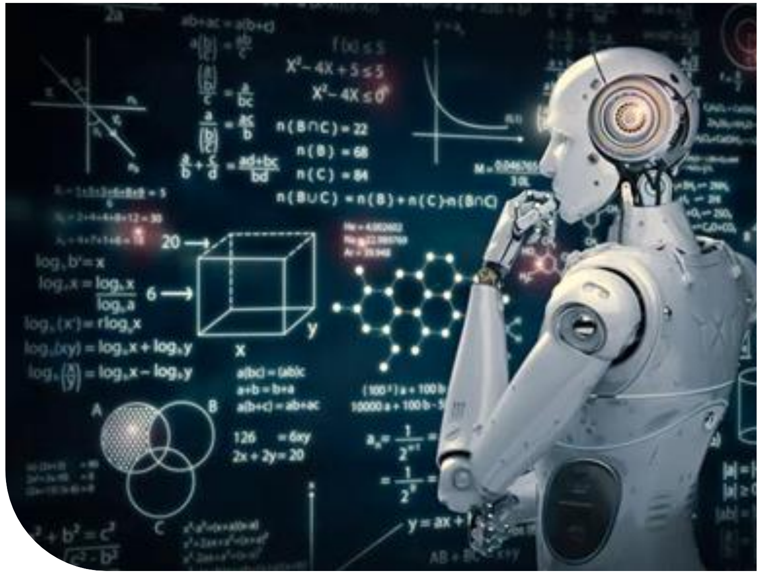
AI-Based Business Information Systems

Generative AI



Prof. Dr. Ulrich Gnewuch

Course Organization

Lecture

AI-Enabled Business Capabilities

- AI-Enabled Innovation
- AI-Enabled Insights & Decisions
- AI-Enabled Engagement
- AI-Enabled Automation

AI Technologies & Trends

- AI Ethics & Ethical AI
- Generative AI
- Explainable AI
- Conversational AI

Foundations

- Introduction to AI in Business & Information Systems
- Design & Management of AI-Based Information Systems

Exercise

- **Exercise 4:** Generative AI & Innovation
- **Exercise 3:** Explainable AI Techniques
- **Exercise 2:** Human-Centered Chatbot Design
- **Exercise 1:** Robotic Process Automation Case Study

Industry Talk
ZF Group

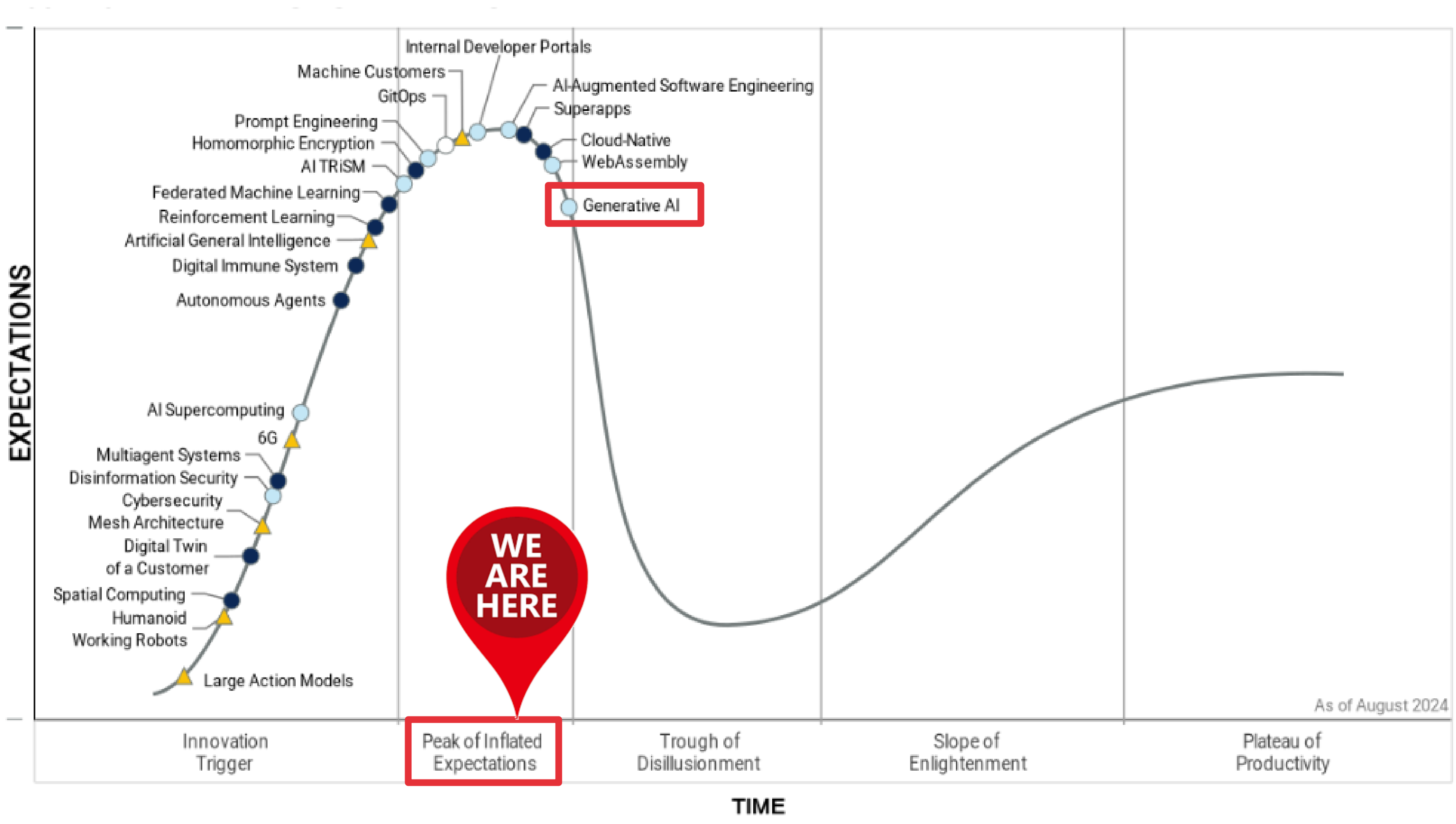
RECAP FROM LAST LECTURE:

- Which innovation type best describes the replacement of human workers with assembly robots in a car factory?
- Which statements about 'creativity' are true?
- What are key challenges associated with the use of AI in innovation?

Learning Goals

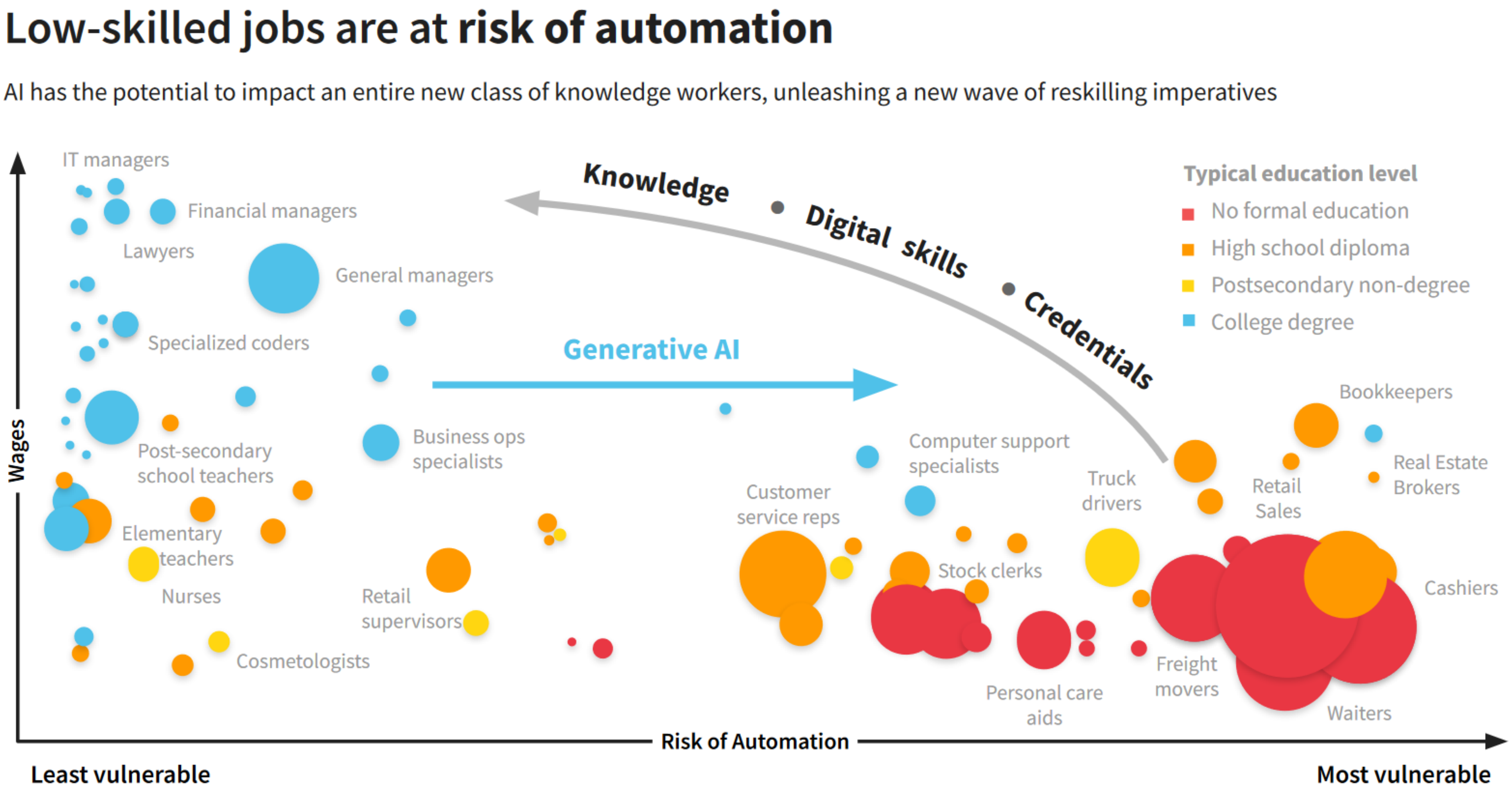
- Explain what generative AI is and how it differs from “traditional” AI
- Describe how foundation models can be adapted to specific tasks and domains
- Explain key prompting strategies and apply them effectively
- Discuss the challenges and risks of generative AI

Generative AI Hype



Gartner 2024

Potential Impact of Generative AI



https://s27.q4cdn.com/928340662/files/doc_financials/2023/q1/COUR_Presentation_Q1-2023.pdf

Definition: Generative AI

UNIVERSITY OF PASSAU

Generative AI is a type of AI that generates seemingly new, meaningful content, such as text, images, audio, or video, from training data.

- Generative AI relies on statistical models created from existing content
- When given a prompt, generative AI uses this statistical model to generate new content

Claude 3
Gemini
Meta AI

Feuerriegel et al. 2024

Generative vs. “Traditional” AI

Generative vs. Traditional AI

AI

Traditional AI

Generative AI

- Generates new data that is similar to the data it was trained on
- Understands the distribution of data and how likely a given example is

Goal: Generate new data instances

Predictive AI

- Typically trained on a dataset of labeled data (→ *supervised ML*)
- Learns the relationship between the features of the data instances and the labels

Goal: Make predictions about future events

Discriminative AI

- Typically trained on a dataset of unlabeled data (→ *unsupervised ML*)
- Learns the boundary between different classes or categories in a dataset

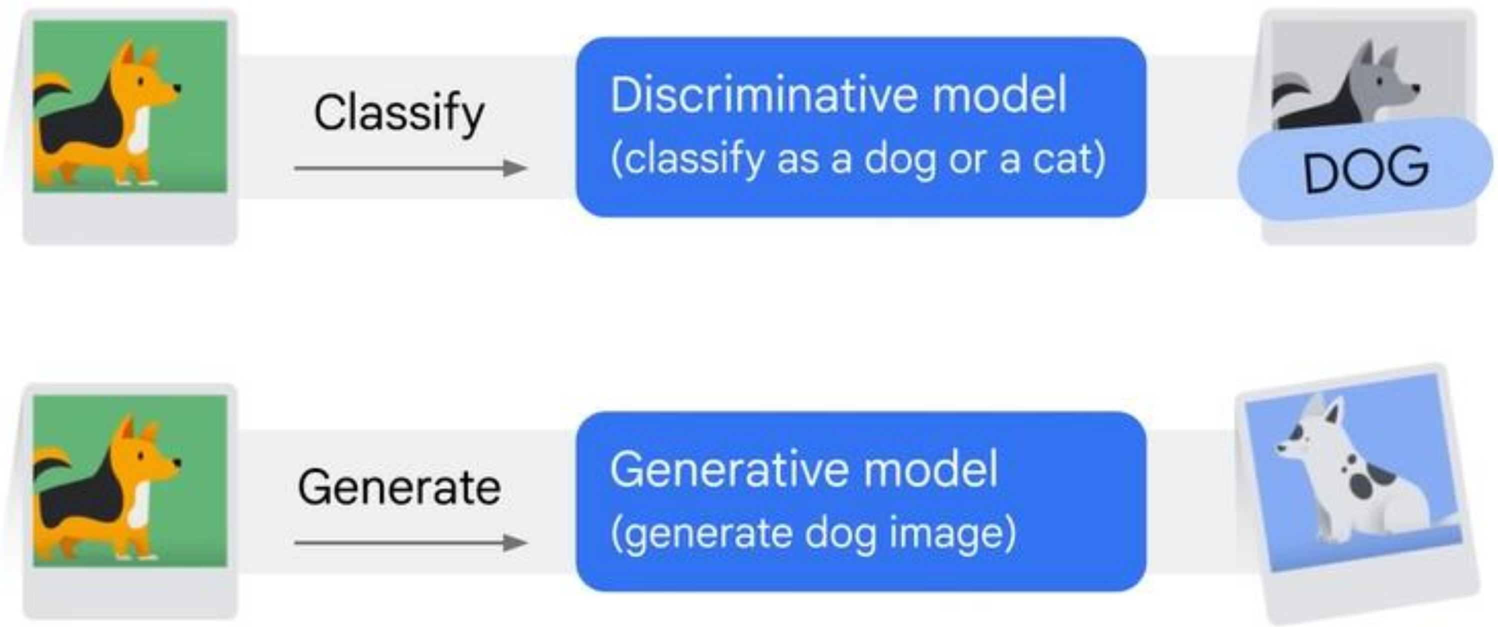
Goal: Discriminate between different kinds of data instances

Feuerriegel et al. 2024

Generative vs. Traditional AI: Example

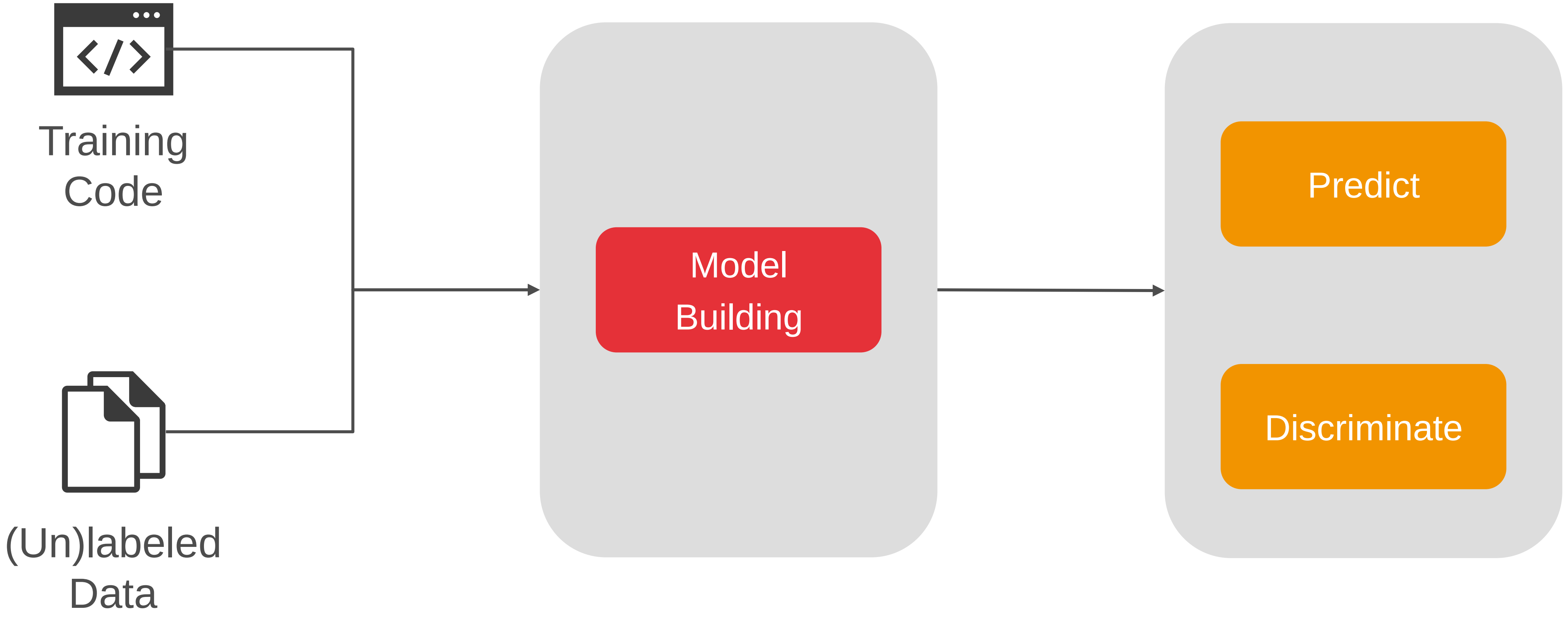
Traditional AI
(Discriminative)

Generative AI



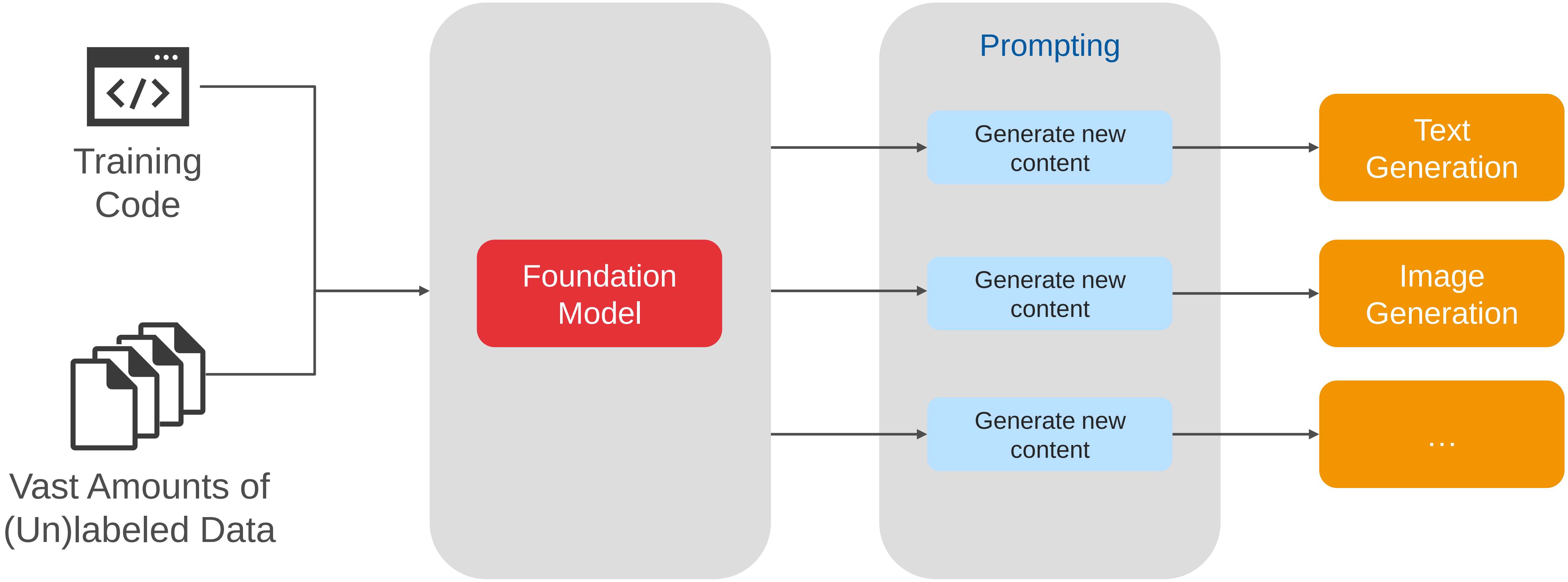
Google 2023

Traditional AI Process



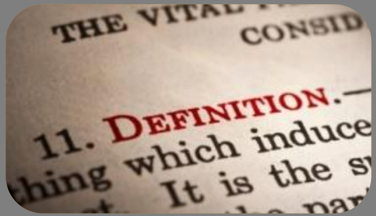
Based on: Google 2023

Generative AI Process



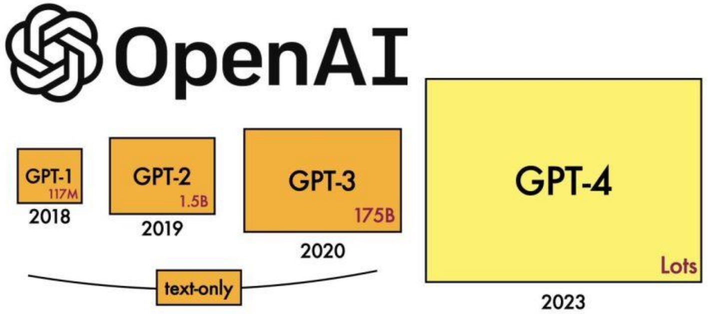
Based on: Google 2023

Foundation Models



A foundation model is an AI model trained with a large amount of data using self-supervision at scale, displays significant generality, and is capable of performing a wide range of distinct tasks.

- These models are sometimes referred to as ‘general-purpose AI (GPAI) models’ or ‘large X models’ (LxM)
- Example: GPT-4 is the fourth generation of OpenAI’s foundation model



EU AI Act (Article 3, 63)

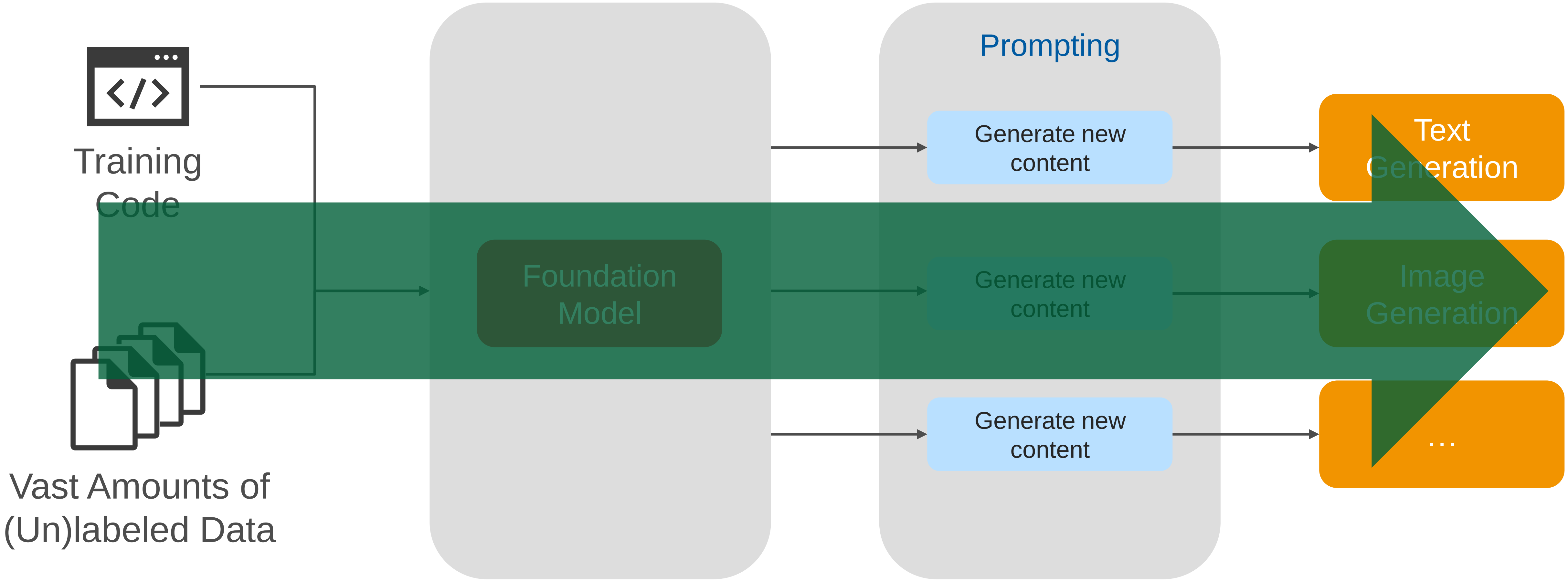
Foundation Models: Examples

Foundation Models

Foundation Language Models	Foundation Image Models	Multimodal Foundation Models
Trained on extensive textual data, learning to predict the next word in a sentence and generate coherent text	Trained on large datasets of visual data, learning to generate and modify images	Trained on multiple modalities simultaneously (e.g., image-text pairs)
Examples: BERT, GPT-3, Llama	Examples: DALL-E, Stable Diffusion	Examples: PaLM-E, GPT-4

Bommasani et al. 2021

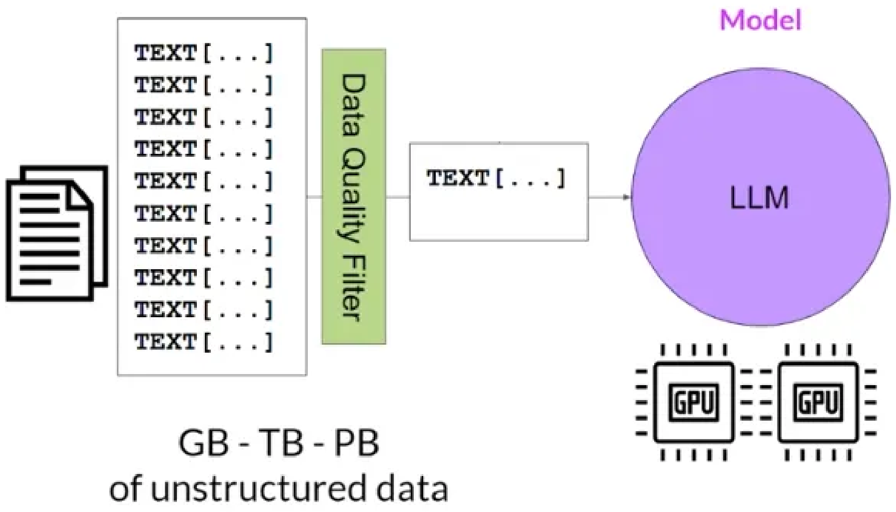
Generative AI: How does it work?



Based on: Google 2023

A Simplified View of Generative AI (1)

- From a very large training dataset, a model learns the **statistical distributions of tokens**
- Tokens can be “chunks” of words, punctuation marks, pixels, etc.



Token String	Token ID	Embedding / Vector Representation
'_The'	37	[-0.0513, -0.0584, 0.0230, ...]
'_teacher'	3145	[-0.0335, 0.0167, 0.0484, ...]
'_teaches'	11749	[-0.0151, -0.0516, 0.0309, ...]
'_the'	8	[-0.0498, -0.0428, 0.0275, ...]
'_student'	1236	[-0.0460, 0.0031, 0.0545, ...]
...	...	...

Vector of probabilities

Faklaris 2023, Stollnitz 2023

A Simplified View of Generative AI (2)

- When given a prompt, the model converts the prompt into tokens and then **analyzes what is likely to come next**, based on the tokens in its own dataset

“Here’s a fragment of text.

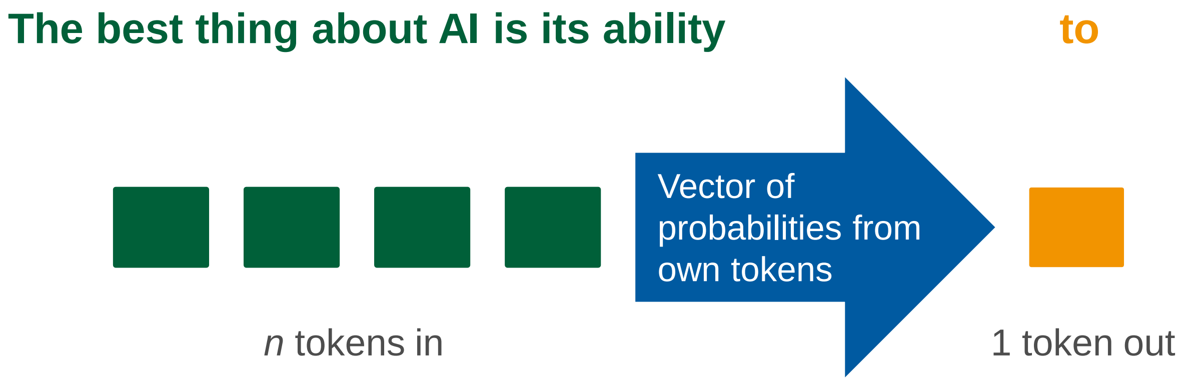
Tell me how this fragment might <continue on in this language, or suggest a particular image>.

According to your model of the statistics of <human language, or human-handled images>, what <words, or pixels> are likely to come next?”

Faklaris 2023, Stollnitz 2023

A Simplified View of Generative AI (3)

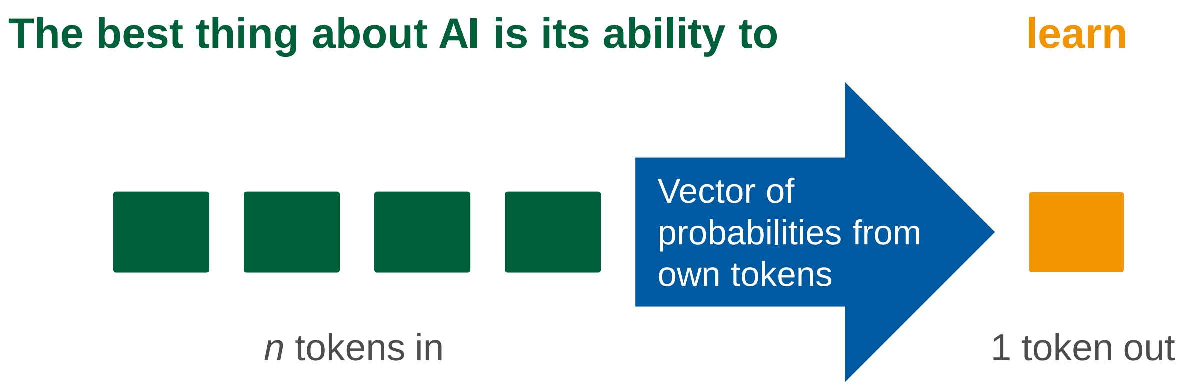
- When given a prompt, the model converts the prompt into tokens and then analyzes what is likely to come next, based on the tokens in its own dataset
- It then **generates a tokenized output**



Faklaris 2023, Stollnitz 2023

A Simplified View of Generative AI (4)

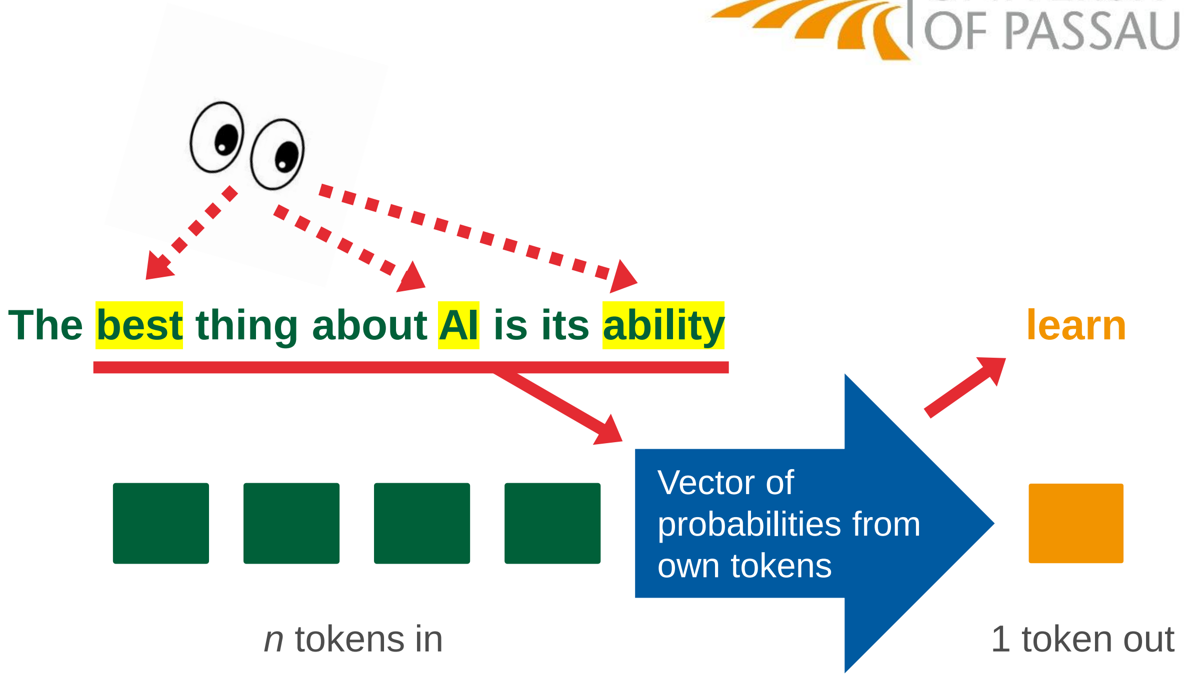
- With each output, the model **keeps reanalyzing** the probabilities to decide the next token



Faklaris 2023, Stollnitz 2023

A Simplified View of Generative AI (5)

- Transformers (the “T in “GPT”) know how to direct attention to specific parts of the input to guide their selection of the next token



- The transformer architecture was a major breakthrough in the field of natural language processing (see paper “*Attention is all you need*”; Vaswani et al. 2017)

Faklaris 2023, Stollnitz 2023

A Simplified View of Generative AI (6)

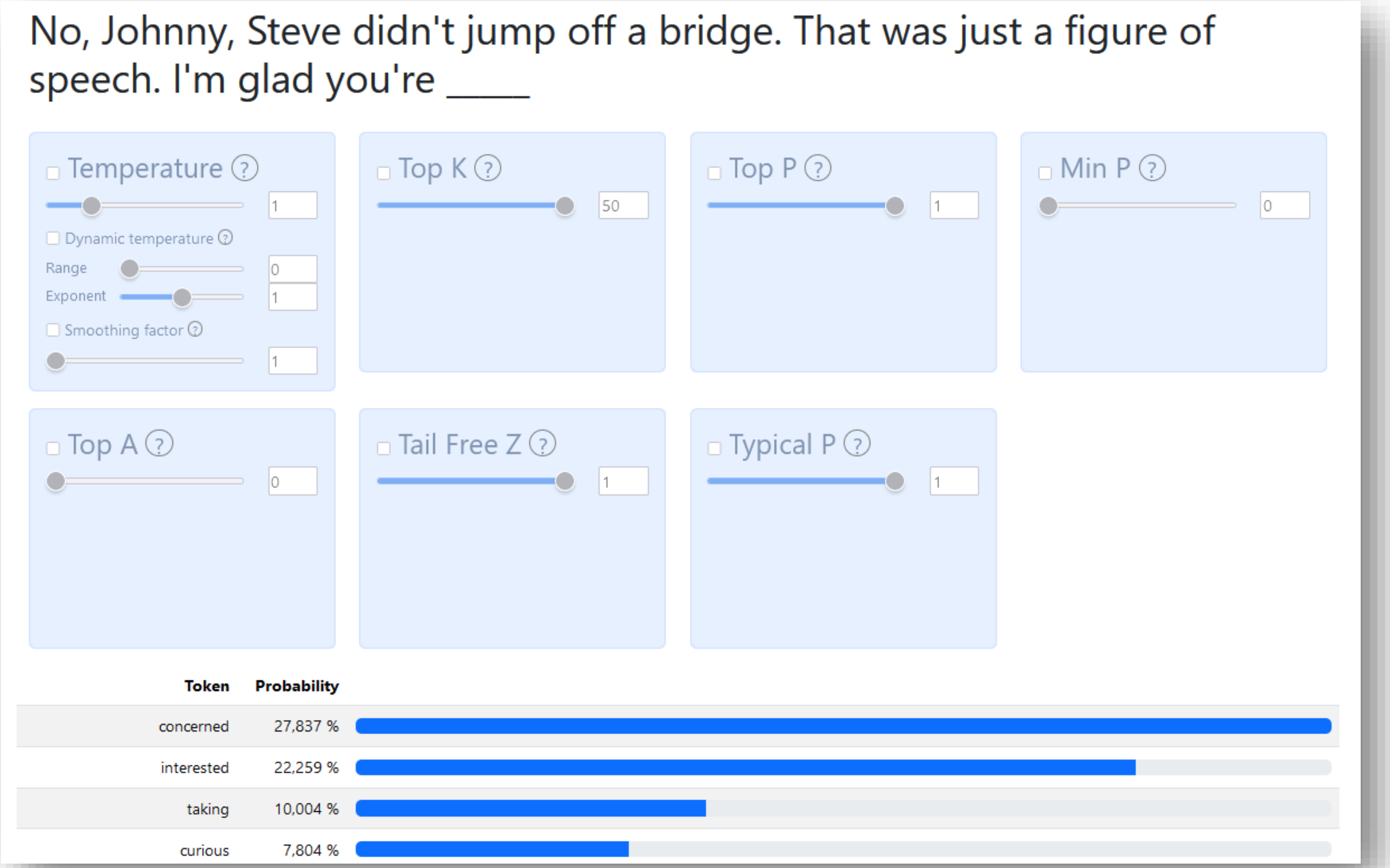
- The model can provide different tokens to the same inputs:

learn	4.5%
predict	3.5%
make	3.2%
understand	3.1%
do	2.9%

The best thing about AI is its ability to **predict**

Faklaris 2023, Stollnitz 2023

Interactive LLM Output Demo



https://chatbotresearch.org/llm-sampling/index.xhtml

- LLMs can use different strategies to choose the next token
- This process is known as “sampling“
- Example based on Meta’s LLaMA model
  - Source: https://github.com/Artefact2/llm-sampling

LLM Demo & Token Sampling

Open the interactive LLM demo, explore how tokens are sampled under different parameter settings, and answer the following questions:

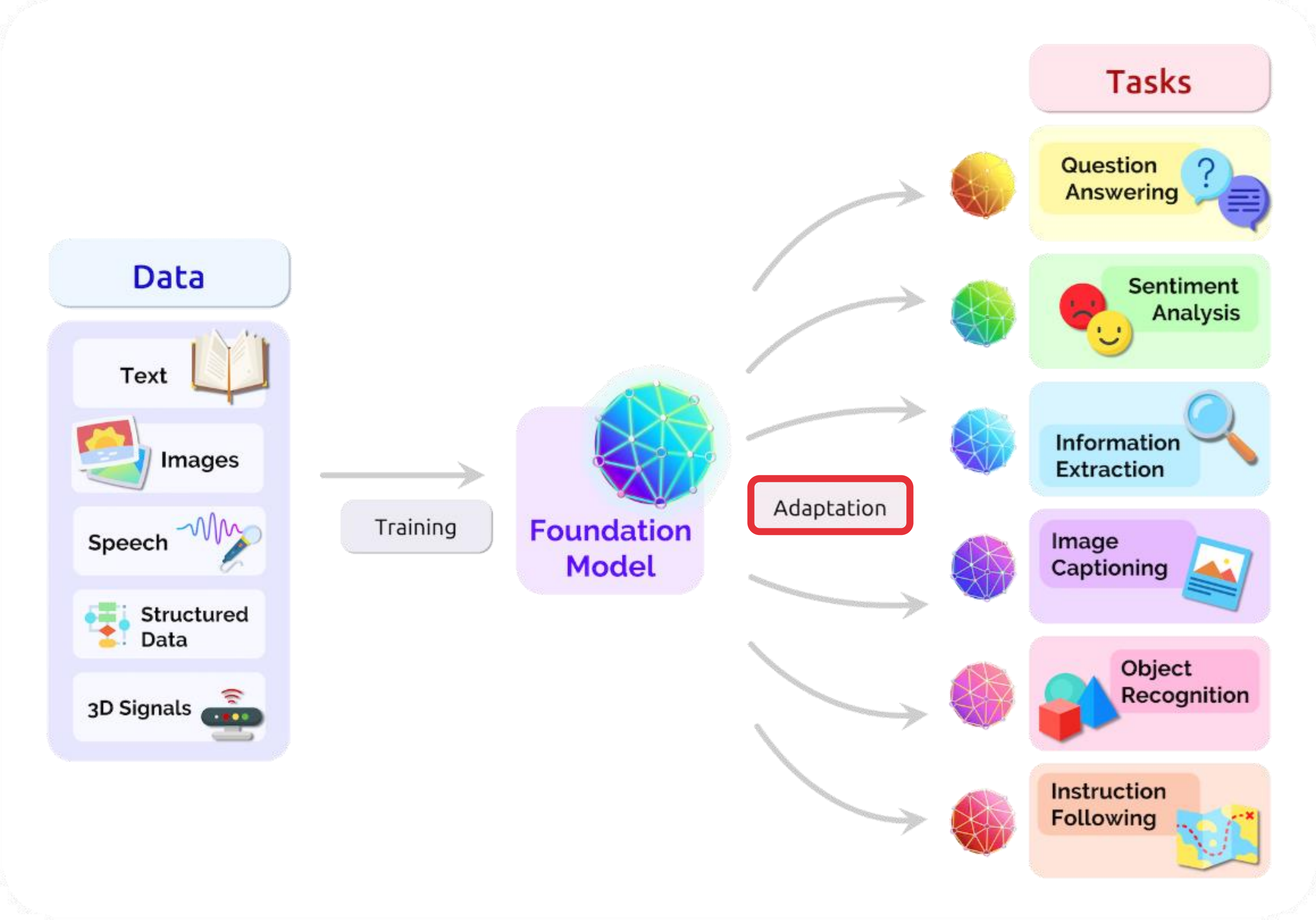
1. What happens when the temperature is zero? What happens when it is set to its maximum value?
2. Why do you think the temperature is typically set to a value between 0 and 1 by OpenAI, Anthropic, and other LLM providers?

→ Discuss these questions with a partner for **~5 minutes** and be ready to share your answers

https://shorturl.at/Midca

Adapting Foundation Models to Specific Tasks and Domains

Building Applications on Top of Foundation Models



Bommasani et al. 2021

Adapting Foundation Models to Specific Tasks and Domains

Adaptation

Fine-tuning	Prompt engineering	Retrieval augmented generation (RAG)
Taking a foundation model and further training it on a smaller, specific dataset to refine its capabilities and improve performance in a specific task or domain	Asking a pre-trained (foundation) model to perform a specific task by giving it a task description and relevant input data (=prompt)	Augmenting a pre-trained (foundation) model with an external knowledge base, giving it the ability to access domain-specific information

- Fine-tuning is a powerful technique but requires sufficiently large annotated training data which is often costly and hard to obtain
- In practice, combinations of different techniques are often used
- New techniques are constantly being developed (e.g., parameter-efficient fine-tuning; PEFT)

Bommasani et al. 2021

Fine-Tuning: Example

Introducing BloombergGPT, Bloomberg's 50-billion parameter large language model, purpose-built from scratch for finance

NEW YORK – Bloomberg today released a research paper detailing the development of BloombergGPTTM, a new large-scale generative artificial intelligence (AI) model. This large language model (LLM) has been specifically trained on a wide range of financial data to support a diverse set of natural language processing (NLP) tasks within the financial industry.

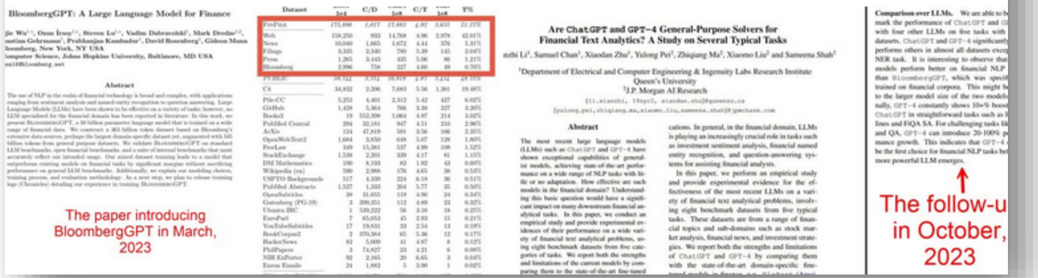
https://www.bloomberg.com/company/press/bloomberggpt-50-billion-parameter-llm-tuned-finance/

Ethan Mollick • 2nd
Associate Professor at The Wharton School. Author of Co-Intelli...
8mo •

This remains one of the most consequential experiments in AI: Bloo[...] over $10M training a GPT-3.5 class AI on their own financial data last y[...]

...only to find that GPT-4 8k, the AI available to billions of people around the world, and without specialized finance training, beat it on almost all finance tasks!

There was a moment that we thought proprietary data would let organizations train specialized AIs that could compete with frontier models. It turns out that probably isn't going to happen. The largest frontier models are just much better at most complex tasks than smaller models.



https://www.linkedin.com/posts/emollick_this-remains-one-of-the-most-consequential-activity-7176398465004896256-Qjx-/

Prompting

Prompting is the process of using carefully crafted phrases or templates to help a pre-trained (foundation) model accomplish a specific, downstream task.

- Targeted prompting can optimize the performance and relevance of model outputs to produce high-quality results
- The quality of the prompt determines the quality of the output!

Brown et al. 2020

Prompting Techniques & Examples

Prompting Technique	Description	Example
Zero-shot promoting	Relying solely on a model's pre-trained information to answer a given prompt	• *Write a summary of the following text: […]* • *Generate ideas for a birthday present.*
Few-shot prompting	Providing examples in the prompt, giving the model more context to improve its performance	*Country: Germany* *Capital city: Berlin* *Country: Argentina* *Capital city:*
Chain-of-thought prompting	Nudging a model to produce intermediate reasoning steps, thus improving the results on complex reasoning tasks	Two ways: • Use few-shot prompting by illustrating examples with detailed answers to questions • Add instructions such as "*Let's think step by step*"

https://huggingface.co/docs/transformers/main/tasks/prompting#advanced-prompting-techniques

Challenges in Prompt Engineering

- Prompting can seem easy, but designing effective prompts is challenging, especially for non-experts

- Common challenges:
  - Expectations stemming from human-to-human communication (e.g., politeness)
  - Tendency to overgeneralize
  - Biases towards giving instruction over depicting examples

Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts

J.D. Zamfirescu-Pereira
zamfi@berkeley.edu
UC Berkeley
Berkeley, CA, USA

Richmond Wong
rwong34@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

Bjoern Hartmann
bjoern@eecs.berkeley.edu
UC Berkeley
Berkeley, CA, USA

Qian Yang
qianyang@cornell.edu
Cornell University
Ithaca, NY, USA

ABSTRACT

Pre-trained large language models ("LLMs") like GPT-3 can engage in fluent, multi-turn instruction-taking out-of-the-box, making them attractive materials for designing natural language interactions. Using natural language to steer LLM outputs ("prompting") has emerged as an important design technique potentially accessible to non-AI-experts. Crafting effective prompts can be challenging, however, and prompt-based interactions are brittle. Here, we explore whether non-AI-experts can successfully engage in "end-user prompt engineering" using a design probe—a prototype LLM-based chatbot design tool supporting development and systematic evaluation of prompting strategies. Ultimately, our probe participants explored prompt designs opportunistically, not systematically, and struggled in ways echoing end-user programming systems and interactive machine learning systems. Expectations stemming from human-to-human instructional experiences, and a tendency to overgeneralize, were barriers to effective prompt design. These findings have implications for non-AI-expert-facing LLM-based tool design and for improving LLM-and-prompt literacy among programmers and the public, and present opportunities for further research.

CCS CONCEPTS

• Human-centered computing → Empirical studies in interaction design; • Computing methodologies → Natural language processing.

KEYWORDS

language models, end-users, design tools

ACM Reference Format:
J.D. Zamfirescu-Pereira, Richmond Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 21 pages. https://doi.org/10.1145/3544548.3581388

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04.
https://doi.org/10.1145/3544548.3581388

1 INTRODUCTION

The idea of instructing computers in natural language has fascinated researchers for decades, as it promises to make the power of computing more customizable and accessible to people without programming training [4]. The combination of pre-trained large language models (LLMs) and prompts brought renewed excitement to this vision. Recent pre-trained LLMs (e.g., GPT-3 [8], ChatGPT [1]) can engage in fluent, multi-turn conversations out-of-the-box, substantially lowering the data and programming-skill barriers to creating passable conversational user experiences [7]. People can improve LLM outputs by prepending *prompts*—textual instructions and examples of their desired interactions—to LLM inputs. Prompts directly bias the model towards generating the desired outputs, raising the ceiling of what conversational UX is achievable for non-AI experts. In the past two years, social media platforms have witnessed an explosion of posts showing the results of lay peoples' experimentation with LLMs for question answering, creative dialogue writing, writing code, and more. This excitement around LLMs and prompting is propelling a rapidly growing set of LLM-powered applications [23] and prompt design tools [3, 20, 32].

Yet despite widespread excitement, surprisingly little is known about how non-experts intuitively approach designing prompts with LLM-and-prompt-based tools, and how effective they are in doing so. While prompting LLMs can appear effortless, designing effective prompting strategies requires identifying the contexts in which these LLMs' errors arise, devising prompting strategies to overcome them, and systematically assessing those strategies' effectiveness. These tasks fall on so-called "prompt engineers"—the designers, domain experts, and any other end-user or professional attempting to improve an LLM's output—and are challenging tasks even for LLM experts, as well as topics of ongoing research in Natural Language Processing (NLP) [7, 30, 42]. Prompt design tools to date have focused on supporting professional programmers [45] and NLP practitioners [42], rather than non-AI experts, non-programmers, and other potential end-users of these systems.

In this work, we investigate how non-AI-experts intuitively approach prompt design when designing LLM-based chatbots, with an eye towards how non-AI-expert-facing design tools might help. Specifically, we investigate these questions in the context of designing an instructional chatbot, that is, a chatbot that walks the user through an activity (e.g., cooking a recipe, fixing a wifi connection) while answering user questions

Zamfirescu-Pereira et al. 2023

Prompt Engineering Guidelines

Company	Guideline	Link
OpenAI	OpenAI's Prompt engineering strategies	https://platform.openai.com/docs/guides/prompt-engineering
Hugging Face	Best practices of LLM prompting	https://huggingface.co/docs/transformers/main/tasks/prompting#best-practices-of-llm-prompting
Google	Tips to enhance your prompt-engineering abilities	https://cloud.google.com/blog/products/application-development/five-best-practices-for-prompt-engineering
Midjourney	Prompting Notes	https://docs.midjourney.com/docs/prompts

Prompt Engineering Guidelines: Example

- Start with a simple and short prompt, and iterate from there.
- Put the instructions at the beginning of the prompt, or at the very end. When working with large context, models apply various optimizations to prevent Attention complexity from scaling quadratically. This may make a model more attentive to the beginning or end of a prompt than the middle.
- Clearly separate instructions from the text they apply to - more on this in the next section.
- Be specific and descriptive about the task and the desired outcome - its format, length, style, language, etc.
- Avoid ambiguous descriptions and instructions.
- Favor instructions that say “what to do” instead of those that say “what not to do”.
- “Lead” the output in the right direction by writing the first word (or even begin the first sentence for the model).
- Use advanced techniques like Few-shot prompting and Chain-of-thought
- Test your prompts with different models to assess their robustness.
- Version and track the performance of your prompts.

Hugging Face

https://huggingface.co/docs/transformers/main/tasks/prompting#best-practices-of-llm-prompting

Bad vs. Good Prompts: Examples

“Create three ideas for social media content”

- Not enough context (e.g., which social media platform)
- Not enough details about the content (e.g., topics, audience)

“Compose a thought-provoking LinkedIn post about the future of remote work. Please write in the first person, highlight the advantages of remote collaboration, mention industry trends, and write for young managers working in the tech sector.”

https://www.plannthat.com/good-vs-bad-ai-prompts/

Prompt Engineering

Here is another example of a bad prompt:
"*Create a social media post about the lecture I attended at university today*"

Please give an example of a better prompt.

Challenges and Risks of Generative AI

A hallucination is when a language model generates seemingly realistic responses that are untrue, nonsensical, or unfaithful to the provided source input.

- LLMs hallucinate because they operate mainly on a probabilistic level by putting one probable word in front of another without any deeper capacity for comprehension → **"Stochastic Parrots"**
- Like a parrot, LLMs excel at regurgitating learned content without knowing the meaning of their responses

Hannigan et al. 2024, Bender et al. 2021

From Hallucinations to “Botshit”

Botshit is AI-generated content that is not grounded in truth (e.g., hallucinations) and is then uncritically used by a human for communication and decision-making tasks.

- Humans might use untrue material created by generative AI in an uncritical and thoughtless way
- This “botshit” makes it harder for people to know what is true and false in the world
- Problems arise when the outputs have important consequences and cannot easily be verified

Hannigan et al. 2024; McCarthy et al. 2024

Botshit: Examples

New York lawyers sanctioned for using fake ChatGPT cases in legal brief

NEW YORK, June 22 (Reuters) - A U.S. judge on Thursday imposed sanctions on two New York lawyers who submitted a legal brief that included six fictitious case citations generated by an artificial intelligence chatbot, ChatGPT.

https://www.reuters.com/legal/new-york-lawyers-sanctioned-using-fake-chatgpt-cases-legal-brief-2023-06-22/

Sports Illustrated Published Articles by Fake, AI-Generated Writers

There was nothing in Drew Ortiz's author biography at *Sports Illustrated* to suggest that he was anything other than human.

"Drew has spent much of his life outdoors, and is excited to guide you through his never-ending list of the best products to keep you from falling to the perils of nature," it read. "Nowadays, there is rarely a weekend that goes by where Drew isn't out camping, hiking, or just back on his parents' farm."

The only problem? Outside of *Sports Illustrated*, Drew Ortiz doesn't seem to exist. He has no social media presence and no publishing history. And even more strangely, his profile photo on *Sports Illustrated* is for sale on a website that sells AI-generated headshots, where he's described as "neutral white young-adult male with short brown hair and blue eyes."

https://futurism.com/sports-illustrated-ai-generated-writers

McCarthy et al. 2024

- The data used to train AI generative AI models often reflect human biases:
  - Cultural biases
  - Gender biases
  - Racial biases
  - Socioeconomic biases
  - …
- Like any AI, these models inherit those biases

Generative AI: UNESCO study reveals alarming evidence of regressive gender stereotypes

Ahead of the International Women's Day, a UNESCO study revealed worrying tendencies in Large Language models (LLM) to produce gender bias, as well as homophobia and racial stereotyping. Women were described as working in domestic roles far more often than men ¬– four times as often by one model – and were frequently associated with words like “home”, “family” and “children”, while male names were linked to “business”, “executive”, “salary”, and “career”.

https://www.unesco.org/en/articles/generative-ai-unesco-study-reveals-alarming-evidence-regressive-gender-stereotypes

Feuerriegel et al. 2024

Copyright Violations

Getty Images lawsuit says Stability AI misused photos to train AI

(Reuters) - Stock photo provider Getty Images has sued artificial intelligence company Stability AI Inc, accusing it in a lawsuit made public on Monday of misusing more than 12 million Getty photos to train its Stable Diffusion AI image-generation system.

The lawsuit, filed in Delaware federal court, follows a separate Getty case against Stability in the United Kingdom and a related class-action complaint filed by artists in California against Stability and other companies in the fast-growing field of generative AI.

https://www.reuters.com/legal/getty-images-lawsuit-says-stability-ai-misused-photos-train-ai-2023-02-06/

The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work

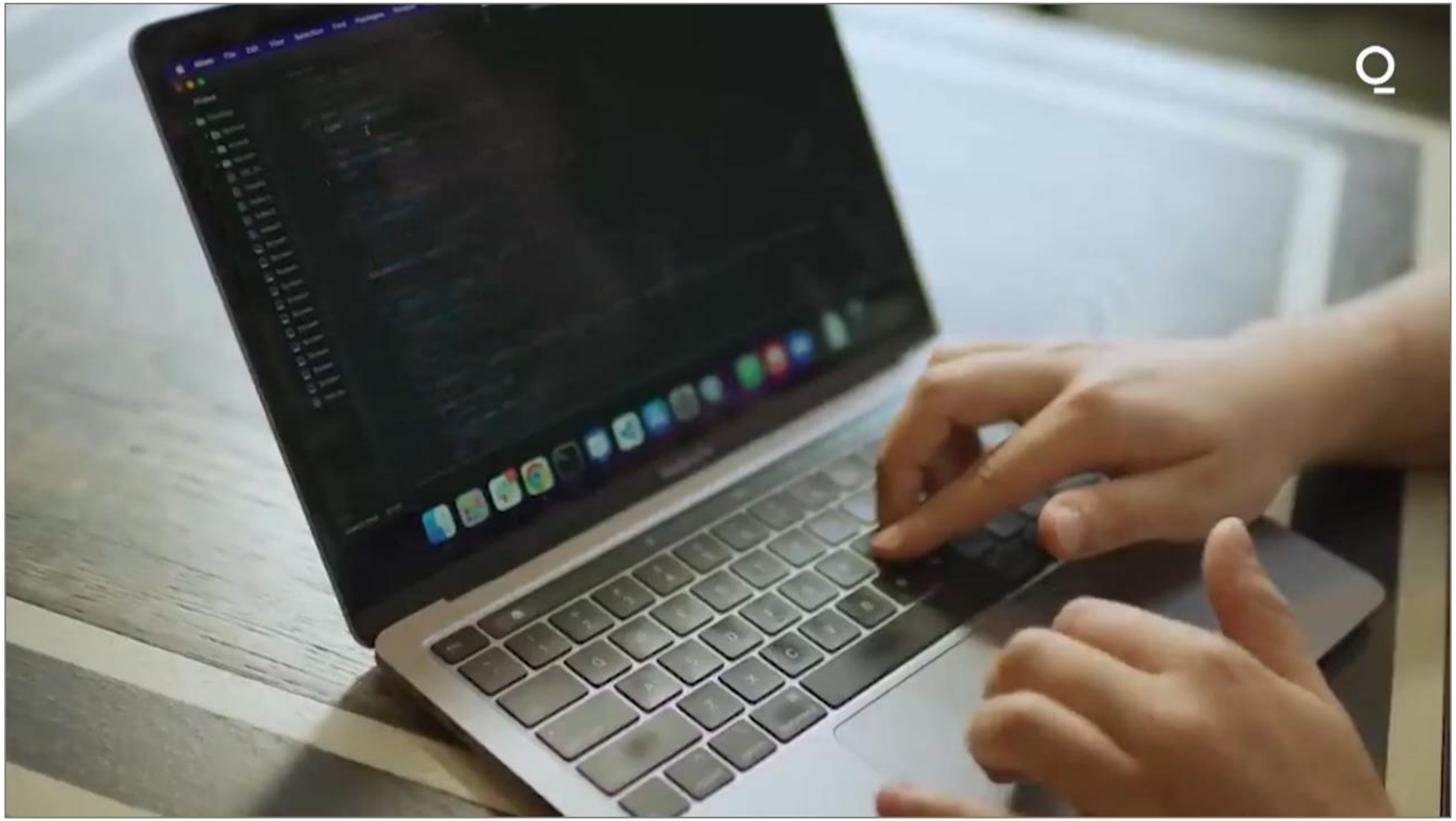
The New York Times sued OpenAI and Microsoft for copyright infringement on Wednesday, opening a new front in the increasingly intense legal battle over the unauthorized use of published work to train artificial intelligence technologies.

The Times is the first major American media organization to sue the companies, the creators of ChatGPT and other popular A.I. platforms, over copyright issues associated with its written works. The lawsuit, filed in Federal District Court in Manhattan, contends that millions of articles published by The Times were used to train automated chatbots that now compete with the news outlet as a source of reliable information.

https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html

Appel et al. 2023

Environmental Impact



https://www.youtube.com/watch?v=LDfbGk9dpWw

When Not to Use Generative AI

"If you have a hammer, everything looks like a nail"

Law of the Instrument or Maslow's Hammer (1966)

Use-case family	Generative models' current usefulness	Example use cases
Prediction/ forecasting	Low	Risk prediction, customer churn prediction, sales/ demand forecasting
Decision intelligence	Low	Decision support, augmentation, automation
Segmentation/ classification	Medium	Clustering, customer segmentation, object classification
Recommendation systems	Medium	Recommendation engine, personalized advice, next best action
Content generation	High	Text generation, image and video generation, synthetic data
Conversational user interfaces	High	Virtual assistant, chatbot, digital worker

Source: Gartner

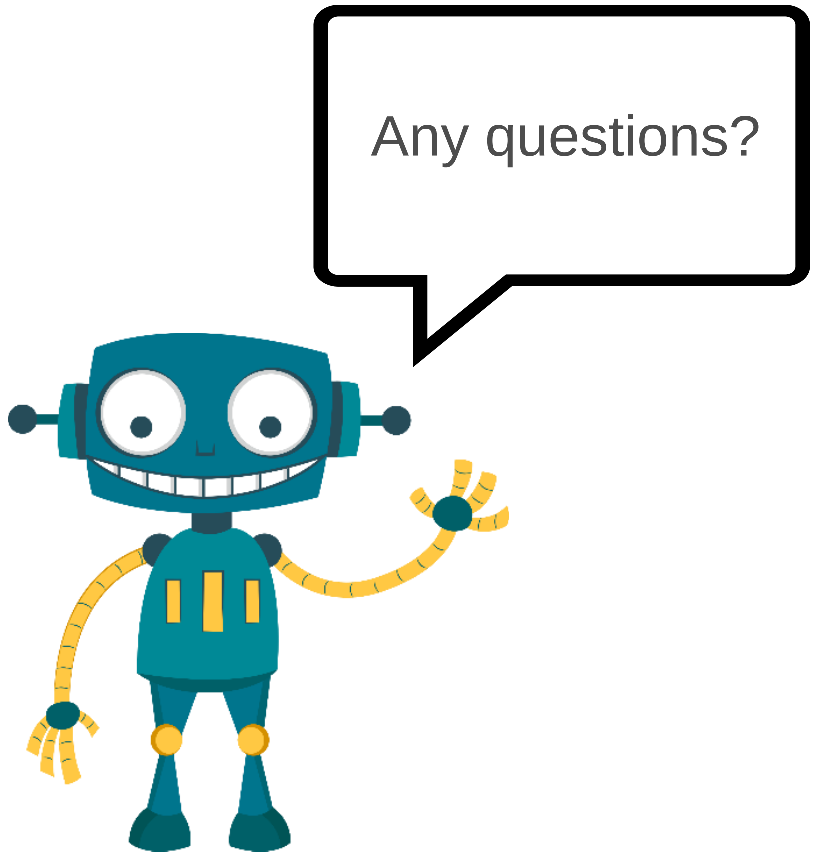
Gartner

https://www.gartner.com/en/articles/when-not-to-use-generative-ai

Key Takeaways From This Lecture

- The primary way in which generative AI differs from traditional AI is that it can generate new data instances (new text, images, videos, etc.)
- Foundation models are at the core of generative AI
- Foundation models can be adapted to a wide range of specific tasks, for example, via fine-tuning or prompting
- There are different prompting techniques (zero-shot, few-shot, chain-of-thought prompting) and a growing number of guidelines for effective prompt design
- Generative AI offers many benefits but also creates several challenges such as hallucinations, biases, and copyright issues

Thank you for your attention!



References

Appel, G., Neelbauer, J., & Schweidel, D. A. (2023). Generative AI has an intellectual property problem. Harvard Business Review. https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big? 🦜. In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency (pp. 610-623).

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33, 1877-1901.

EU AI Act. (2024). http://data.europa.eu/eli/reg/2024/1689/oj

Faklaris, Cori. (2023). An Introduction to Generative AI. https://corifaklaris.com/files/genAIintro.pdf

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. Business & Information Systems Engineering. 66, 111–126 (2024). https://doi.org/10.1007/s12599-023-00834-7

Gartner 2024. https://www.gartner.com/en/newsroom/press-releases/2024-08-21-gartner-2024-hype-cycle-for-emerging-technologies-highlights-developer-productivity-total-experience-ai-and-security

Google. (2023). Introduction to Generative AI. https://www.youtube.com/watch?v=G2fqAlgmoPo

Hannigan, T., McCarthy, I. P., & Spicer, A. (2023). Beware of botshit: How to manage the epistemic risks of generative chatbots. Business Horizons, 67(5), 471-486.

Maslow, A. H. (1966). The Psychology of Science. Harper & Row.

McCarthy, I. P., Hannigan, T., & Spicer, A. (2023). The Risks of Botshit. Harvard Business Review. https://hbr.org/2024/07/the-risks-of-botshit

Stollnitz, Bea. (2023). https://bea.stollnitz.com/blog/how-gpt-works/

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

Zamfirescu-Pereira, J. D., Wong, R. Y., Hartmann, B., & Yang, Q. (2023, April). Why Johnny can't prompt: how non-AI experts try (and fail) to design LLM prompts. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (pp. 1-21).