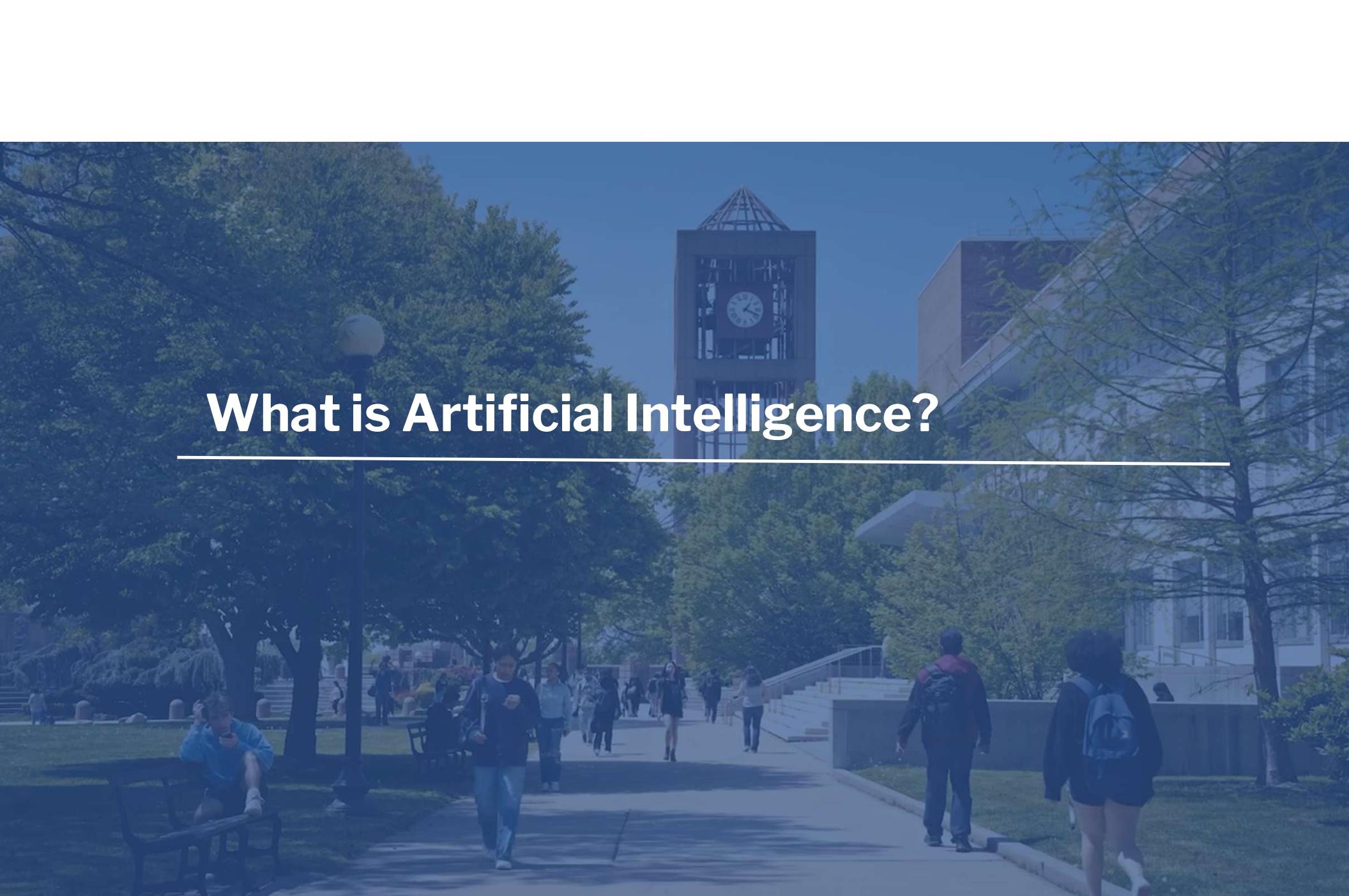
Our World
in Data

This data is expressed in US dollars, adjusted for inflation.

Cost (constant 2023 US\$; plotted on a logarithmic axis)

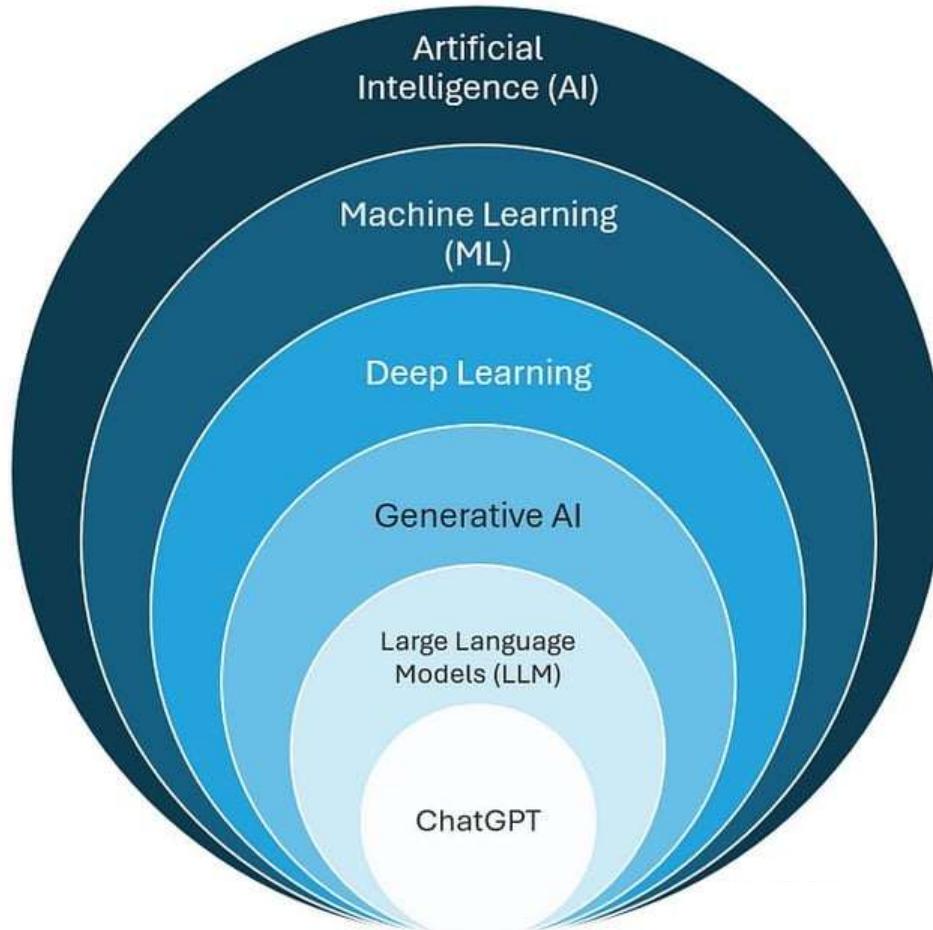


Source: Our world in data



What is Artificial Intelligence?

What is Artificial Intelligence?



What is intelligence?

"A fundamental problem in artificial intelligence is that nobody really knows what intelligence is."

-Legg & Hutter (2007)

"... the ability to solve hard problems." M. Minsky

"The ability to use memory, knowledge, experience, understanding, reasoning, imagination and judgement in order to solve problems and adapt to new situations." AllWords Dictionary, 2006

"The capacity to acquire and apply knowledge." The American Heritage Dictionary, fourth edition, 2000

"Intelligence is not a single, unitary ability, but rather a composite of several functions. The term denotes that combination of abilities required for survival and advancement within a particular culture." A. Anastasi

"Intelligence is the ability for an information processing system to adapt to its environment with insufficient knowledge and resources." P. Wang

"... the ability to plan and structure one's behavior with an end in view." J. P. Das

"... doing well at a broad range of tasks is an empirical definition of 'intelligence'" H. Masum

"... the ability of a system to act appropriately in an uncertain environment, where appropriate action is that which increases the probability of success, and success is the achievement of behavioral subgoals that support the system's ultimate goal." J. S. Albu

"Intelligence measures an agent's ability to achieve goals in a wide range of environments."

S. Legg and M. Hutter

Source: A Collection of Definitions of Intelligence, Shane Legg, Marcus Hutter, 2007 arXiv:0706.3639

A blue-tinted photograph of a university campus. In the background, a tall clock tower with a glass facade and a pyramid-shaped roof stands prominently. To the right, a modern building with large windows and a staircase is visible. In the foreground, several students are walking on a paved path. One student in a blue hoodie is sitting on a bench on the left. The scene is set against a clear blue sky.

What is Machine Learning?

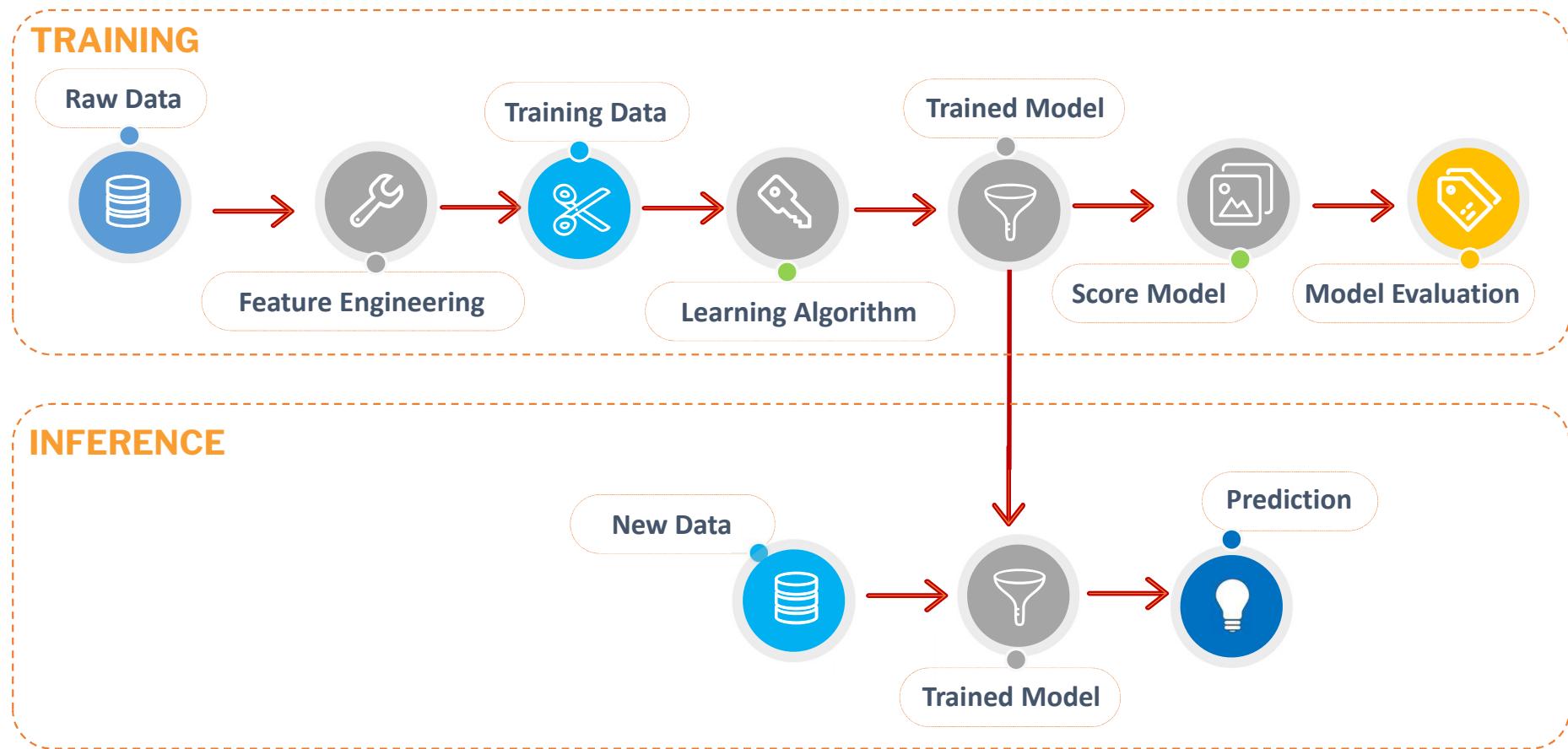
What is Data Science?

“At a high level, data science is a set of fundamental principles that guide the extraction of knowledge from data.”

Principles can be statistical, computational, algorithmic, visual, etc.

Source: NVIDIA DLI Accelerated Data Science

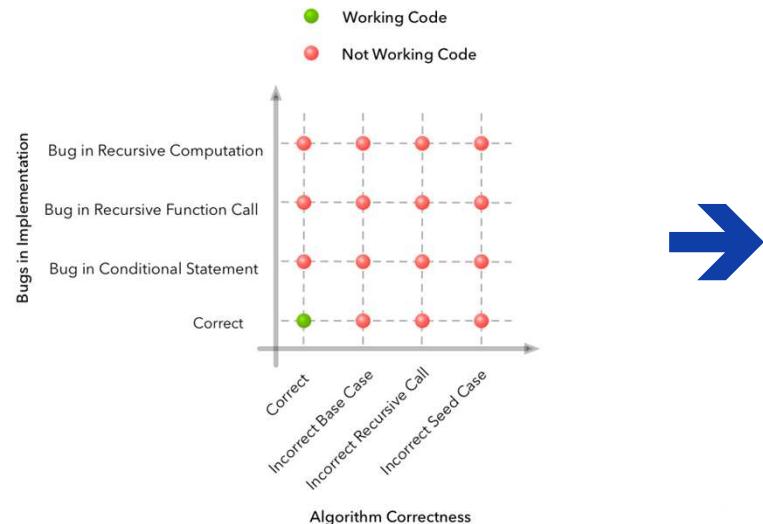
How does Machine Learning work?



Why is Data Science hard?

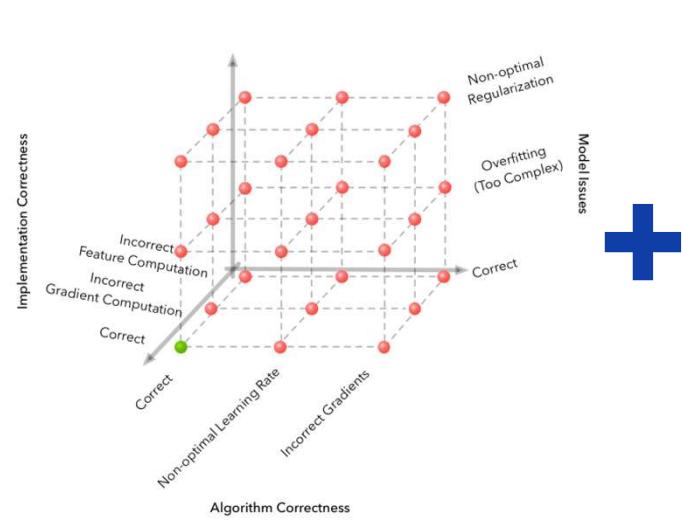
Software Development is a Two-dimensional Problem

Algorithm selection and Software implementation



Machine Learning is a Four-dimensional Problem

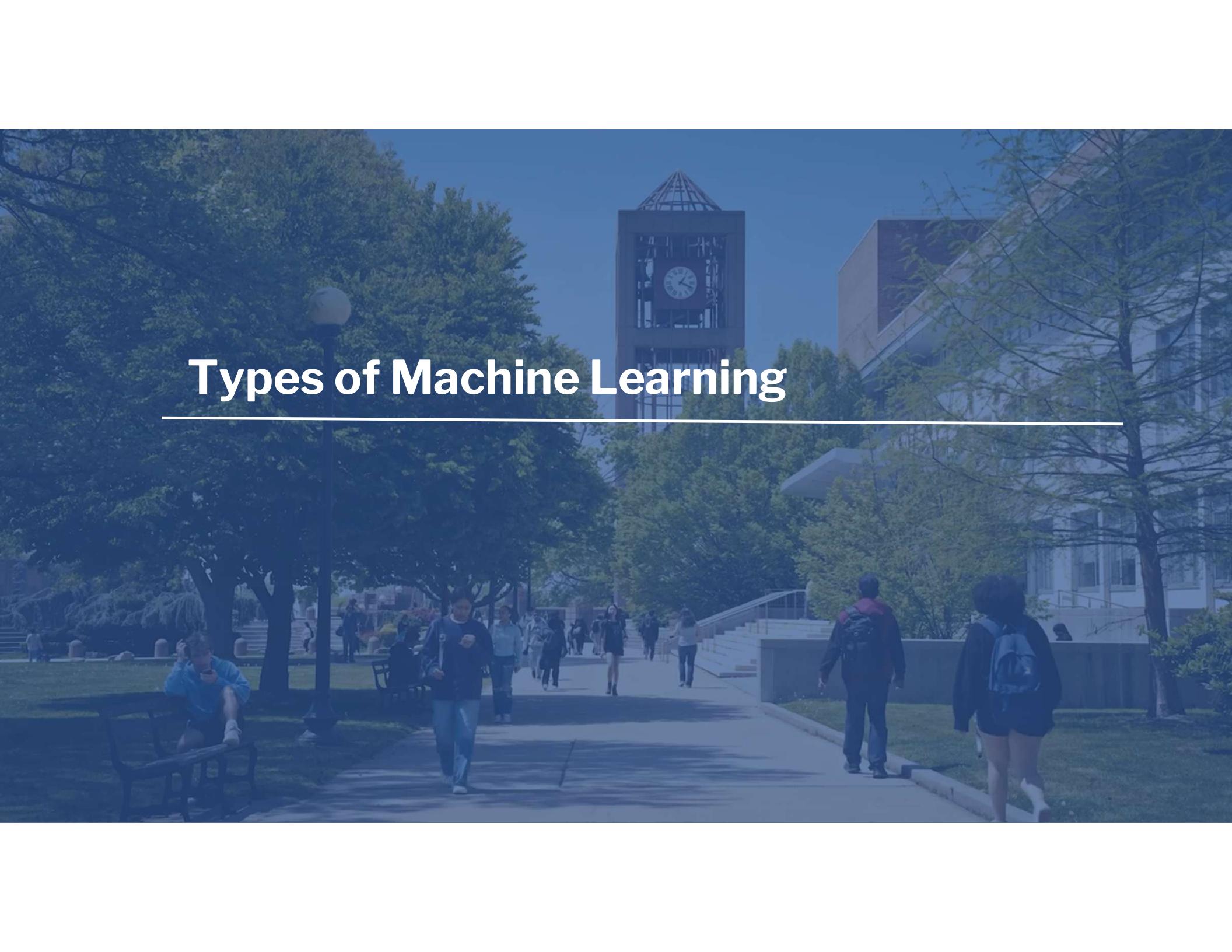
Software development + Model Implementation +Data



Source: <https://ai.stanford.edu/~zayd/why-is-machine-learning-hard.html>

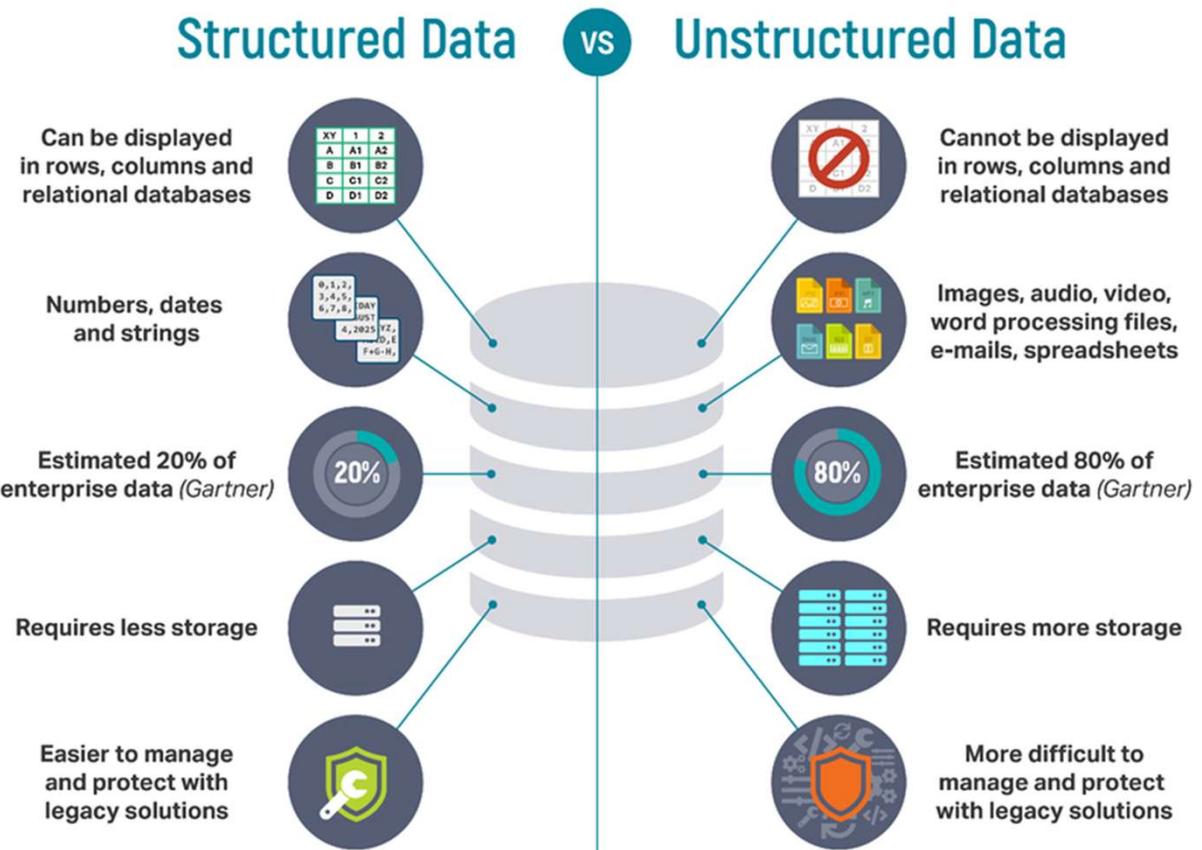
Why is Data Science complex?

- 1. Data science is interdisciplinary**
- 2. Applicable to many domains (e.g., sciences, finance, healthcare, etc.)**
- 3. Big Data**
- 4. Analysis of data**
- 5. Statistics (a pillar of data science)**
- 6. Experimentation**
- 7. Constant Evolution**
- 8. Data Wrangling**

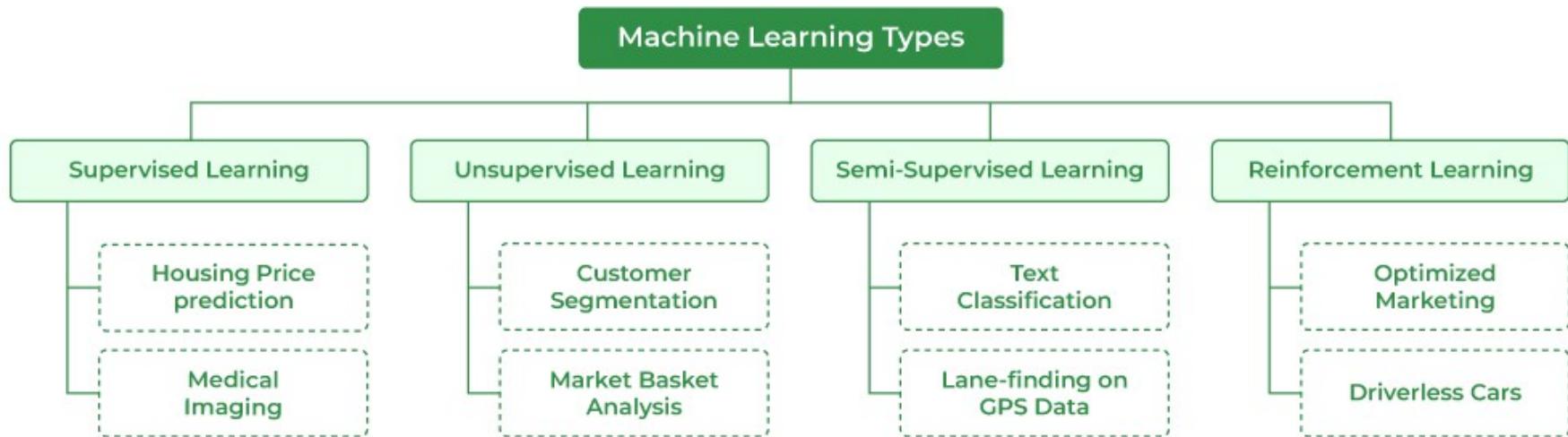
A blue-tinted photograph of a university campus. In the background, a tall clock tower with a glass facade and a pyramid-shaped roof stands prominently. To the right, a modern building with large windows and a staircase is visible. In the foreground, several students are walking along a paved path. One student is sitting on a bench on the left side.

Types of Machine Learning

Types of data



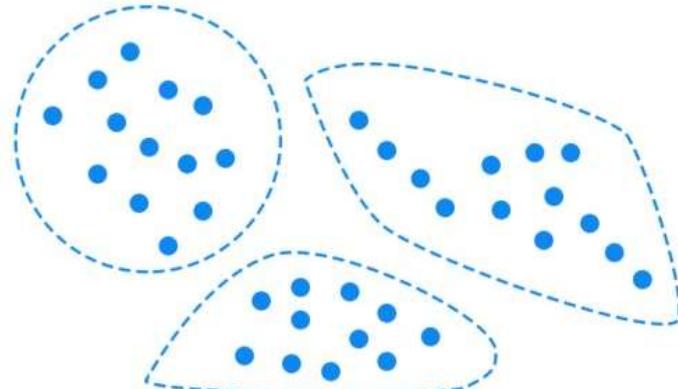
Types of Machine Learning



Supervised vs Unsupervised Learning

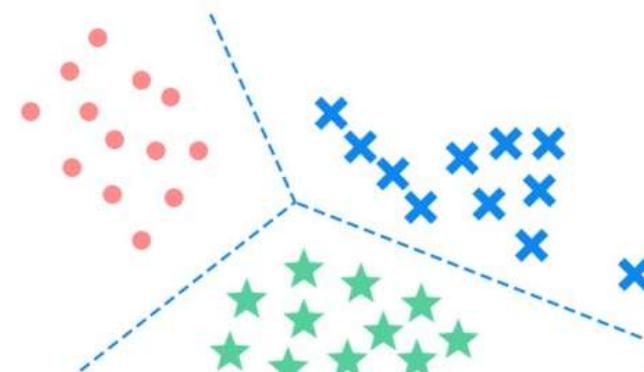
UNSUPERVISED

Clustering



SUPERVISED

Classification



Business use-cases:

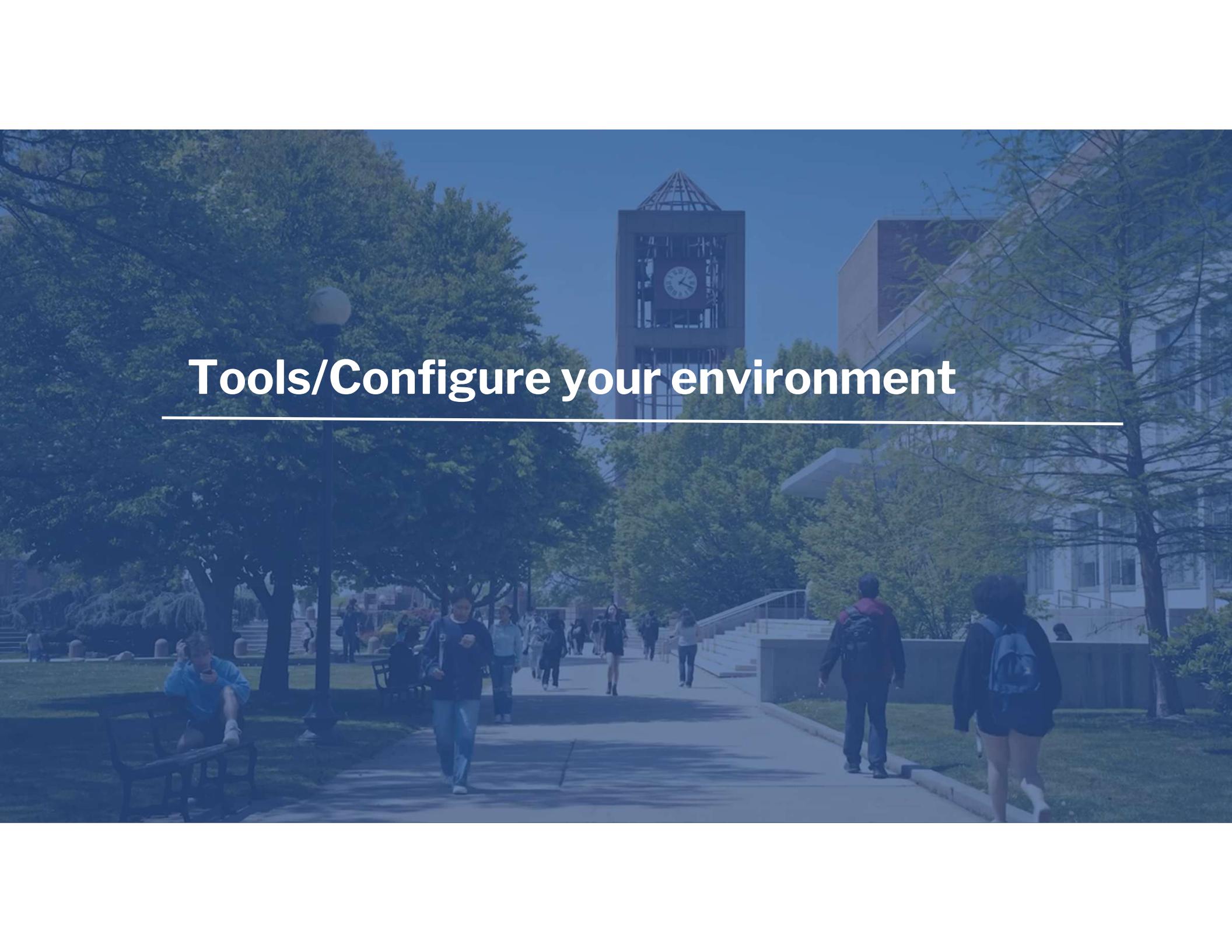
- Customer segmentation
- Cross- selling
- Anomaly detection e.g. Expenses
- Document clustering
- Recommendation & personalization
- Data Reporting (dimensionality Reduction)

Business use-cases:

- Loan approval, credit risk assessment
- Credit Card Fraud Detection
- Sentiment analysis, text classification
- Speech recognition, Image classification
- Spam Detection
- Stock Price Prediction

Types of Algorithms



A blue-tinted photograph of a university campus. In the background, a tall clock tower with a glass facade and a pyramid-shaped roof stands prominently. To the right, a modern building with large windows and a glass-enclosed staircase is visible. In the foreground, several students are walking along a paved path. One student in a blue hoodie is sitting on a bench on the left. The scene is set against a backdrop of green trees and a clear sky.

Tools/Configure your environment

Register for free tools

1. Colab

- Go to <https://colab.research.google.com/signup>
- Create a Google account (if you don't have one)
- Sign up for **Colab Pro for Education** (free for students)

2. GitHub Copilot

- Go to <https://github.com/education>
- Click on **Join GitHub Education**
- Create a GitHub account (if you don't have one)
- Click on Start an Application to get access

3. Google Gemini

- Go to <https://gemini.google/us/students>
- Fill out the verification form

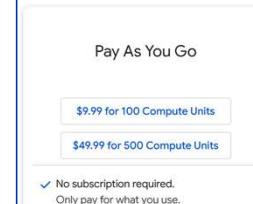
4. Microsoft Copilot 365

- Free with your Office 365 access (via CUNY)

Choose the Colab plan that's right for you

Whether you're a student, a hobbyist, or a ML researcher, Colab has you covered. Colab is always free of charge to use, but as your computing needs grow there are paid options to meet them.

[Restrictions apply, learn more here](#)



University students get Gemini in Google AI Pro for **1 year for free**

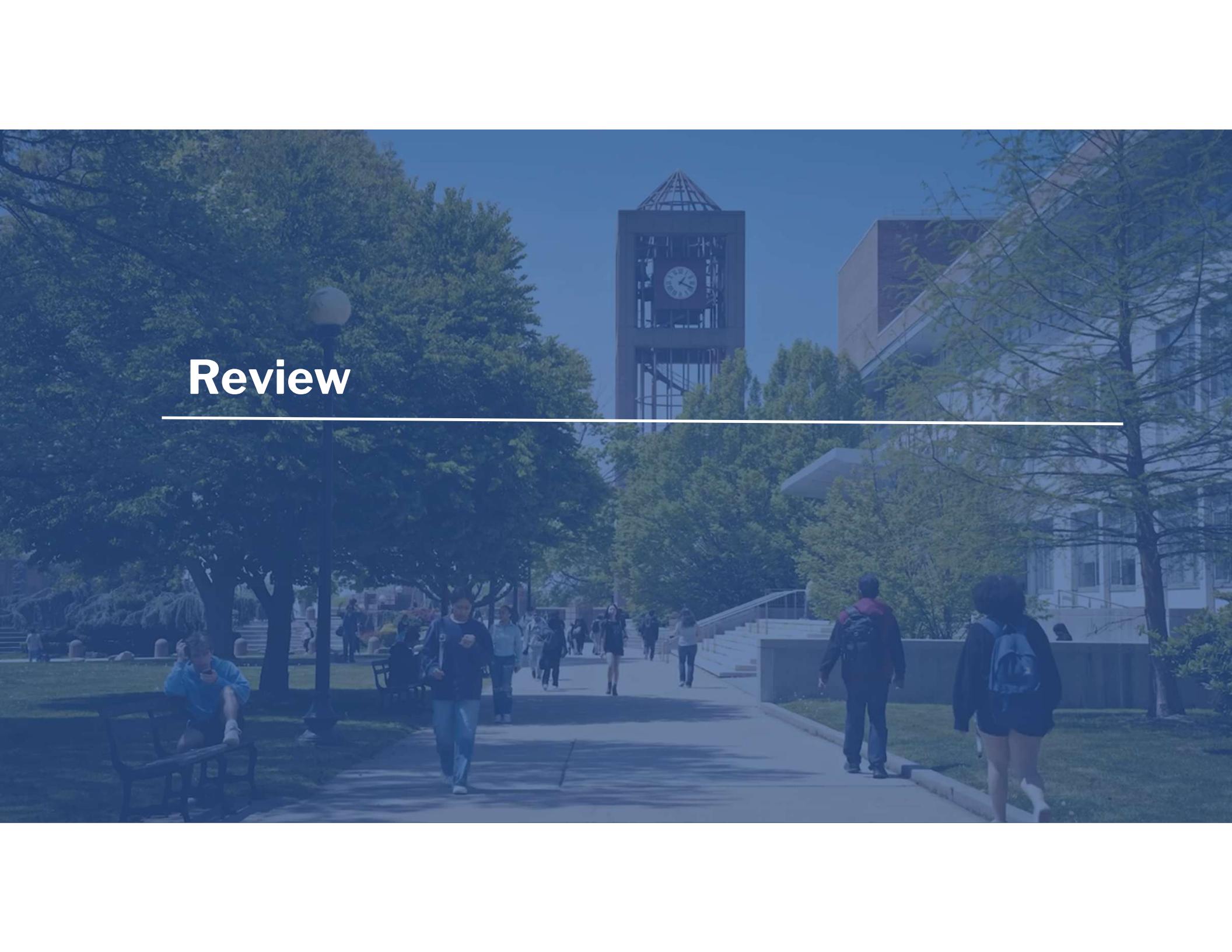
\$19.99/mo \$0/mo for 12 months

[Verify eligibility](#)

Cancel anytime. By subscribing, you agree to [Google One](#), [AI credits](#) and [offer](#) terms. See how [Google handles data](#).

Featured Gemini benefits

- Homework help & exam prep
Analyze entire textbooks up to 1,500 pages

A scenic view of a university campus under a clear blue sky. In the center, a tall, dark rectangular tower with a glass-enclosed clock face stands prominently. To its right is a large, light-colored building with a glass facade and a staircase leading up to it. In the foreground, a paved walkway leads towards the buildings, lined with lush green trees and bushes. Several students are walking along the path; one is sitting on a bench on the left, while others are strolling or carrying backpacks. The overall atmosphere is bright and sunny.

Review

This week we covered

1. Supervised & Unsupervised Learning

Define the basic types of machine learning (supervised, unsupervised) using examples from real-world applications.

2. Business Role

Explain the role of machine learning in modern business contexts and decision-making.

3. Tooling

Set up and configure Python (or R) environments for machine learning tasks.

4. ML Algorithms

Identify appropriate ML techniques for structured and unstructured business data.

The background image shows a panoramic view of the New York City skyline at sunset or sunrise. The Brooklyn Bridge is prominent in the lower-left foreground, stretching across the East River. The Manhattan skyline, with its numerous skyscrapers, rises in the background under a clear blue sky.

CU | School of
NY | Professional Studies