AI FOR DATA ANALYSIS: PRIVACY AND CODING IN DIGITAL FINANCE

Dr. Luca Di Grazia

MSCA Digital Finance

Berner Fachhochschule

October 10, 2025

MEET THE TEACHER

Actual Position:

- Luca DI Grazia, PhD
- Great Minds Fellow
- School of Computer Science (SCS)
- University of St. Gallen
- www.lucadigrazia.com
- work@lucadigrazia.com

Previously:

- CH University of Lugano (USI), Postdoc in Software Engineering
- DE Stuttgart Universität, PhD in Computer Science
- NL Uber, Research Internship in Generative AI for Code
- IT Politecnico di Torino, Bachelor and Master in Computer Engineering

EXPLAINABLE AI - TRAINING WEEK

- The course provides a comprehensive introduction to Explainable Artificial Intelligence (XAI), emphasizing the methodologies and practical applications of cutting-edge models such as LIME, SHAP, deep learning XAI, time series-based XAI methods and others.
- https://www.digital-finance-msca.com/event-details-registration/explainable-ai-training-week

TODAY - AI FOR DATA ANALYSIS: PRIVACY AND CODING IN DIGITAL FINANCE

Time	Instruction mode	Topic	Description
9:00-10:00	Lecture	Introduction, Basics concepts and Setup	Overview of AI privacy concerns, coding tools and local LLMs.
10:00-11:30	Project & Coaching	Build a simple prototype	Build a basic app interface connected to a private on-device AI assistant.
11:30-12:00	Tutorial	Working with Financial Data	Data analysis with AI for company revenue data; discuss privacy, IP, and compliance.
13:00-16:30	Project & Coaching	Applying the protype built to a case study	Students work in groups to extend their prototype to address a specific problem in the digital finance domain.
16:30-17:45	Presentation session	Group Presentations and Feedback formulation	Each group presents their findings (10 minutes), with constructive feedback from peers and the instructor.
17:45-18:00	Wrap-up	Discussion: Lessons Learned and Pathways Forward	Wrap-up discussion, and implications for digital finance.

TOPIC AND COURSE INTRODUCTION

INTRODUCTION

Time	Instruction mode	Topic	Description
9:00-10:00	Lecture	Introduction, Basics concepts and Setup	Overview of AI privacy concerns, coding tools and local LLMs.

COURSE PHILOSOPHY

- LLMs are changing the way we develop software

- It is too early to create a structured course on this topic, but:
  - Many organizations just started experimenting with AI for Coding
  - AI tools are also evolving fast ("*prompt engineering is dead*")

- But it is not too early to start experimenting with AI tools for coding to understand how they help for data analysis, and what their limits are.

- This course is a combination of frontal lectures, tutorials, and project work to experiment
  ith AI-supported coding for data analysis.

PREREQUISITES

- The prerequisites are basic programming skill in any ‘high level’ programming language (e.g., Python)

- Example:
  - You have written and run Python programs
  - You are familiar with the process of turning a specification/requirement in a list of implementation tasks
  - You know how to organize a software project in modules (e.g., functions) that interact with each other

LEARNING OBJECTIVES

- Explain the role of artificial intelligence in reshaping software development and data analysis
- Identify and critically assess privacy, ethical, and regulatory challenges in AI for digital finance.
- Apply basic Python programming skills and use an IDE (e.g., Visual Studio Code) to perform a financial data analysis.
- Use a local large language model (LLM) to analyze data securely and reflect on its advantages and limitations.
- Collaborate in groups to design simple prototypes to analyze data and critically evaluate their societal and regulatory impacts.

EXPERIMENTING WITH AI

- What is AI in "Coding with AI"?

- You are free to experiment.
- In this lecture we often refer to conversational AI as in GPT
- We recommend to experiment with different conversational models (e.g., Copilot)
- Also, experiment with different interaction modes (e.g., AI integrated into an IDE)

EXAMINATION FORMAT

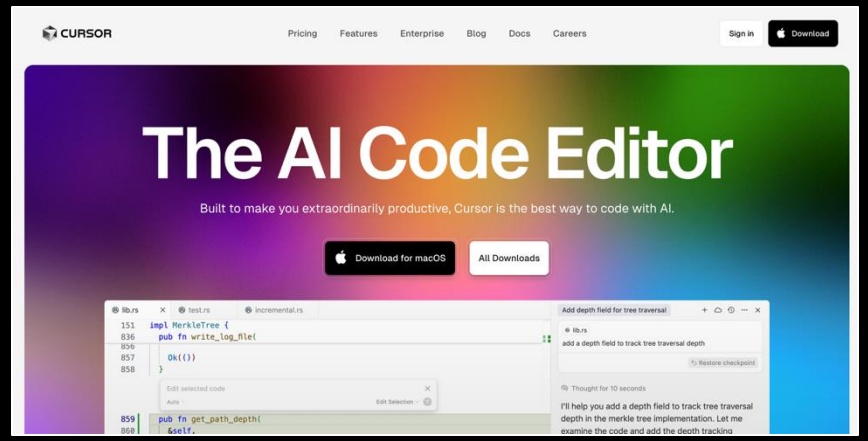
- **Project (presentation)**
- Group work (2-3 people), individual evaluation
- Pass/fail setting

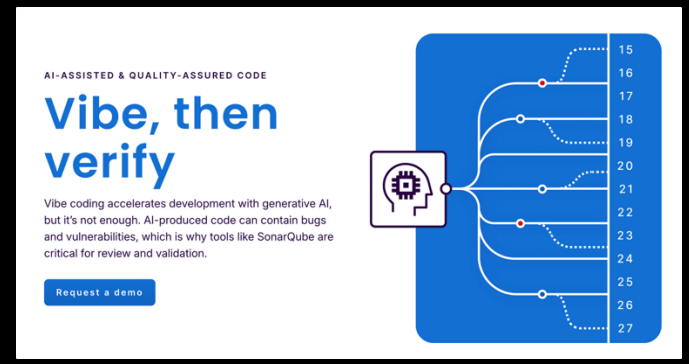
Time	Instruction mode	Topic	Description
16:30-17:45	Presentation session	Group Presentations and Feedback formulation	Each group presents their findings (10 minutes), with constructive feedback from peers and the instructor.

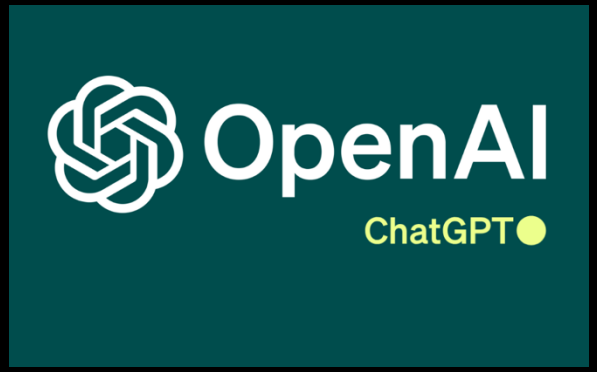
AI IN GENERAL: BASIC CONCEPTS AND EXAMPLES

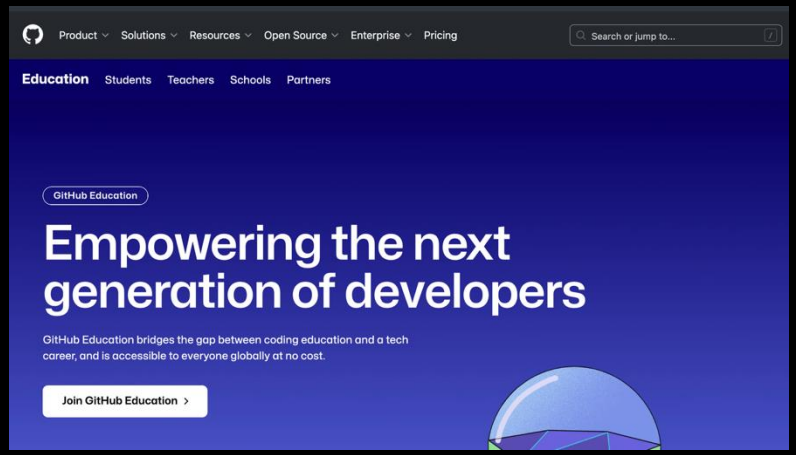
WHAT ARE AI TOOLS?

- You can **choose** the tools you like
- You are highly encouraged to use tools with **different interaction modes** (e.g., prompting, code completion)
- Please **share** interesting tools with your peers!









A VERY SIMPLE AI

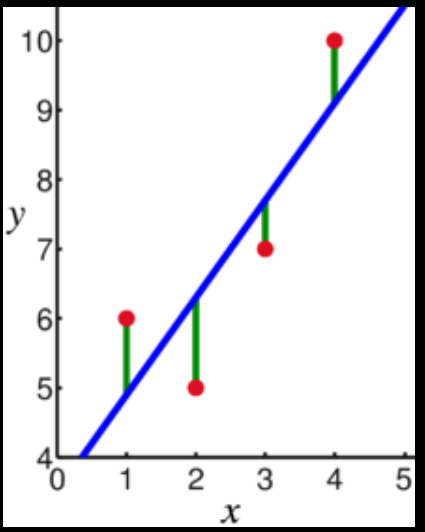
- Linear regression: A **linear** approach to modelling the relationship between a scalar response and one or more explanatory variables

  $y = ax + b$

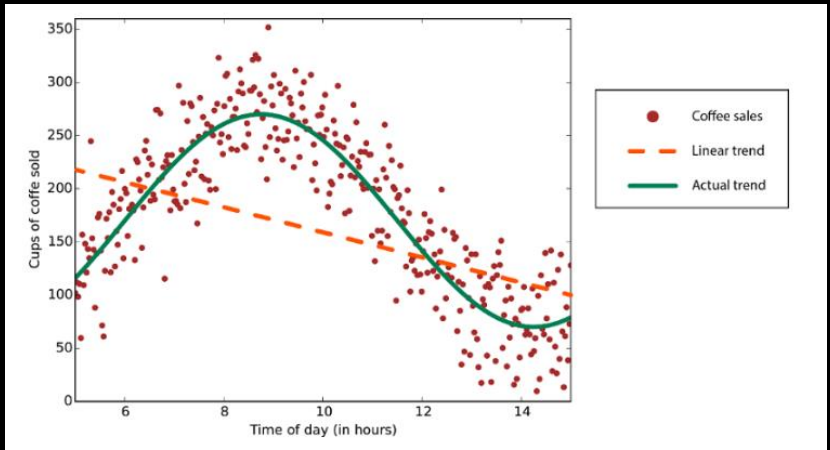
  Find a and b such that the error with the data is minimized

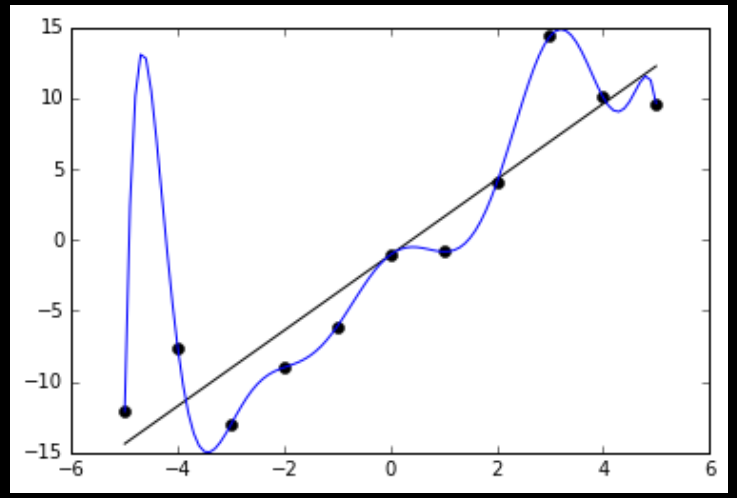
  Error: sum of the squares

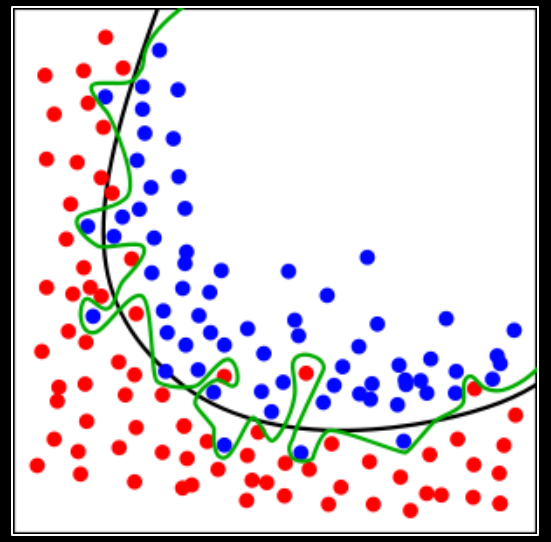
- Some applications:
  - Find data that is far from the "expected" values
  - Value prediction



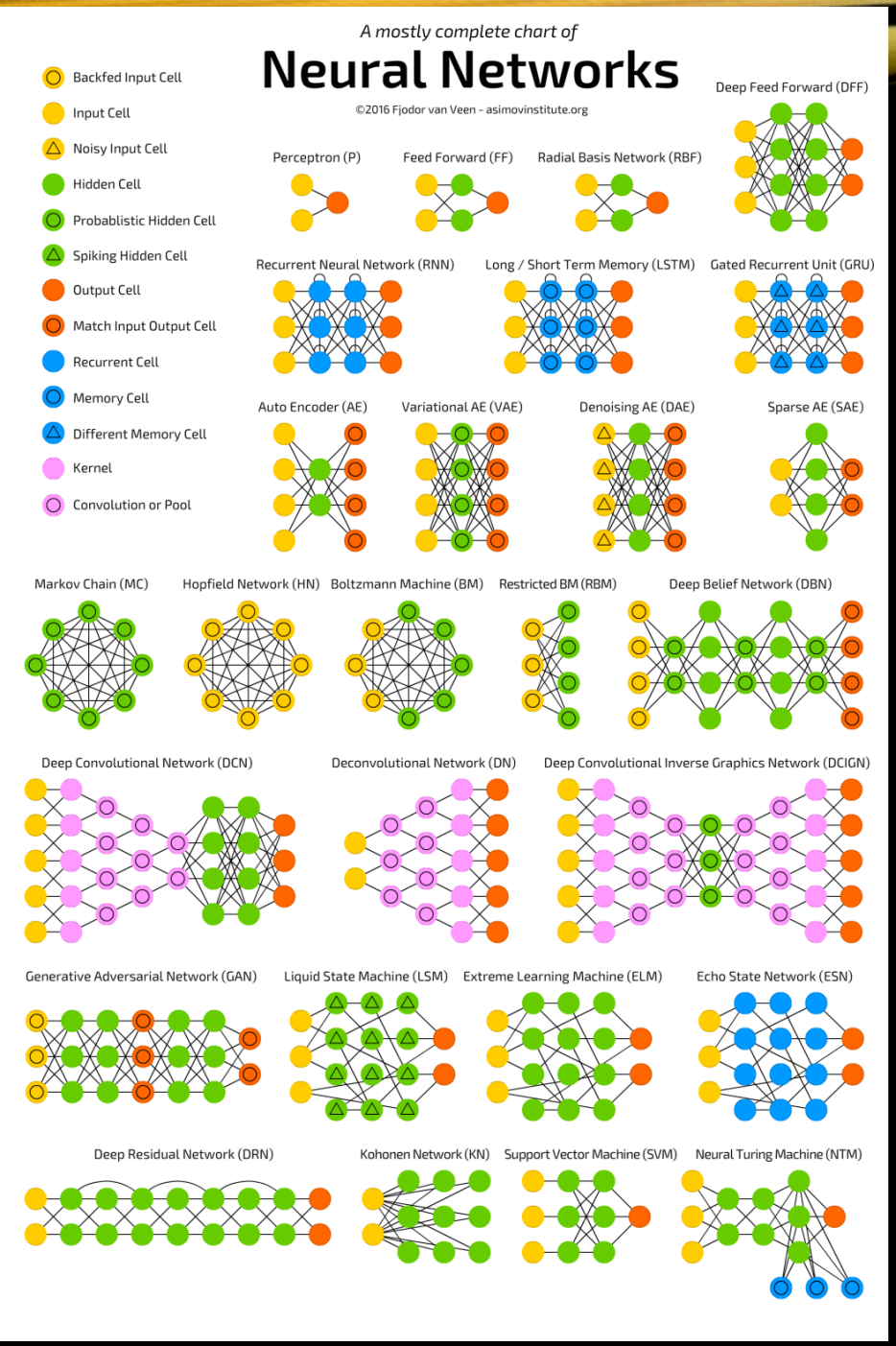
LEARNING MORE COMPLEX PATTERNS

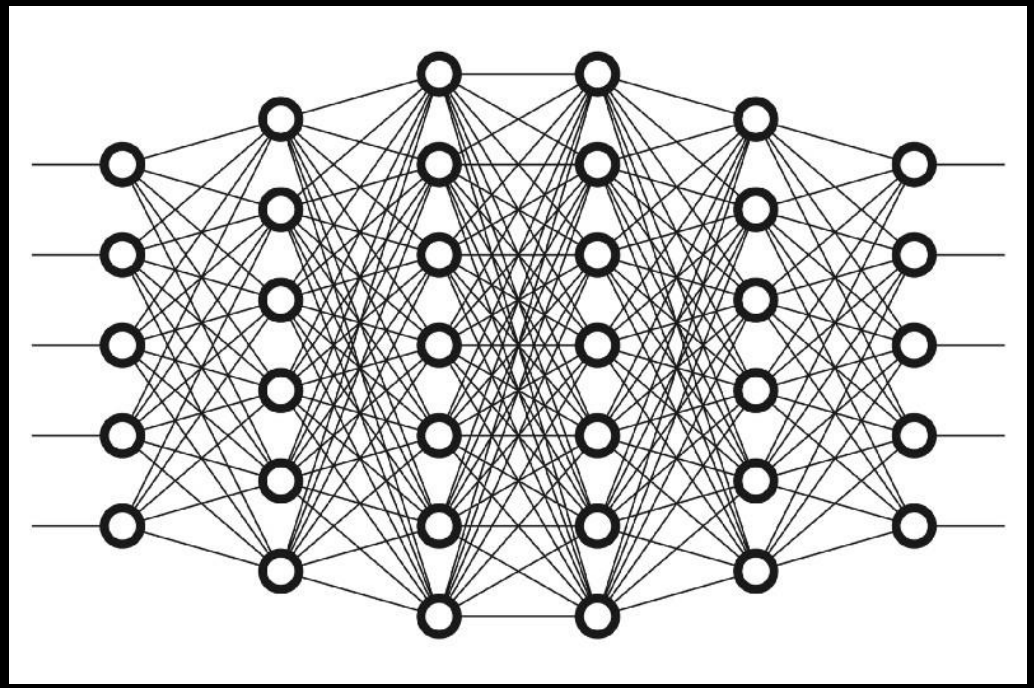




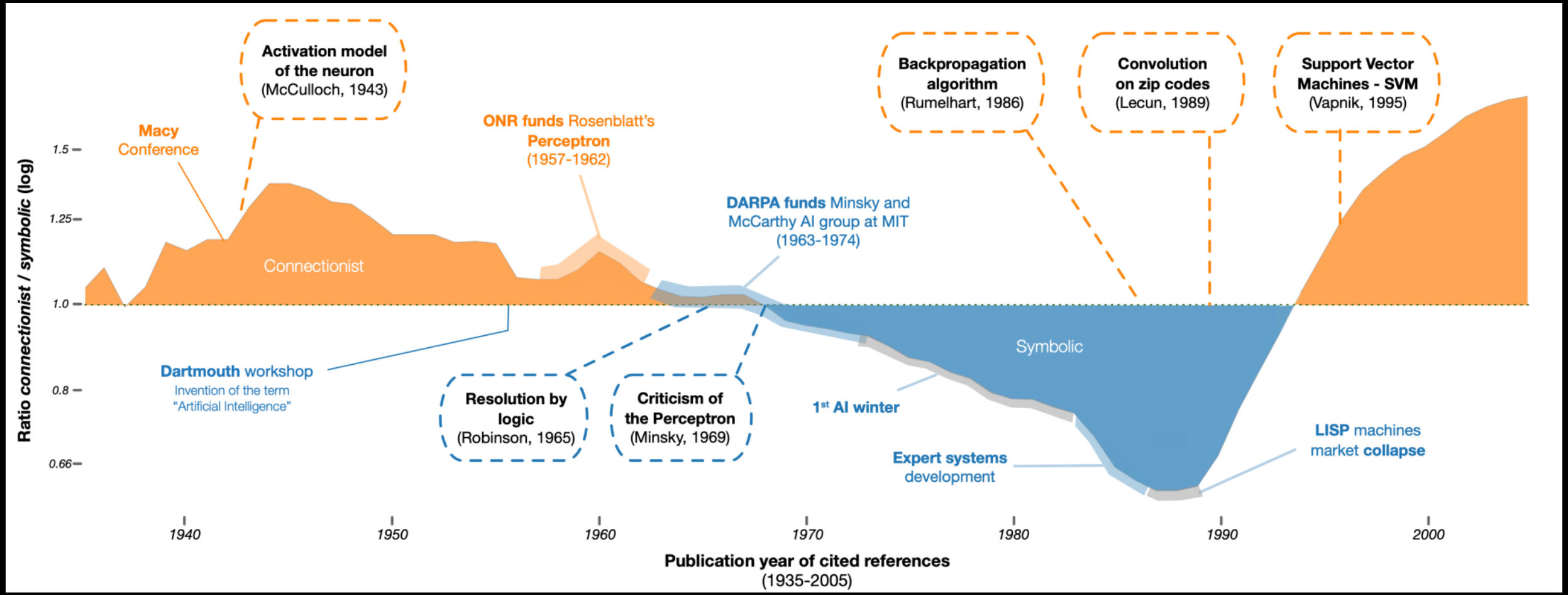


LEARNING VERY COMPLEX PATTERNS





WHY NOW?

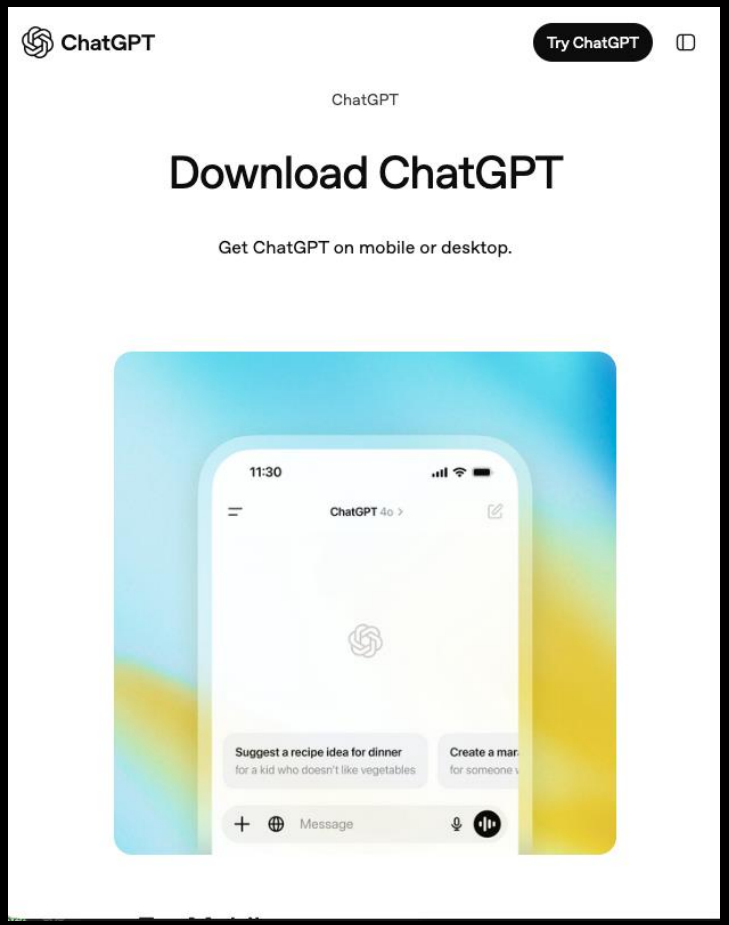

Activation model of the neuron (McCulloch, 1943)
Macy Conference
ONR funds Rosenblatt's Perceptron (1957-1962)
Backpropagation algorithm (Rumelhart, 1986)
Convolution on zip codes (Lecun, 1989)
Support Vector Machines - SVM (Vapnik, 1995)
DARPA funds Minsky and McCarthy AI group at MIT (1963-1974)
Connectionist
Symbolic
Dartmouth workshop
Invention of the term "Artificial Intelligence"
Resolution by logic (Robinson, 1965)
Criticism of the Perceptron (Minsky, 1969)
1st AI winter
Expert systems development
LISP machines market collapse
Ratio connectionist / symbolic (log)
1.5
1.25
1.0
0.8
0.66
1940
1950
1960
1970
1980
1990
2000
Publication year of cited references (1935-2005)

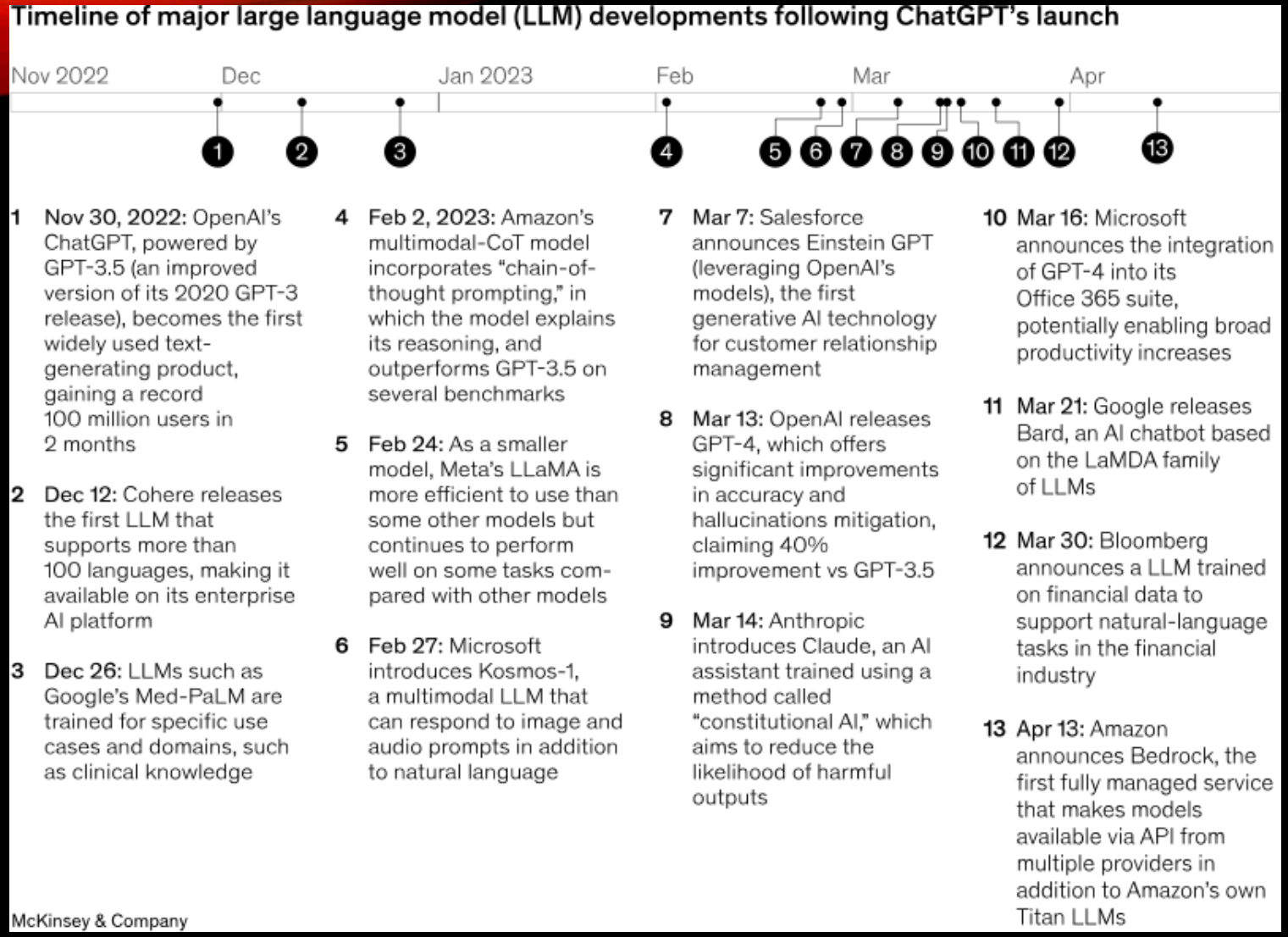
EXPLORING OUR USE OF GENAI

- Reflection Time (2 minutes):
  - Take a moment to reflect individually on the following question:
    - How have you used Generative AI models, such as Large Language Models (LLMs), so far?
- Speak Your Voice (1 minute per person):
  - Each person will have 1 minute to share their thoughts without interruption.
- Open Discussion (5-6 minutes):
  - After everyone has spoken, the floor is open for a discussion (questions, feedback, thoughts, …)

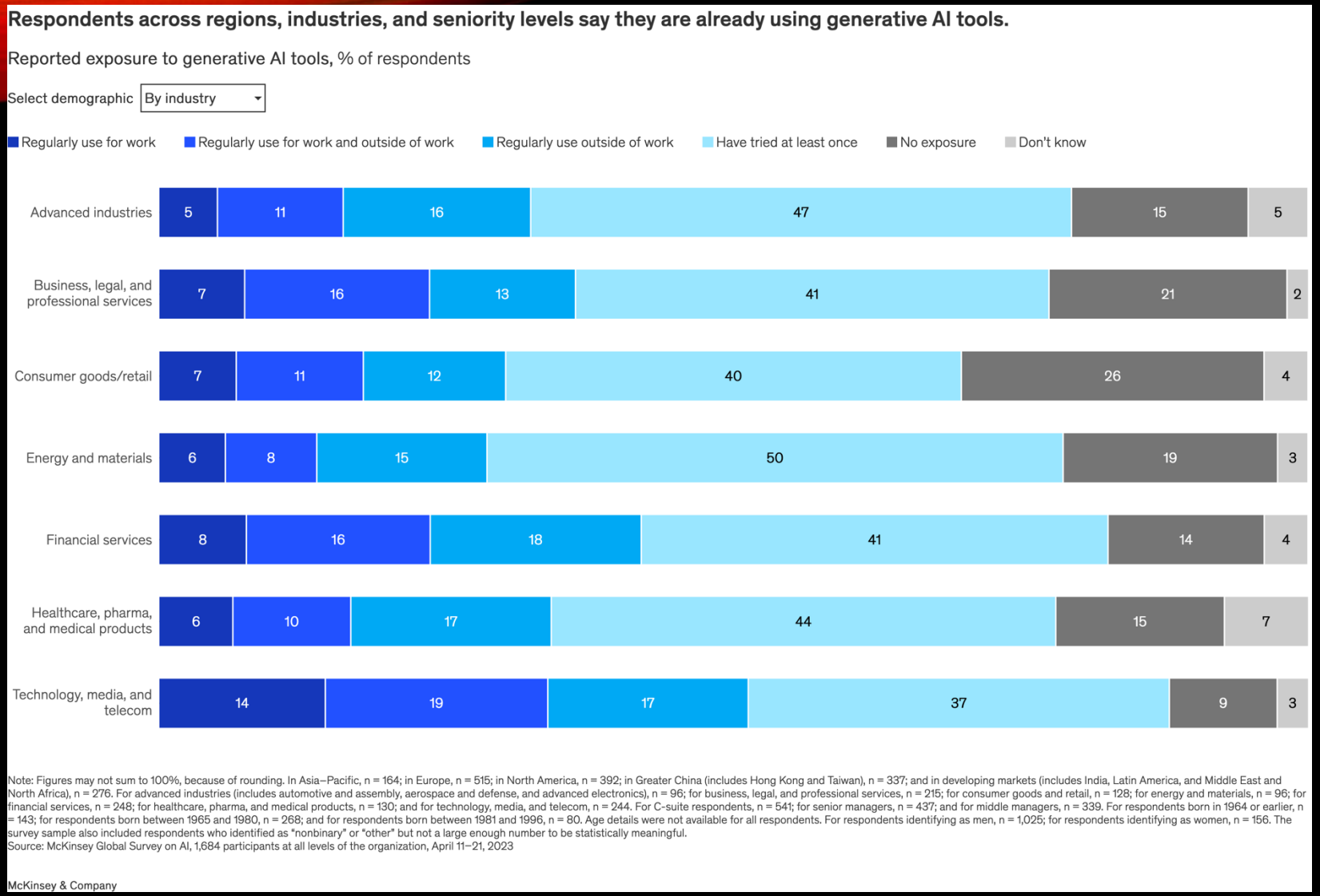
GENERATIVE AI: THE WHY

- Generative AI, exemplified by ChatGPT's rise in 2022, has rapidly accelerated AI investment, innovation, and adoption worldwide.





…
ACCELERATED AI INVESTMENT, INNOVATION, …



… ACCELERATED AI … ADOPTION WORLDWIDE

GENERATIVE AI: THE WHY

- Generative AI, exemplified by ChatGPT's rise in 2022, has rapidly accelerated AI interest, innovation, and adoption worldwide.
- **Business Impact:**
  - Adoption is widespread, with one-third of organizations using generative AI in at least one business function (McKinsey).
  - Gartner projects over 80% of organizations will deploy generative AI applications by 2026.
  - (expected) significant productivity benefits for individuals and organizations.
- **Future Potential**:
  - Leading AI researchers acknowledge the rapid approach of AI systems potentially surpassing human intelligence.

“The recent advances suggest that even the future where we know how to build superintelligent AIs (smarter than humans across the board) is closer than most people expected just a year ago.”

— Yoshua Bengio, ACM Turing Awardee

“I have suddenly switched my views on whether these things are going to be more intelligent than us. I think they’re very close to it now and they will be much more intelligent than us in the future.”

— Geoffrey Hinton, ACM Turing Awardee

GENERATIVE AI: THE WHAT

- Definition:
  - A generative AI model is a type of AI that can create new content, such as text, images, audio, or video, based on the data it has been trained on.
- User Interaction:
  - Users provide input or prompts (like a question or a request), and the model generates an appropriate response or output.
- Examples:
  - Generating written articles, creating artwork, composing music, or even engaging in human-like conversations.

GENERATIVE AI: THE HOW

Training

Purpose: Creating a foundation model.

Description: The AI is trained on vast datasets to build a general-purpose model that serves as the base for multiple generative AI applications.

Fine-Tuning

Purpose: Tailoring the foundation model.

Description: The foundation model is adjusted or fine-tuned for a specific application, aligning it with goals or tasks.

Generation and Evaluation

Purpose: Assessing and improving outputs.

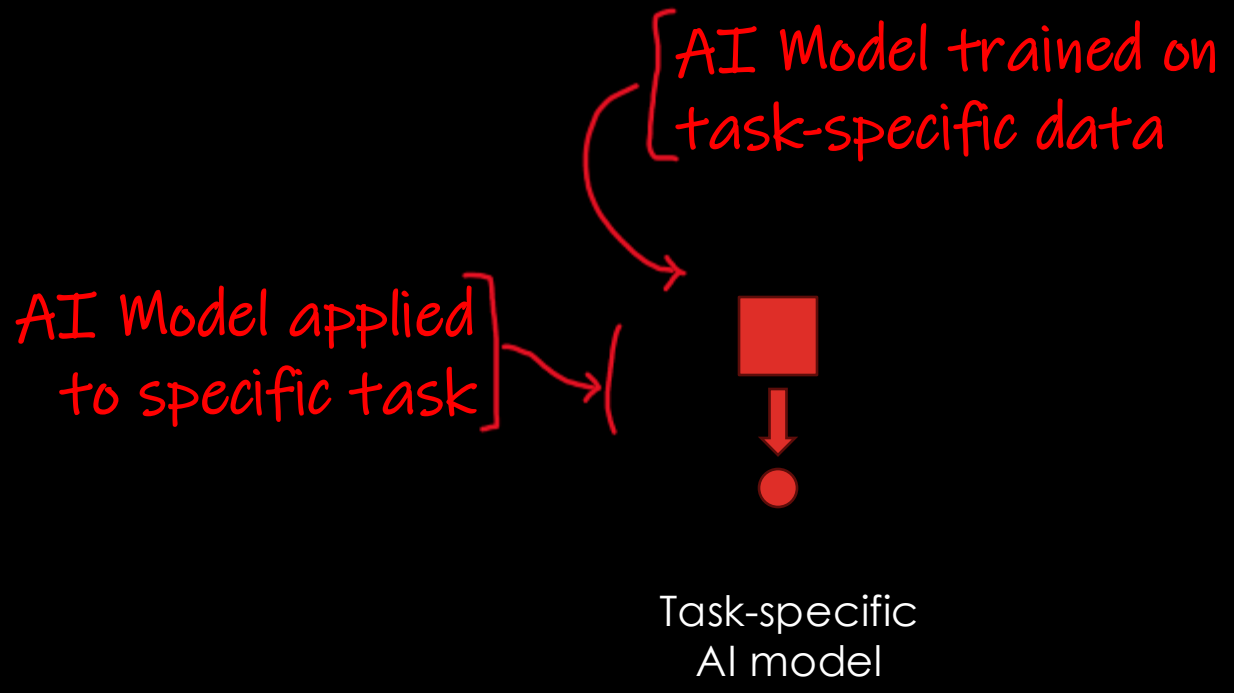
Description: The AI generates content, which is then evaluated. Based on this evaluation, the model is retuned to enhance the quality of its outputs.

GENERATIVE AI: THE HOW

Training

Purpose: Creating a foundation model.

Description: The AI is trained on vast datasets to build a general-purpose model that serves as the base for multiple generative AI applications.

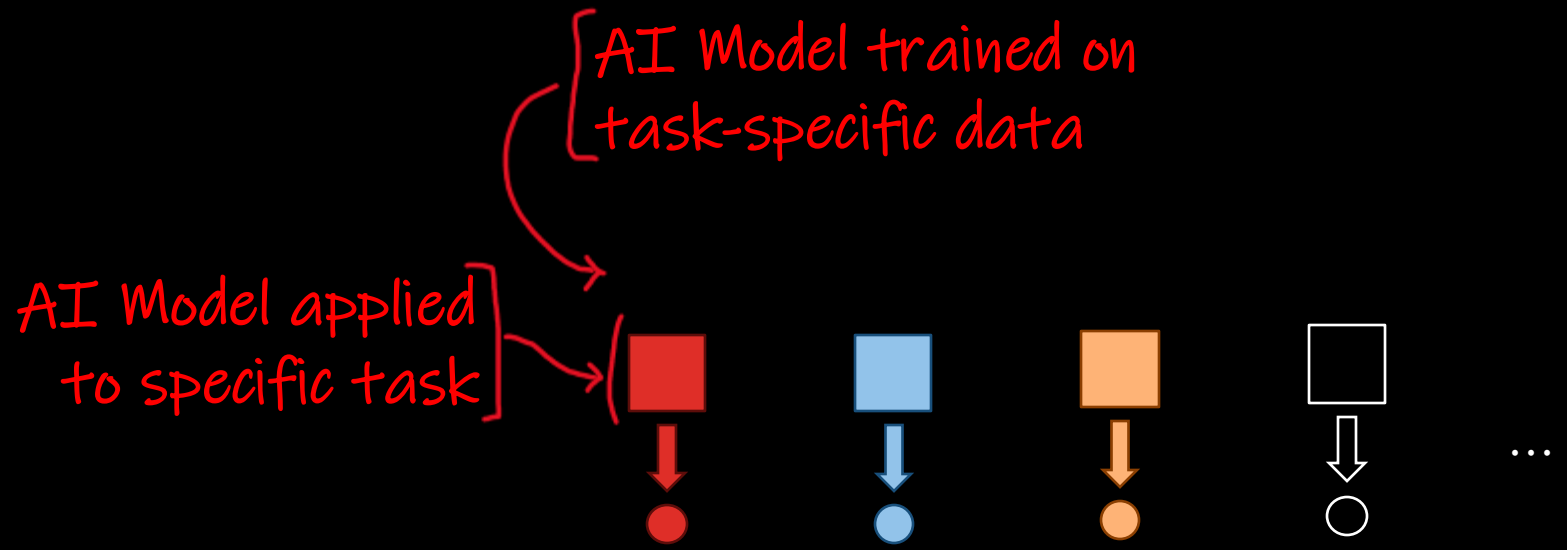


GENERATIVE AI: THE HOW

Training

Purpose: Creating a foundation model.

Description: The AI is trained on vast datasets to build a general-purpose model that serves as the base for multiple generative AI applications.

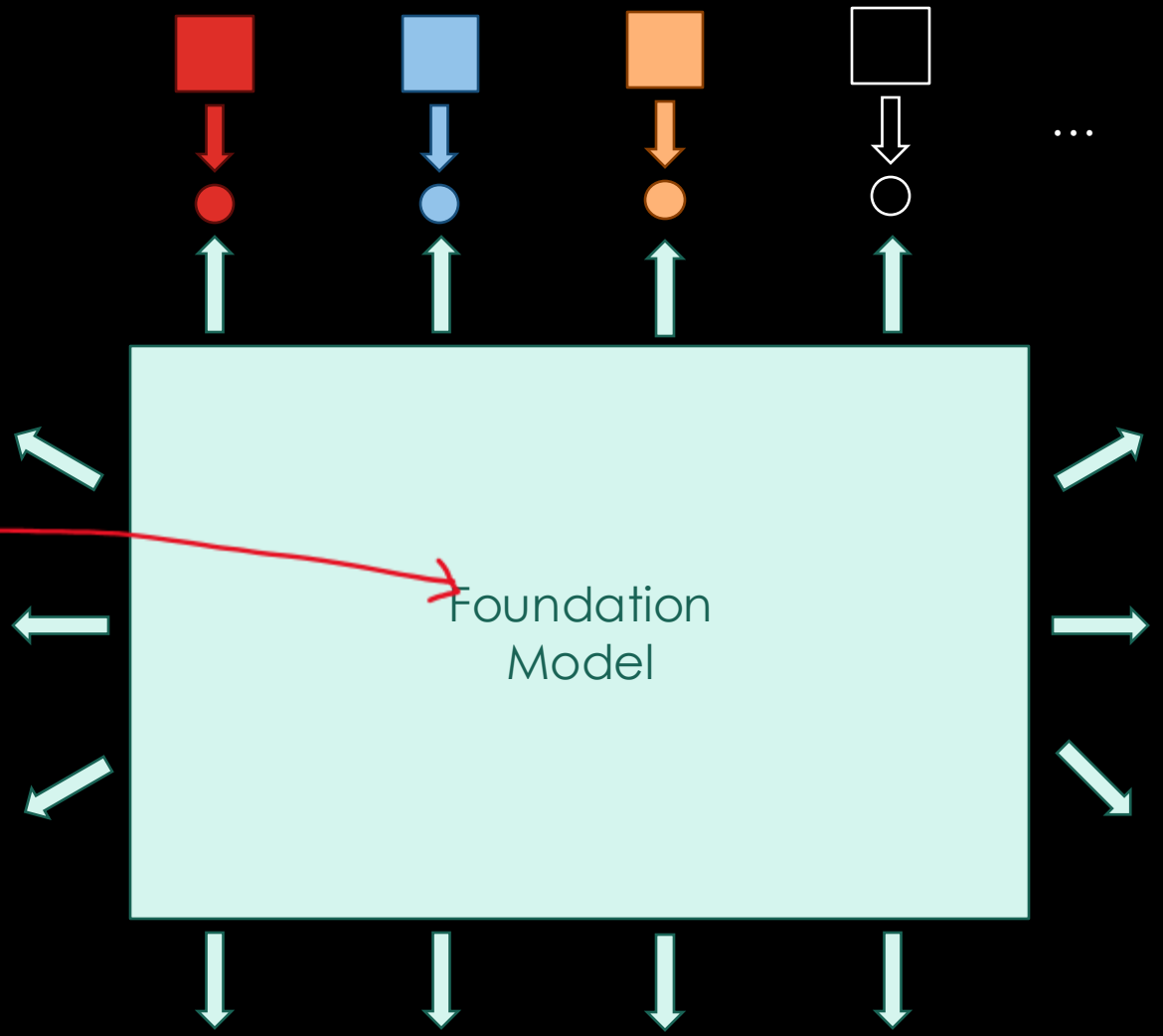


GENERATIVE AI: THE HOW

Training

Purpose: Creating a foundation model.

Description: The AI is trained on vast datasets to build a general-purpose model that serves as the base for multiple generative AI applications.



GENERATIVE AI: THE HOW

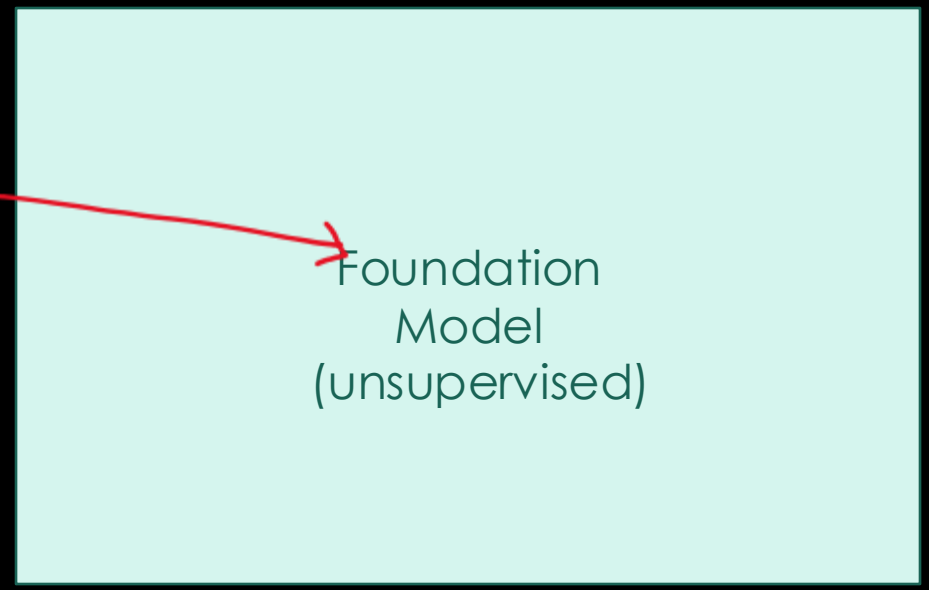
Training

Purpose: Creating a foundation model.

Description: The AI is trained on vast datasets to build a general-purpose model that serves as the base for multiple generative AI applications.

`Would you like coffee or _____`

`Bellinzona is the capital of the canton of _____`



GENERATIVE AI: THE HOW

Training

Purpose: Creating a foundation model.

Description: The AI is trained on vast datasets to build a general-purpose model that serves as the base for multiple generative AI applications.

- **Data**: Massive, raw, unstructured datasets (e.g., internet data).
- **Process**: The model learns by predicting and refining outputs of 'fill in the blank' exercises.
- **Outcome**: Creates a neural network capable of autonomous content generation.
- **Resources**: Requires significant computing power, time, and cost.

GENERATIVE AI: THE HOW

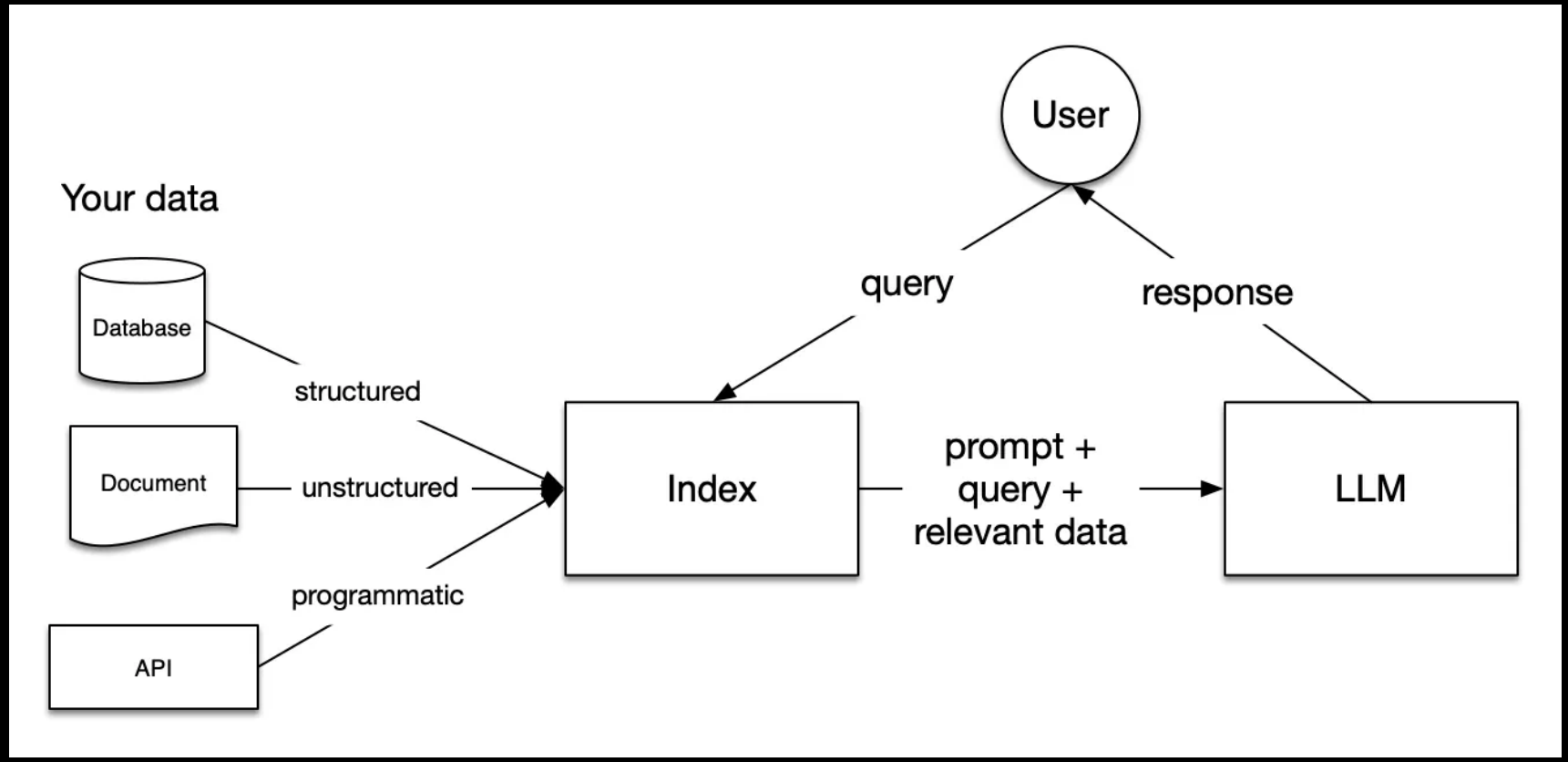
Generation and Evaluation

Purpose: Assessing and improving outputs.

Description: The AI generates content, which is then evaluated. Based on this evaluation, the model is retuned to enhance the quality of its outputs.

- **Continuous Improvement**: Regularly assess and fine-tune the model for accuracy and relevance (e.g., weekly).
- **Foundation Model Updates**: Less frequent, updated every 12-18 months.
- **Retrieval Augmented Generation (RAG):**
  - **Purpose**: Enhance performance by incorporating external, up-to-date sources.
  - **Process**: Extends the foundation model to use relevant data beyond its original training set.
  - **Benefits**: Provides transparency and ensures access to current information.

RAG: RETRIEVE → AUGMENT → GENERATE



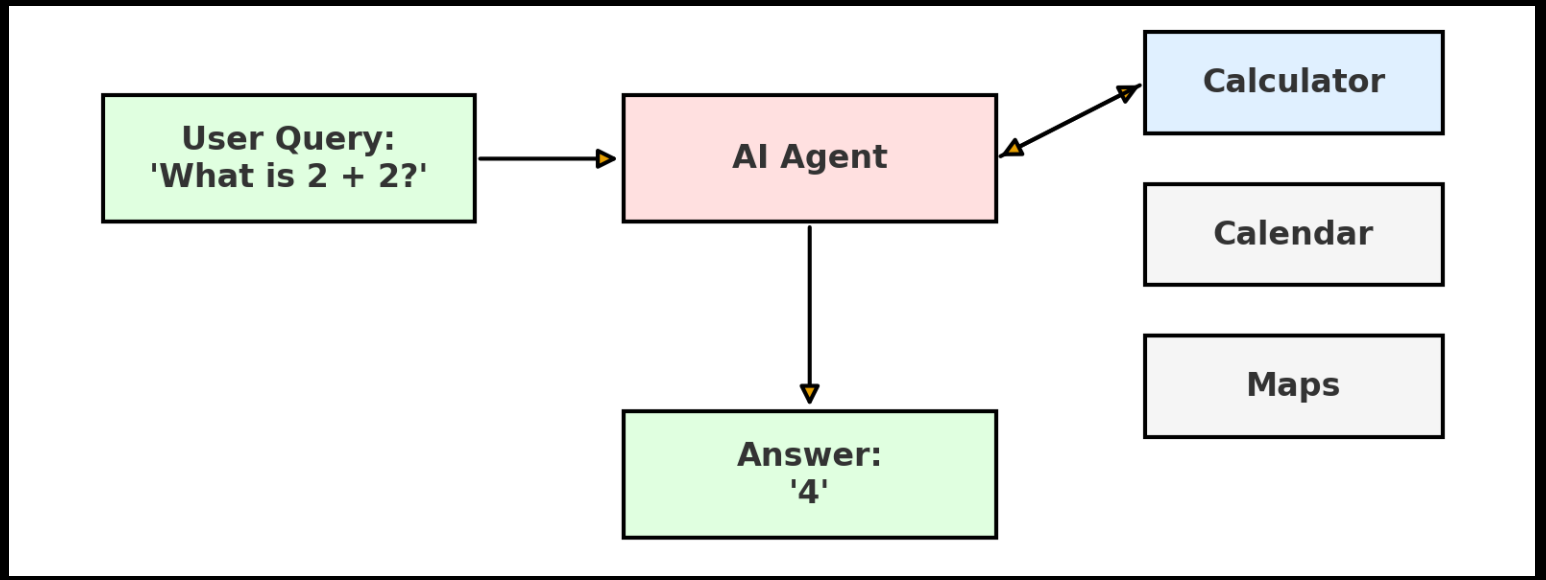
Reference: https://medium.com/@tejaswi_kashyap/rag-processing-using-llamaindex-43d9786f9d8e

AI AGENTS IN THE AGE OF LLMS

An AI Agent can:

- Receive a user request
- Decide which tool/module to use (e.g., calculator, calendar, maps)
- Use the tool to get accurate info
- Return the answer

Example below: The agent uses the Calculator to solve 2 + 2.



GENERATIVE AI USES

Generative AI use cases, nonexhaustive
McKinsey & Company

Modality	Application	Example use cases
Text	Content writing	Marketing: creating personalized emails and posts Talent: drafting interview questions, job descriptions
	Chatbots or assistants	Customer service: using chatbots to boost conversion on websites
	Search	Making more natural web search Corporate knowledge: enhancing internal search tools
	Analysis and synthesis	Sales: analyzing customer interactions to extract insights Risk and legal: summarizing regulatory documents
Code	Code generation	IT: accelerating application development and quality with automatic code recommendations
	Application prototype and design	IT: quickly generating user interface designs
	Data set generation	Generating synthetic data sets to improve AI models' quality
Image	Stock image generator	Marketing and sales: generating unique media
	Image editor	Marketing and sales: personalizing content quickly
Audio	Text to voice generation	Trainings: creating educational voiceover
	Sound creation	Entertainment: making custom sounds without copyright violations
	Audio editing	Entertainment: editing podcast in post without having to rerecord
3-D or other	3-D object generation	Video games: writing scenes, characters Digital representation: creating interior-design mockups and virtual staging for architecture design
	Product design and discovery	Manufacturing: optimizing material design Drug discovery: accelerating R&D process
Video	Video creation	Entertainment: generating short-form videos for TikTok Training or learning: creating video lessons or corporate presentations using AI avatars
	Video editing	Entertainment: shortening videos for social media E-commerce: adding personalization to generic videos Entertainment: removing background images and background noise in post
	Voice translation and adjustments	Video dubbing: translating into new languages using AI-generated or original-speaker voices Live translation: for corporate meetings, video conferencing Voice cloning: replicating actor voice or changing for studio effect such as aging
	Face swaps and adjustments	Virtual effects: enabling rapid high-end aging; de-aging; cosmetic, wig, and prosthetic fixes Lip syncing or "visual" dubbing in postproduction: editing footage to achieve release in multiple ratings or languages Face swapping and deep-fake visual effects Video conferencing: real-time gaze correction

CREDITS, RESOURCES, AND REFERENCES

- https://programming-group.com/
- Wharton's Ethan Mollick & Stefano Puntoni – AI Horizons: Co-Intelligence
- Why Large Language Models Hallucinate
- What Is AI Bias? | IBM
- What is Generative AI? | IBM
- Evaluating Large Language Models Using “Counterfactual Tasks”
- Co-Intelligence: Living and Working with AI
- Co-Intelligence: AI in the Classroom with Ethan Mollick | ASU+GSV 2024
- How to Become a Top 1% Student using AI (it's not cheating!)
- Math problems with GPT-4o
- How do we know how smart AI systems are? | Science

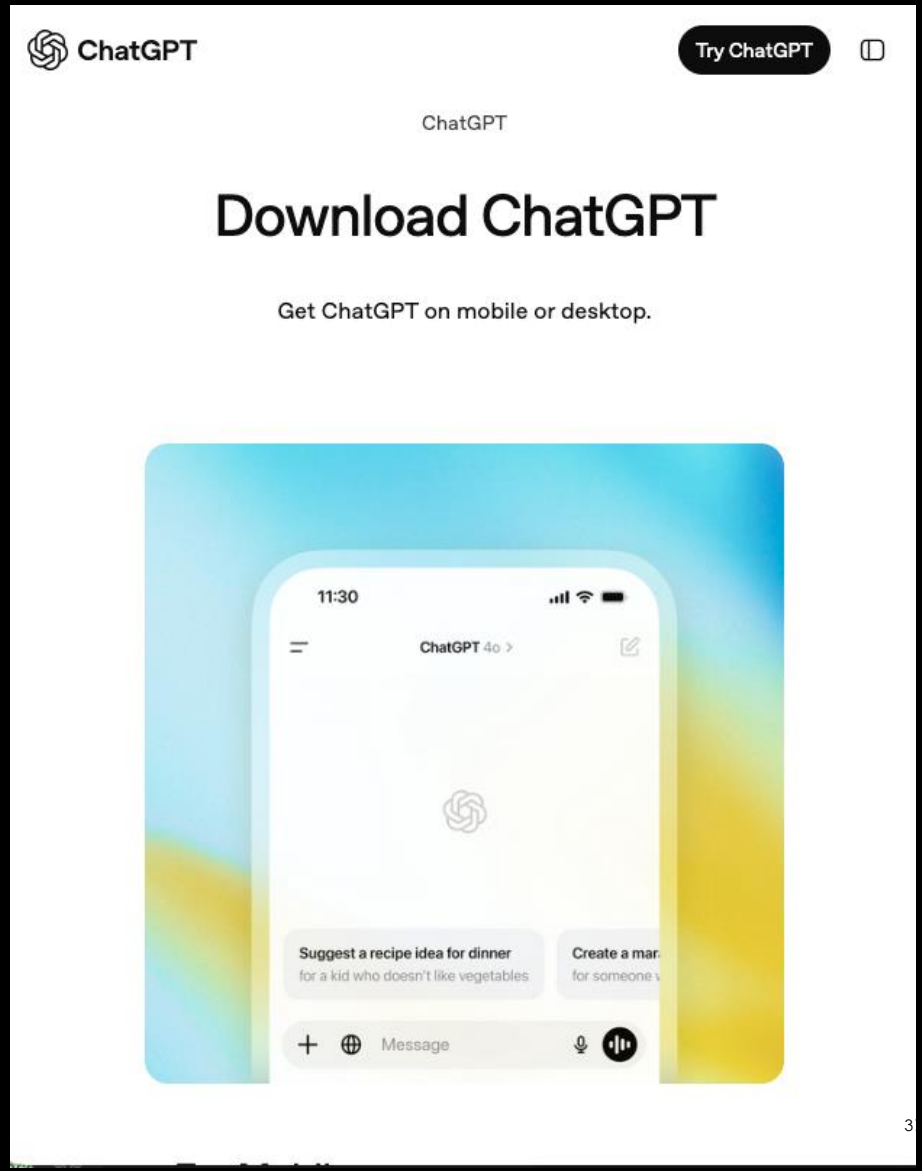
SOFTWARE SETUP

- We will do some group work with Python.
- Now we will:
  - Install ChatGPT
  - Install Python
  - Install VSCode
  - Install Copilot
  - Install Ollama
  - Run a simple program with LLM in background

CHATGPT

Download from:

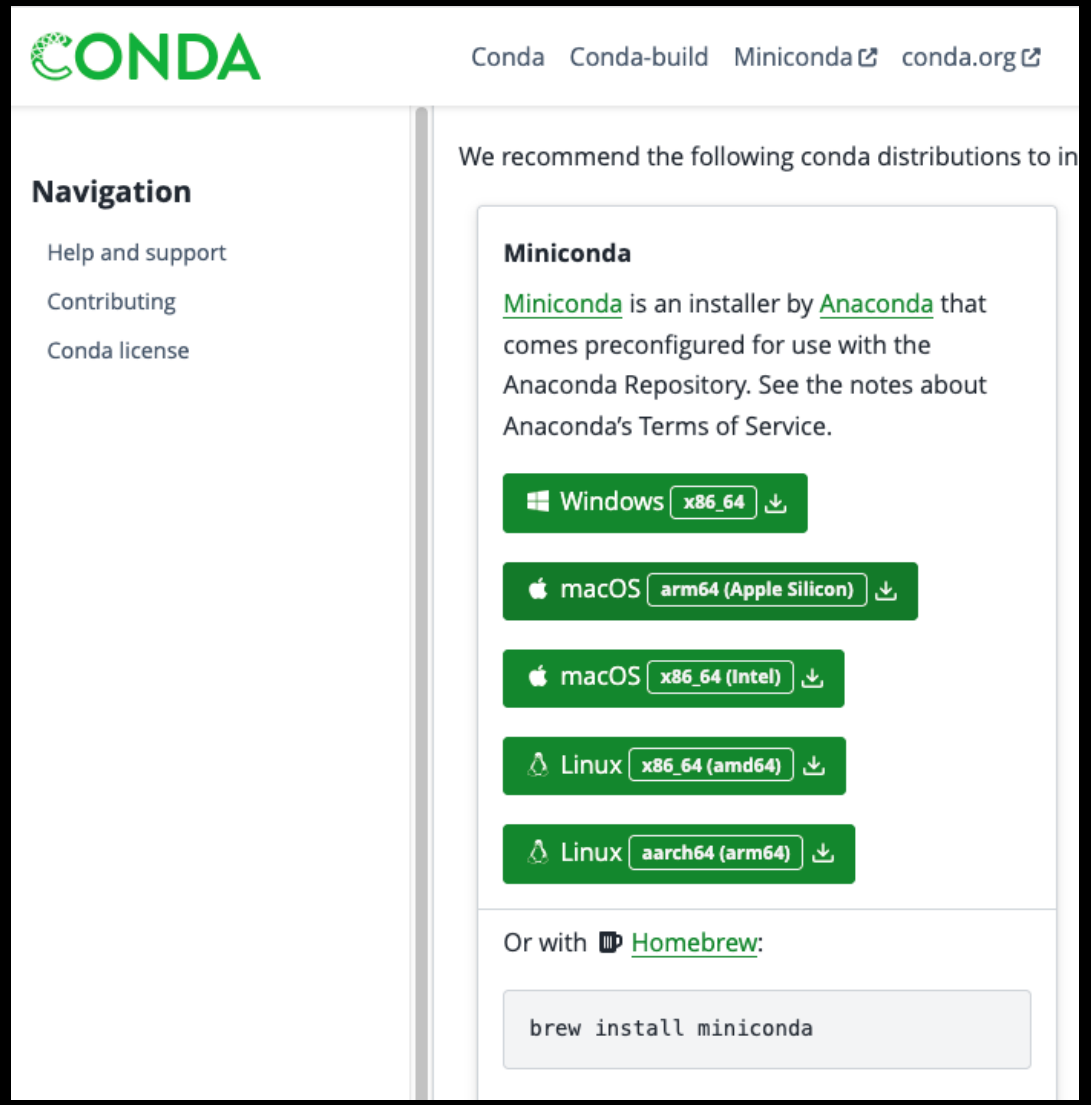
- https://chatgpt.com/download/



CONDA

Download from it:

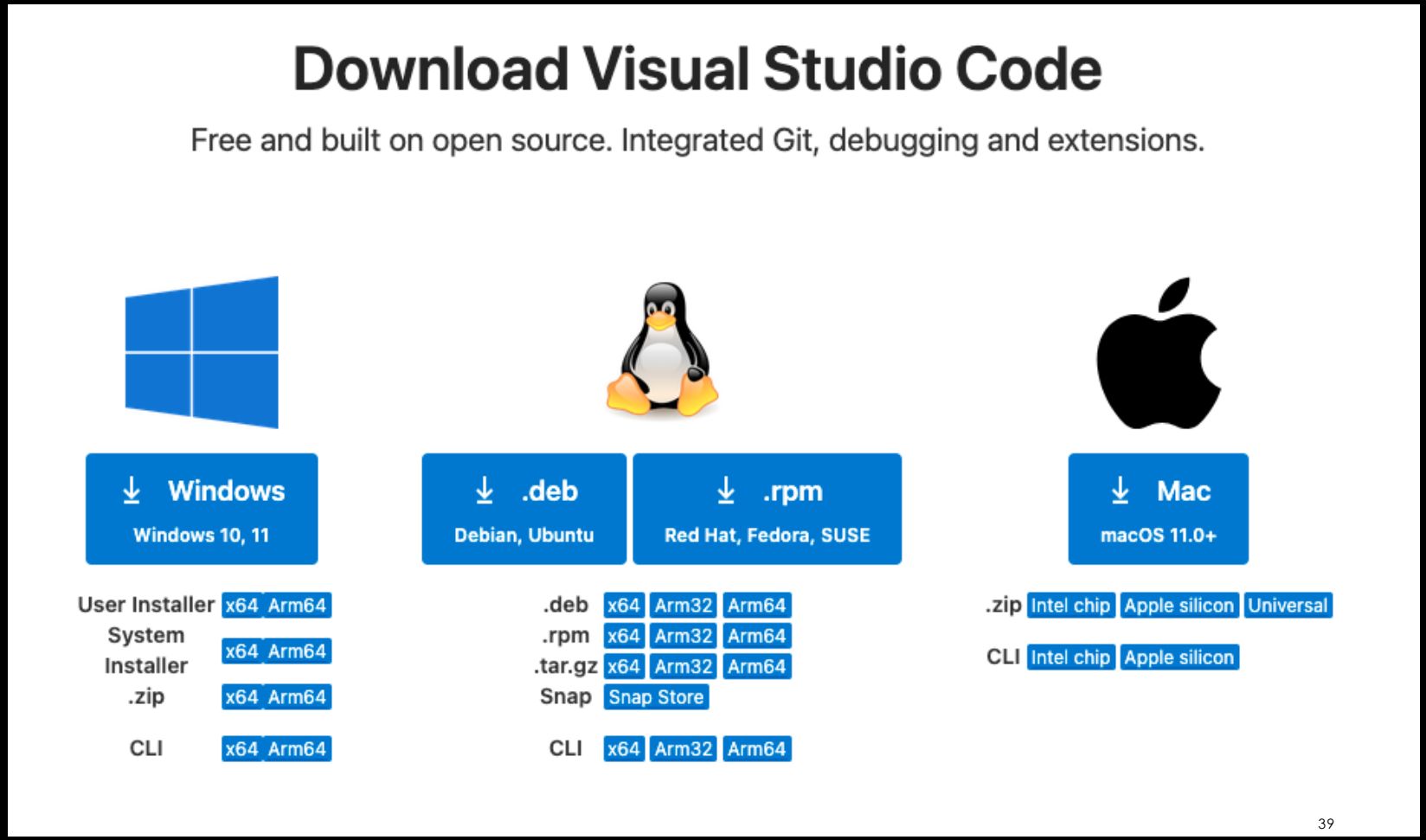
- https://docs.conda.io/en/latest/



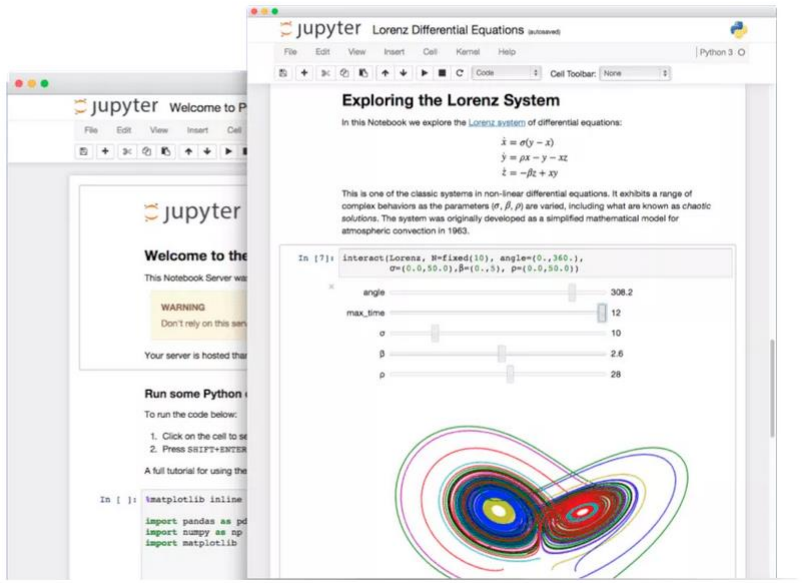
VSCODE

Download from it:

- https://code.visualstudio.com/download



JUPITER NOTEBOOK



Jupyter Notebook: The Classic Notebook Interface

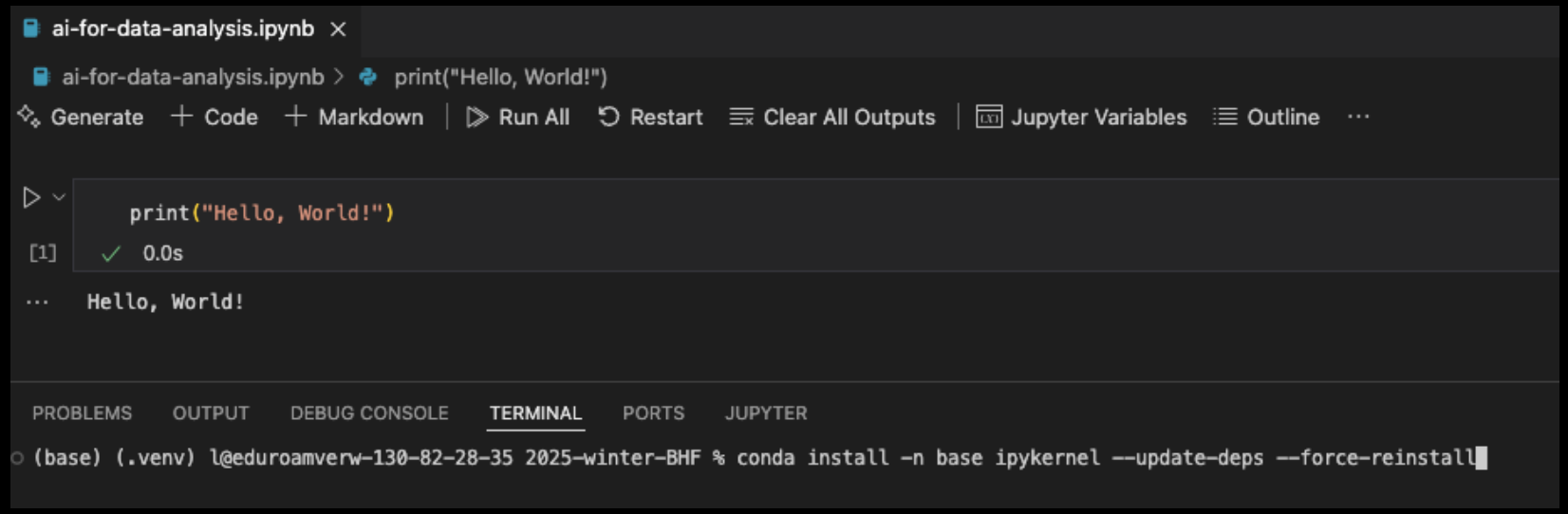
The Jupyter Notebook is the original web application for creating and sharing computational documents. It offers a simple, streamlined, document-centric experience.

Try it in your browser | Install the Notebook

RUN “HELLO WORLD”

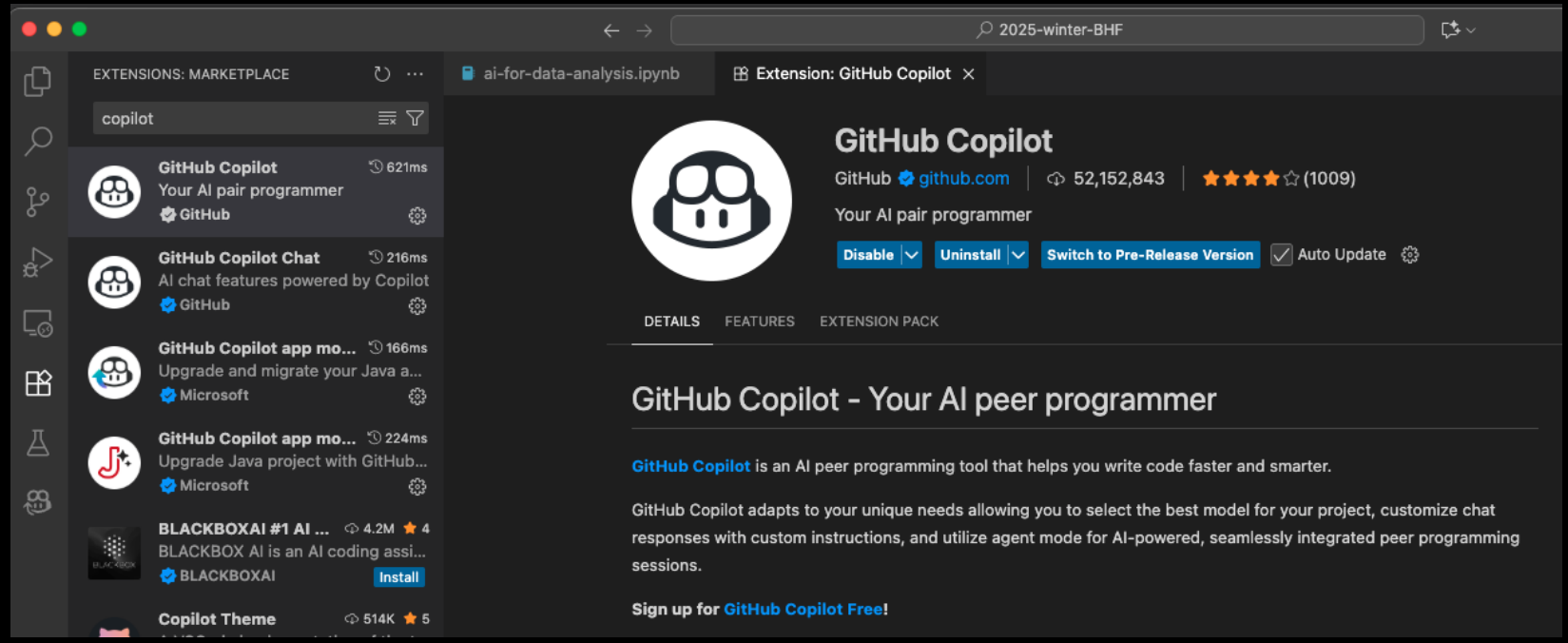
Download the ipykernel:

- conda install -n base ipykernel --update-deps --force-reinstall



(OPTIONAL) INSTALL COPILOT

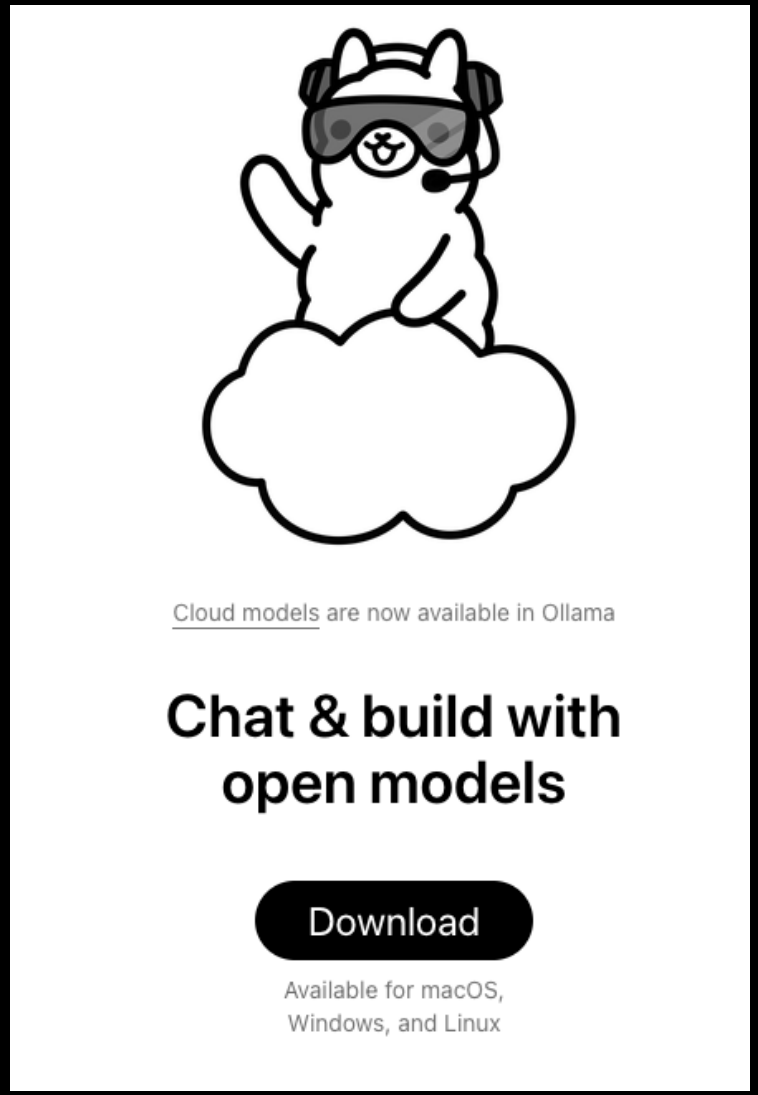
From Vscode plugin. After the installation you have to create an account. It is free for students.



OLLAMA

Download from it:

- https://ollama.com/download



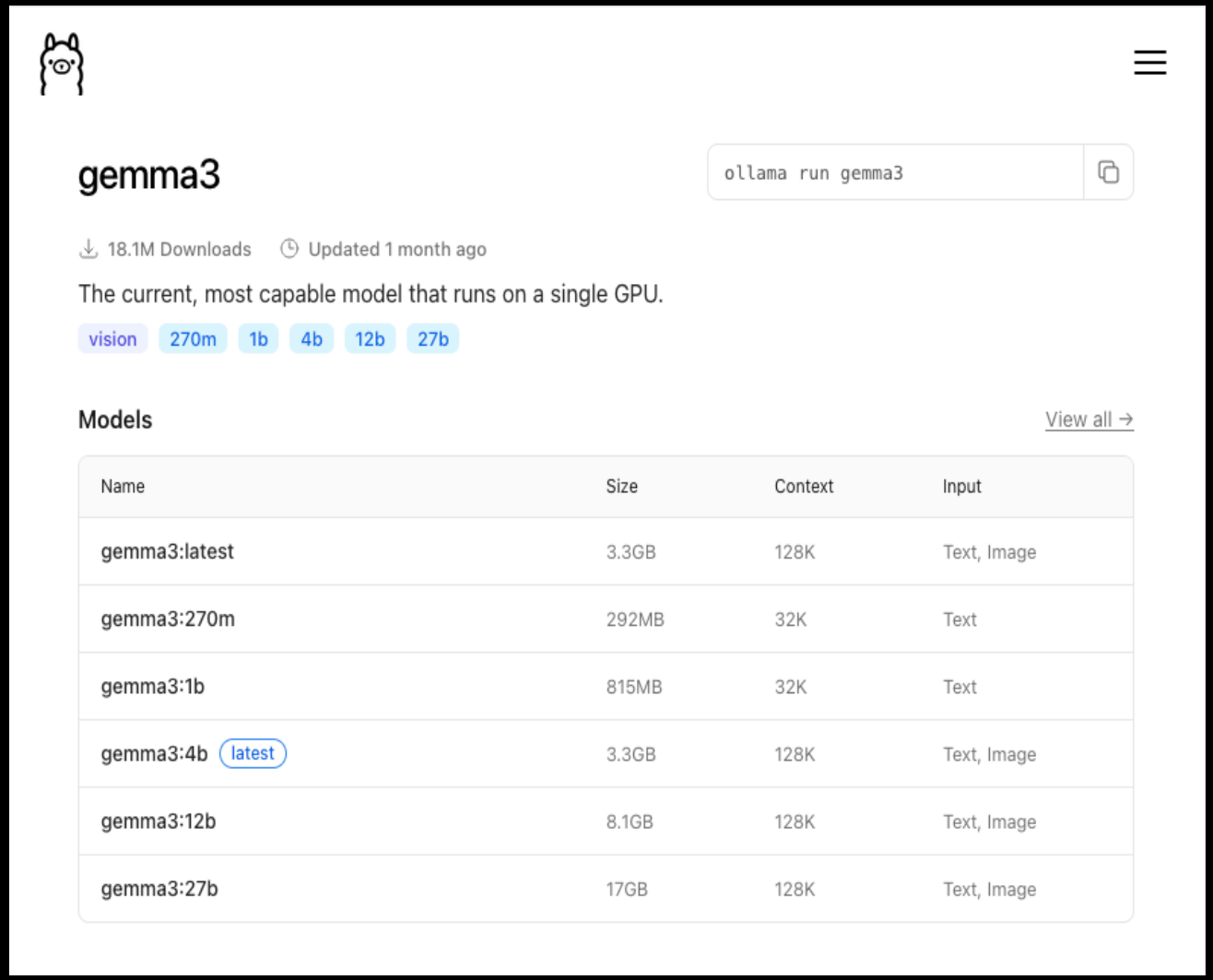
LLM WITH OLLAMA

Download Gemma Model:

- https://ollama.com/library/gemma3

Download Gemma Embeddings:

- https://ollama.com/library/embeddinggemma



AI SETUP AND PRIVACY CONCERNS

WHY PRIVACY MATTERS WITH AI IN FINANCE

- Financial data = highly sensitive (revenues, forecasts, client records)
- Sending data to external servers = potential compliance risk (GDPR, banking secrecy)
- Cloud-based LLMs (like ChatGPT) process data on third-party infrastructure
- Risks:
  - Data leakage or misuse
  - Regulatory violations
  - Loss of client trust

BENEFITS OF RUNNING LLMS LOCALLY

- Data processed on company servers or devices → stays private
- No exposure to external providers = lower compliance risk
- Greater control over data flow and storage
- Tailored models: can fine-tune on internal data without risk of leaks
- Supports secure scenarios: audits, client contracts, strategy reports

LOCAL LLMS IN CORPORATE PRACTICE

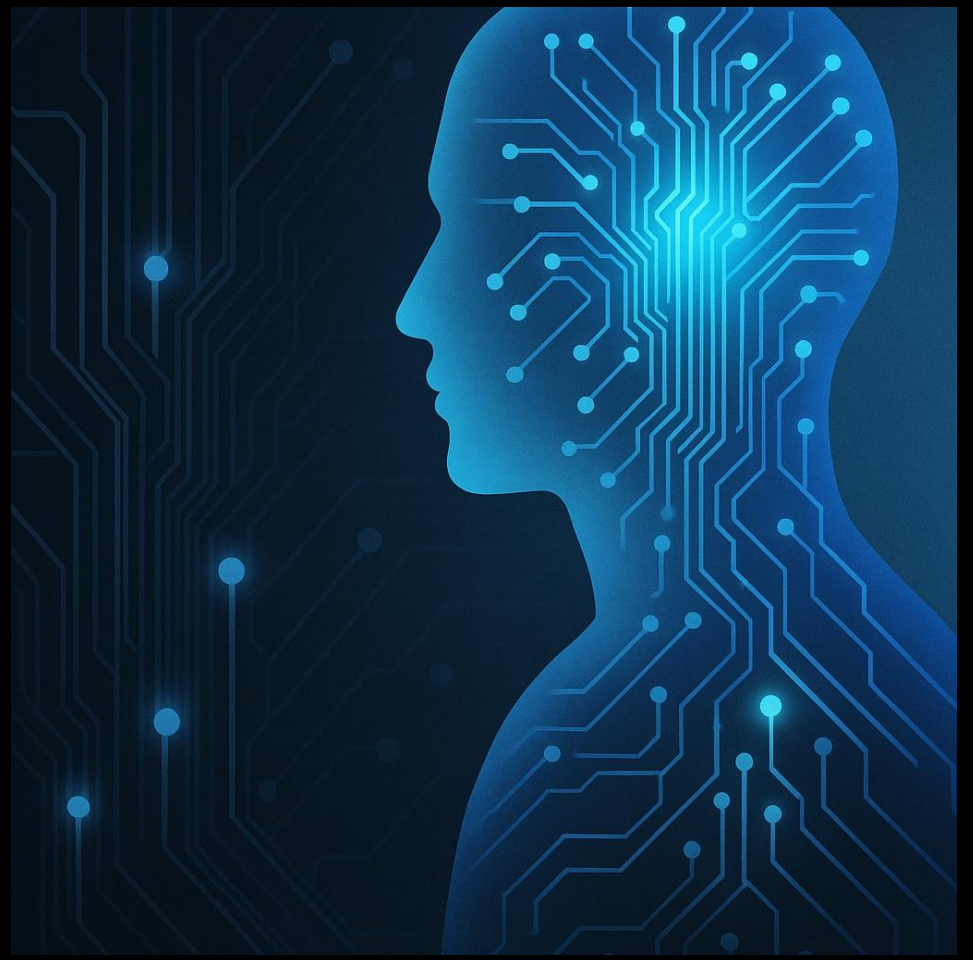
- Privacy & Security: protect intellectual property and client data
- Regulatory Compliance: align with EU AI Act, GDPR, and sector rules
- Cost Predictability: avoid pay-per-token costs of cloud APIs
- Customization: adapt models to internal language, KPIs, and style
- Trust: Analysts and managers can use AI without fear of exposing sensitive information

DATA PRIVACY & FUNDAMENTAL RIGHTS

- AI Act focuses on safe technical development while the GDPR is a fundamental rights law; the two work hand-in-hand to protect individuals' personal data.
- GDPR's broad definition of personal data makes it hard to separate personal from non-personal information, so AI systems often process personal data.
- Large-scale data collection for AI training and use can clash with data-minimisation principles, underscoring the need for robust privacy management.

GOVERNANCE, TRANSPARENCY & HUMANS

- High-risk AI must satisfy data governance, transparency and human oversight obligations.
- High-risk systems must be designed so natural persons can effectively oversee decisions (human-in-the-loop).
- Limited-risk and generative AI systems face transparency obligations: users must know when they are interacting with AI and AI-generated content (e.g. deepfakes) must be clearly labelled.



BALANCING INNOVATION & COMPLIANCE

- The AI Act strives to ensure AI serves people and aligns with European values while fostering innovation and competitiveness.
- It applies to all AI systems marketed or used in the EU, including those operated outside the EU when their outputs are used within the Union.
- Finding the right balance is challenging companies must comply with ethical requirements without stifling competition and innovation.

THE PROJECT AND THE FINAL EVALUATION

THE PROJECT

- AI-Assisted Financial Data Analysis Project
- Group project (2–3 students per team)
- Deliverable: Jupiter Notebook + Final Presentation

Time	Instruction mode	Topic	Description
13:00-16:30	Project & Coaching	Applying the protype built to a case study	Students work in groups to extend their prototype to address a specific problem in the digital finance domain.
16:30-17:45	Presentation session	Group Presentations and Feedback formulation	Each group presents their findings (10 minutes), with constructive feedback from peers and the instructor.

AI FOR DATA ANALYSIS

- Finance meets AI: analyze a company's performance using data and AI tools
- Combine quantitative analysis (CSV) with qualitative insights (PDF)
- Experiment with local LLMs (Ollama) and ChatGPT
- Reflect on privacy, accuracy, and usability of AI in finance

THE DATASET

- Delta Air Lines 2020 earnings release
  - https://ir.delta.com/news/news-details/2021/Delta-Air-Lines-Announces-December-Quarter-and-Full-Year-2020-Financial-Results/default.aspx

- Together: combine narrative insights (PDF) with structured data (CSV)
- Public, academic-style datasets: safe for student use

PROJECT WORKFLOW

1. Extract insights from PDF
2. Build structured data with CSV
3. Use Ollama with Gemma-1B and another local model
4. Repeat tasks with ChatGPT
5. Compare results (quality, privacy, usability)
6. Present findings

STEP 1: PDF ANALYSIS

- Parse the PDF with an LLM and ask to extract a CSV file from it
- Identify key financial figures (revenue, loss, strategy)
- Summarize text with local LLMs
- Check correctness against the report

STEP 2: CSV ANALYSIS

- Load the csv file generated by the LLM
- Compute year-over-year changes for key metrics
- Calculate ratios: profit margin, revenue growth, expenses
- Create at least 2 clear visualizations (line chart, bar chart, etc.)

STEP 3: LOCAL LLMS WITH OLLAMA

- Use Gemma:1B locally for text summarization & Q&A
- Try another model (e.g., a larger one or from a different family) for comparison
- Document: response quality, speed, ease of prompting
- Note: all processing stays on your laptop (privacy advantage)

STEP 4: CHATGPT ANALYSIS

- Run the same Q&A tasks with ChatGPT
- Ask for summaries, explanations of charts, financial context
- Compare clarity, detail, and correctness
- Note: cloud processing → potential privacy trade-off

MODEL COMPARISON

- Compare ChatGPT vs Local LLMs on:
  - Accuracy of financial facts
  - Depth of explanation
  - Privacy & data security
  - Ease of use & interactivity
  - Document examples of strong/weak answers

CODING TASKS

- Pre-written starter notebook provided
- Students can:
  - Add code blocks for analysis
  - Add code for plots, stats, and markdown explanations
  - Show AI prompts and outputs
  - Notebook is part of the final submission

EVALUATION

1. Jupyter Notebook (completed, clean, reproducible)
2. Group Presentation (10 min)
   - Data analysis results
   - AI outputs & comparisons
   - Critical reflections on privacy/usability

PRESENTATION GUIDELINES

- 10 minutes total, all members must present
- Use clear visuals (charts, screenshots of AI outputs)
- Divide slides evenly (each presents ~3–4 min)
- Focus on insights, not just process
- Anticipate questions from instructor & peers

EVALUATION CRITERIA

- Data Analysis: Clarity and meaningful plots
- AI Integration: Use of both Ollama & ChatGPT, documented outputs
- Critical Comparison: Accuracy, privacy, usability reflection
- Presentation: Clarity, equal participation, design quality
- Teamwork & Creativity: Original insights, balanced roles

TIPS FOR SUCCESS

- Verify AI answers against data!
- Use clear, readable charts (labels, titles, units)
- Divide work early, rehearse presentation
- Keep track of prompts & outputs for evidence
- Don’t be afraid to show AI mistakes: it is part of the learning

COMMON PITFALLS TO AVOID

- Overloading slides with text
- Relying only on ChatGPT (must compare!)
- 1 Slide = 1 Minute of talk
- Less than 30 words per slide
- Uneven team contribution

LET'S START THE PROJECT

- Goal: Practice financial analysis with AI support
- Balance coding, data visualization, and AI experiments
- Learn trade-offs: privacy vs convenience, detail vs simplicity
- Deliver as a team: notebook + presentation
- You can ask me any question, do not be shy!

Time	Instruction mode	Topic	Description
13:00-16:30	Project & Coaching	Applying the protype built to a case study	Students work in groups to extend their prototype to address a specific problem in the digital finance domain.
16:30-17:45	Presentation session	Group Presentations and Feedback formulation	Each group presents their findings (10 minutes), with constructive feedback from peers and the instructor.

THANKS AND CONCLUSION

Contact:

- Luca Di Grazia
- www.lucadigrazia.com
- work@lucadigrazia.com

WRAP-UP SLIDE

- Let's built the final slide together. What did we learn today?