JOHNS HOPKINS
CAREY BUSINESS SCHOOL

Lecture 1

BU.330.760 Generative AI for Business

Minghong Xu, PhD.
Associate Professor

Today's Agenda

- Course introduction
  - Requirements and evaluation
- Introduction to generative AI
- Deep learning and NLP review

Instructor Bio

- Minghong Xu, Ph.D.
- Email: xu.minghong@jhu.edu
  - Best way to connect

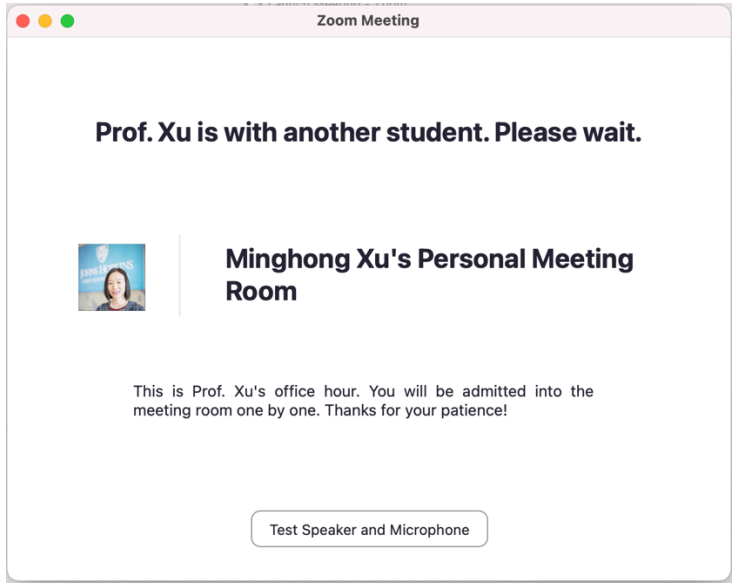
- Affiliation: Center for Digital Health and Artificial Intelligence (CDHAI)
- Research and teaching: Artificial Intelligence and Big Data

Office Hours

- Office hours (in EST):
  - Wednesdays 11:30am-1:30pm, and by appointment
- Virtually via Zoom:

https://jhucarey.zoom.us/j/4658557490?pwd=Y2NvL0M0RjdFb3RpUjlVOFBSSkFLZz09

- Waiting room enabled, admit one by one
- First come first served
  - unless book ahead of time
- Do not hesitate to reach out



Course Overview

- Nexus award-winning
  - https://hub.jhu.edu/2024/05/21/nexus-awards-program-hopkins-washington-dc/
- Section 51
  - Primarily reserved for BARM, no deep learning background assumed
- Section 52
  - Primarily reserved for IS, or have learnt deep learning
  - Continuation of AI Essentials course

Experimental Course

- Generative AI technologies are still evolving rapidly
- Experimentation with new materials in this course
  - Please be flexible as and when topics or materials are revised or modified
- Hands-on learning
- Both technical and business
  - Newest developments
  - Strategies for managing Generative AI, exploring questions such as:
    - *What can we learn from Google?*
    - *How are GenAI products similar to or different from other product?*
    - *How far away are we from AGI?*

Two Examples Today

- Technical perspective
  - Nvidia GTC 2025, March 17-21, 2025
- Business perspective
  - The state of AI, McKinsey, March 12, 2025

Classroom Policy

- **Academic Integrity (Honor code is enforced)**
  - No cheat, no copy
- No cellphones, or excessive talking
  - Questions during lecture should be addressed to me
- Avoid inappropriate content
- Rectifying your score
  - Discuss with TA ASAP, within 1 week after scores are posted
  - Request after 1 week may not be entertained

Requirements

Assignment	Weight
Attendance and participation in class discussion	5%
Homework	40%
Project	30%
Final Exam	25%
Total	100%

Class Participation (5%)

- Class participation is an important part of learning
- Highly interactive
- Discussion questions every week
- Key to success: participate!

Assignments (40%)

- 4 weekly assignments

- Week 2: attention mechanism
- Week 3: prompt engineering
- Week 5: chatbot using AWS
- Week 6: image generation

Project (30%)

- Agentic AI business solution
- 4 students per group
- Each group need to work with a collaborator
  - Collaborator provide you the business case
  - MBA students for negotiation bots
  - You can also find your own collaborator
- Rubrics on Canvas

Final Exam (25%)

- In week 8, closed-book
- Administered via **Respondus LockDown Browser**

- Install LockDown Browser from https://download.respondus.com/lockdown/download.php?id=123533816

Tentative Course Calendar

Week	Weekly Objectives/Topics	Hands-on Learning	Assignments
1	Introduction to Generative AI Deep Learning and NLP Review		
2	Foundations of Text Generation Generative AI Value Chain	Attention Mechanism	HW 1 release
3	Large Language Models and Strategies: Prompt Engineering and “Reasoning”	Prompt Engineering	HW 2 release
4	Agentic AI and Business Cases LLM Ethical Issues		Business case kickoff
5	LLM Applications in Business Domains: BloombergGPT and RAG	Chatbot using AWS	HW 3 release
6	Foundations of Image Generation Dark Side of Gen AI	Image Generation	HW 4 release
7	Responsible Gen AI and Looking Ahead Student Business Case Presentation		
8	Final exam		

Final Grades

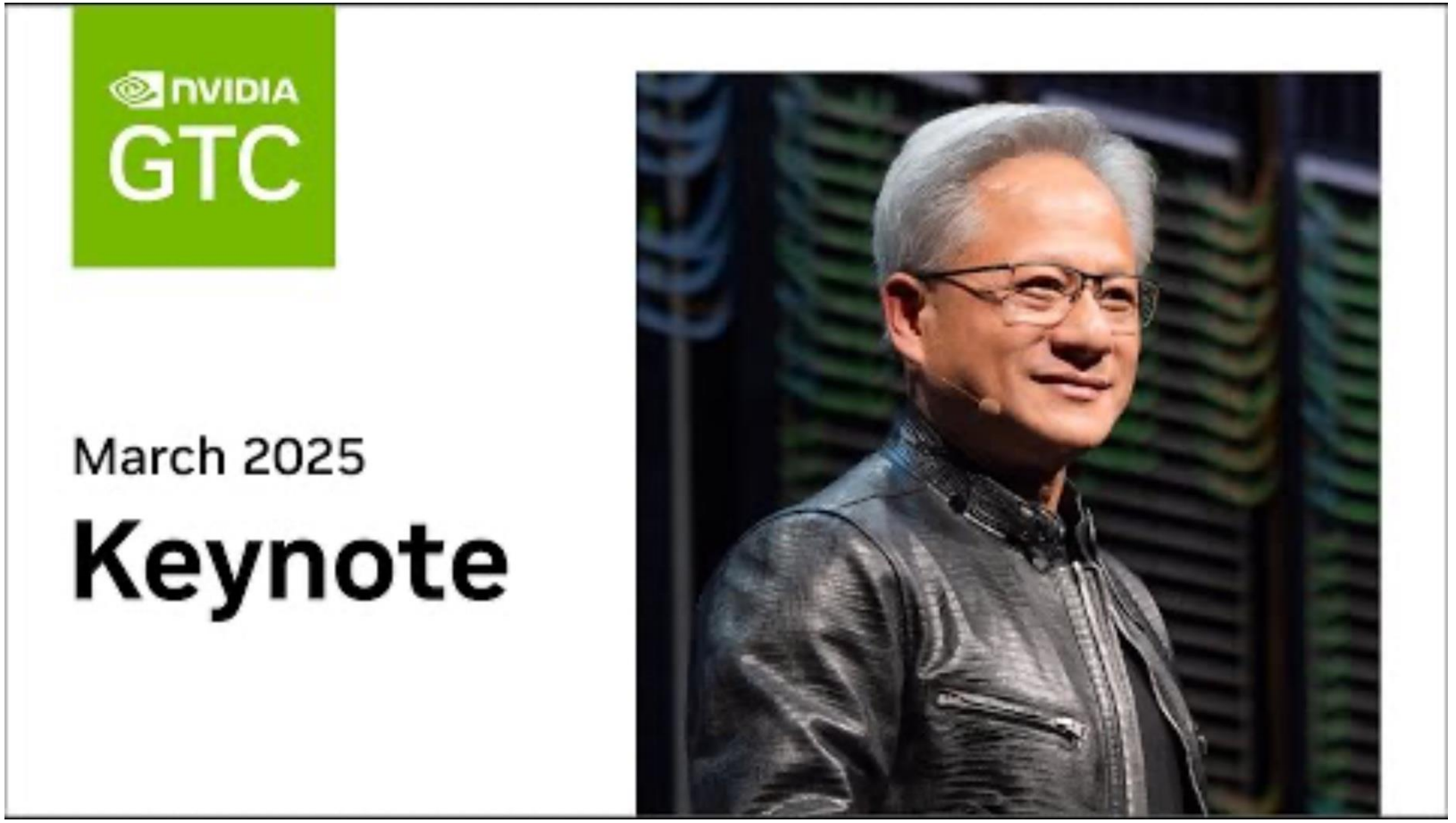
- I may curve up or curve down at the end

- Sample schema
- A/A-: 20%
- B+: 60%
- B and below: 20%

JOHNS HOPKINS
CAREY BUSINESS SCHOOL

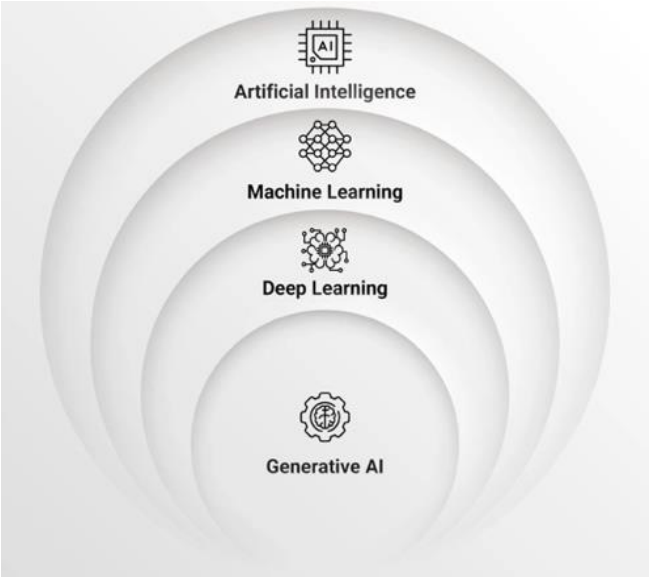
Introduction to Generative AI

GTC 2025



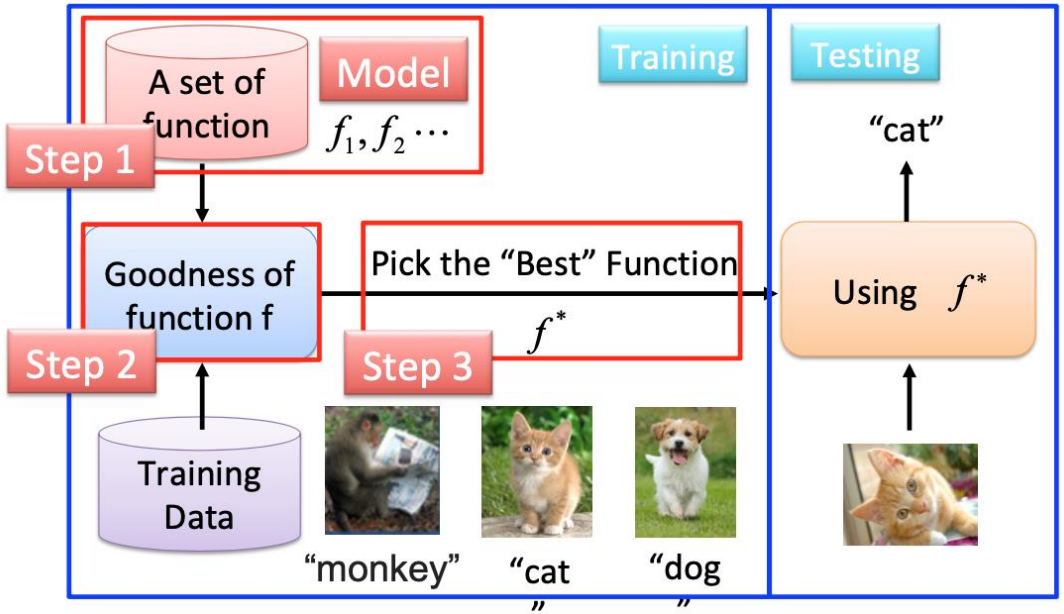
AI Paradigm

- **Artificial Intelligence**: broad field of developing machines that can replicate human behaviors
  - Perceiving, reasoning, learning, problem solving,…
- **Machine Learning**: methodology of teaching machines to learn patterns in data
- **Deep Learning**: branch of ML that uses neural network models
- **Generative AI**: branch of DL that produces new data that is similar to given dataset
  - Text, image, audio, video,…

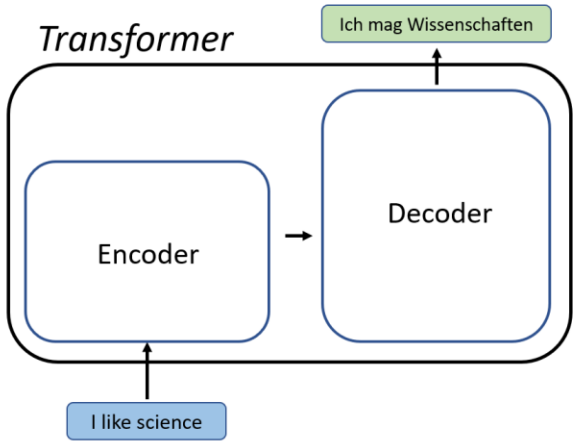


Deep Learning: Discriminative vs Generative

- Discriminative Modeling: make predictions/discriminations

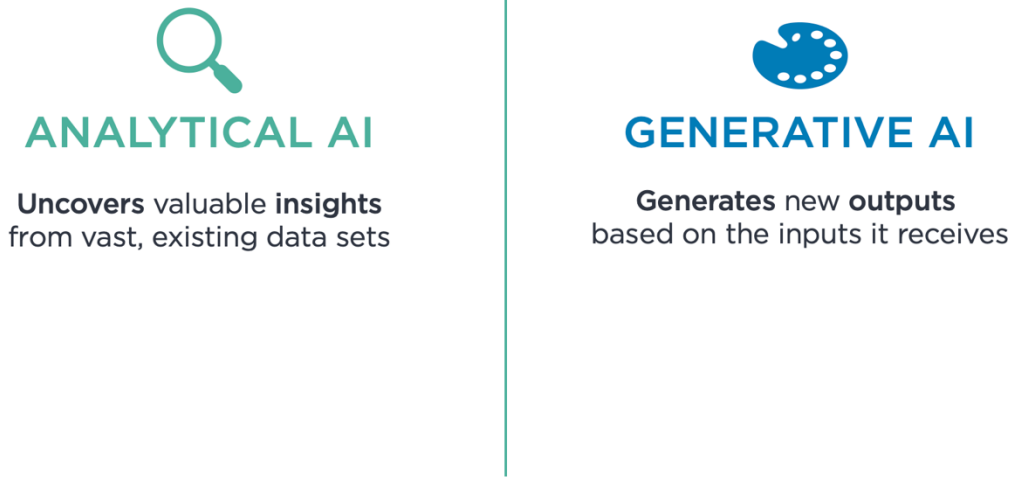


- Generative Modeling: foundation models with small variations from Google's Transformer model in 2017 paper "*Attention is all you need*"



AI Becomes: Analytical vs Generative

- Analytical AI: focus on analyzing existing data to extract insights and make predictions, also called predictive AI
- Generative AI: focus on creating new content, such as images, text, or code, based on patterns learned from data

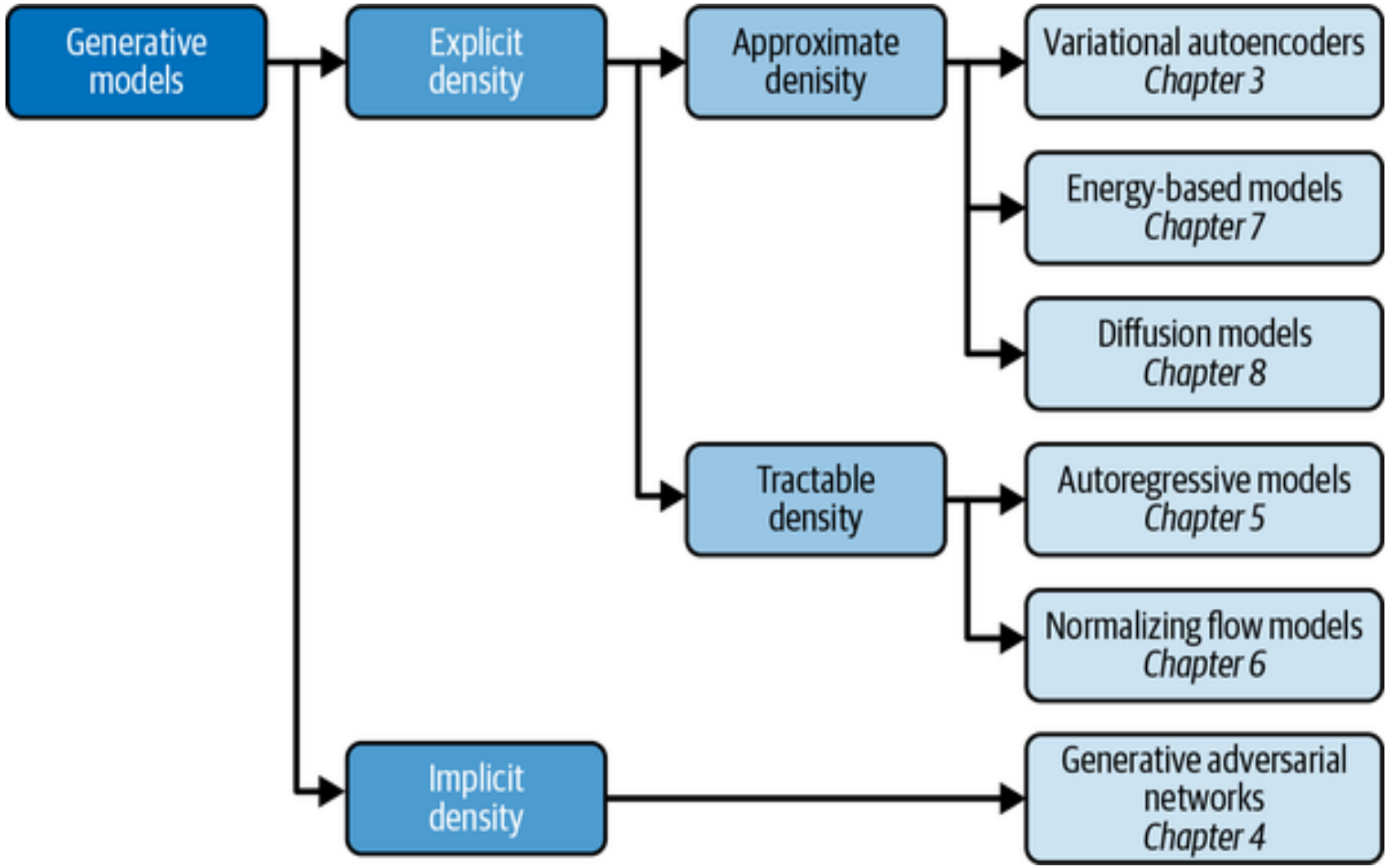


Idea of “Generate”

- *How do you generate a random number?*
- Distribution!
- Discriminative Modeling: model the probability of a label *y*
- Generative Modeling: model the probability of observation *x*
- *That's how hallucinations may occur…*

Generative Modeling Approaches (Optional)

- Explicitly model the density function, but constrain the model so that the density function is tractable
- Explicitly model a tractable approximation of the density function
- Implicitly model the density function through a stochastic process that directly generates data



Foster, David. *Generative deep learning*. " O'Reilly Media, Inc.", 2022.

Timeline of Generative AI

- 2014–2017: The VAE and GAN era
- 2018–2019: The Transformer era
- 2020–2022: The Big Model era
- 2023–?: *What's next?*

Brief History

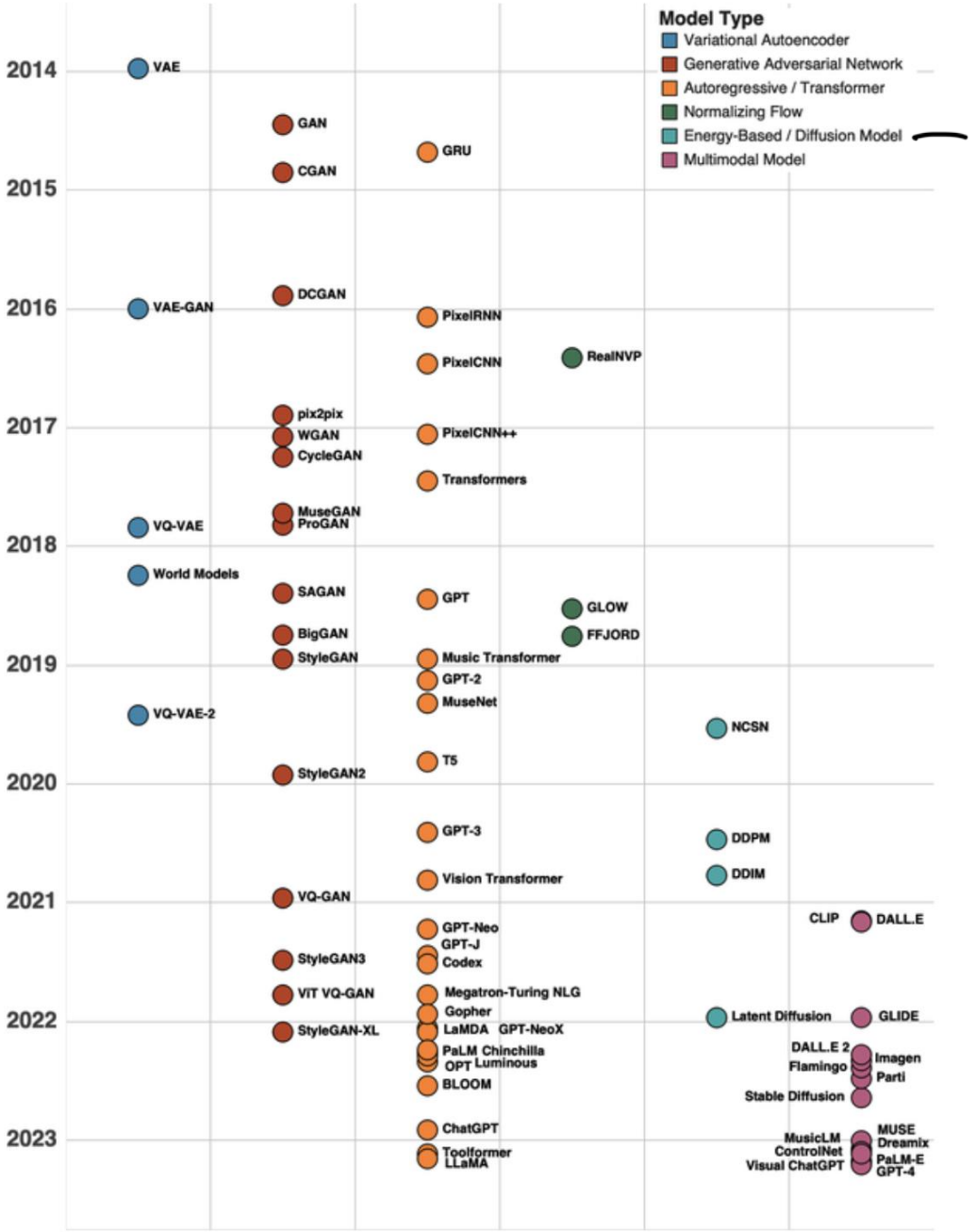


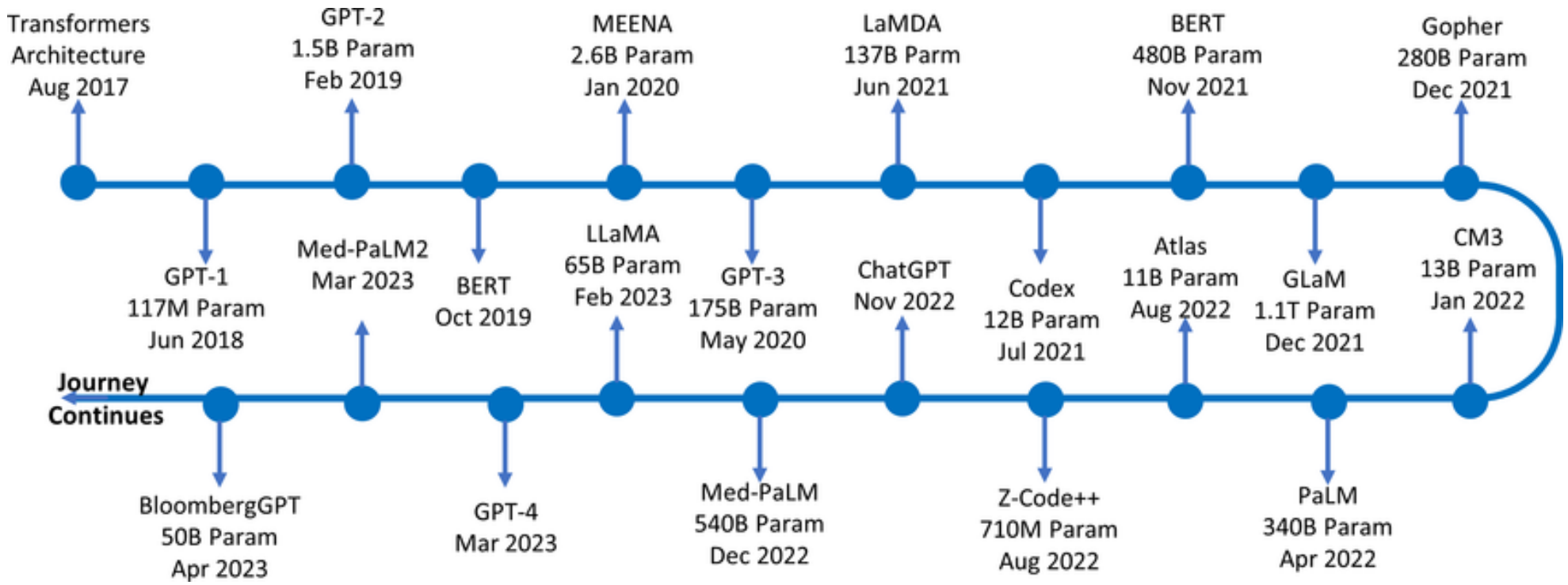
Figure 14-1. A brief history of generative AI from 2014 to 2023 (note: some important developments such as LSTMs and early energy-based models [e.g., Boltzmann machines] precede this timeline)

Model Types

- Depend on the input/output types

- Audio and speech
- Image generation
- Text generation
- Multi-modality
  - Multiple types of input and output data

Text Model Timeline



Mohamadi, Salman & Mujtaba, Ghulam & Le, Ngan & Doretto, Gianfranco & Adjeroh, Donald. (2023). ChatGPT in the Age of Generative AI and Large Language Models: A Concise Survey. 10.48550/arXiv.2307.04251.

What's after Text Model, and Why?

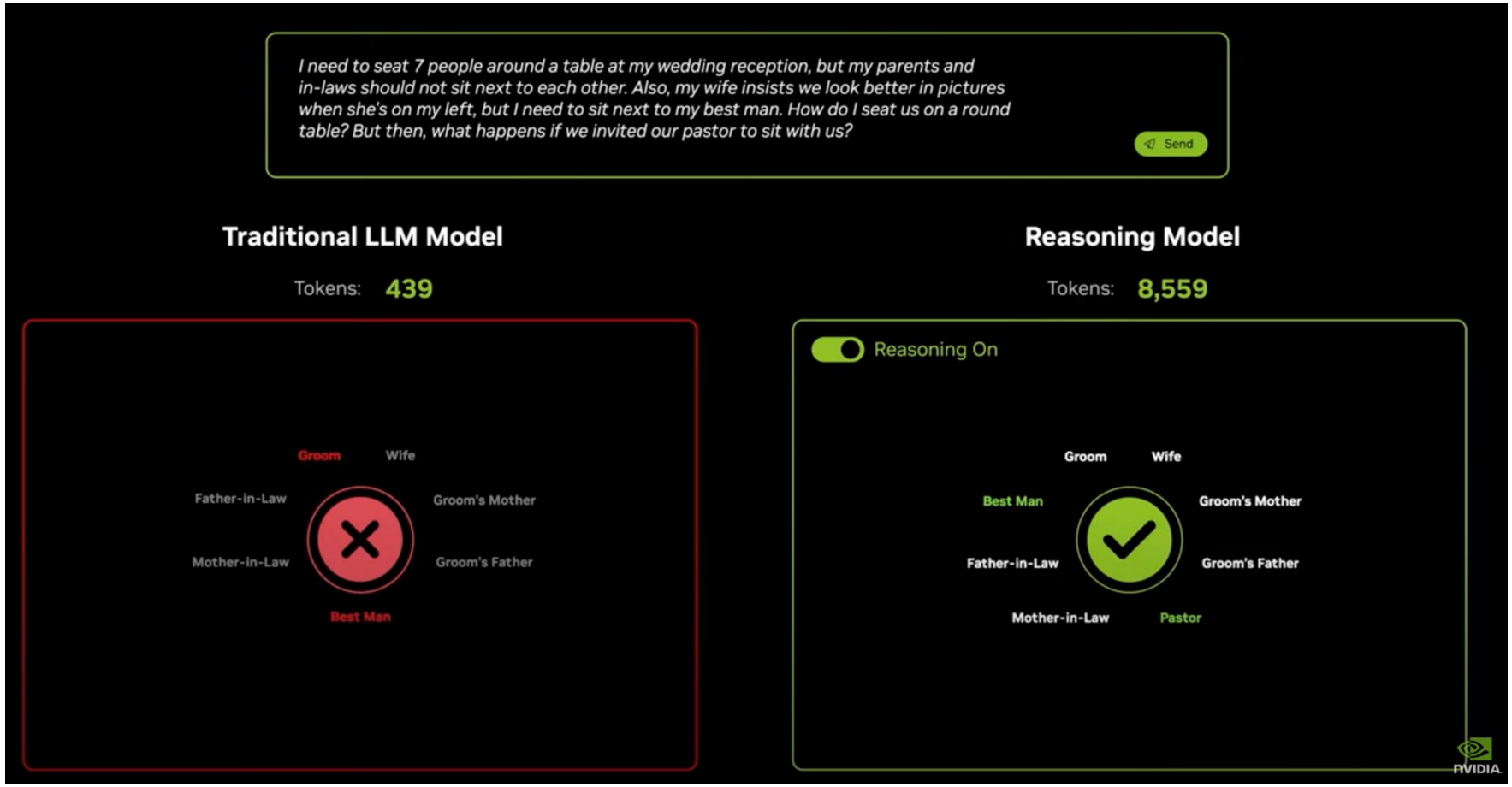
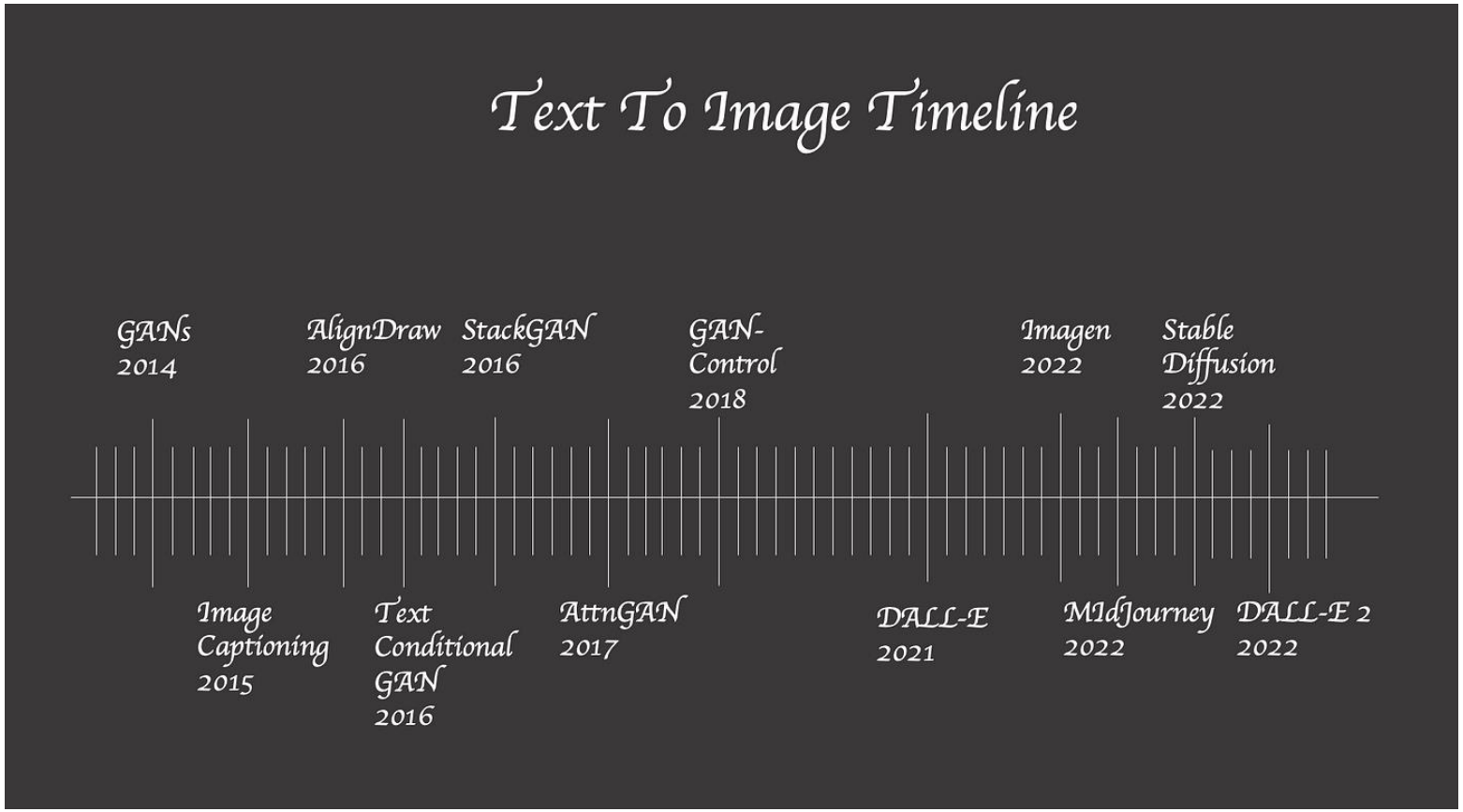
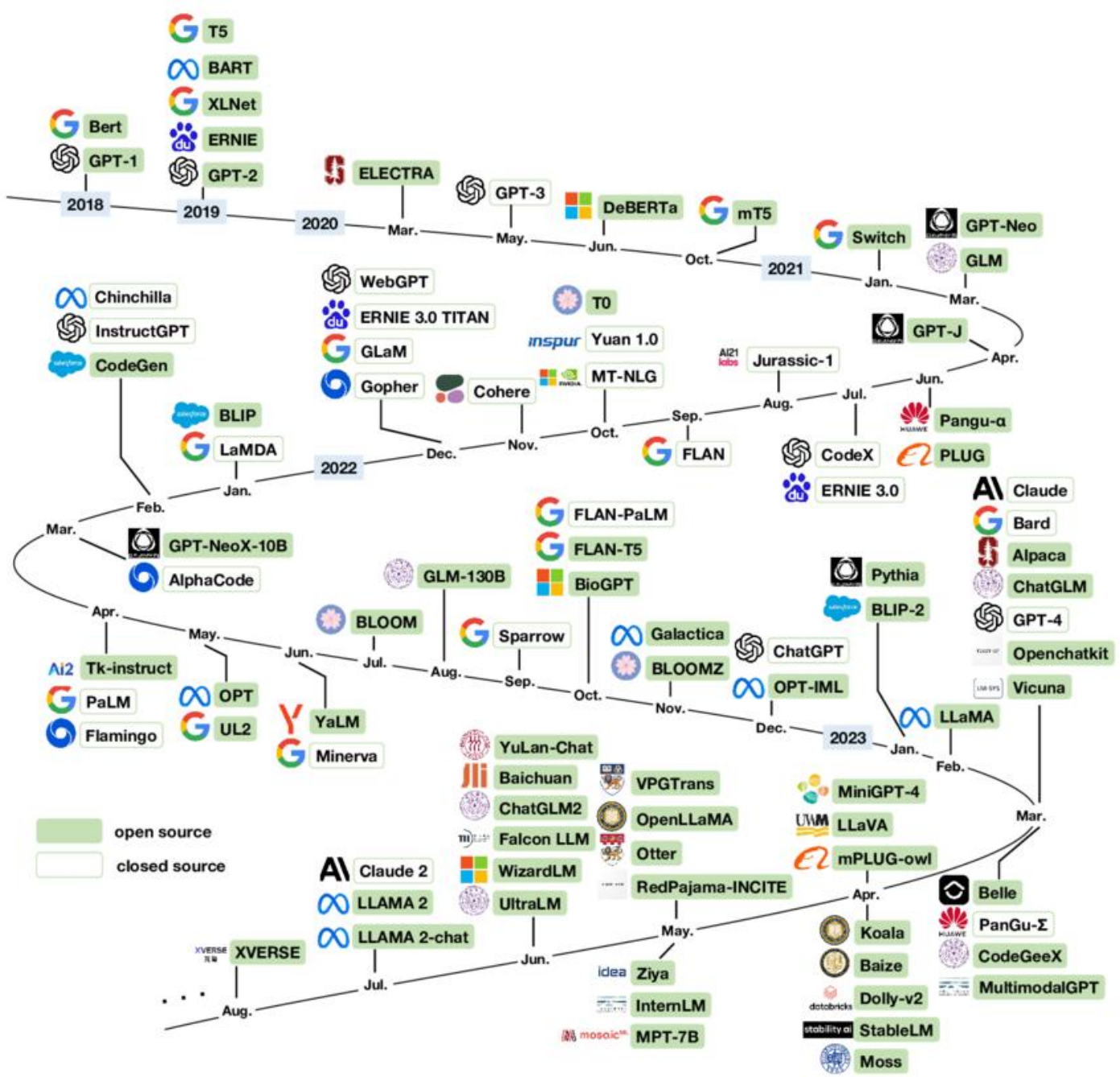


Image Model Timeline



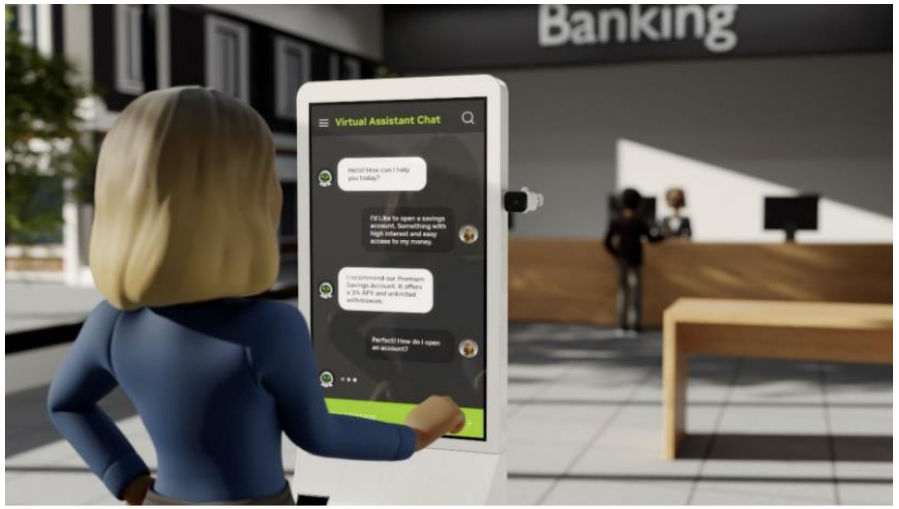
Credit: Serop Baghdadlian
Published in Artificial Intelligence in Plain English



Gao, Kaiyuan & He, Sunan & He, Zhenyu & Lin, Jiacheng & Pei, Qizhi & Shao, Jie & Zhang, Wei. (2023). Examining User-Friendly and Open-Sourced Large GPT Models: A Survey on Language, Multimodal, and Scientific GPT Models. 10.48550/arXiv.2308.14149.

Enhance Customer Experience

- Generative AI can be used to create personalized ad campaigns that target specific demographics based on their browsing and purchase history
- Personal branding, logo design
- Chatbots and virtual assistants



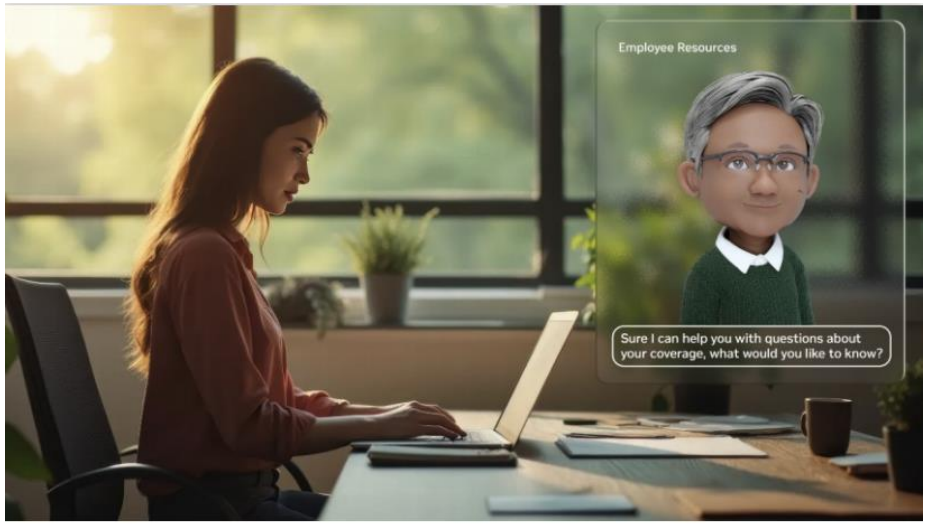
Customer Experiences

Learn how to transform customer experiences with AI-powered agents, digital humans, recommendation systems, and personalized solutions revolutionizing call centers and beyond.

GTC 2025, Nvidia

Boost Productivity and Optimize Processes

- Fashion: create unique and diverse clothing designs, taking into account trends and wearer preferences
- Automotive: design and develop new vehicle models and automatically find interesting variations on a particular design
- Pharmaceutical research/healthcare: generate new drug compounds, aid in the development of new treatments



Productivity and Process Automation

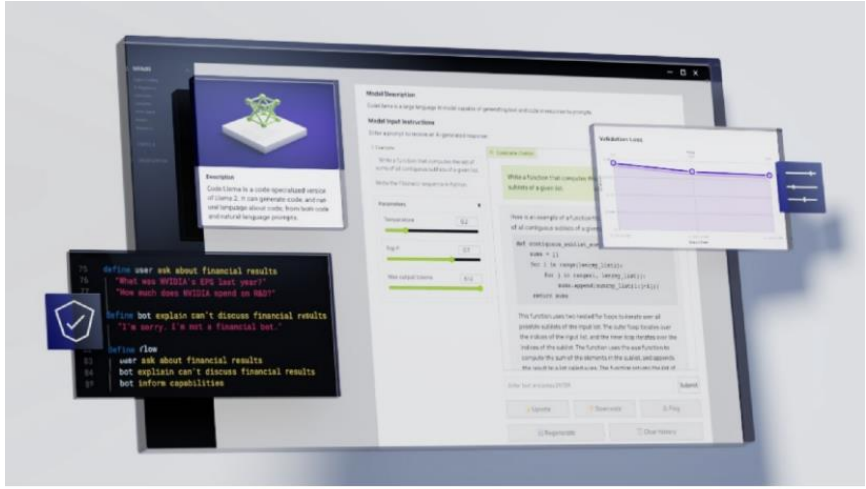
Boost productivity and streamline processes with AI-driven automation, decision-making, inventory management, analytics, and business insights from data.

GTC 2025, Nvidia


Generate left turn in urban environment during winter, wipers on
Reason: The vehicle is traveling on a two-lane road, with snow and ice visible on the surface. Potential hazards include the icy road conditions, which could cause reduced traction and longer stopping distances. The traffic light ahead is green, allowing the vehicle to continue without stopping. The ego vehicle plans to continue turning, maintaining its current speed and lane position.
Cosmos

Content Creation

- Music: compose and produce new music tracks
- Film and video production: create special effects and animations, as well as to generate dialogue for entire scenes or storylines
- Game design: design and develop new game levels and content, creating an infinite variety of gameplay experiences
- Digital design: create original digital art and animations, as well as to design and develop new user interfaces and web designs

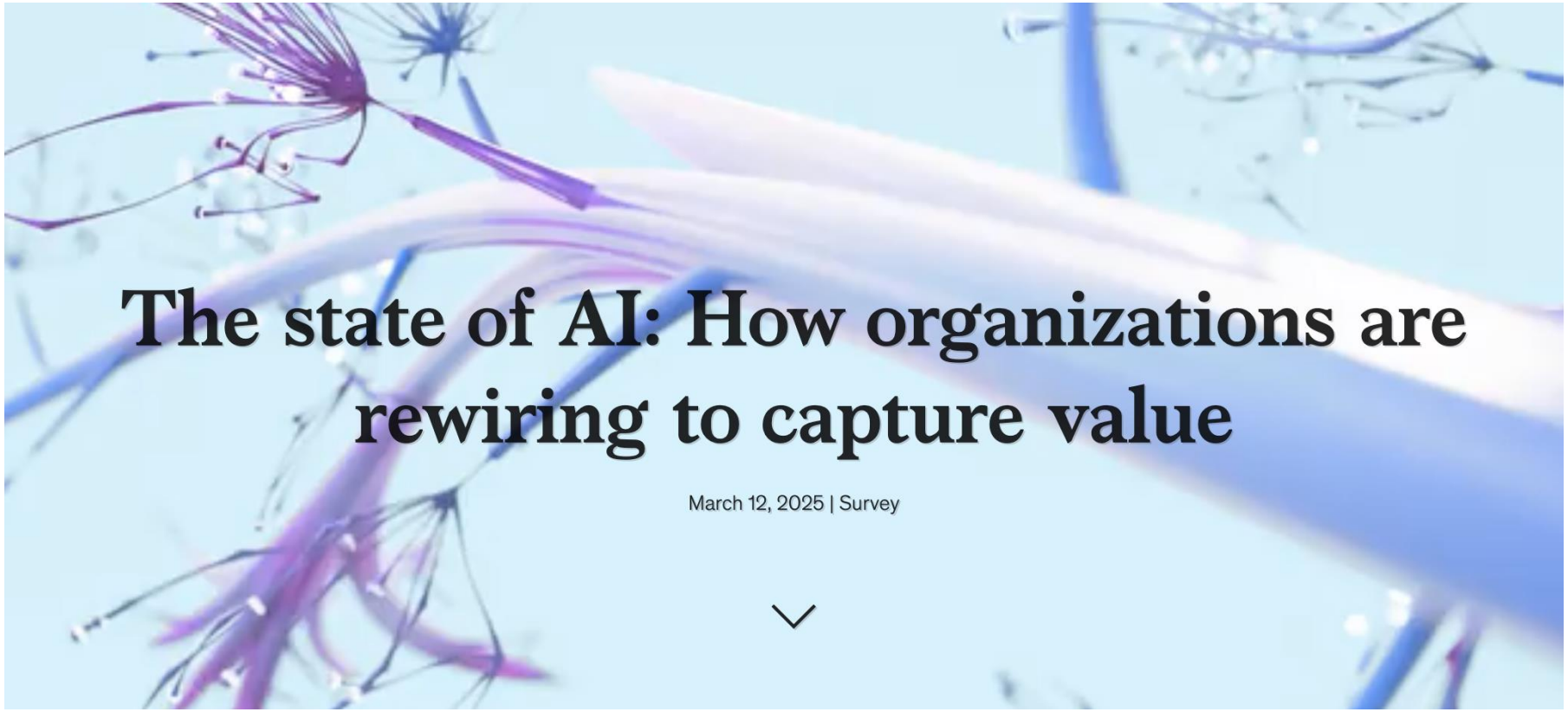


Tools and Training for Developers

Immerse yourself in technical sessions exploring generative AI, performance optimization, programming languages, coding assistants, and CUDA designed to elevate your skills and hands-on

GTC 2025, Nvidia


NVIDIA

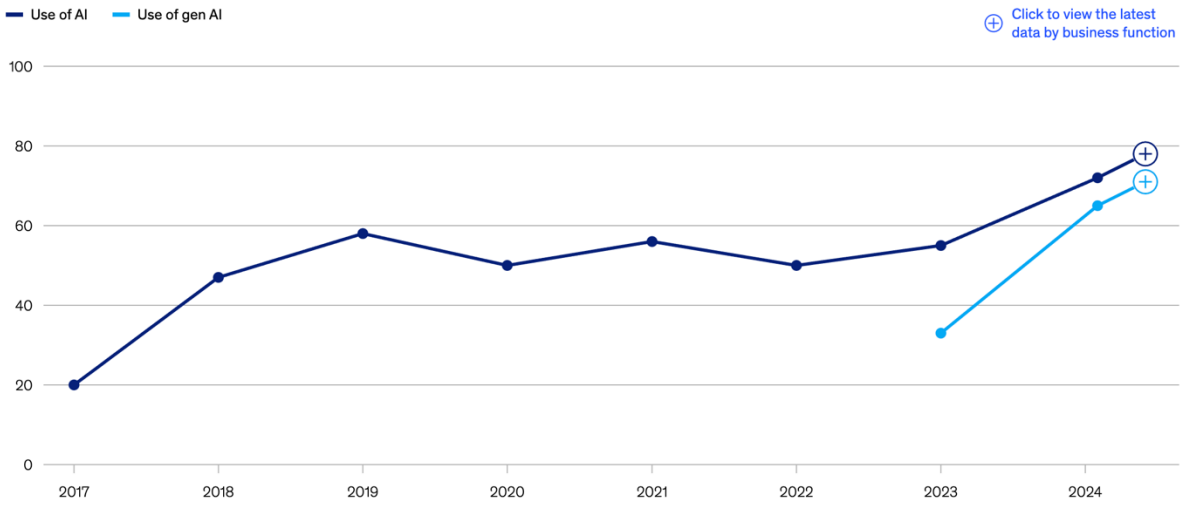

The state of AI: How organizations are rewiring to capture value
March 12, 2025 | Survey

AI Adoption

Exhibit 8

Organizations' use of AI has accelerated markedly in the past year, after years of little meaningful change.

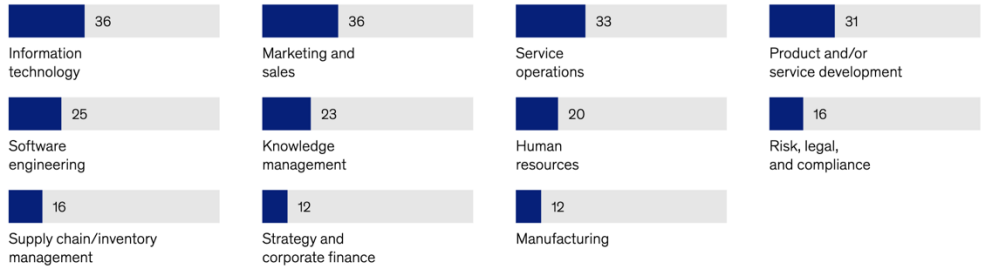
Organizations that use AI in at least 1 business function,[1] % of respondents



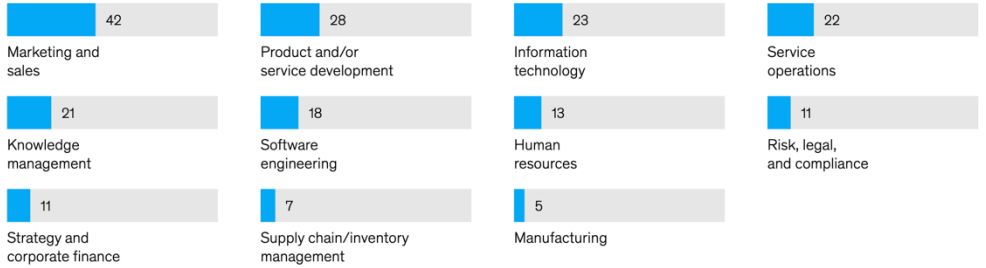
[1]In 2017, the definition for AI use was using AI in a core part of the organization's business or at scale. In 2018–2019, the definition was embedding at least 1 AI capability in business processes or products. Since 2020, the definition has been that the organization has adopted AI in at least 1 function.

Source: McKinsey Global Surveys on the state of AI

Use of AI by business function, % of respondents



Use of gen AI by business function, % of respondents



Current Status

- “Organizations are beginning to create the structures and processes that lead to meaningful value from gen AI. While still in early days, companies are redesigning workflows, elevating governance, and mitigating more risks.”

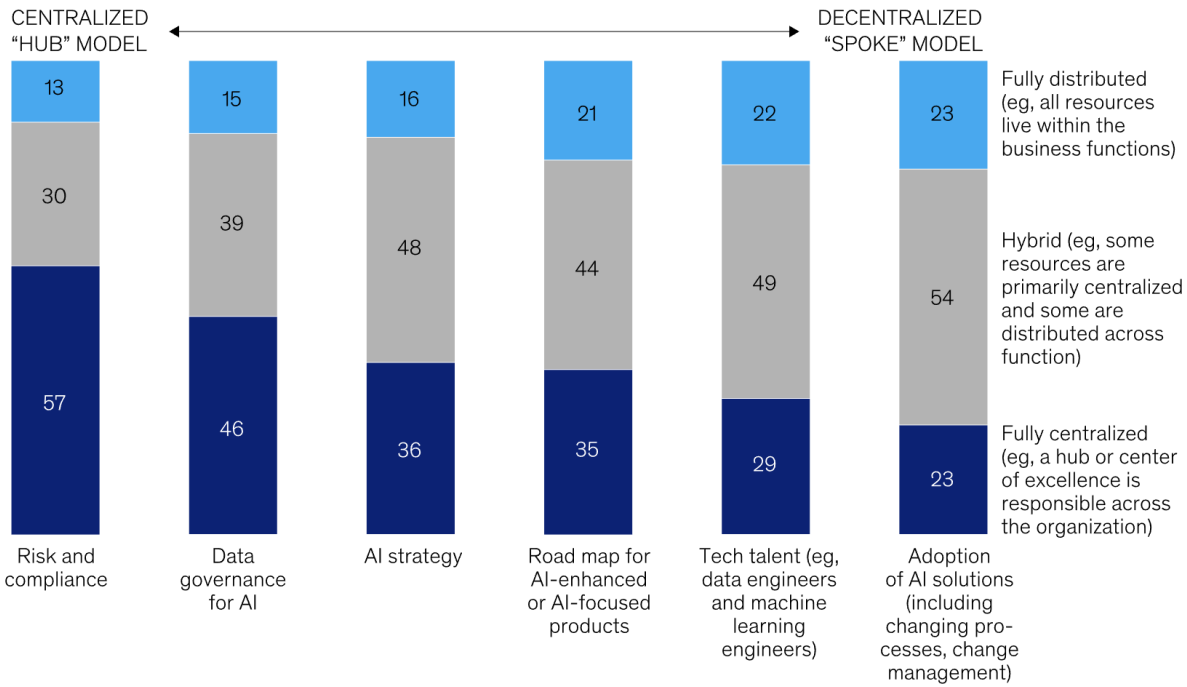
- “(Organizations) are hiring for new AI-related roles while they retrain employees to participate in AI deployment.”

Centralized vs Decentralized

Exhibit 1

Risk and data governance are two of the most centralized elements of deploying AI solutions, whereas tech talent is often hybrid.

Degree of centralization of AI deployment,[1] % of respondents



[1]Question was asked only of respondents whose organizations use AI in at least 1 function, n = 1,229. Figures were calculated after removing the share who said "don't know/not applicable."
Source: McKinsey Global Survey on the state of AI, 1,491 participants at all levels of the organization, July 16–31, 2024

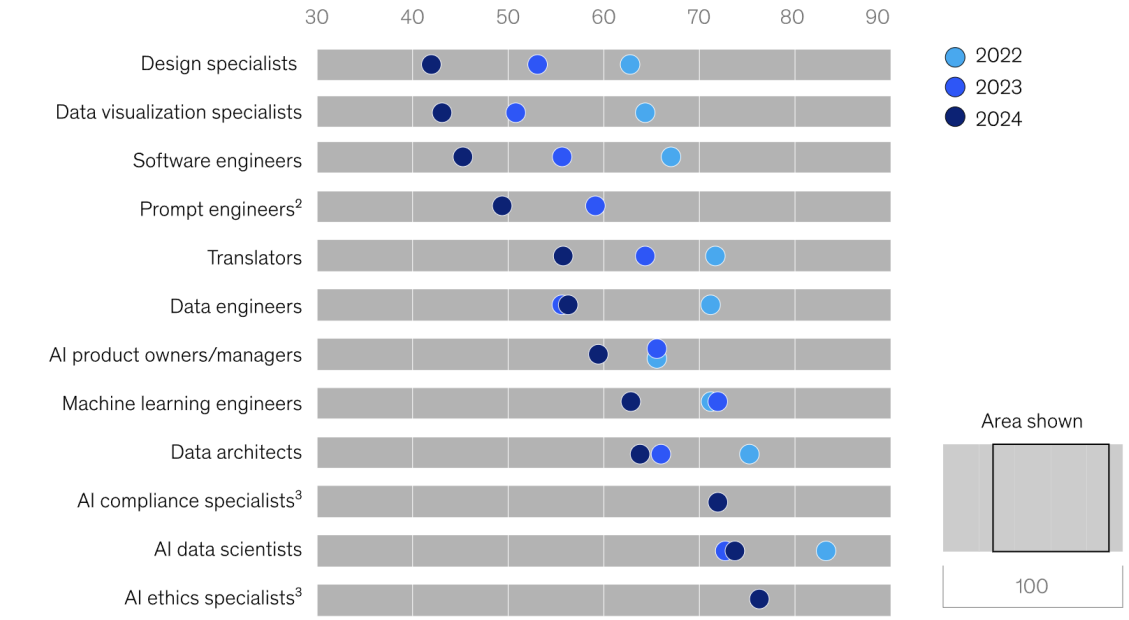
McKinsey & Company

AI Talent Demand

Exhibit 5

Smaller shares of respondents report difficulty in hiring for AI-related roles, compared with previous years.

Share of respondents reporting difficulty in organizations' hiring of AI-related roles,[1] % of respondents



[1]Only asked of respondents who said their organizations use AI in at least 1 function and who said their organization hired the given role in the past 12 months. Figures were calculated after removing the share who said "don't know." Respondents who described hiring for given role as "easy" or "neither difficult nor easy" are not shown.
[2]Not asked of respondents in 2022.
[3]Not asked of respondents in 2022 or 2023.
Source: McKinsey Global Surveys on the state of AI, 2022–24

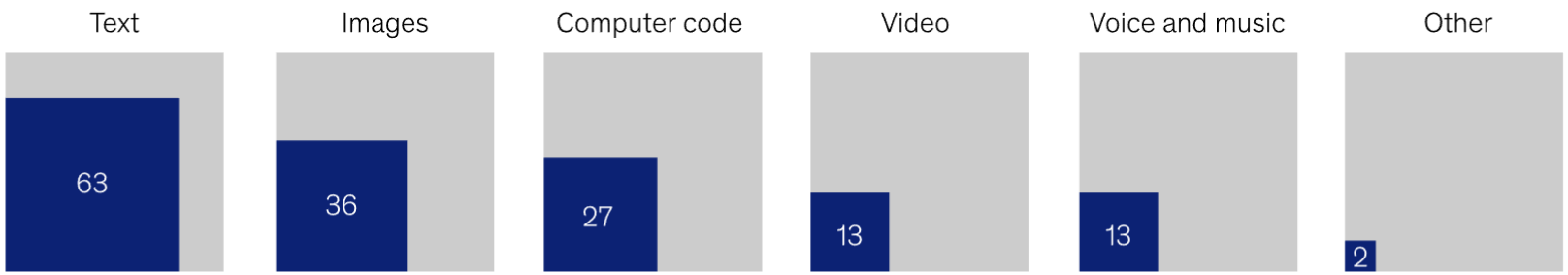
McKinsey & Company

Modality

Exhibit 11

While text is the type of content that organizations are most commonly creating with gen AI, they are also experimenting with other modalities.

Types of content generated by gen AI at respondents' organizations,[1] % of respondents



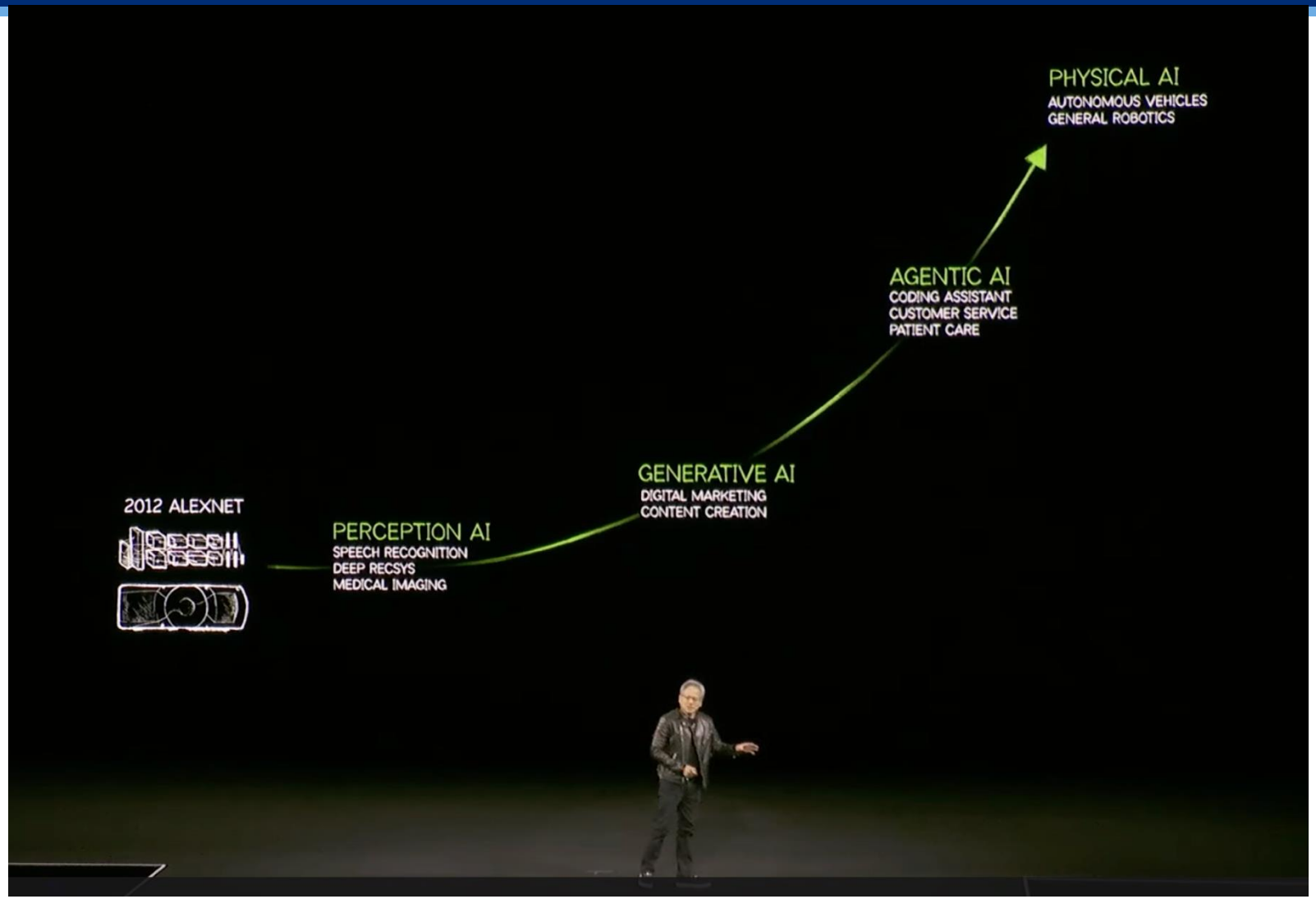
[1]Only asked of respondents whose organizations regularly use gen AI in at least one function. Figures were calculated after removing the respondents who said "don't know."
Source: McKinsey Global Survey on the state of AI, 1,491 participants at all levels of the organization, July 16–31, 2024

McKinsey & Company

Next Frontier

Organizations have been experimenting with gen AI tools. Use continues to surge, but from a value capture standpoint, these are still early days—few are experiencing meaningful bottom-line impacts. Larger companies are doing more organizationally to help realize that value. They invest more heavily in AI talent. They mitigate more gen-AI-related risks. We have seen organizations move since early last year, and the technology also continues to evolve, with a view toward agentic AI as the next frontier for AI innovation. It will be interesting to see what happens when more companies begin to follow the road map for successful gen AI implementation in 2025 and beyond.

So, What's Next?



Boston Dynamics

Next Week

- Encoder-Decoder Architecture
- Attention Mechanism and Transformer
- GenAI Ecosystem

References

- https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai
- https://www.nvidia.com/gtc/