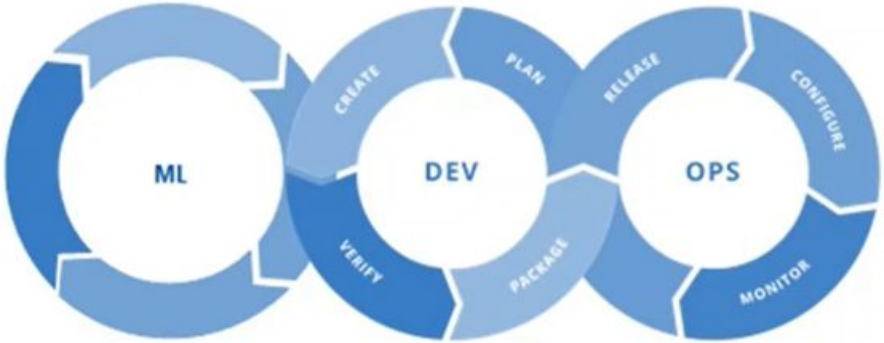
Module 9

AI Hardware

- AWS EC2
- GOOGLE CLOUD
- MICROSOFT AZURE
- JARVIS LABS

From Colab to Deployment

MLOps = ML + DEV + OPS

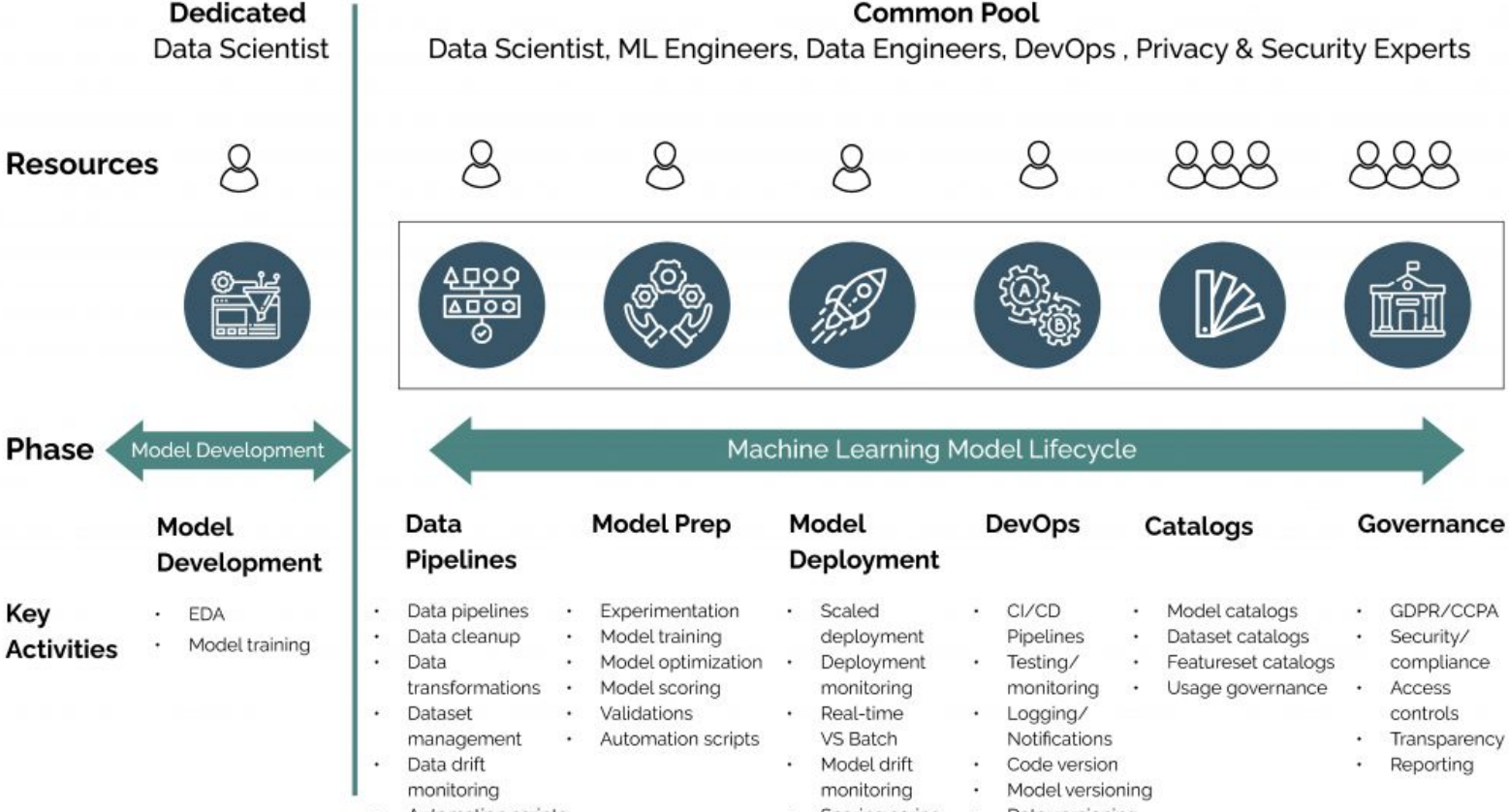


Experiment
Data Acquisition
Business Understanding
Initial Modeling

Develop
Modeling + Testing
Continuous Integration
Continuous Deployment

Operate
Continuous Delivery
Data Feedback Loop
System + Model Monitoring

Machine Learning Operations (MLOps)



Hardware

Software Tools:

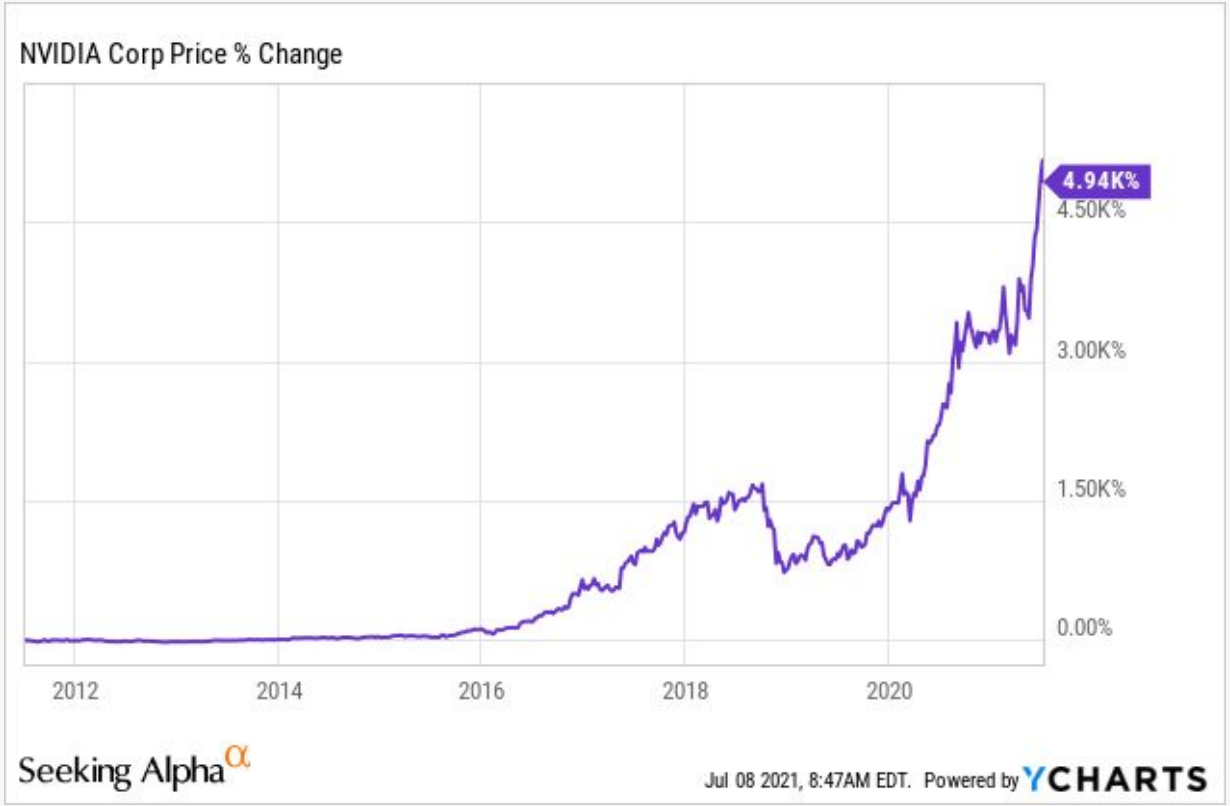
- Libraries like Keras, NumPy, SK-Learn
- Tools like Jupyter notebooks, Google Colab
- Opensource projects like Llama2
- Communities like HF, Kaggle, Github

Hardware:

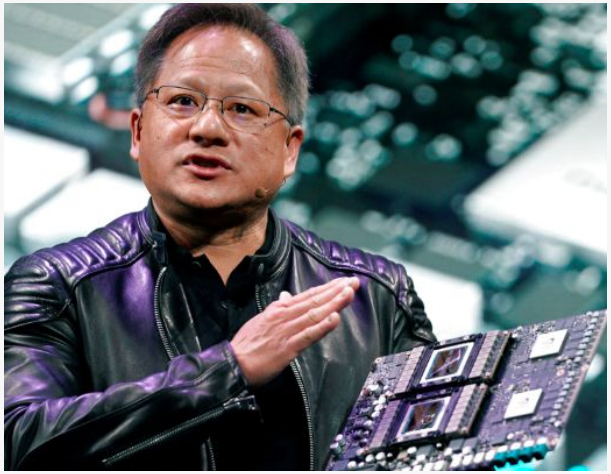
- GPU???

NVIDIA GPU

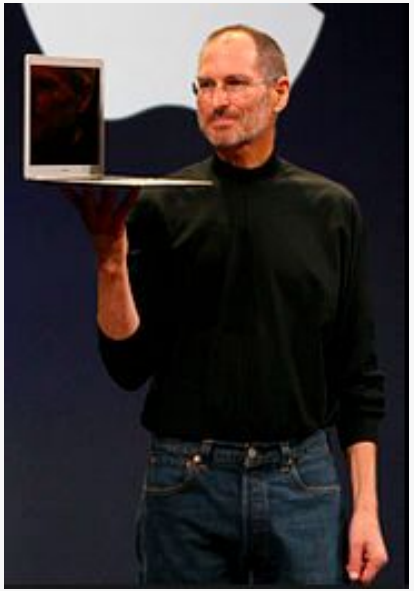
NVIDIA



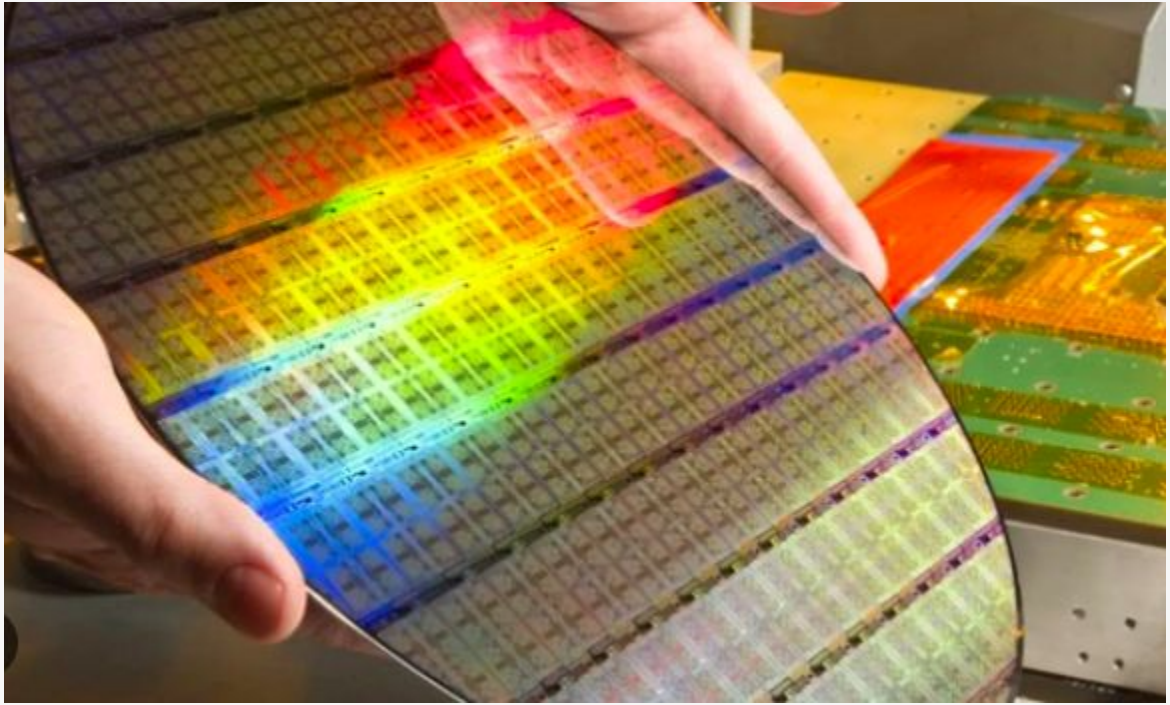
Jensen Huang, CEO of Nvidia Corp



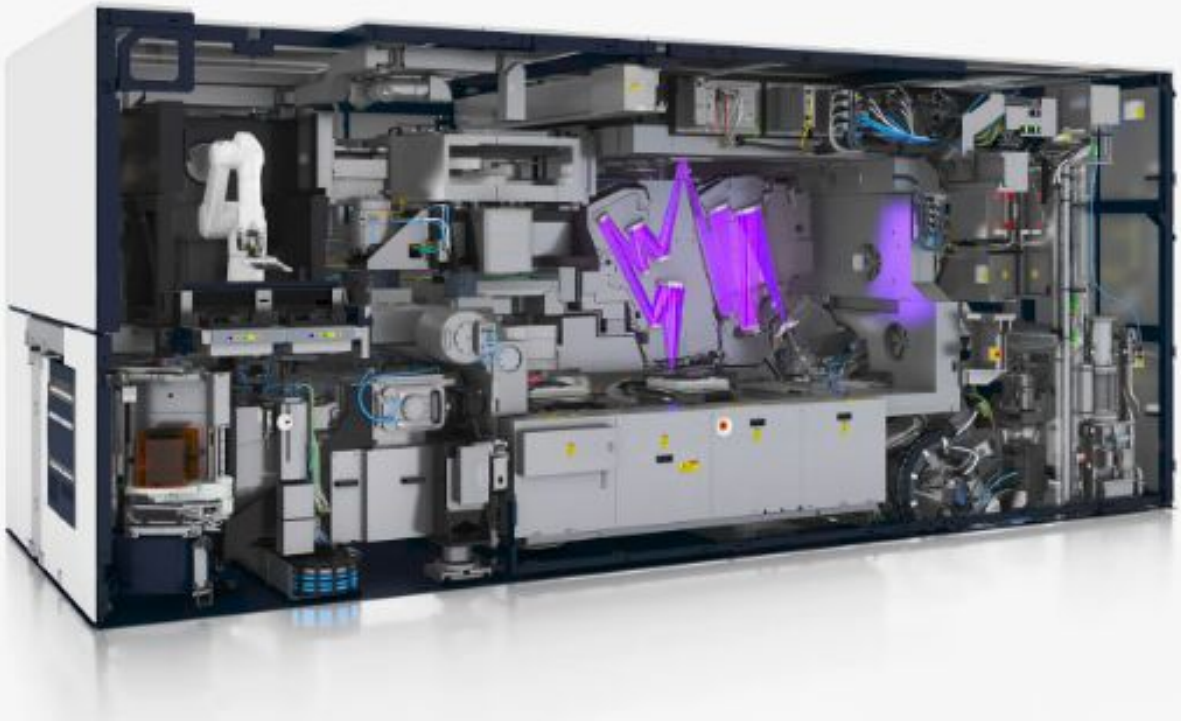
Silicon era heroes



TSMC and Chip Fabrication



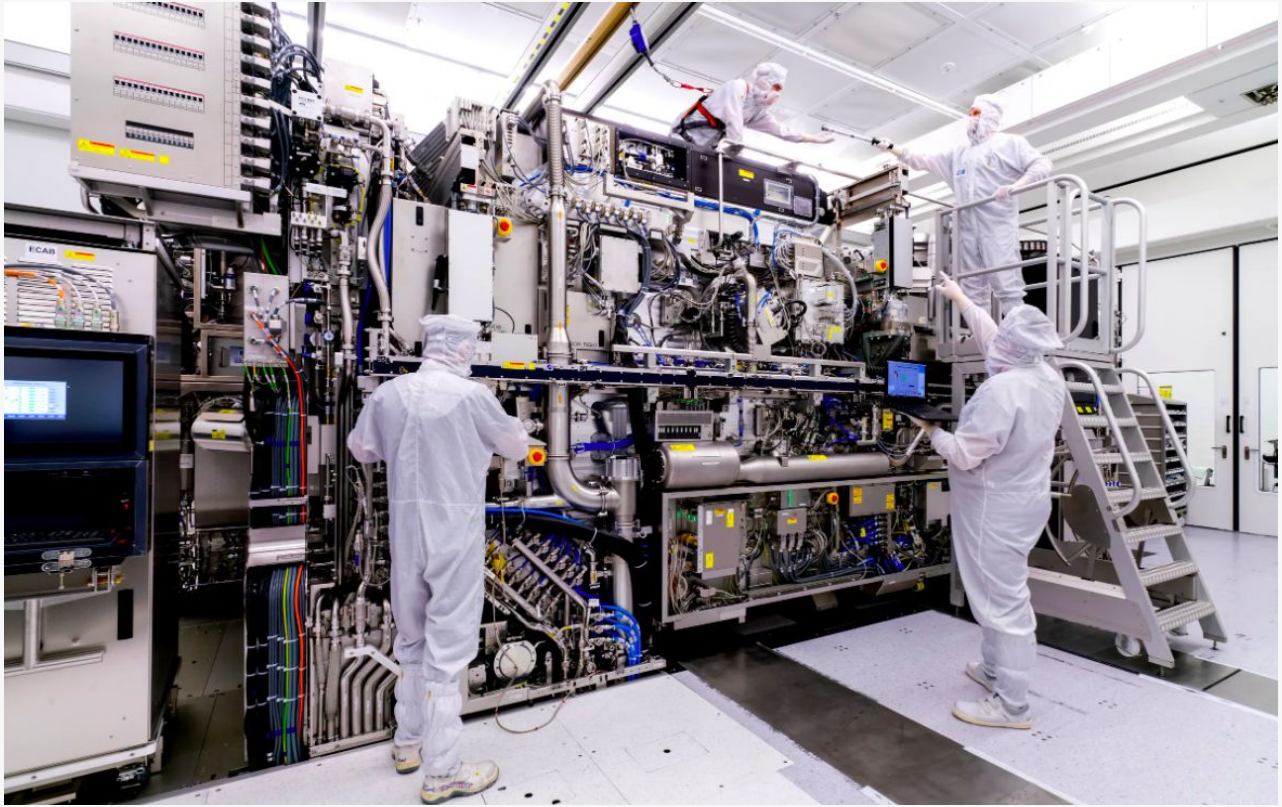
EUV - Lithography machine



A ball of tin is pulverized by incredibly powerful lasers and explodes into a plasma measuring several times hotter than the surface of the sun.

ASML

ASML - Lithography machine



The GPU revolution of AI

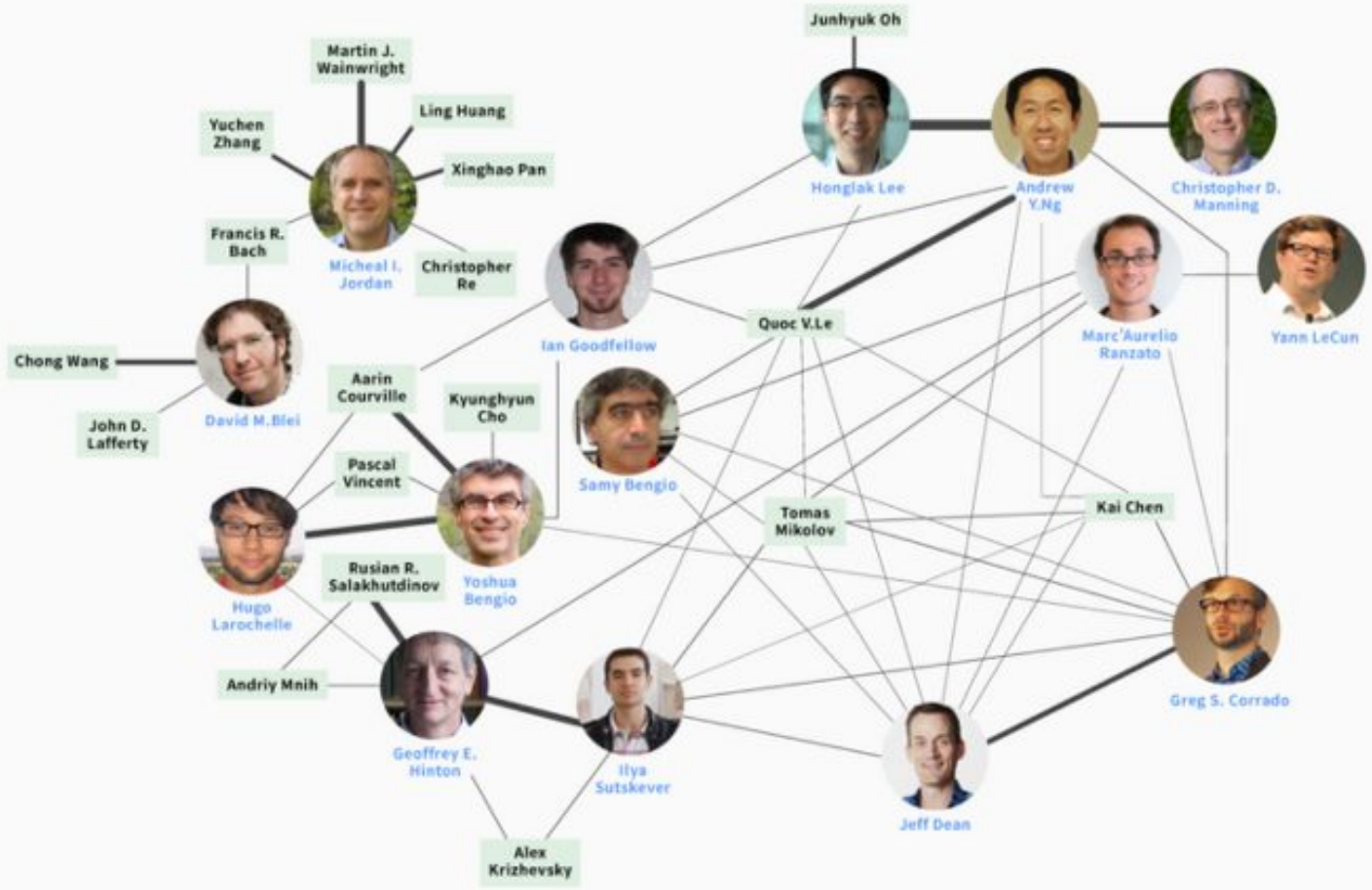
- NVIDIA and CUDA
- TSMC
- ASML
- INTEL & AMD
- ARM PROCESSORS
- APPLE M-Series CHIPS

Chip War Article: Understanding the GPU revolution of AI

9A. Research Landscape

GOFAI research
Seminal Papers in AI
Arxiv
Research Community - Academia, Industry Labs
Key People and ideas
Conferences
Twitter and debates
How to read ML papers
Bringing your research ideas to life - experimentation and publishing work
Github / Huggingface

Who knows the AI Researchers ?


Junhyuk Oh
Martin J. Wainwright
Ling Huang
Yuchen Zhang
Xinghao Pan
Francis R. Bach
Micheal I. Jordan
Christopher Re
Honglak Lee
Andrew Y.Ng
Christopher D. Manning
Ian Goodfellow
Quoc V.Le
Marc'Aurelio Ranzato
Yann LeCun
Chong Wang
David M.Blei
John D. Lafferty
Aarin Courville
Kyunghyun Cho
Samy Bengio
Pascal Vincent
Tomas Mikolov
Kai Chen
Hugo Larochelle
Rusian R. Salakhutdinov
Yoshua Bengio
Andriy Mnih
Geoffrey E. Hinton
Ilya Sutskever
Jeff Dean
Greg S. Corrado
Alex Krizhevsky

AI Conferences

Neural Information Processing Systems (**NeurIPS**): One of the largest and most prestigious conferences in machine learning and AI. It covers a wide range of topics including deep learning, reinforcement learning, computer vision, and more.

Conference on Computer Vision and Pattern Recognition (**CVPR**): A premier conference for computer vision research, covering topics such as image and video analysis, object recognition, and scene understanding.

International Conference on Machine Learning (**ICML**): A leading conference in machine learning that covers a broad spectrum of topics, including deep learning, optimization, theoretical foundations, and more.

International Joint Conference on Artificial Intelligence (**IJCAI**): A major AI conference that covers a wide range of AI-related topics, including machine learning, robotics, natural language processing, and more.

Association for Computational Linguistics (**ACL**): A key conference for natural language processing and computational linguistics, covering areas like language understanding, generation, and translation.

International Conference on Learning Representations (**ICLR**): A conference focused on the study of deep learning representations and optimization, with an emphasis on novel architectures and techniques.

Labs

Google AI

DeepMind

OpenAI

Microsoft
Research

AI2
Allen Institute for AI

Tencent AI Lab

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

9B. Startups

Core:

Gen AI Startups - Current trends

Silicon Valley vs Bengaluru

Creating a product - Levels

Other:

Building a team

Fundraising fundamentals

The startup mindset and the inevitable grind

GenAI startup opportunities

Module 10

Course Recap

- What have we learnt?
- Student flash presentations
- How can we improve the course?
- How to scale the community?

Capstone Project

Course credits

Ashwin

Gautham

Anuj

Shivansh

Sai

Karthik

Vishnu

Pranav

Smit

Vig