Berkeley
UNIVERSITY OF CALIFORNIA

Responsible GenAI and Decentralized Intelligence

CS294/194-196 Fall 2023

Instructors

Dawn Song

Matei Zaharia

https://rdi.berkeley.edu/responsible-genai/f23

Teaching Staff

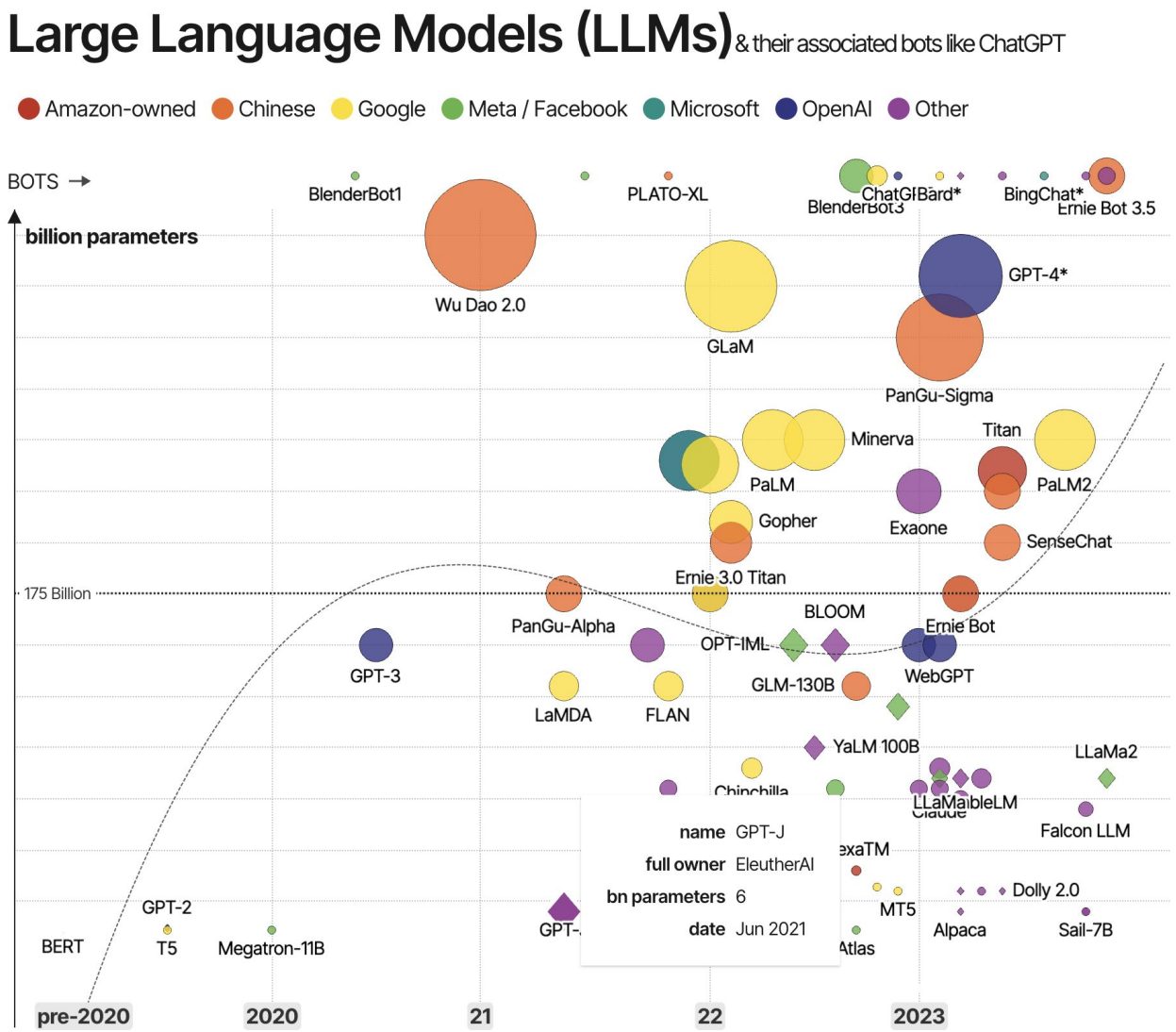
Instructors: **Prof. Dawn Song** and **Prof. Matei Zaharia**

Guest lecturer & project mentor: **John Whaley**

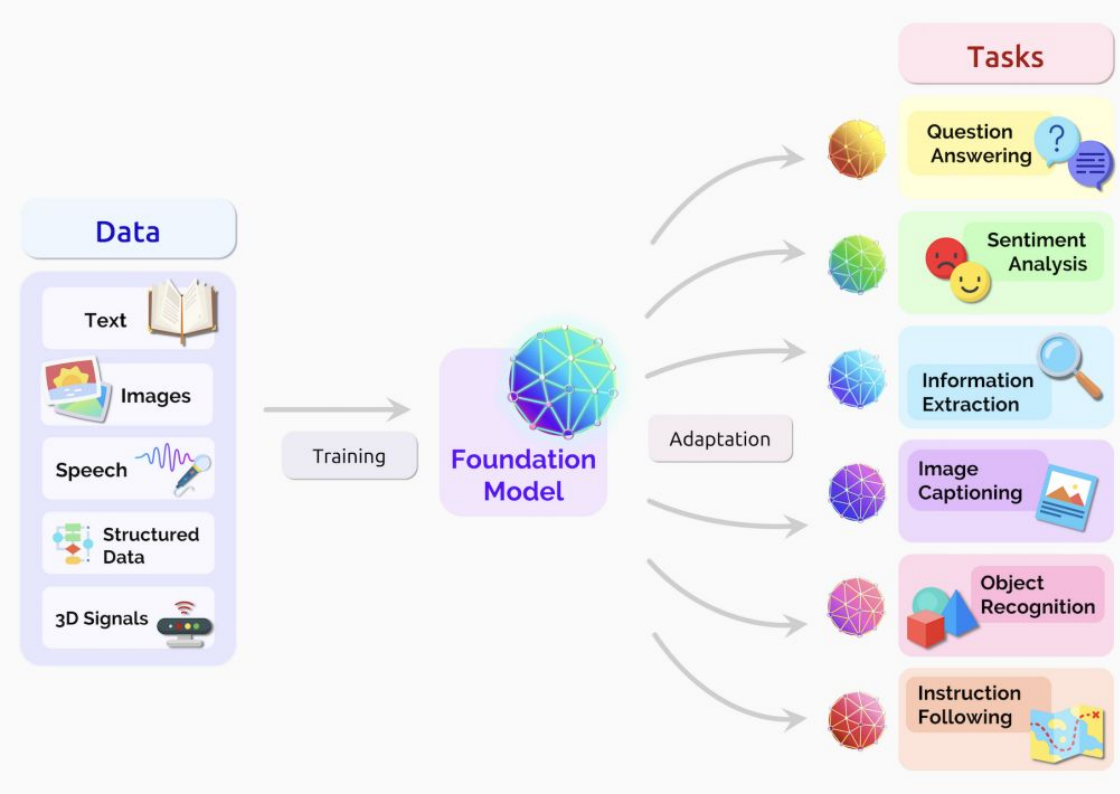
GSI: Yu Gai

Readers: Shiladitya Dutta and Bill Zheng

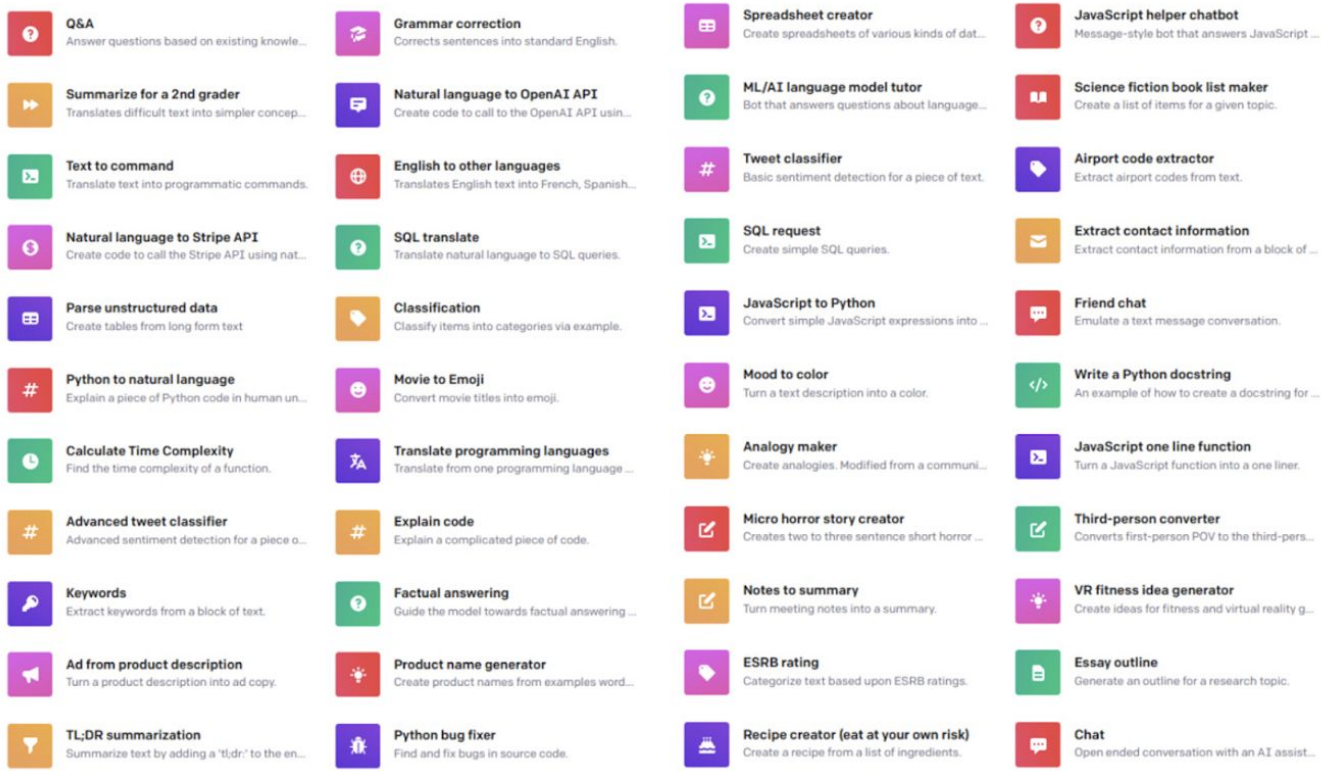
Exponential Growth in LLMs



Powering Rich New Capabilities

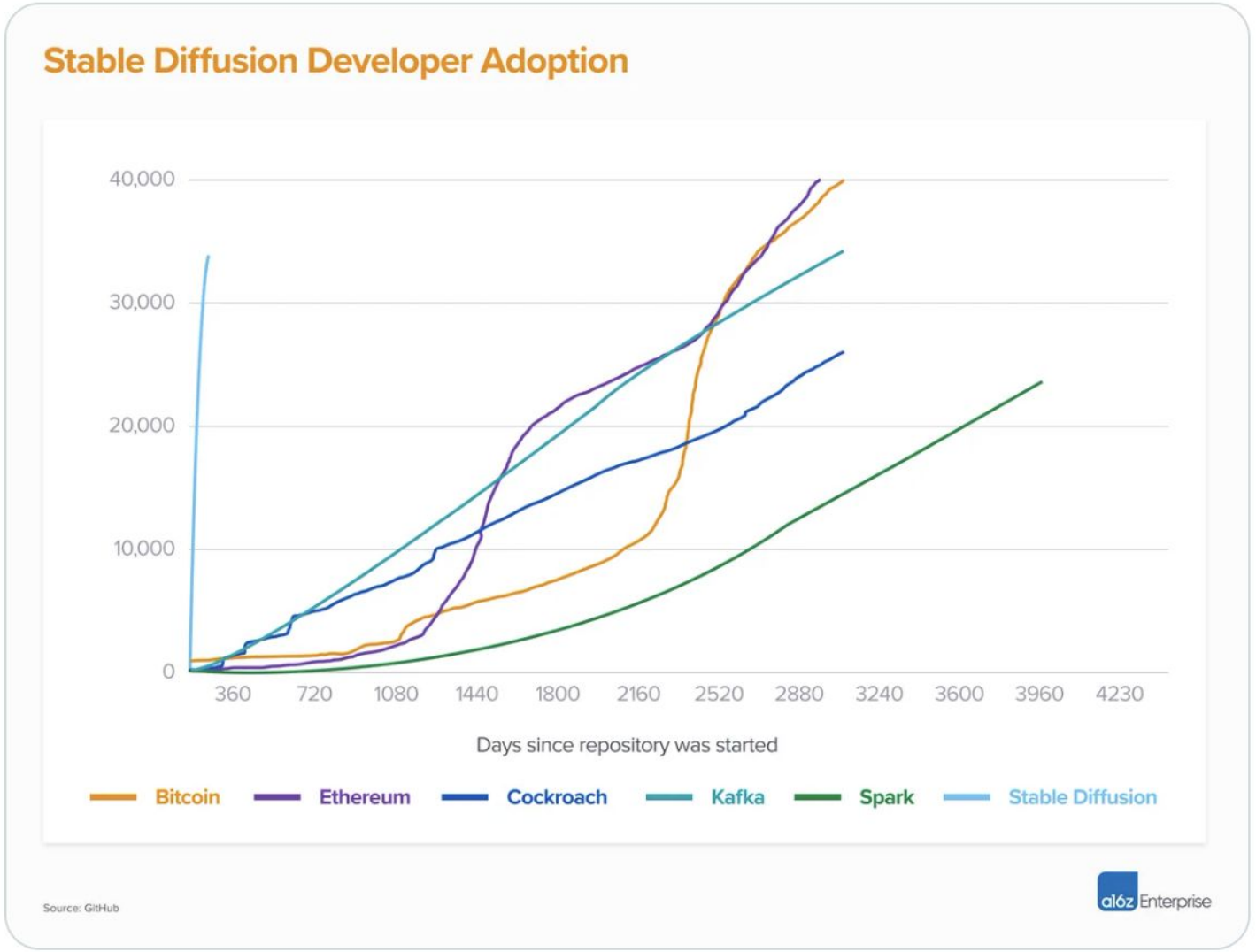


https://arxiv.org/pdf/2108.07258.pdf

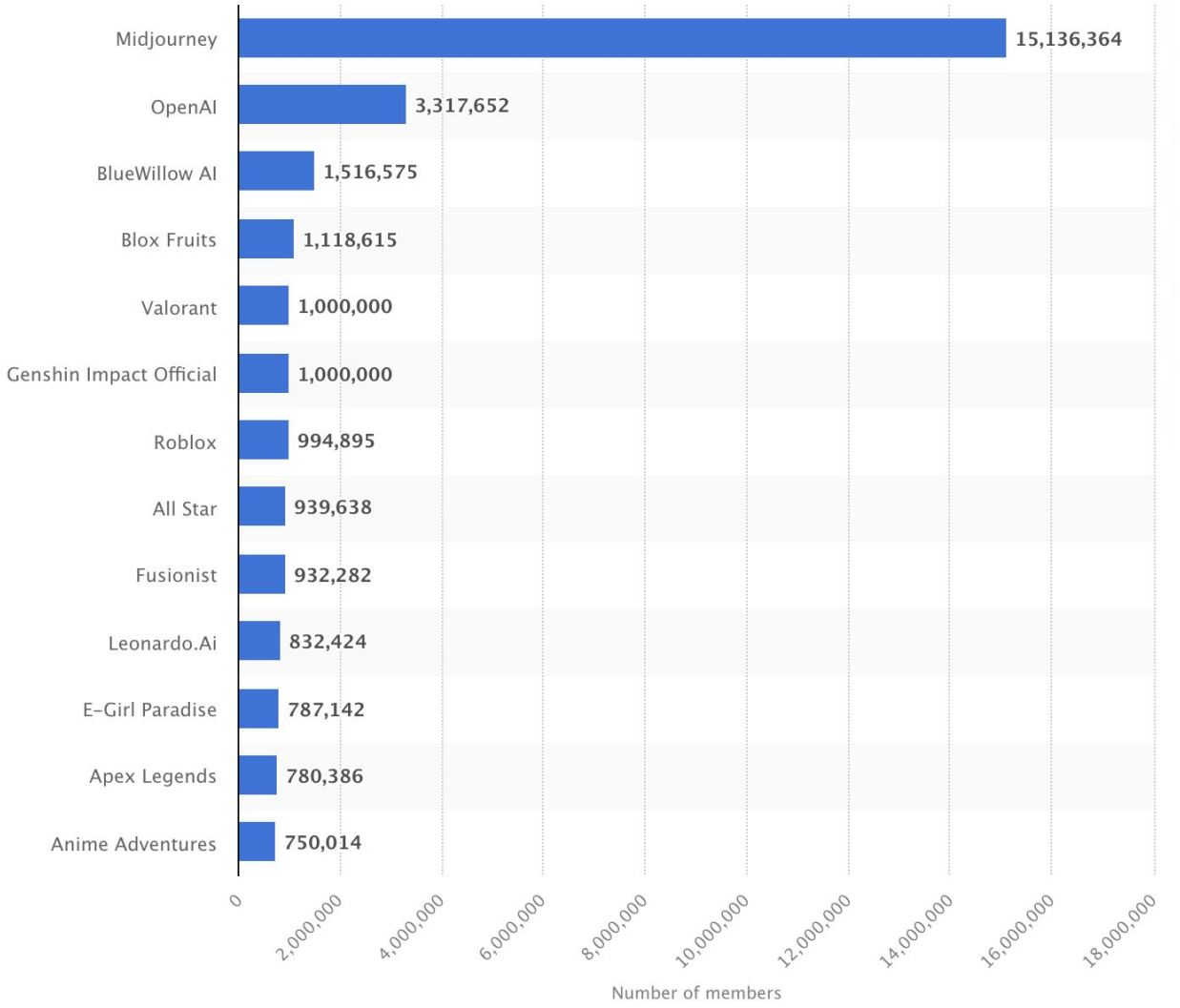


Source: openai

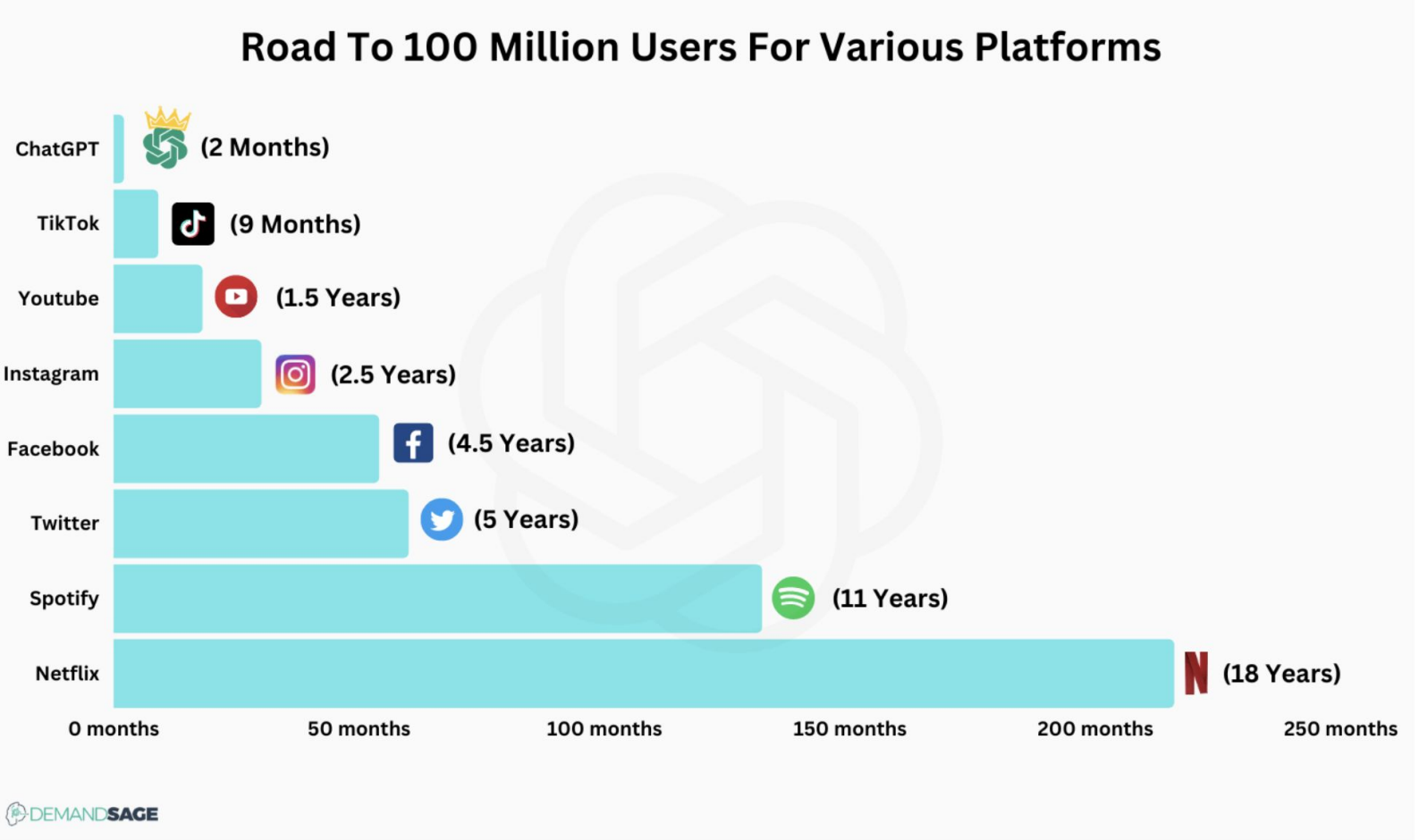
Stable Diffusion: Fastest Repo to Reach 35K Stars on GitHub



Midjourney: The Largest Discord



ChatGPT: Fastest Platform to 100M Users



Many Risks & Open Challenges for Responsible AI

- Who controls AI?
  - centralized vs. decentralized control; open vs. closed source
- Trustworthiness
  - Robustness
    - Adversarial robustness
    - Out-of-distribution robustness
    - Test-time attacks vs. training-time attacks
  - Privacy
  - Fairness
  - Toxicity
  - Stereotype
  - Machine ethics
- AI Safety
  - Misuse/abuse of AI
  - Super intelligence

Importance of Mitigating Risk of Extinction from AI

Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war.

Center for AI Safety

The New York Times

A.I. Poses 'Risk of Extinction,' Industry Leaders Warn

Leaders from OpenAI, Google DeepMind, Anthropic and other A.I. labs warn that future systems could be as deadly as pandemics and nuclear weapons.

Signatories:

☑ AI Scientists ☑ Other Notable Figures

Geoffrey Hinton
Emeritus Professor of Computer Science, University of Toronto

Yoshua Bengio
Professor of Computer Science, U. Montreal / Mila

Demis Hassabis
CEO, Google DeepMind

Sam Altman
CEO, OpenAI

Dario Amodei
CEO, Anthropic

Dawn Song
Professor of Computer Science, UC Berkeley

Ted Lieu
Congressman, US House of Representatives

Bill Gates
Gates Ventures

Ya-Qin Zhang
Professor and Dean, AIR, Tsinghua University

Ilya Sutskever
Co-Founder and Chief Scientist, OpenAI

Shane Legg
Chief AGI Scientist and Co-Founder, Google DeepMind

Future of Responsible AI with Decentralization & Democratization

- Can we build a full decentralized, open-source stack for AI/ML:
  - Open-source decentralized AI/ML infrastructure for training and inference, with provenance, integrity, privacy guarantees
  - Open-source models and tooling
  - Personalized AI with privacy and trustworthiness
  - Decentralized, cooperative AI with incentives and social welfare
  - Democratic, decentralized process for AI governance & alignment

Open Challenges

- Is this technically feasible?
  - Is it possible to close the gap btw open source & closed source models?
  - Is it possible to scale decentralized training to large scale models?
  - Is it possible to build autonomous cooperative, decentralized agents?
- How to design proper incentive to maximize societal benefits?
- Can such an open source, decentralized system lead to more misuse/abuse? Is this at odds with AI safety guarantees?
- What are the different alternatives and possibilities that should be considered?

Date	Topic
Aug 27	Join The Future of Decentralization, AI, and Computing Summit!
Aug 29	No class
Sep 5	Intro & Foundations of LLM
Sep 12	Infrastructure Layer I: Training and Inference, Performance Optimization, Scalability
Sep 19	Infrastructure Layer II: Retrieval, Vector Databases, Search
Sep 26	App Development Layer: Prompt Engineering, Chains, Tools
Oct 3	Application Domains I: Software Engineering/Code Generation, Data Science
Oct 10	Application Domains II: Security, Education
Oct 17	Agents: RPA, Virtual Assistants
Oct 24	Trustworthiness: Privacy, Hallucinations, Adversarial Attacks
Oct 31	Decentralized Training and Inference, Open-Source Models
Nov 7	Decentralized Decision Making
Nov 14	Ethics and Fairness, Safety, Alignment
Nov 21	Edge compute; Federated learning; Open source data
Nov 28	Project Demos

Grading

	1 unit	2 units	3/4 units
Participation	50%	20%	10%
Article	50%		
Lab		20%	10%
Project			
Proposal		10%	10%
Milestone		10%	10%
Presentation		25%	25%
Report		15%	15%
Implementation			20%

Tasks for Each Number of Units

1 unit: lecture participation + summary article

2 units: lecture participation + lab assignment + project (no implementation required)

3 units: lecture participation + lab assignment + project with implementation

4 units: lecture participation + lab assignment + project with significant implementation and end-to-end working demo

5-6 students per project group; each group can only contain students with the same number of units

Lab and project timeline

	Released	Due
Project group formation	Sep 5	Sep 19
Project proposal	Sep 12	Oct 3
Lab	Sep 19	Oct 17
Project milestone	Oct 3	Oct 31
Project presentation	Oct 31	Nov 28
Project final report	Oct 31	Dec 12

Project Ideas

- Use GenAI to build an app of your choice
- Create new LLM benchmark unlikely to overlap with training
- Analyze the quality of open source GenAI training datasets and their impact on models
- Build on GenAI research projects happening at Berkeley (e.g., Chatbot Arena, DSPy, FrugalGPT, GenAI Security, DecodingTrust)