Introduction to Deep Learning

The Team

Lecturer



Prof. Angela Dai

TAs



Jiahui Fan



Haoxuan Li



Chandan Yeshwanth

Student Tutors



Amine Mnif



Barbora Novosadova



Hadar Sleiman



Pablo Díaz

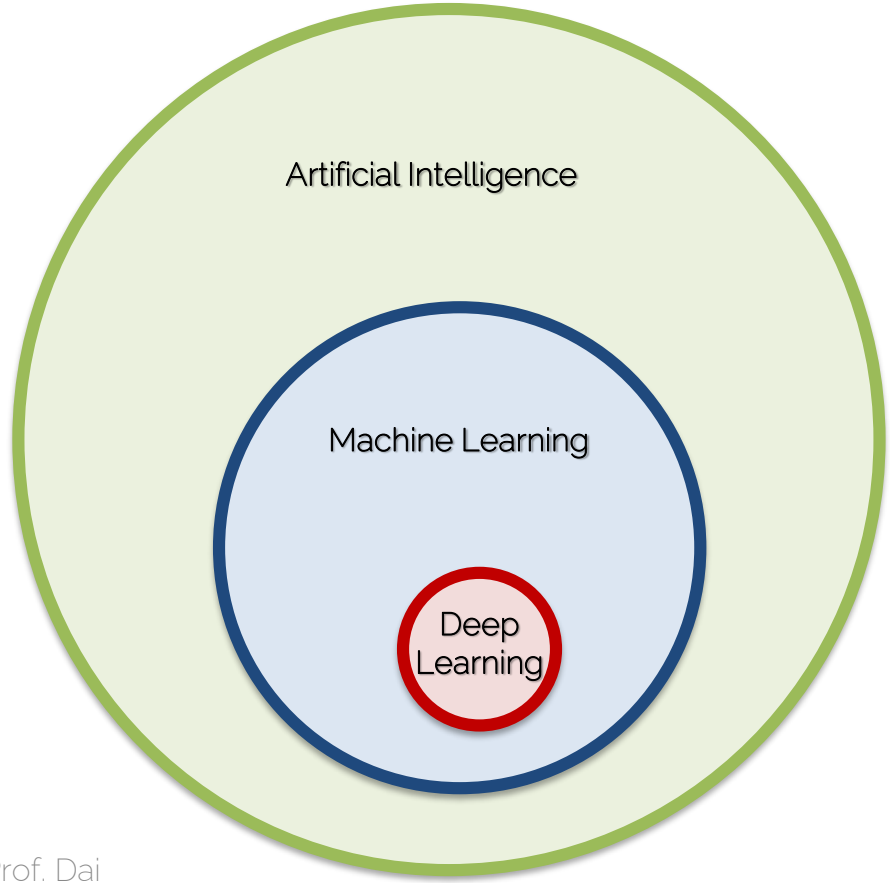


Shiyuan Zhang

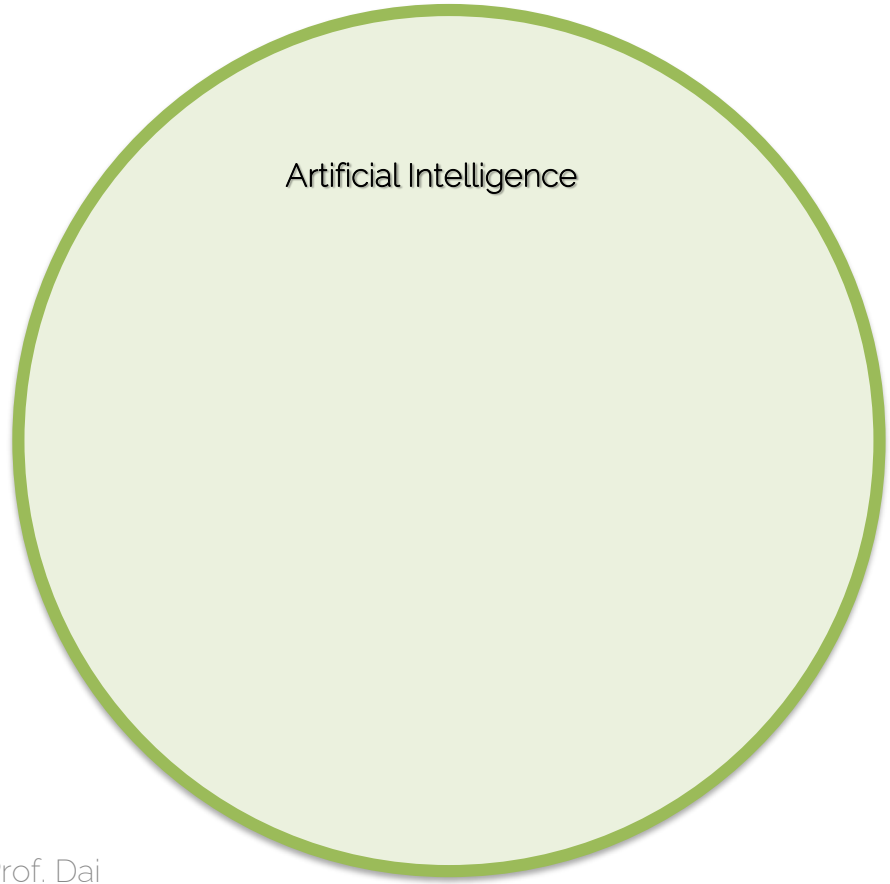


Ting Li

What is this lecture about?



What is this lecture about?

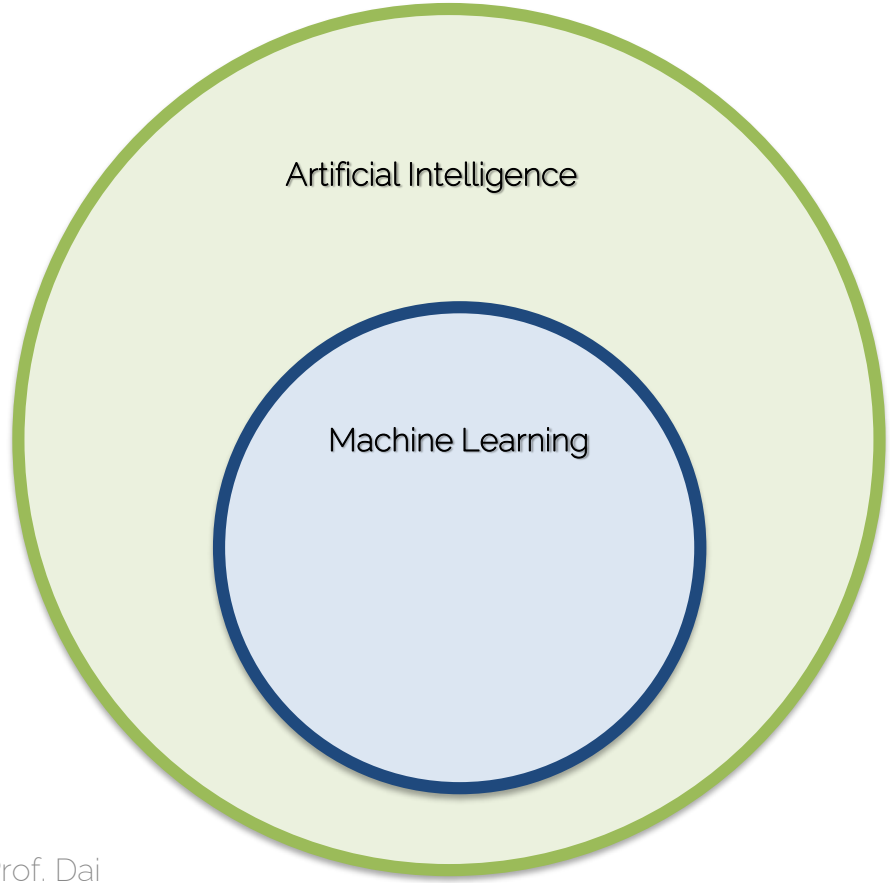


- AI Methods
  - Broad definition!
    - "if" statements"

```
if (cold)
    turn_up_heat();
```

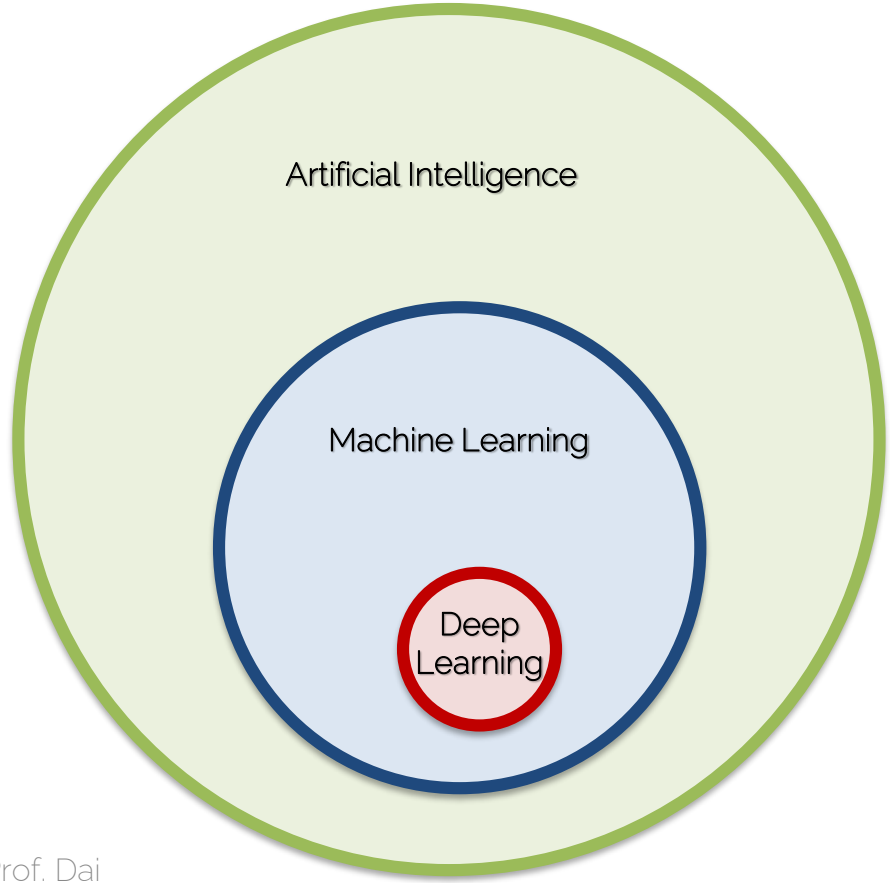
    - Binary Search
    - Dijkstra, A*, …
    - Prim, Kruskal, …
    - Logic algorithms, etc.
  - ….

What is this lecture about?



- ML Methods
  - Linear/logistic regression
  - Support Vector Machines
  - Random Decision Trees, Forests, Jungles, …
  - …

What is this lecture about?



- Deep Learning
  - ML-methods leveraging neural networks
    - Multi-layer perceptrons
    - Convolutional neural networks
    - Recurrent neural networks
    - Transformers
    - Generative models, etc…

Computer Vision

What's in the image?



What does a machine see?

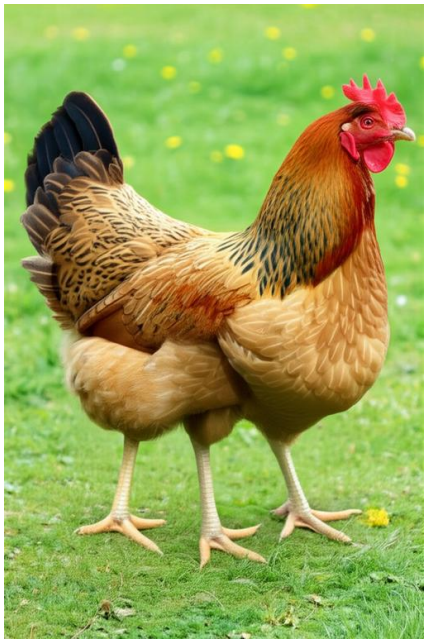


Wikimedia Commons

What does a machine see?



What's in the image?



How many legs?



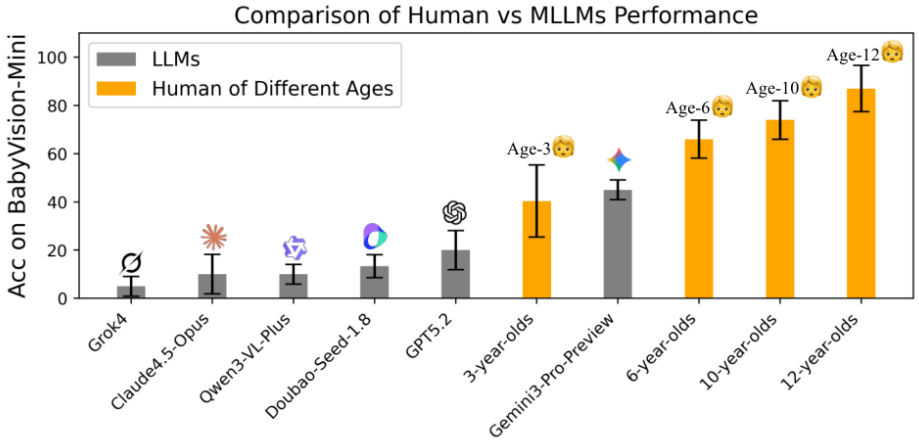
How many rings in the logo?

[Vo et al., ICML'25] Vision Language Models are Biased

What's in the image?



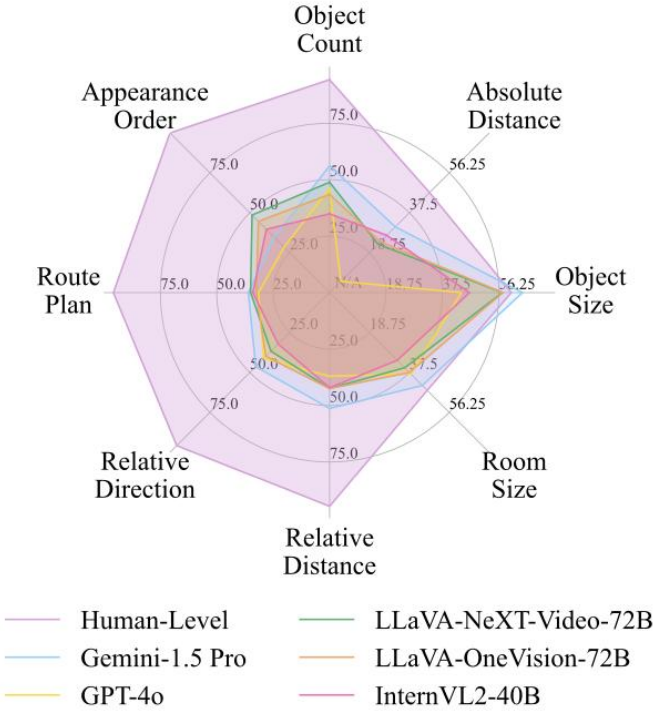
Which line leads to which cup?

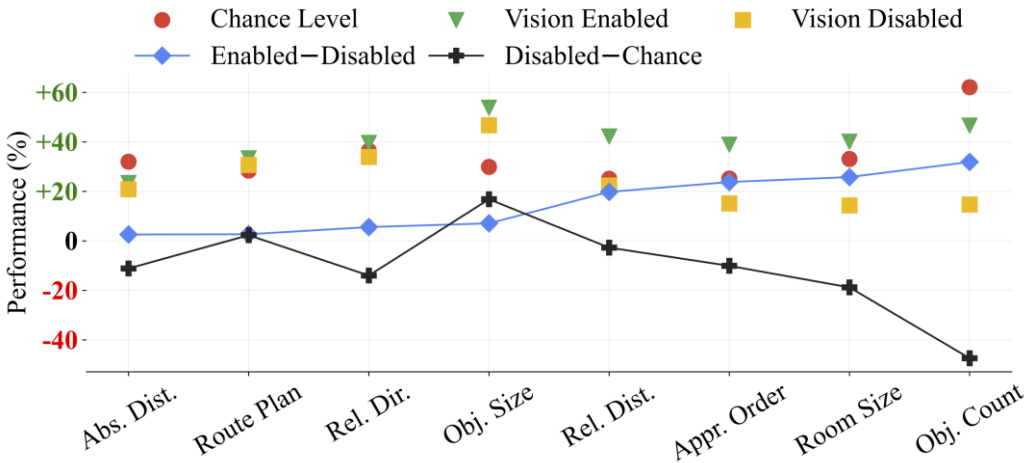


Full benchmark: best model 49.7%, adults 94.1%, age 6 ≈ 70%

[Chen et al., arXiv'26] BabyVision

What's in the scene?





Video questions about a room: humans 79%, best model 49%

[Yang et al., CVPR'25] Thinking in Space

What is Computer Vision?

- First defined in the 60s in artificial intelligence groups
- “Mimic the human visual system”
- Center block of robotic intelligence



MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

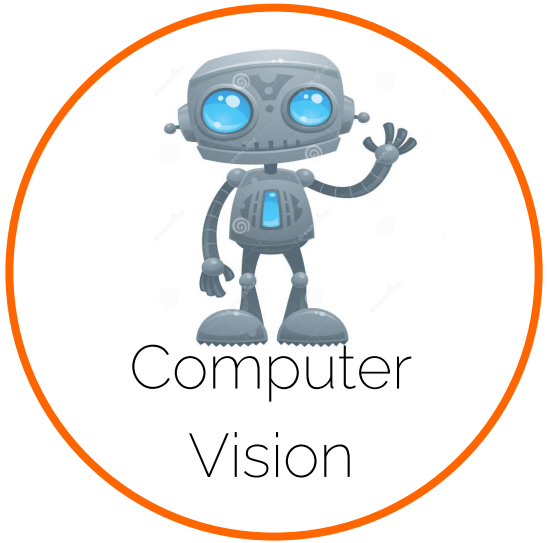
July 7, 1966

THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

A Few Decades Later…



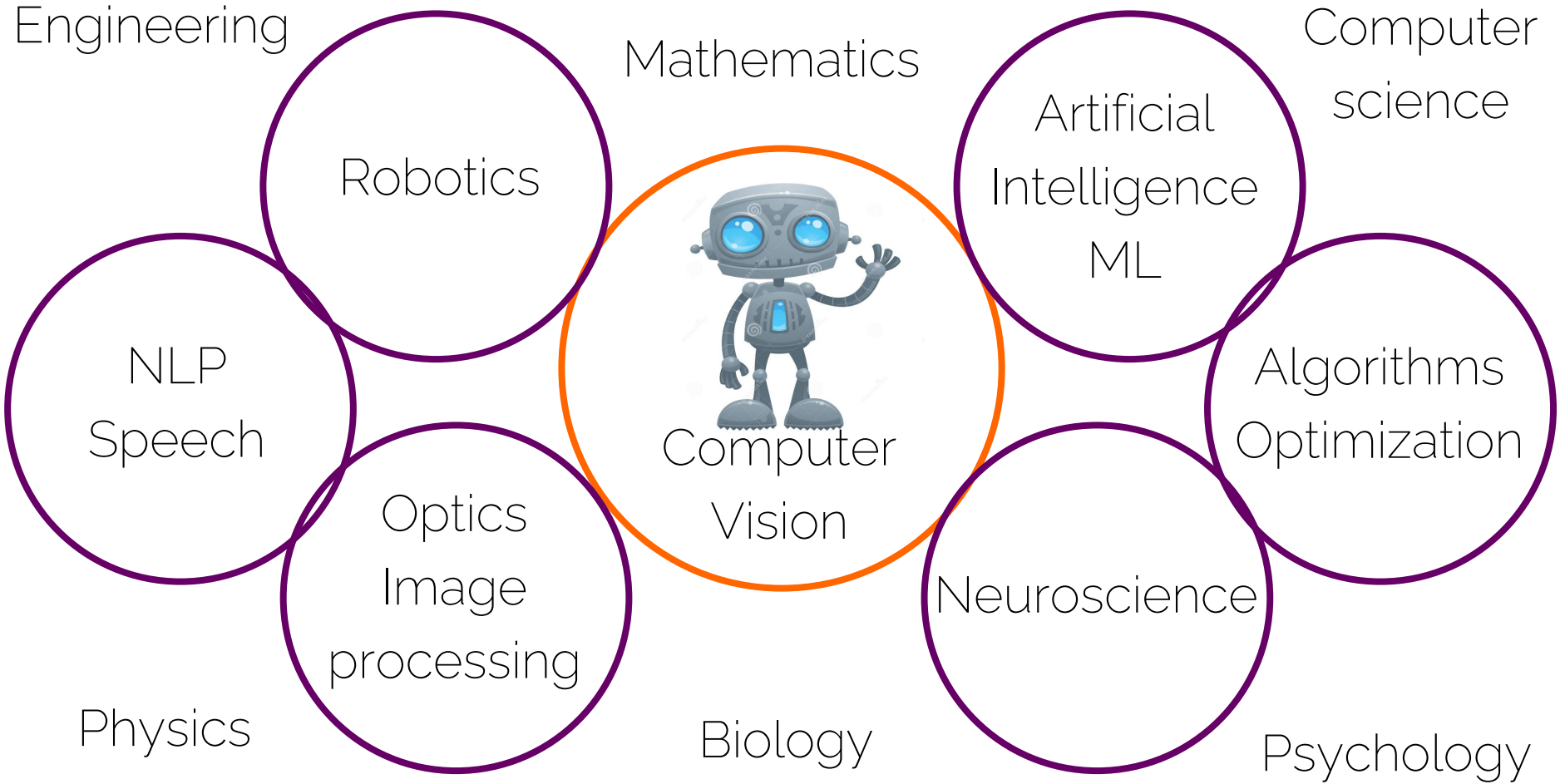

Engineering
Mathematics
Computer science
Robotics
Artificial Intelligence ML
NLP Speech
Computer Vision
Algorithms Optimization
Optics Image processing
Neuroscience
Physics
Biology
Psychology

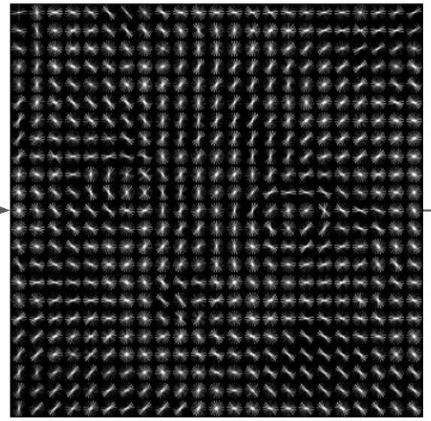
Image Classification



Input image



Preprocessing



Features : HAAR, HOG, SIFT, SURF

Learning Algorithm

SVM, Random Forests, ANN

Label Assignment

Cat or Background



Image Classification



Input image

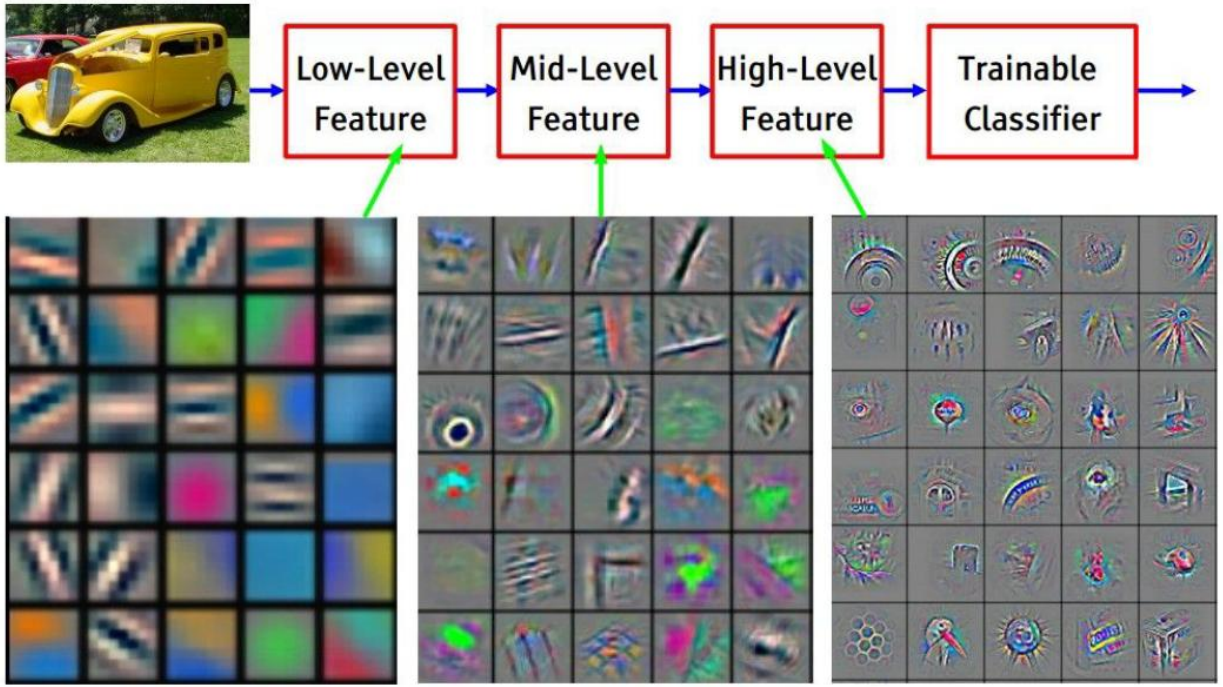
Deep Neural Network

Label Assignment

Cat or Background

Post 2012

Features are Learned

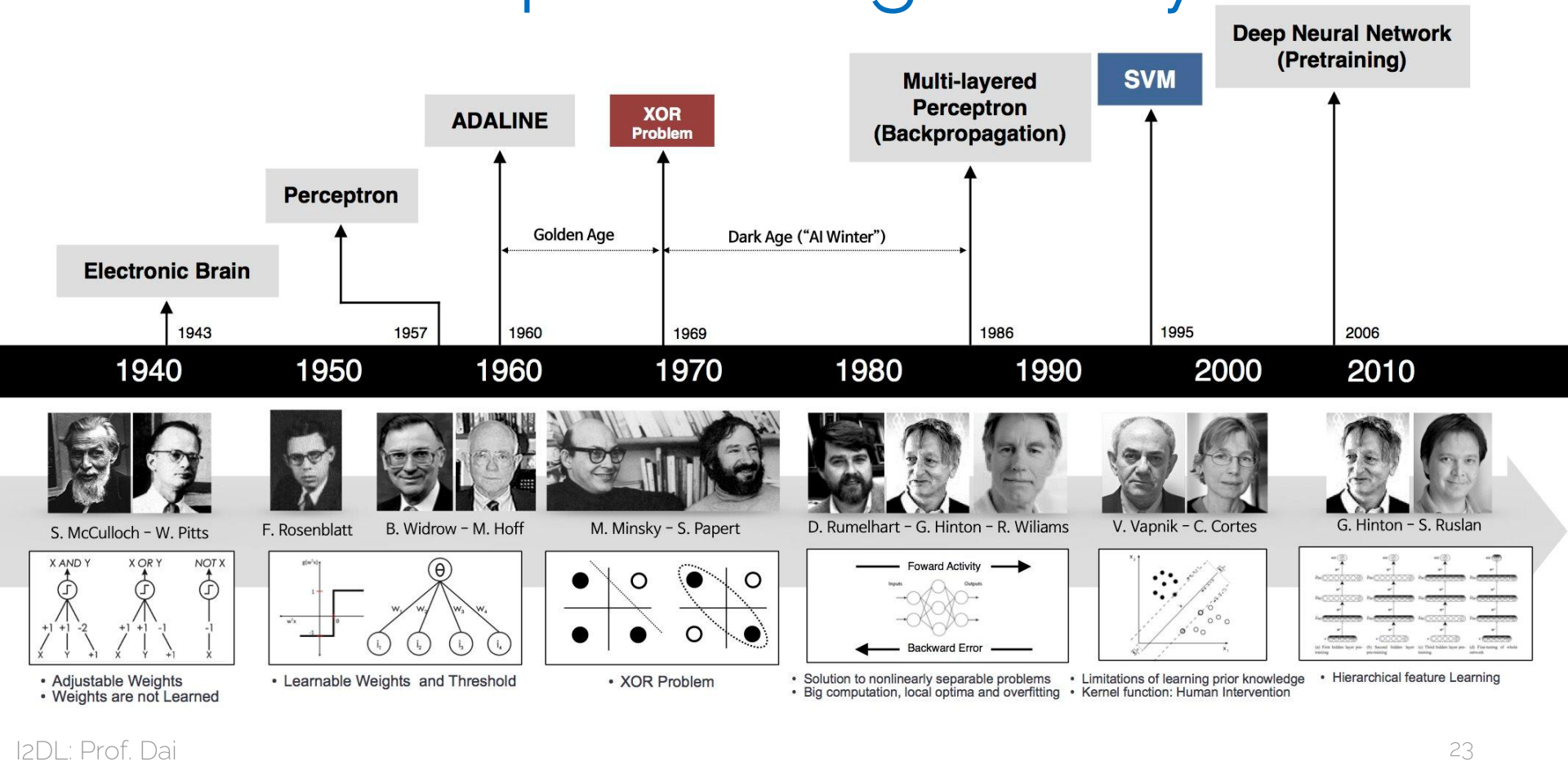


Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

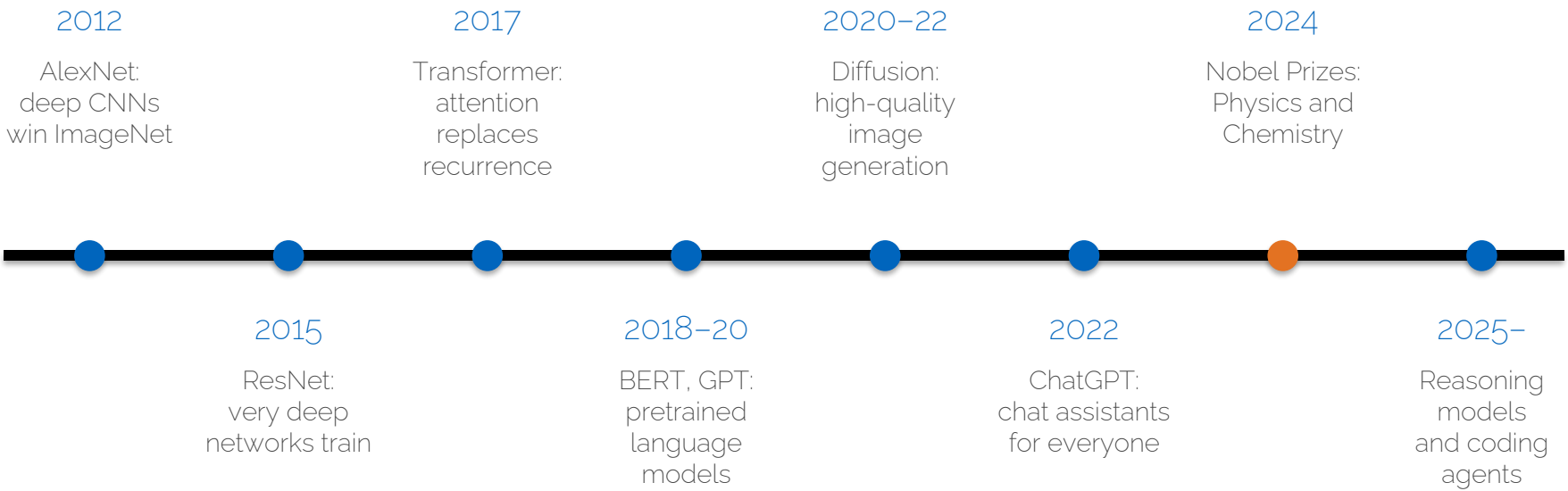
[Zeiler & Fergus, ECCV'14]

Why Deep Learning?

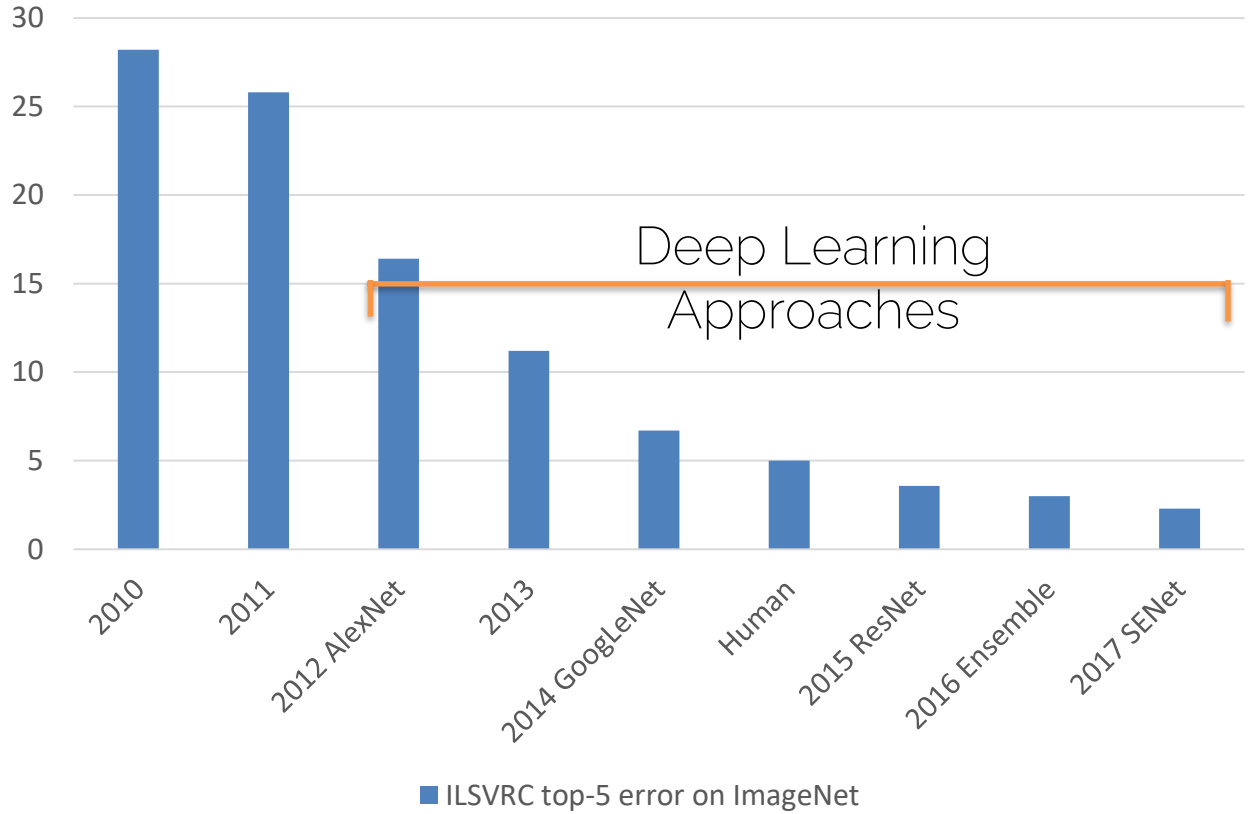
Deep Learning History



Deep Learning History (2012–)

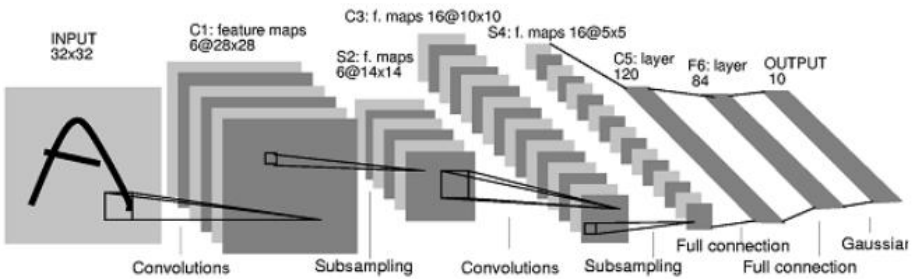


Modern Deep Learning in Computer Vision



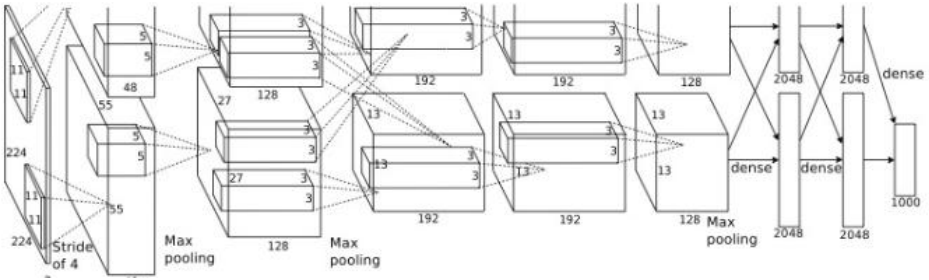
What has Changed?

1998
LeCun
et al.



- MNIST digit recognition dataset
- 10^7 pixels used in training

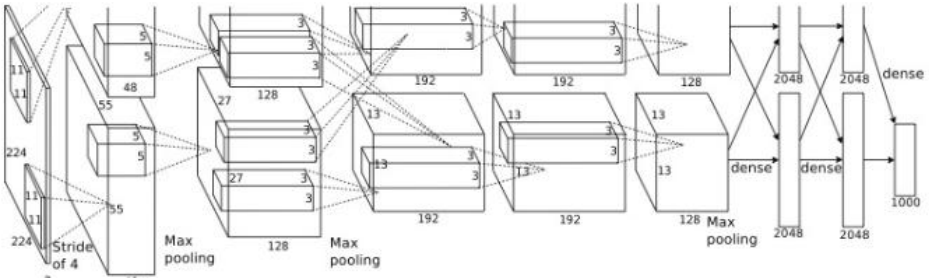
2012
Krizhevsky
et al.



- ImageNet image recognition dataset
- 10^{11} pixels used in training

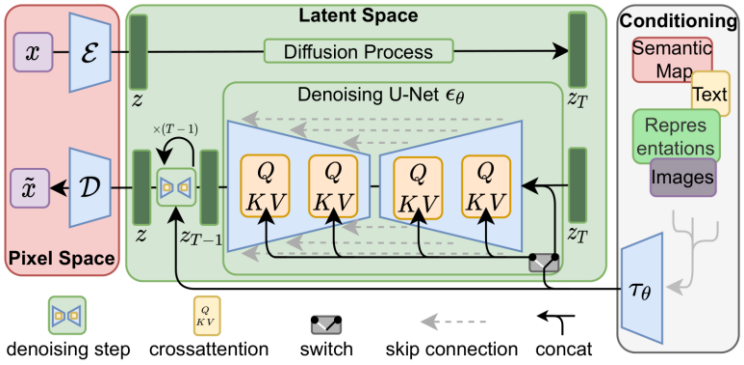
What has Changed?

2012
Krizhevsky
et al.



- ImageNet image recognition dataset
- 10^{11} pixels used in training

2022
Rombach
et al.



- LAION-5B web image dataset
- 10^{15} pixels used in training

What has Changed?

	Est. GPUs	Est. compute (FLOP)	Est. cost of final run
2012 AlexNet	2 × GTX 580	5×10^{17}	$16
2020 GPT-3	10,000 V100	3×10^{23}	$2M
2023 GPT-4	25,000 A100	2×10^{25}	$40M
2024 DeepSeek-V3	2,048 H800	3×10^{24}	$5.6M*
2025 GPT-4.5	H100, count n/a	4×10^{26}	$370M
2026 GPT-6 Astra	100,000+ GB200	10^{27}	n/a

* self-reported by DeepSeek Data: Epoch AI (Oct 2026)

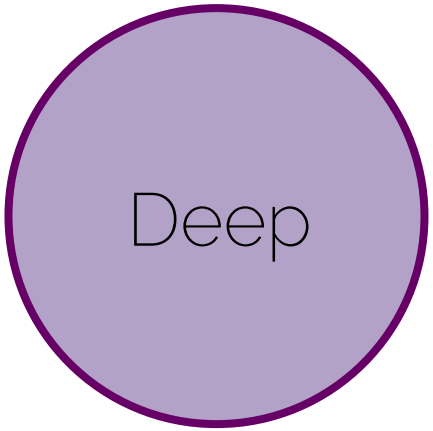
What Made this Possible?



Models know where to learn from



Models are trainable



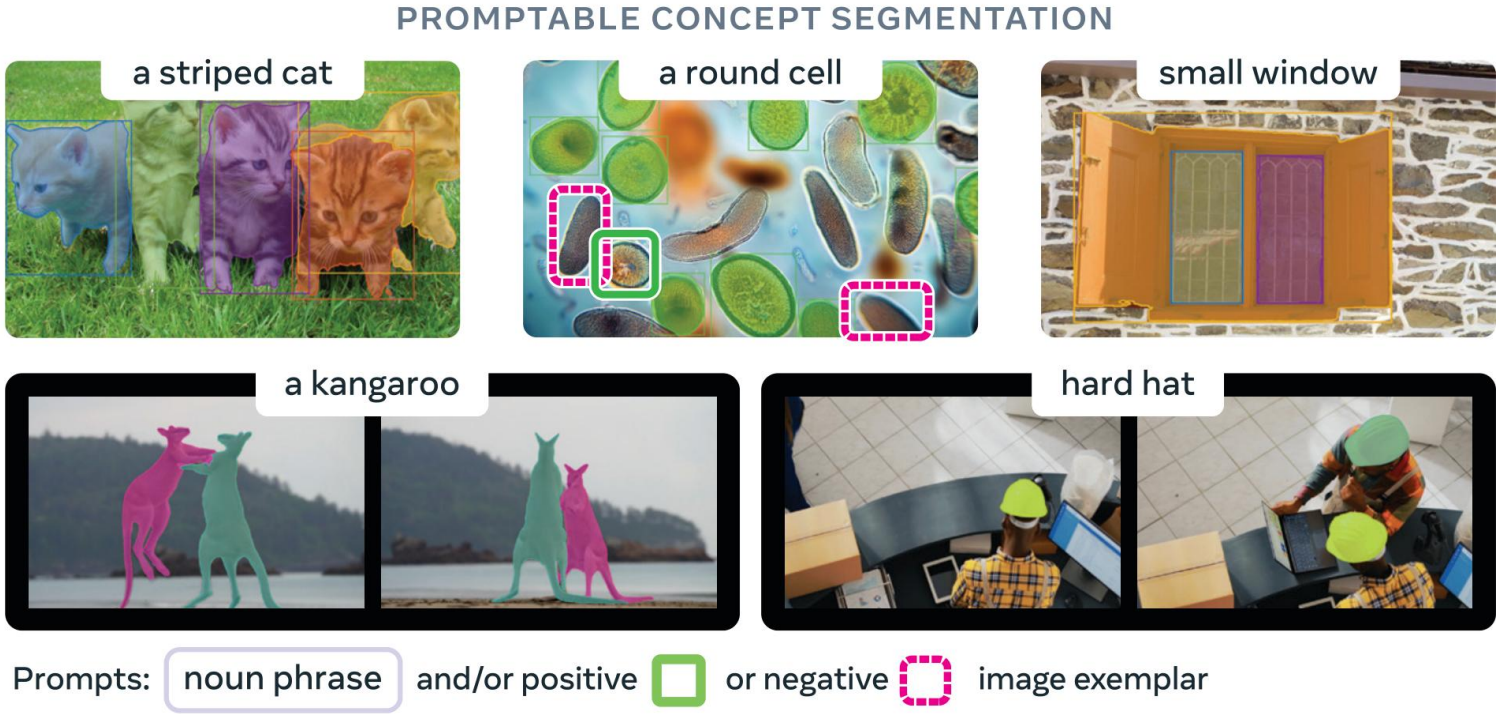
Models are complex

Deep Learning Today



Object Detection

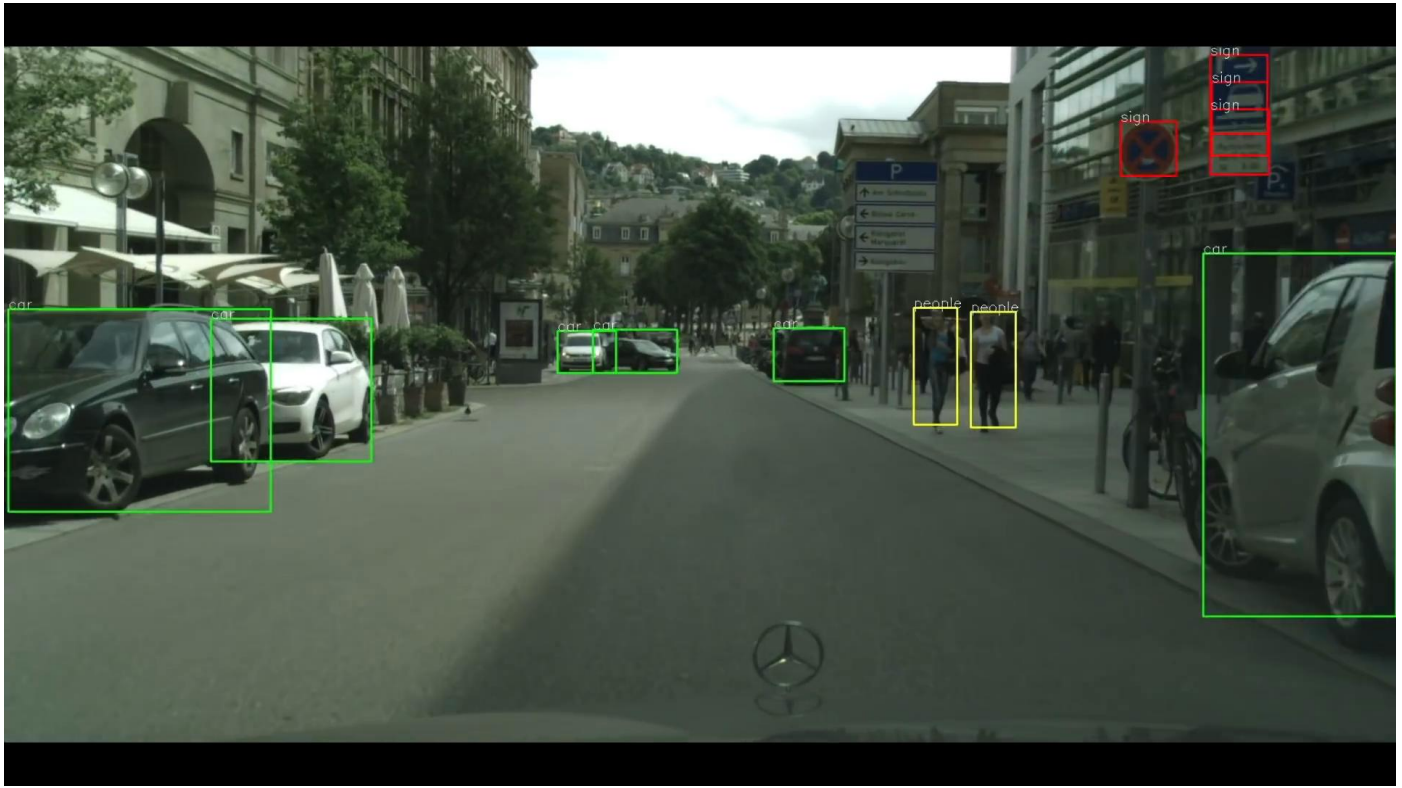
Deep Learning Today



Segment everything that matches a text prompt, in images and video

[Meta, 2025] SAM 3

Deep Learning Today

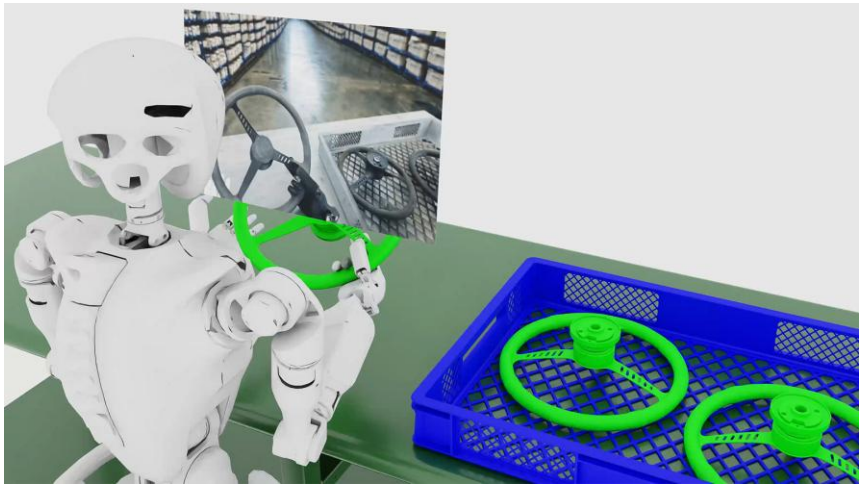


Self-driving cars

Deep Learning Today



Robots that act from vision and language

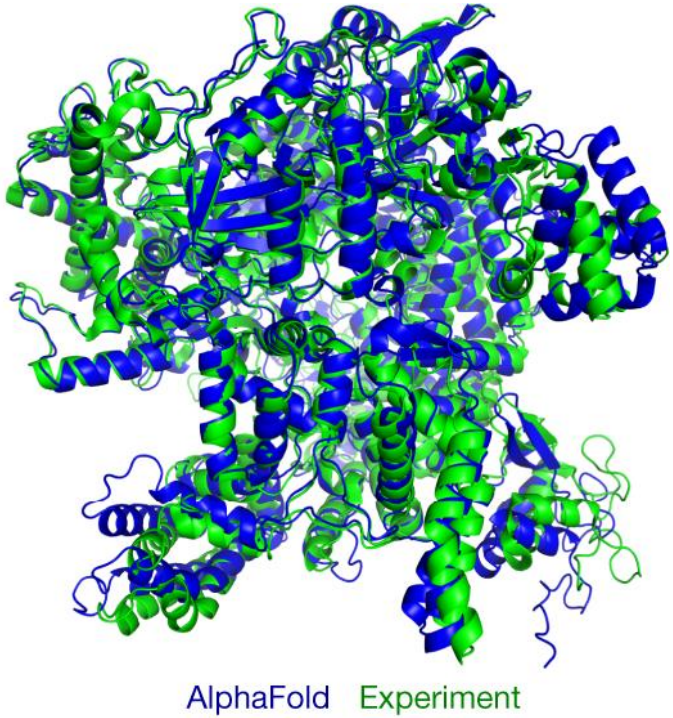


World models that simulate training scenes

Physical AI

[Google DeepMind, 2025] Gemini Robotics 1.5 · [NVIDIA, 2025] Cosmos

Deep Learning Today

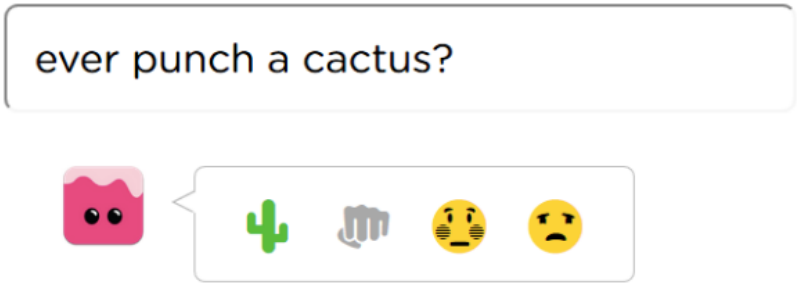


Biological research

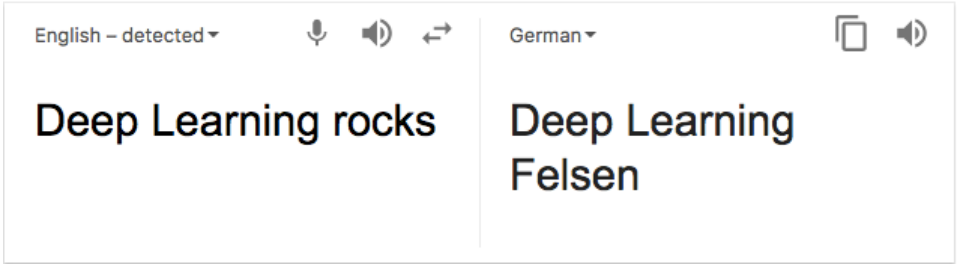
Deep Learning Today



AlphaGo

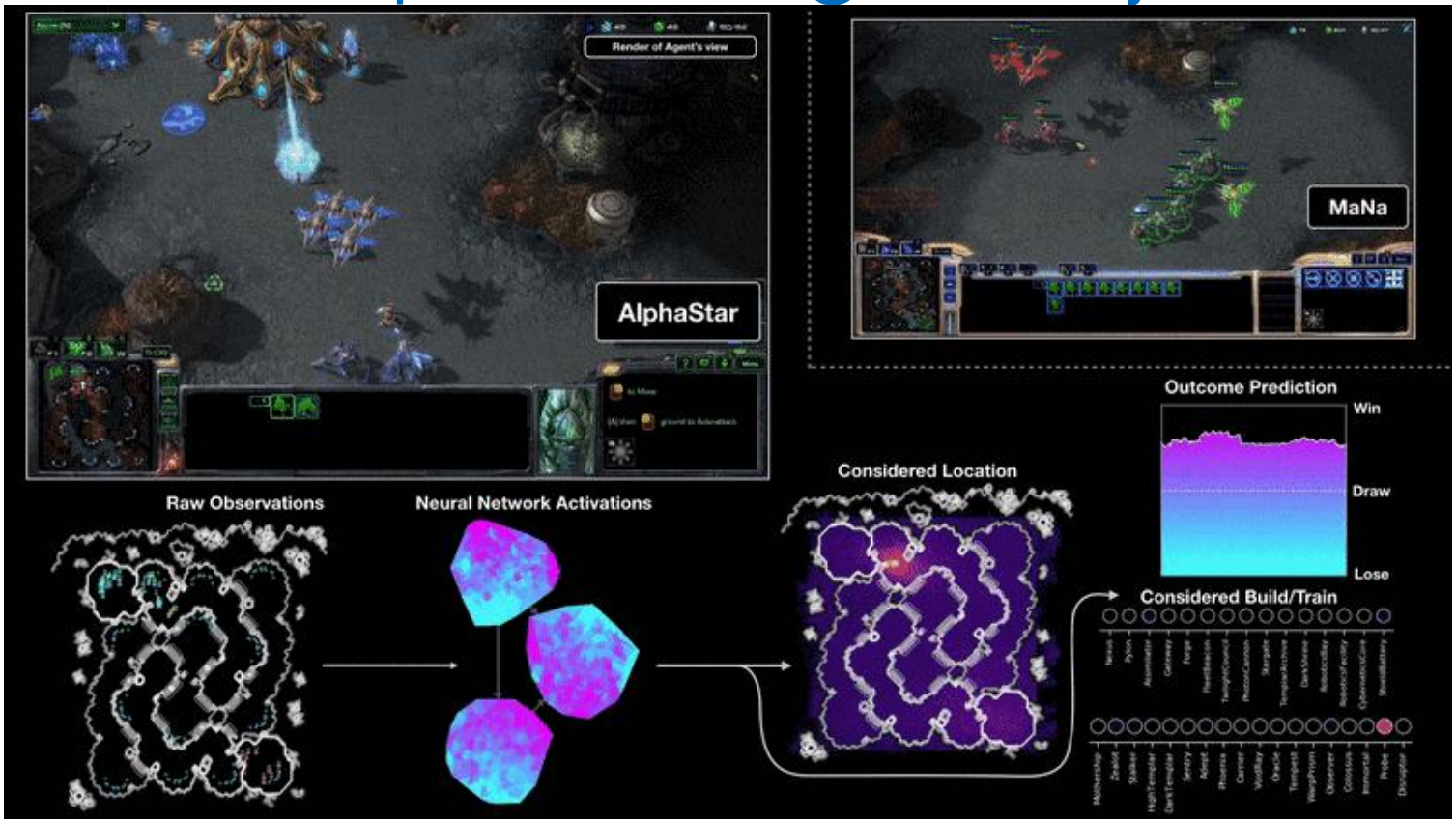


Emoticon suggestion



Machine translation

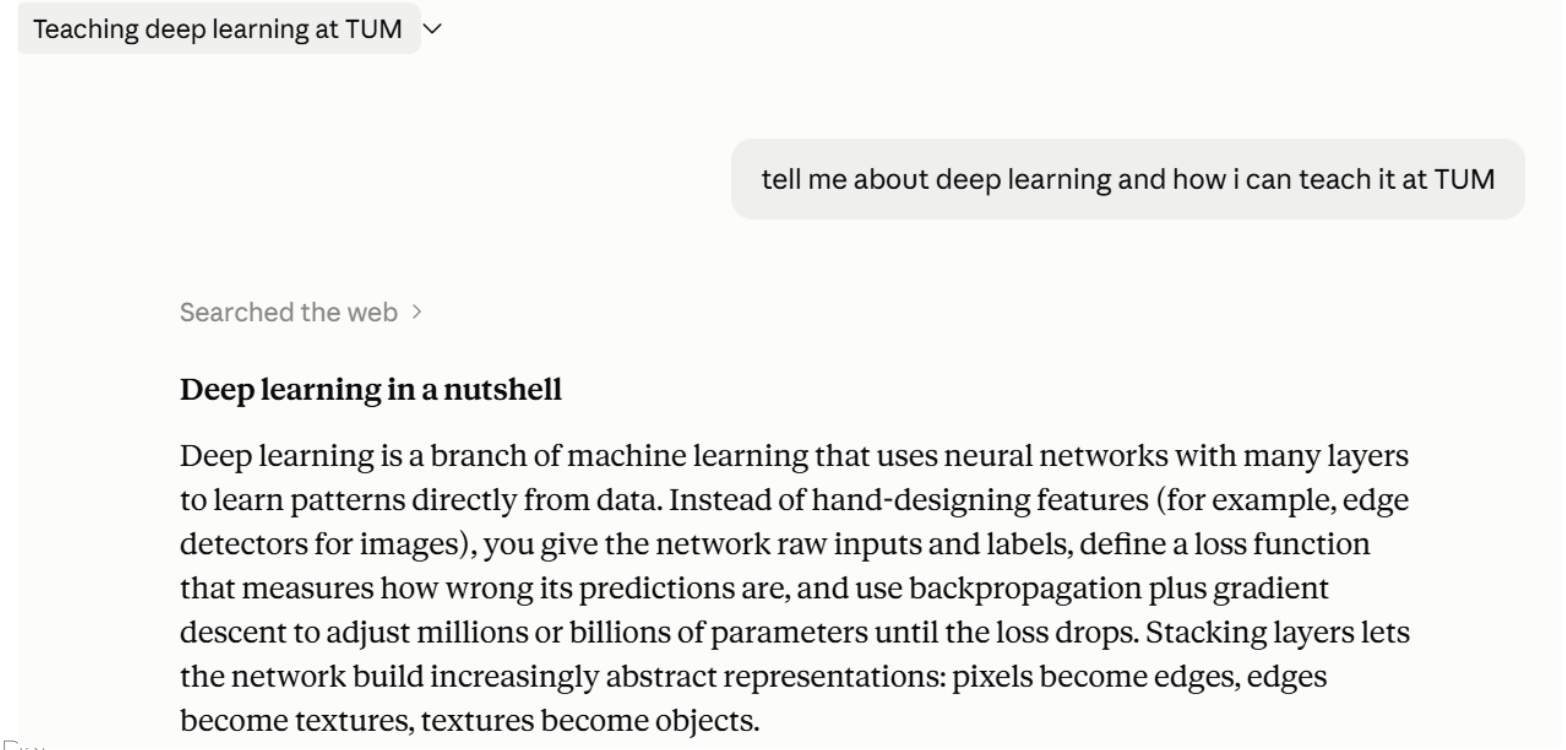
Deep Learning Today



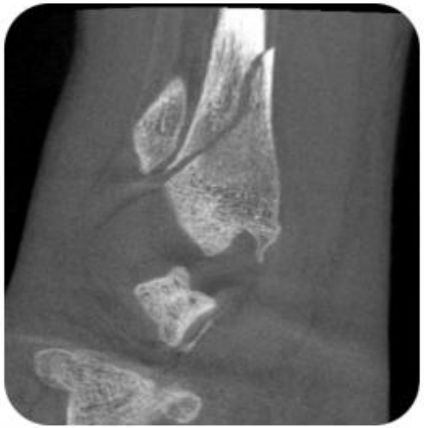
AlphaStar

Deep Learning Today

- Large Language Models (LLMs), e.g., ChatGPT, Claude



Deep Learning Today



This is a scan of my ankle, showing the fibula. Anything wrong? Short answer only; need additional assessment to radiologist.

Short answer: **Yes, this looks abnormal.**

There appears to be a **fracture of the distal fibula** with **displacement of bone fragments**, which can indicate an unstable injury.

Deep Learning Today



"a robot painting a colorful sunset on a canvas in a futuristic art studio"



"a majestic dragon made of water flying through a colorful, abstract city"

https://huggingface.co/black-forest-labs/FLUX.1-dev

Deep Learning Today



OpenAI Sora

Deep Learning Today



Trained jointly on images, video, sound



Fine-tuned to predict robot actions

[Black Forest Labs, 2026] FLUX 3 · FLUX 3 Action

Deep Learning Today



Text to video with sound, open weights



+ orbit around a photo: fine-tuned on 28 clips

[MiniMax, 2026] MiniMax H3 · 360° orbit LoRA: P. Dawson (Hugging Face, 2026)

Deep Learning Today

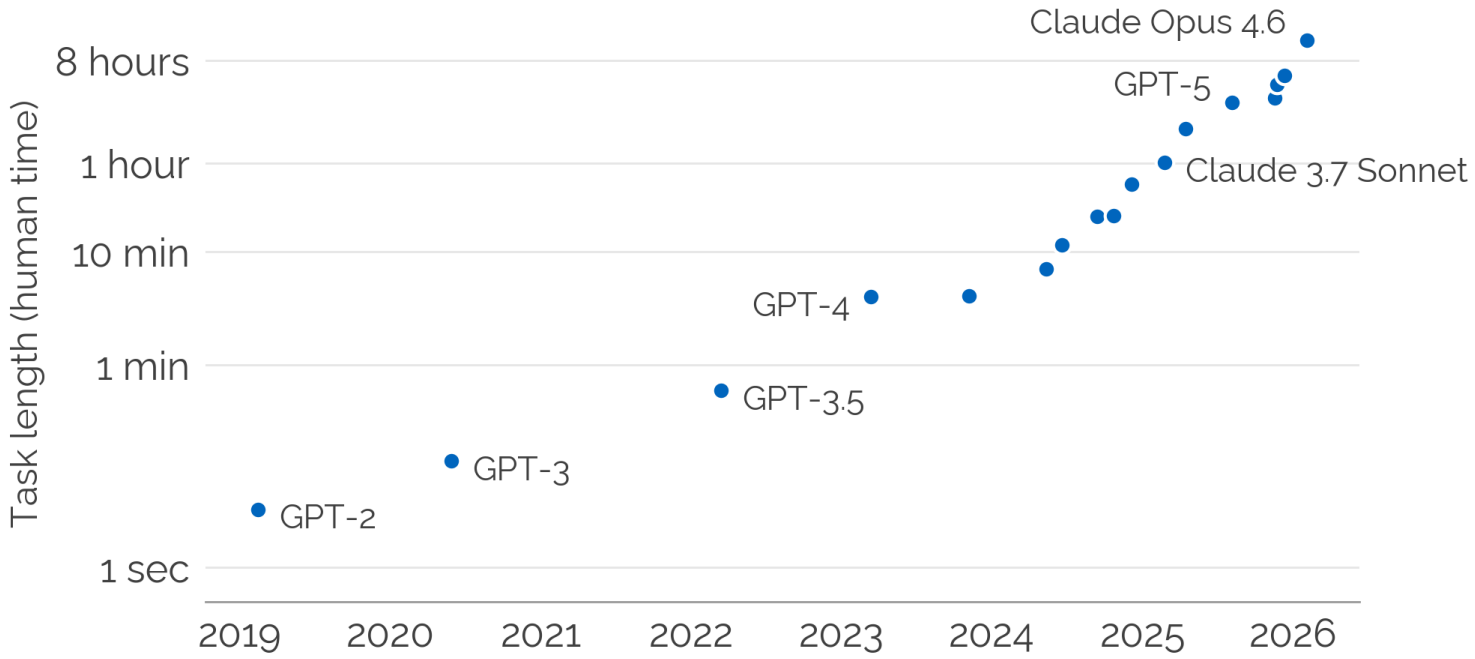


Deep Learning Today



Image to 3D World

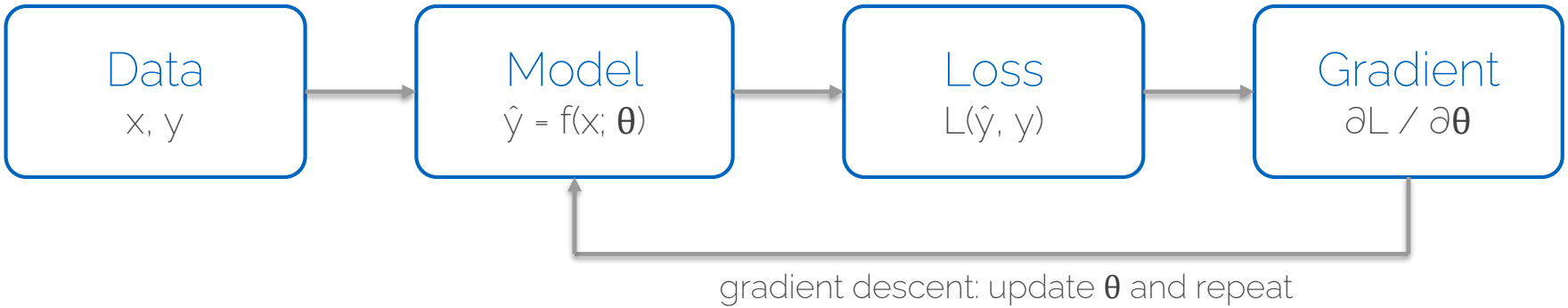
Deep Learning Today



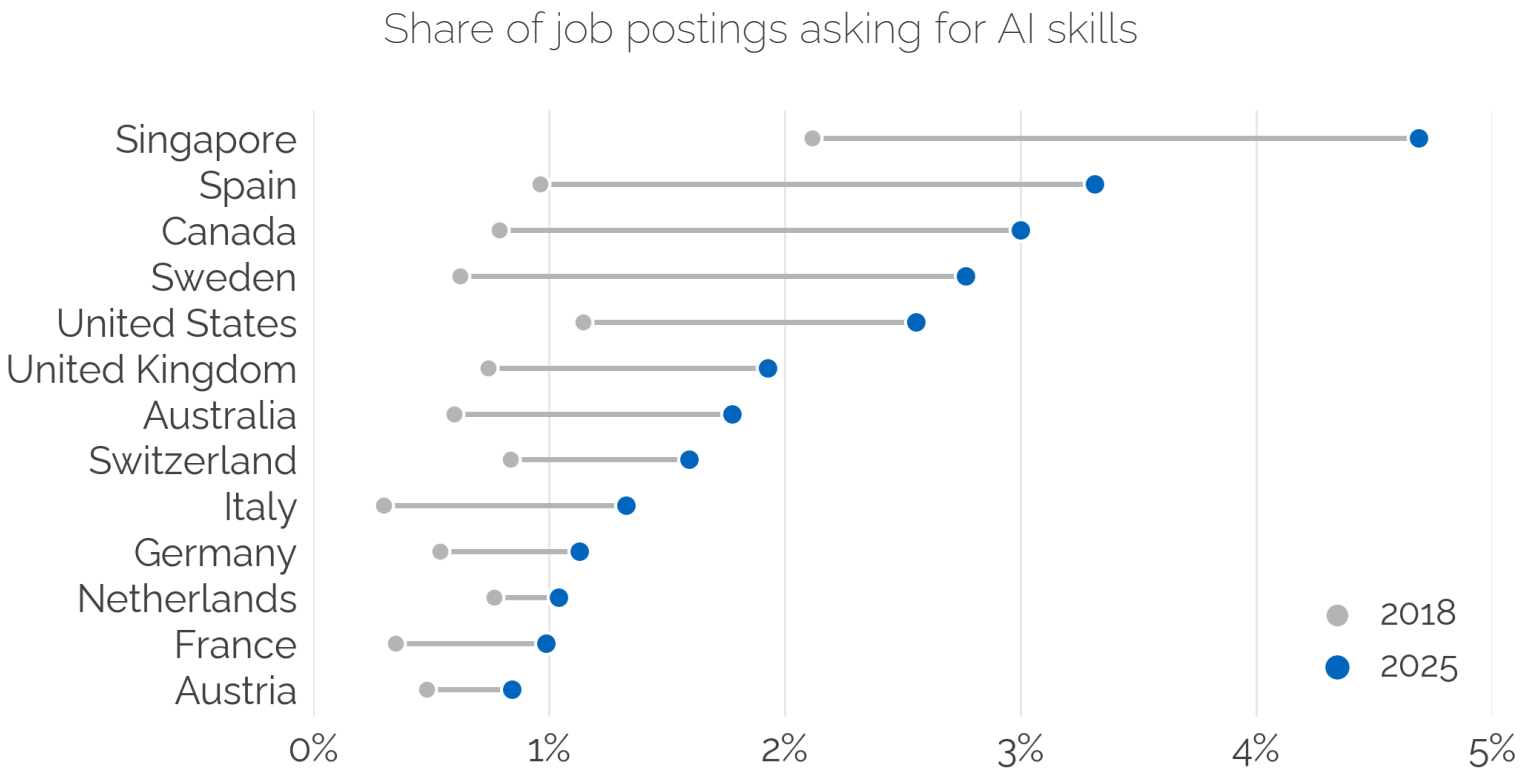
AI coding agents: how long a software task they can finish (50% success)

[METR, 2026] Task-completion time horizons

How do these Models Learn?



Deep Learning Job Perspective



Data: Lightcast via AI Index 2026

But also Challenging!

- High-level understanding is not enough
  - Lots of consumers, few creators
  - Need proper theory background
  - Need proper practical skillsets

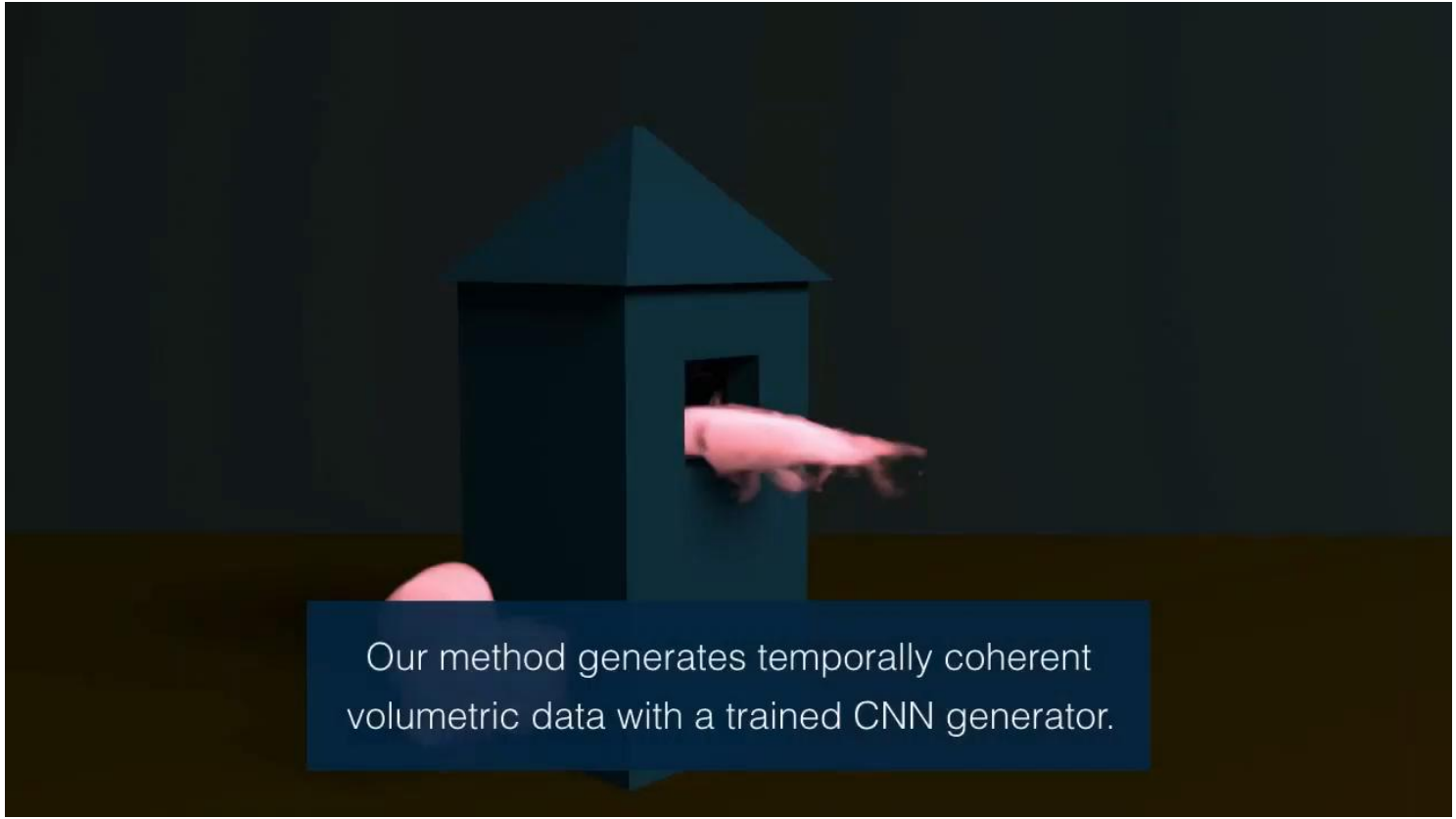
- Can be competitive!
  - Many good people
  - Education gap widens
  - Running scripts / implementation also by AI ☺

Deep Learning at TUM

Many TUM Research Labs use DL

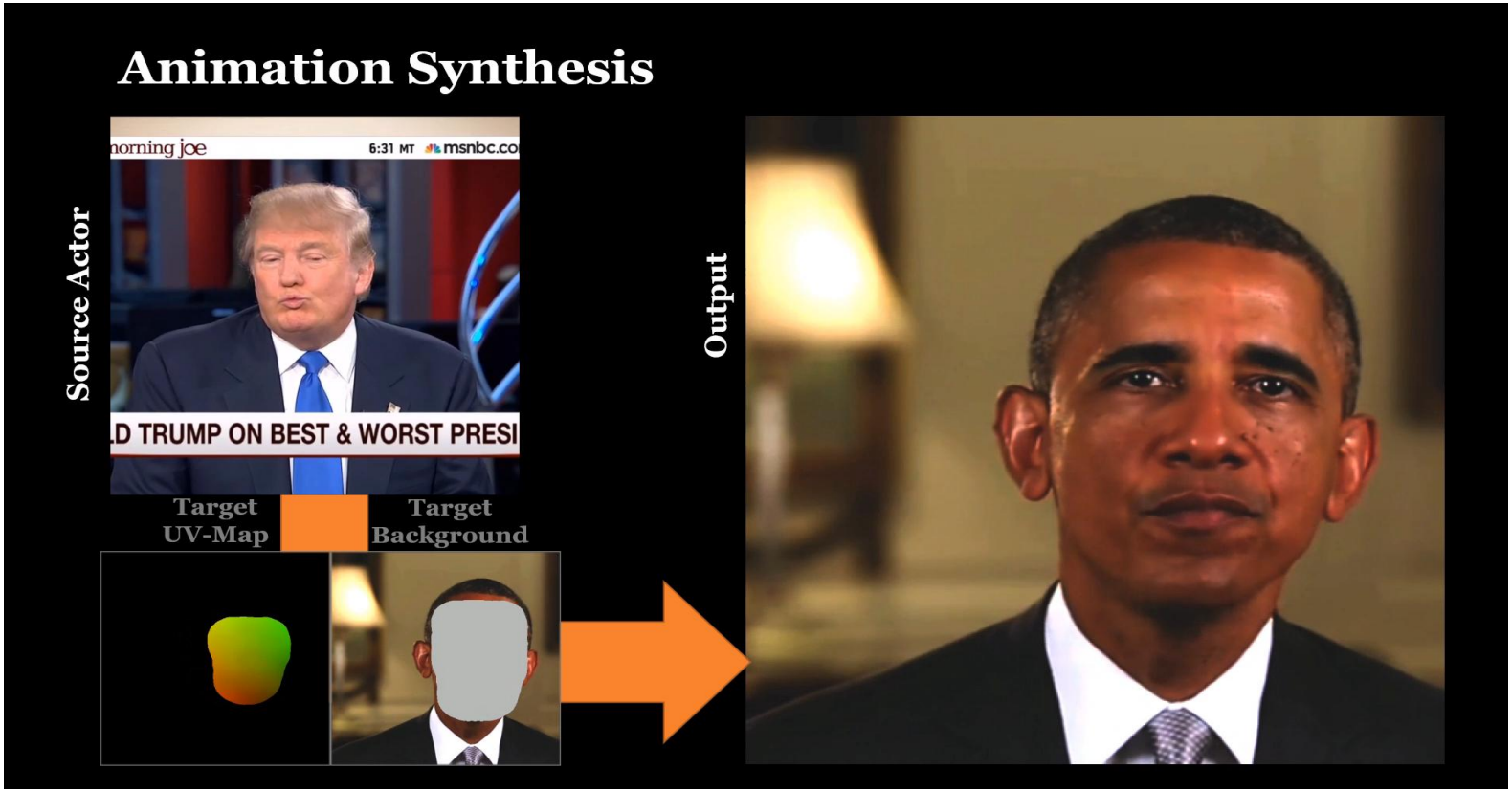
- 3D AI Lab (Prof. Dai)
  - Research in 3D perception, 3D scene understanding
- Visual Computing & AI (Prof. Niessner)
  - Research in computer vision, graphics, and machine learning
- Interpretable and Reliable Machine Learning (Prof. Akata-Schulz)
  - Research in explainable AI
- Robotics and Artificial Intelligence (Prof. Schoellig)
  - Research in Robotics and machine learning
- Computer Vision Group (Prof. Cremers)
  - Research in computer vision, machine learning and robotics
- Data Mining and Analytics Lab (Prof. Günnemann)
  - Research methods for robust machine learning
- Computer Aided Medical Procedures (Prof. Navab)
  - Research in machine learning for medical applications
- And many more ☺

Deep Learning at TUM



[Xie et al., SIGGRAPH'18] tempoGAN

Deep Learning at TUM



[Thies et al., SIGGRAPH'19] Neural Textures

Deep Learning at TUM



[Kirschstein et al., ToG'23] NeRSemble

Deep Learning at TUM



ScanNet Stats:
-Kinect-style RGB-D sensors
-1513 scans of 3D environments
-2.5 Mio RGB-D frames
-Dense 3D, crowd-source MTurk labels
-Annotations projected to 2D frames

[Dai et al., CVPR'17] ScanNet

Deep Learning at TUM



ScanNet++:
-1006 high-fidelity Scenes
- 1mm laser scans
-DSLR images
-iPhone RGB-D
-semantics

[Yeshwanth & Liu et al., ICCV'23] ScanNet++

Deep Learning at TUM



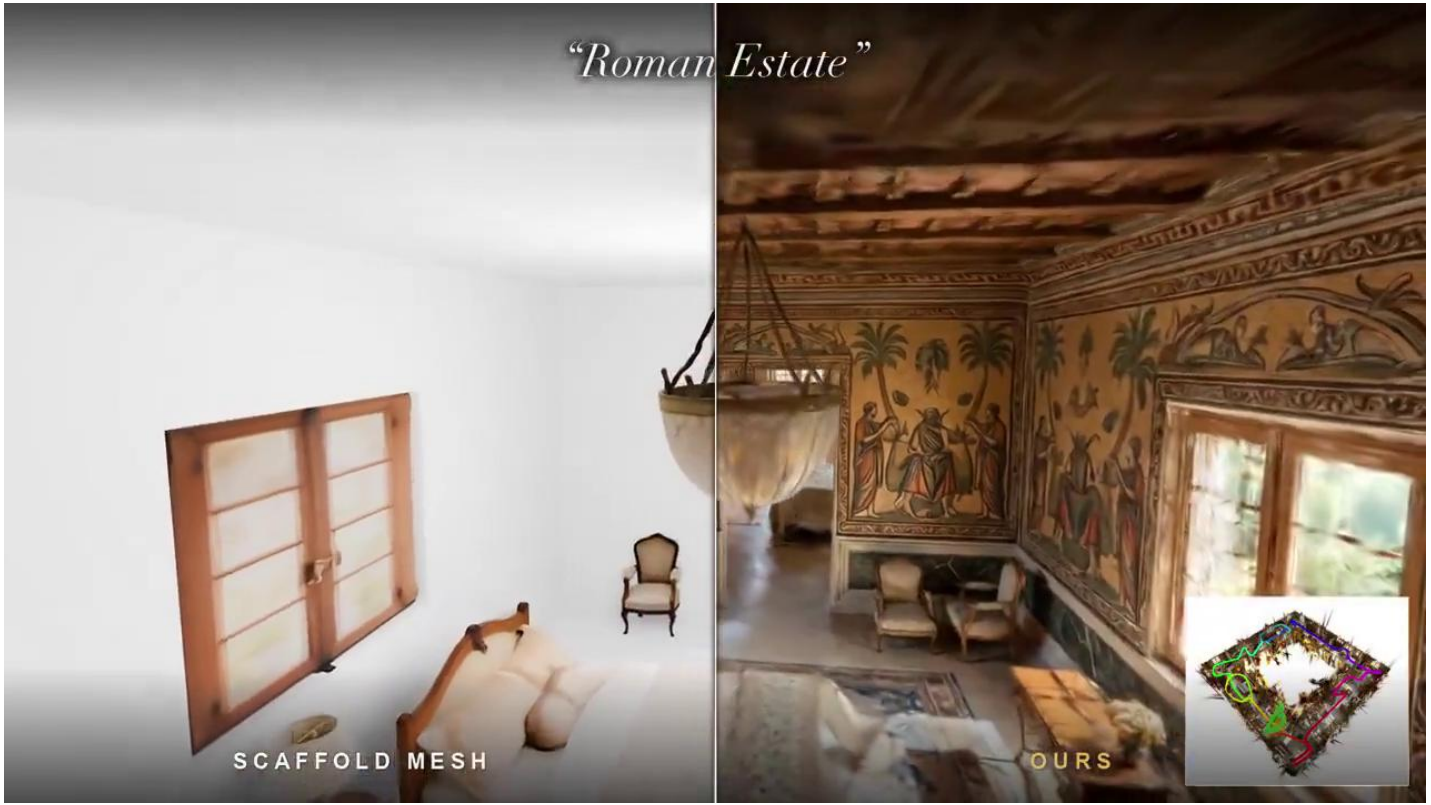
[Siddiqui et al., CVPR'24] MeshGPT

Deep Learning at TUM



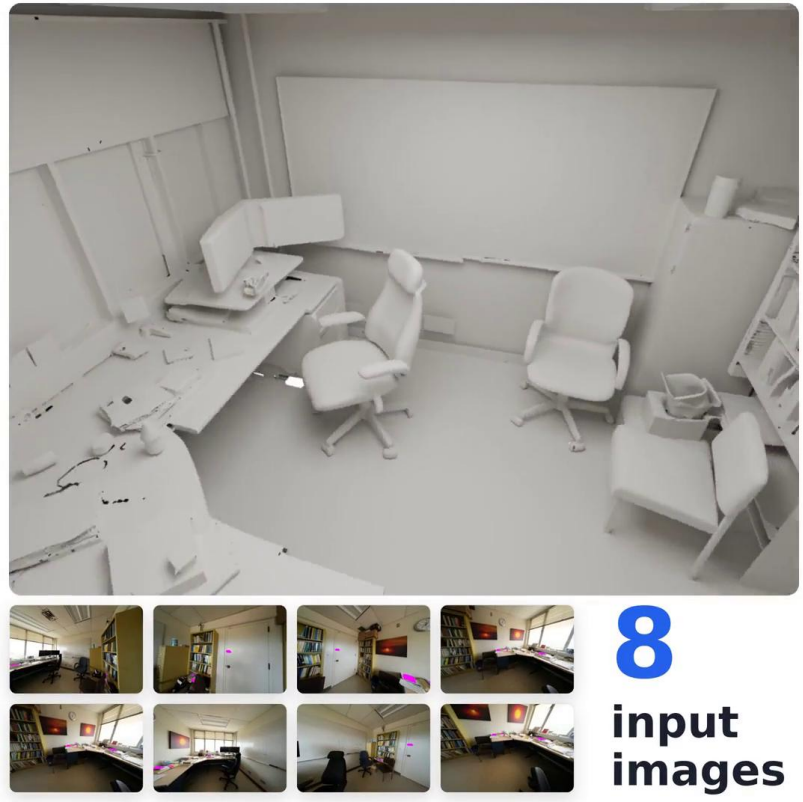
[von Lützow et al., arXiv'26] GaussianGPT

Deep Learning at TUM



[Schneider & Dai, ECCV'26] WorldMesh

Deep Learning at TUM



[Schmid et al., NeurIPS'26] GenRecon

Deep Learning at TUM



[Boudjoghra et al., NeurIPS'26] UnfoldArt

Deep Learning at TUM



[Li & Dai, CVPR'24] GenZI

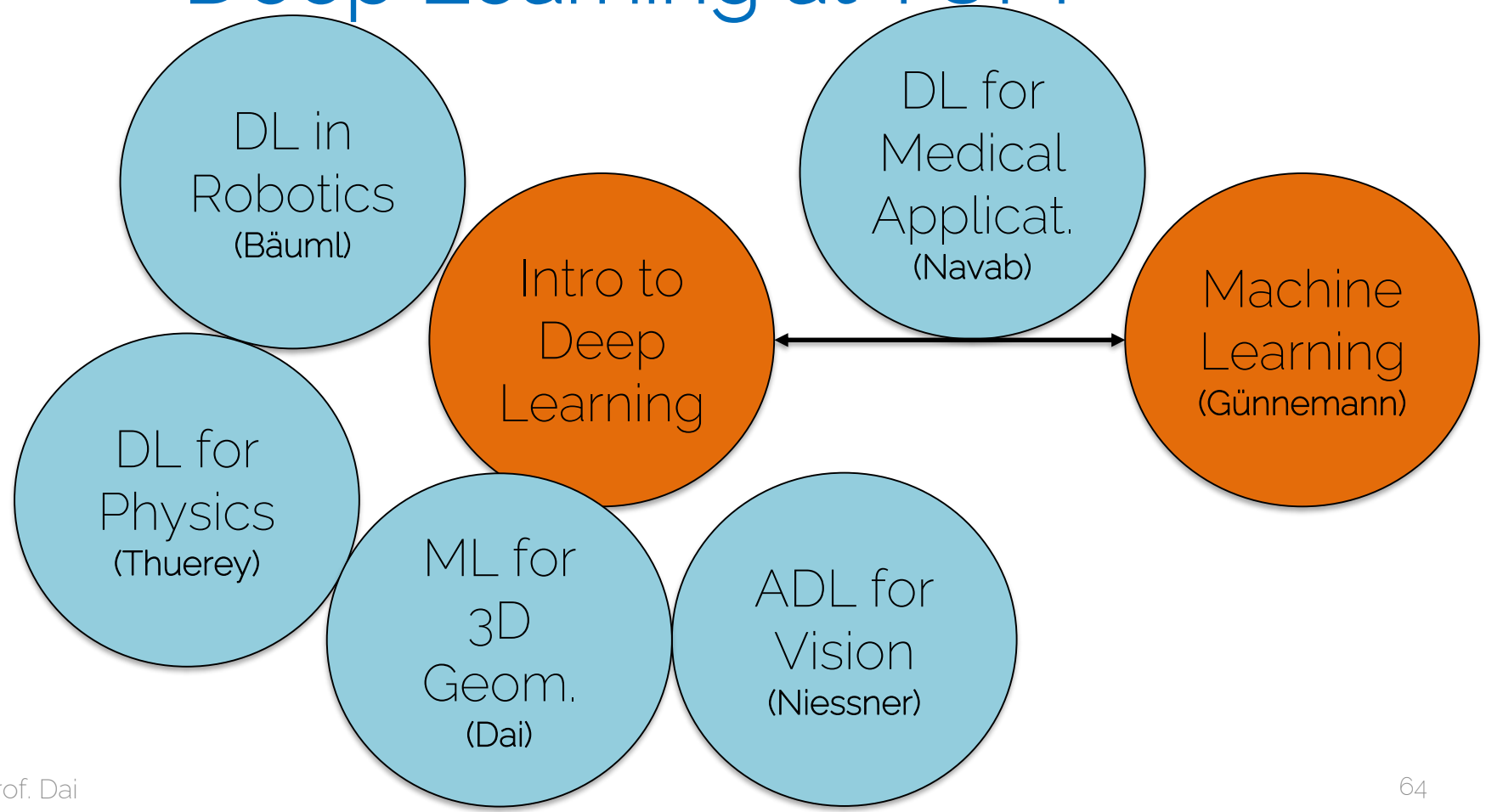
Deep Learning at TUM



[Taouil et al., CoRL'26] MotionDisco

Relation to other lectures at TUM

Deep Learning at TUM



Deep Learning at TUM

- I2DL is critical for the curriculum:
  - often a pre-requisite for many other lectures
  - many lectures build on top of it

- Follow up lectures such as Machine Learning for 3D Geometry, Advanced Deep Learning for Visual Computing
- Take I2DL first -> ML3D/AdvancedDL -> practical/GR/IDP

Introduction to Deep Learning

Logistics

About the Lecture

- Course webpage: https://www.3dunderstanding.org/i2dl-w26/
- Theory lectures (every Tuesday at 14:15)
  - In-person, live-streamed

- Tutorials and exercises (every Thursday at 10:00)
  - Tutorial: Online videos posted to Moodle and the webpage
  - Practical exercises

Preliminary Syllabus

Lecture 1: Introduction to the lecture, Deep Learning, Machine Learning
Lecture 2: Machine Learning Basics, Linear Regression, Maximum Likelihood
Lecture 3: Introduction to Neural Networks, Computational Graphs
Lecture 4: Optimization and Backpropagation
Lecture 5: Scaling Optimization to Large Data, Stochastic Gradient Descent
Lecture 6: Training Neural Networks
Lecture 7: Losses, Activations, Initialization
Lecture 8: Convolutional Neural Networks
Lecture 9: CNN Architectures
Lecture 10: Attention
Lecture 11: Transformers
Lecture 12: Generative Models
Lecture 13: Highlights (TBD)

Moodle

- Announcements via Moodle - IMPORTANT!
  - We will share common information (e.g., regarding exam)
- The Q&A Forum
  - Ask and discuss questions
    - Tutors will monitor and answer questions
    - You are very welcome to actively participate
  - Please do not post solutions of the exercises
- Private Questions: email i2dl@vc.in.tum.de

Email

- Email list:

  i2dl@vc.in.tum.de

- Do NOT email us personally!
  - Cannot handle so many emails / hence will be ignored

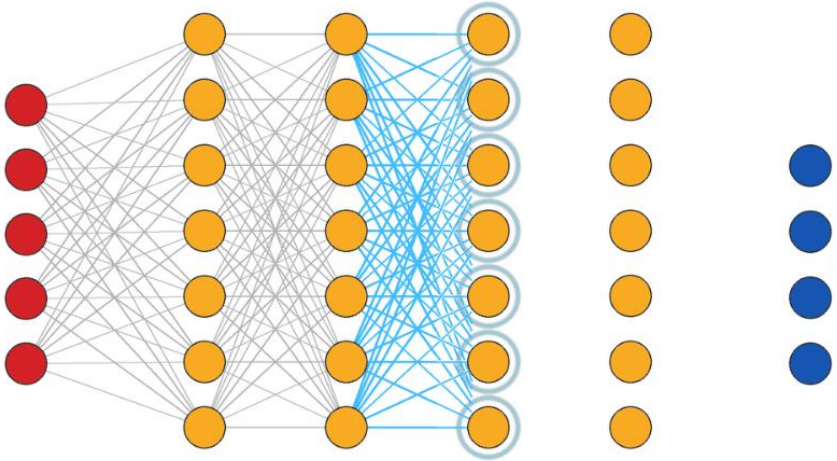
- Email list for organizational questions only!
  - Content questions -> Moodle or Office Hours

Website

3D AI Lab
Prof. Dr. Angela Dai

Home Publications Teaching Openings Team Contact

Introduction to Deep Learning (I2DL) (IN2346)



https://www.3dunderstanding.org/i2dl-w26/

(Virtual) Office Hours

- We will have dedicated office hours regarding
  - Theoretical help (e.g., specific lecture questions)
  - Help on exercises

- More info in the first tutorial session
- Video links will be posted on Moodle

Exam FAQ

- Final Exam: 08.02.2027, 17:00-18:30
- Content: Lecture & exercises
- Important: No retake exam (I2DL is taught every semester)

- Grade Bonus:
  - Solve 8 out of 9 “non-optional” programming exercises
  - Bonus 0.3 on a passed final exam
  - Bonus can be conditionally carried forward to next semester

Other Administrative

- "External" students welcome (LMU, TUM PhD)
  - You should naturally have the Moodle access.
  - If not, fill out registration form and we will add you to the course
  - Will get Certificate / Schein at the end

- Again:
  - Check announcements on Moodle

  - Check content on website:
    https://www.3dunderstanding.org/i2dl-w26/

Practical Exercises

Exercise – Goal

- Hands-on programming experience (learning by doing)
- Reimplement basic building blocks
- Introduction to common libraries
- (Get grade bonus)
- Ultimately: Gather enough experience to start your own individual (research) deep learning project

Exercise – Format

- Tutorial:
  - Video only
    Posted on Moodle and course website
  - Video length
    Ex02: Full lecture with written exercise
    Ex03-12: Short (~30min) video and coding exercises
- Programming:
  - Interactive coding notebooks (~4h each)
- Start time: Thursdays 10:00
- Working Time: 1-2 weeks
- Deadline: Wednesdays, 23:59

Upcoming Lecture

- Important:
  - Next lecture is on 20.10
- Next Lecture: Lecture 2: Machine Learning basics
- Thursday: Tutorial 1 and Exercise 1

See you next time ☺