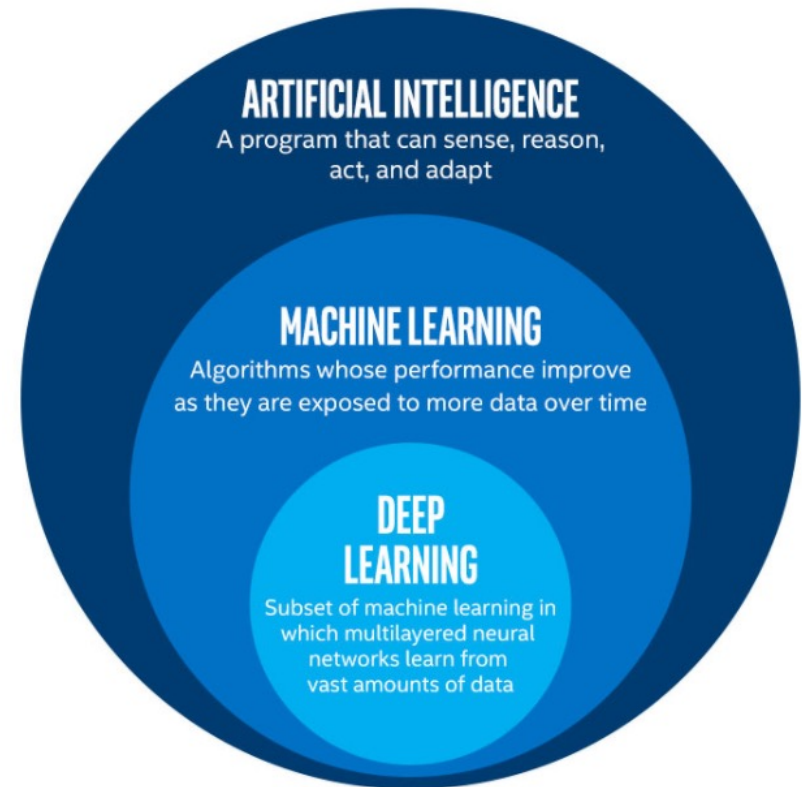
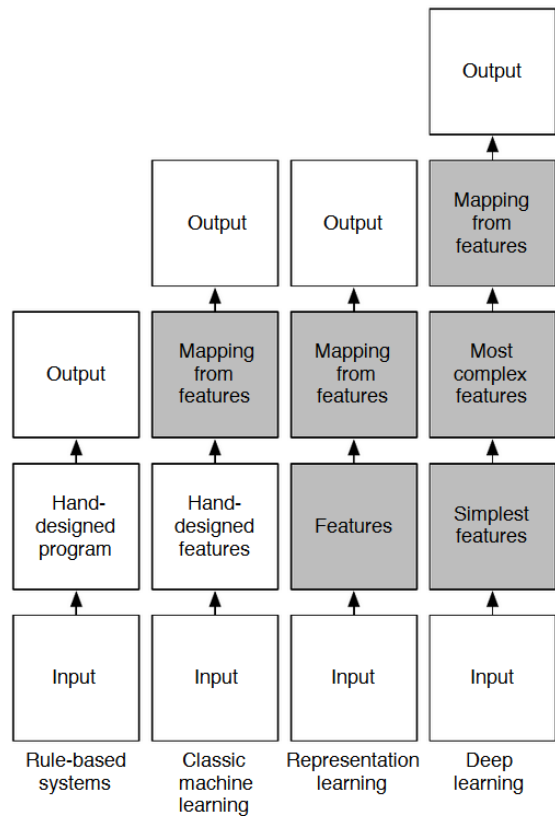


430.457

Introduction to Intelligent Systems

Prof. Songhwai Oh
ECE, SNU



DEEP LEARNING

ImageNet Large-Scale Visual Recognition Challenge, 2012

Tasks:

- Decide whether a given image contains a particular type of object or not. For example, a contestant might decide that there are cars in this image but no tigers.
- Find a particular object and draw a box around it. For example, a contestant might decide that there is a screwdriver at a certain position with a width of 50 pixels and a height of 30 pixels.



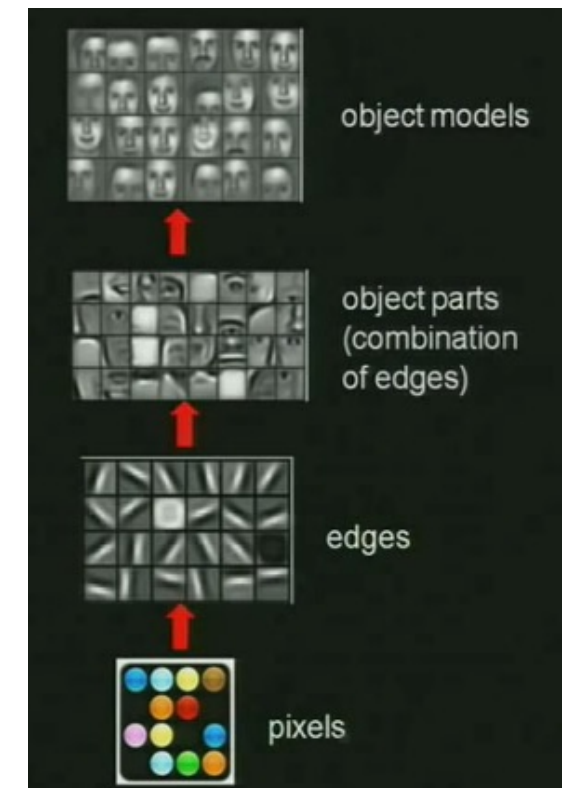
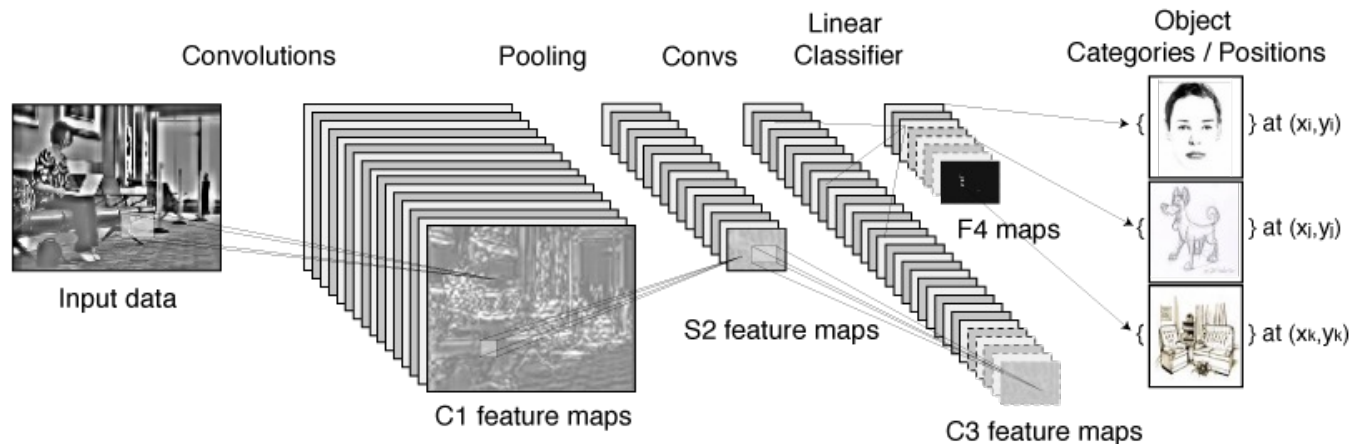
- 1000 different categories
- Over 1 million images
- Training set: 456,567 images

Year	Winning Error Rate (top 5)
2010	28.2%
2011	25.8%
2012	16.4% (2 nd 25.2%)
2013	11.2%
2014	6.7%
2015	3.57%
Human	About 5.1%

ImageNet Large Scale Visual Recognition Challenge. Russakovsky et al. *arXiv preprint arXiv:1409.0575*. URL: <http://arxiv.org/abs/1409.0575v1>

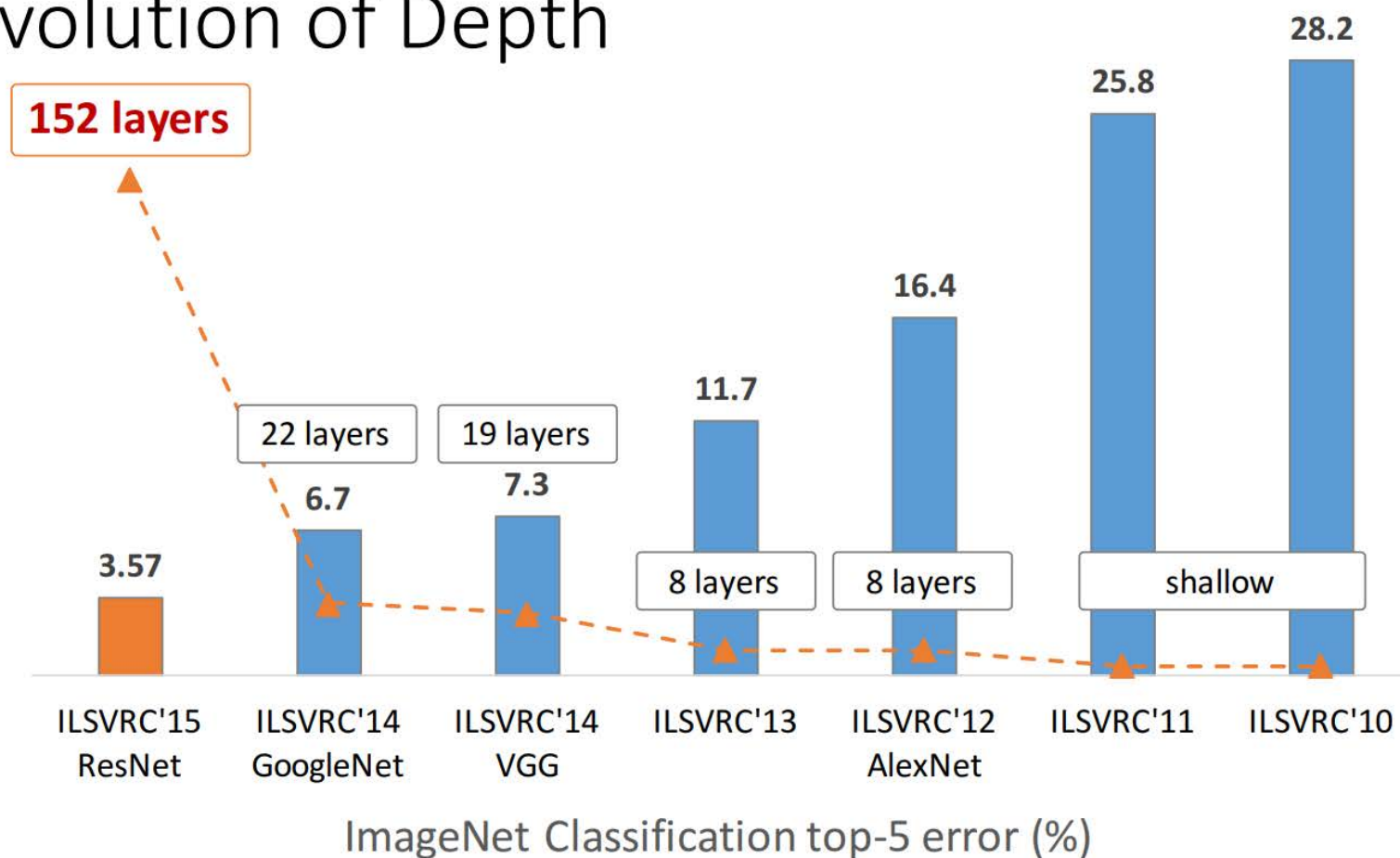
Convolutional Neural Networks (CNNs)

- SuperVision (2012)
 - Deep convolutional neural network
 - 650,000 neurons
 - 5 convolutional layers
 - Over 60 million parameters
- Clarifai (2013)
- GoogLeNet (2014) – 22 layers
- ResNet (2015) – 152 layers



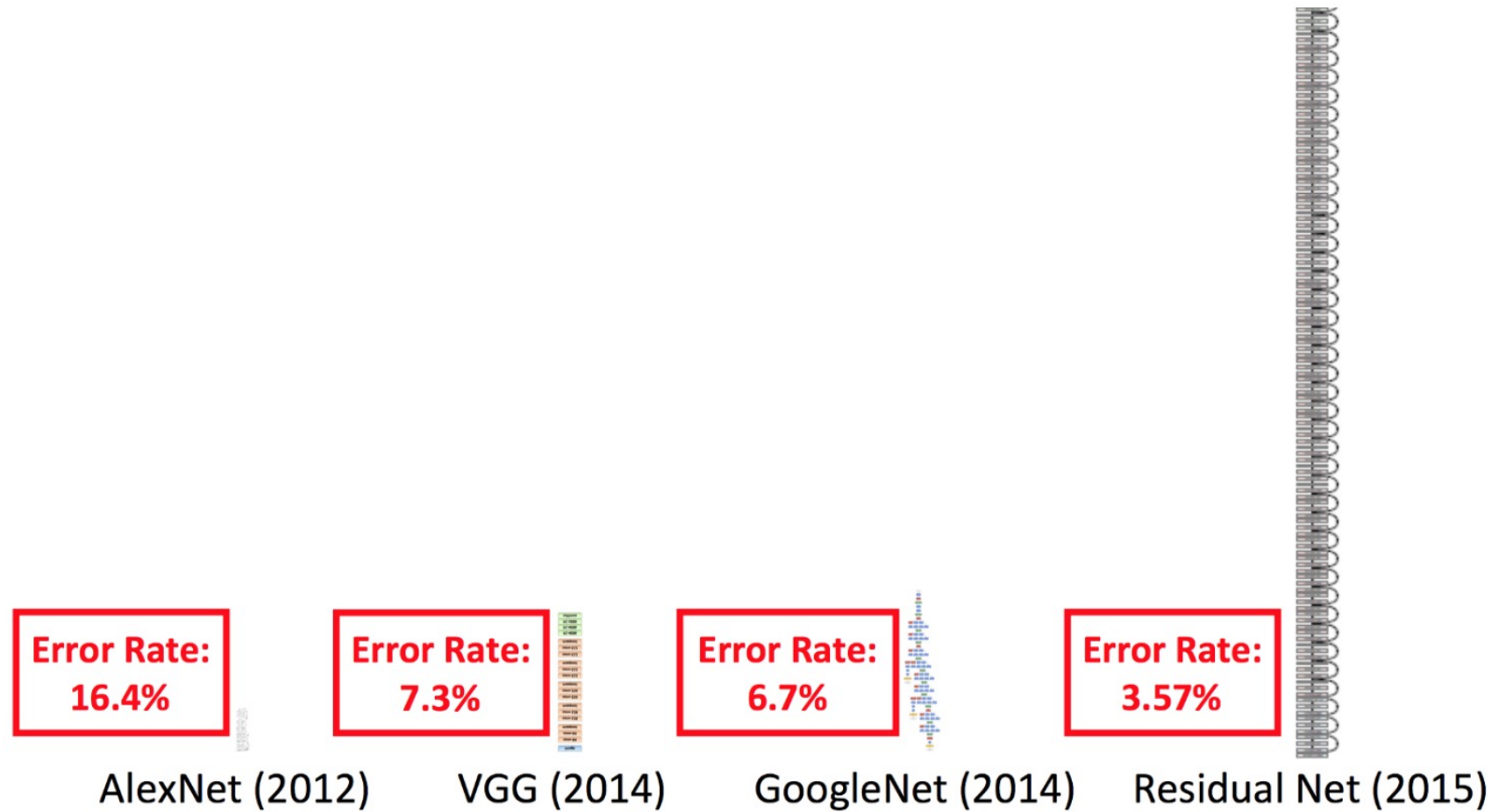
ImageNet Challenge

Revolution of Depth



K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. CVPR, 2016.

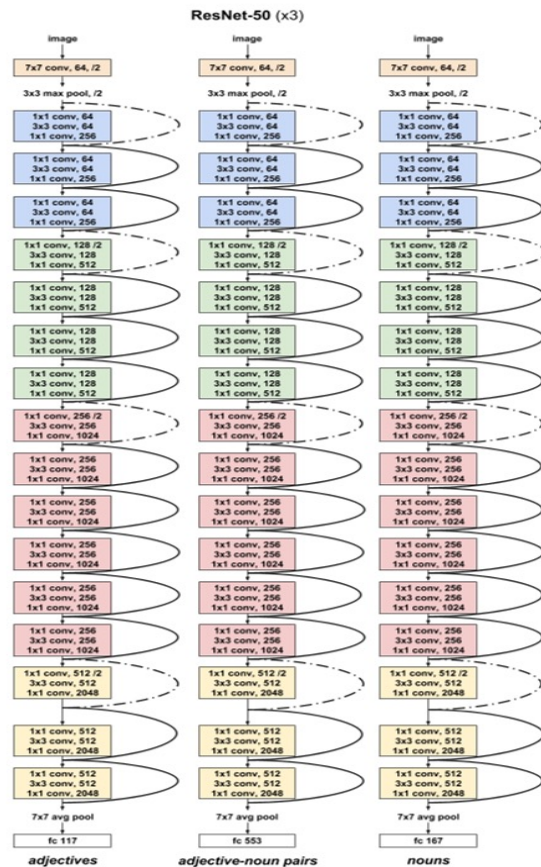
The Trend



<https://medium.com/@k3083518729/tensorflow-image-recognition-58b0ac77c263>

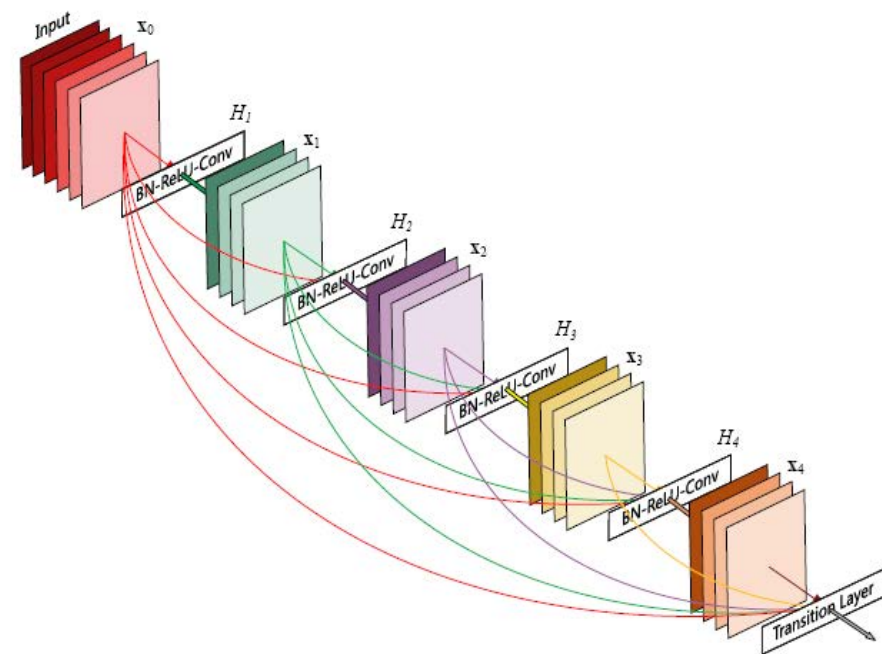
The Trend

Going deeper



ResNet (CVPR'16 Best Paper)

Going denser



DenseNet (CVPR'17 Best Paper)

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. CVPR, 2016.

G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten. Densely connected convolutional networks. CVPR, 2017.

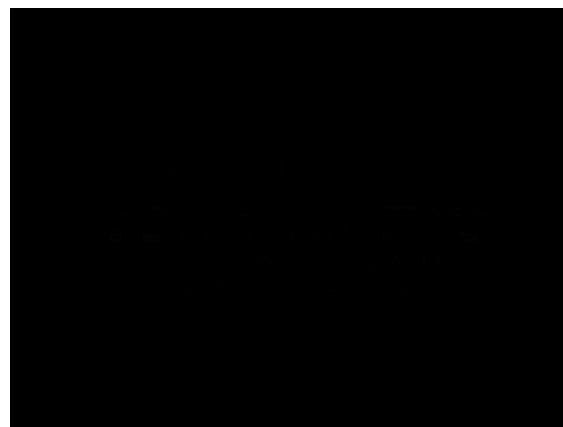
DEEP REINFORCEMENT LEARNING

Deep Q-Network (DQN), 2013

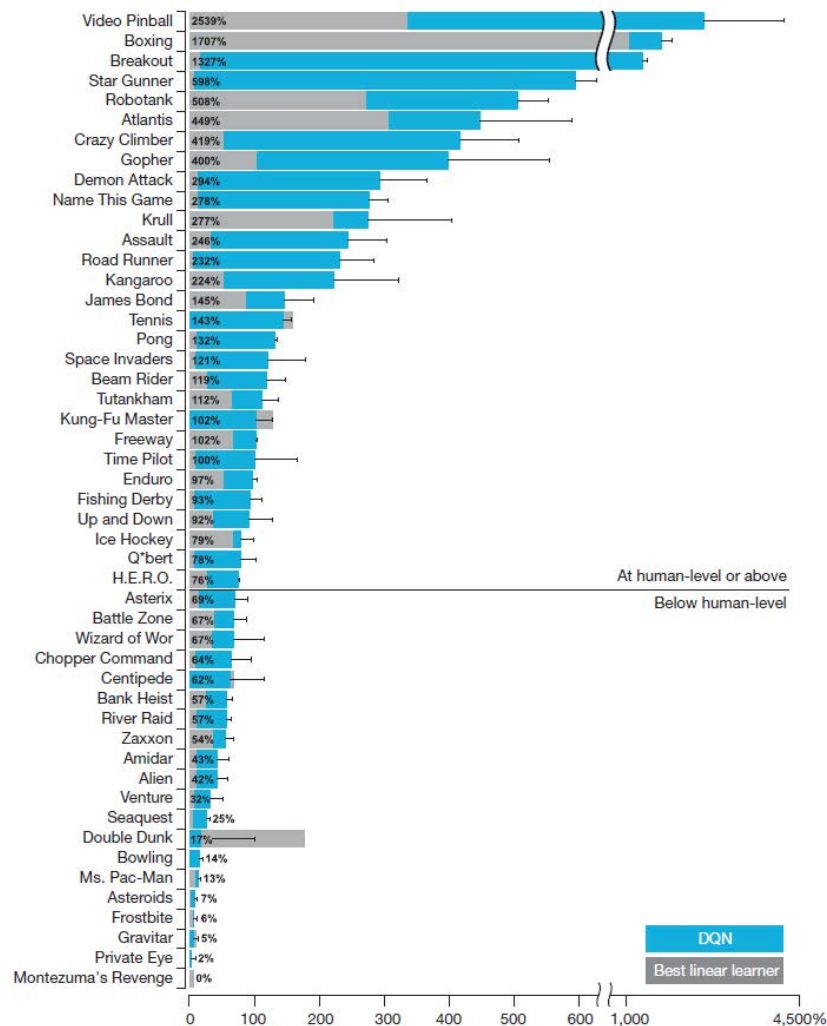
- Playing Atari games
- Input: Game screen shots
- Output: Control (left, right, shoot, ...)
- Convolutional neural networks (CNN)
- Reinforcement learning: Q-learning



Breakout



Space Invaders



AlphaGo, 2016

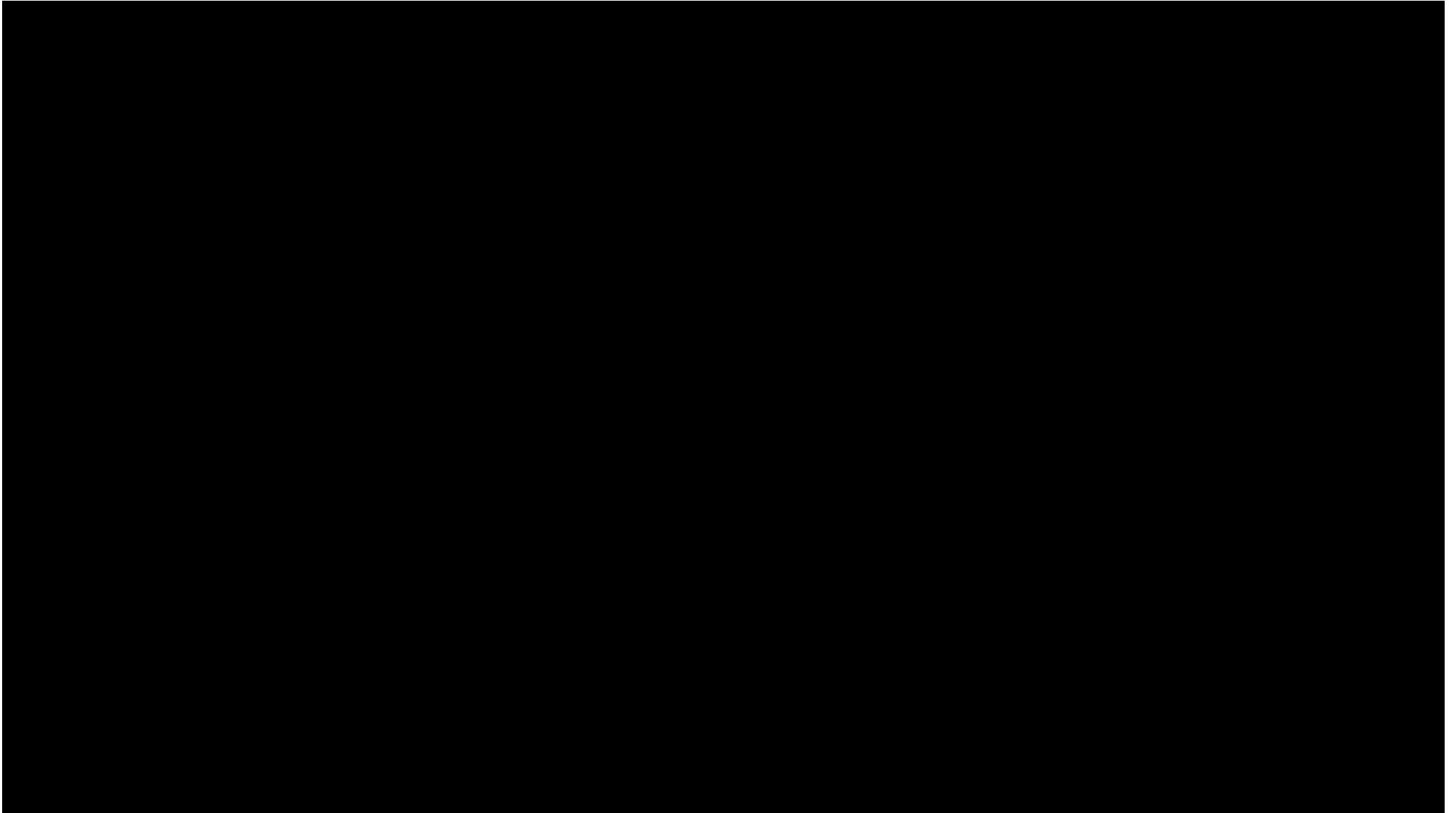
Google DeepMind's AlphaGo vs. Lee Sedol, March 2016

- Possible board positions of Go: 10^{170}
 - cf. Chess: 10^{47}
- Monte Carlo tree search
- Deep neural networks:
 - Value network
 - Policy network
- Reinforcement learning
- Trained from
 - 30 million human moves
 - Playing against itself
- 1,202 CPUs, 176 GPUs



Game	Date	Black	White	Result	Moves
1	9 March 2016	Lee Sedol	AlphaGo	Lee Sedol resigned	186 Game 1
2	10 March 2016	AlphaGo	Lee Sedol	Lee Sedol resigned	211 Game 2
3	12 March 2016	Lee Sedol	AlphaGo	Lee Sedol resigned	176 Game 3
4	13 March 2016	AlphaGo	Lee Sedol	AlphaGo resigned	180 Game 4
5	15 March 2016	Lee Sedol ^[note 1]	AlphaGo	Lee Sedol resigned	280 Game 5
Result: AlphaGo 4 – 1 Lee Sedol					

Robotics (OpenAI)

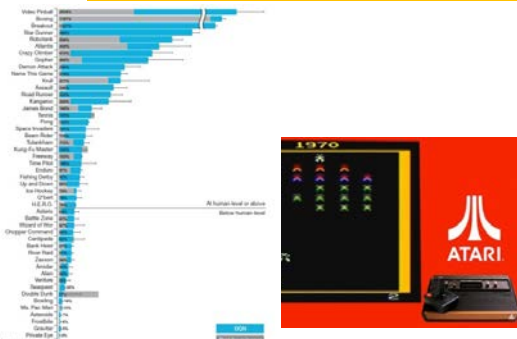


Major Milestones in AI

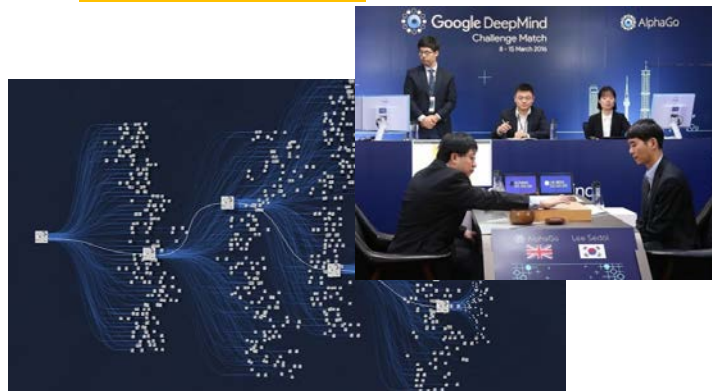
ImageNet Challenge (2012)

Image classification	
Easiest classes	
 red fox (100)	 hen-of-the-woods (100)
 ibex (100)	 goldfinch (100)
 flat-coated retriever (100)	
 tiger (100)	 hamster (100)
 porcupine (100)	 squirrel (100)
 Blenheim spaniel (100)	
Hardest classes	
 muzzle (71)	 hatchet (68)
 water bottle (68)	 velvet (68)
 loupe (66)	
 hook (66)	 spotlight (66)
 ladle (65)	 restaurant (64)
 letter opener (59)	
Year	Winning Error Rate
2010	28.2%
2011	25.8%
2012	16.4% (2 nd 25.2%)
2013	11.2%
2014	6.7%
2015	3.57%
Human	About 5.1%

Atari Games, DQN (2013)



AlphaGo (2016)

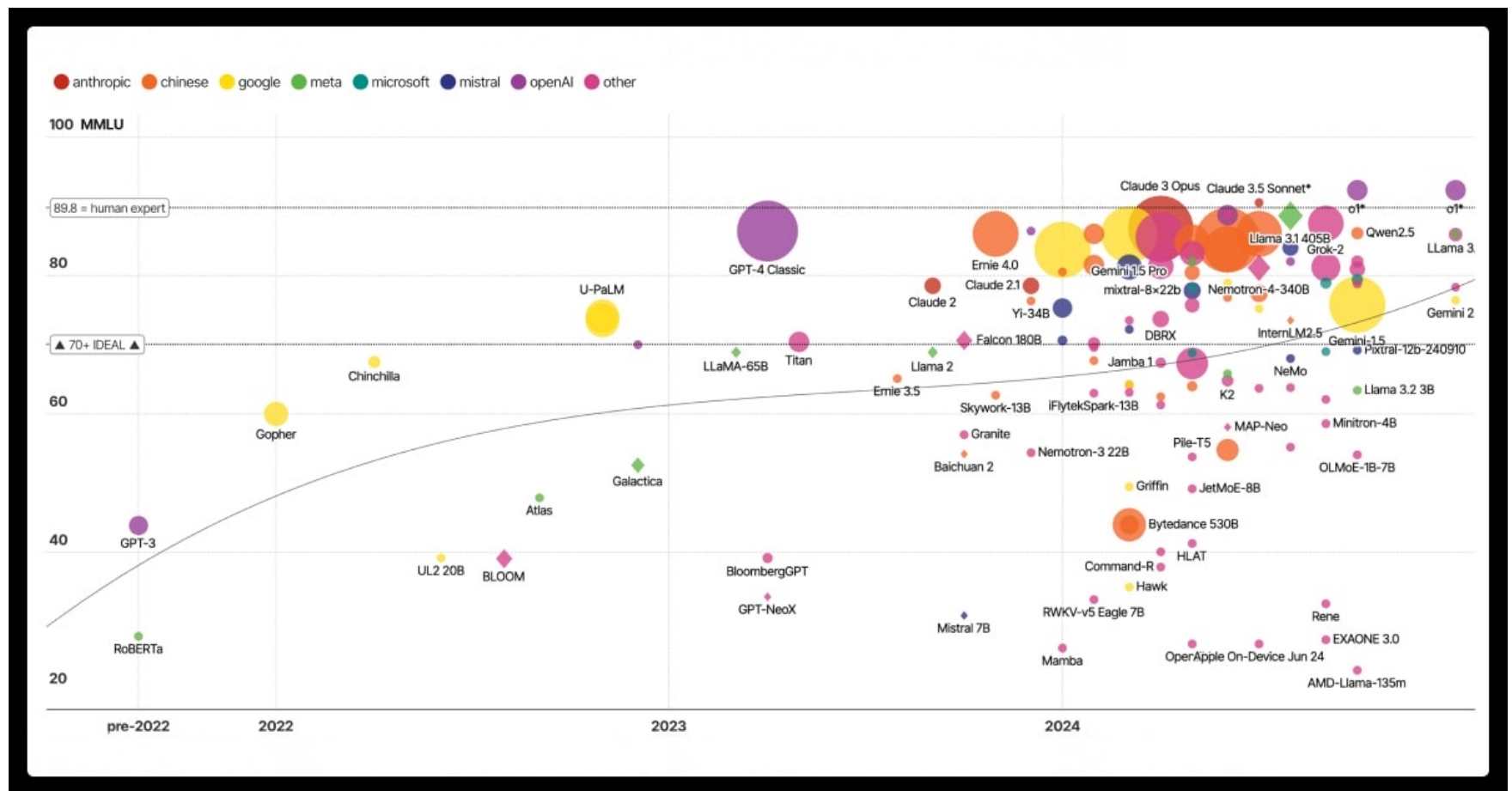


StarCraft, AlphaStar (2019)



Large AI Models

- Large language models (LLM)
- Vision language models (VLM)
- Generative models

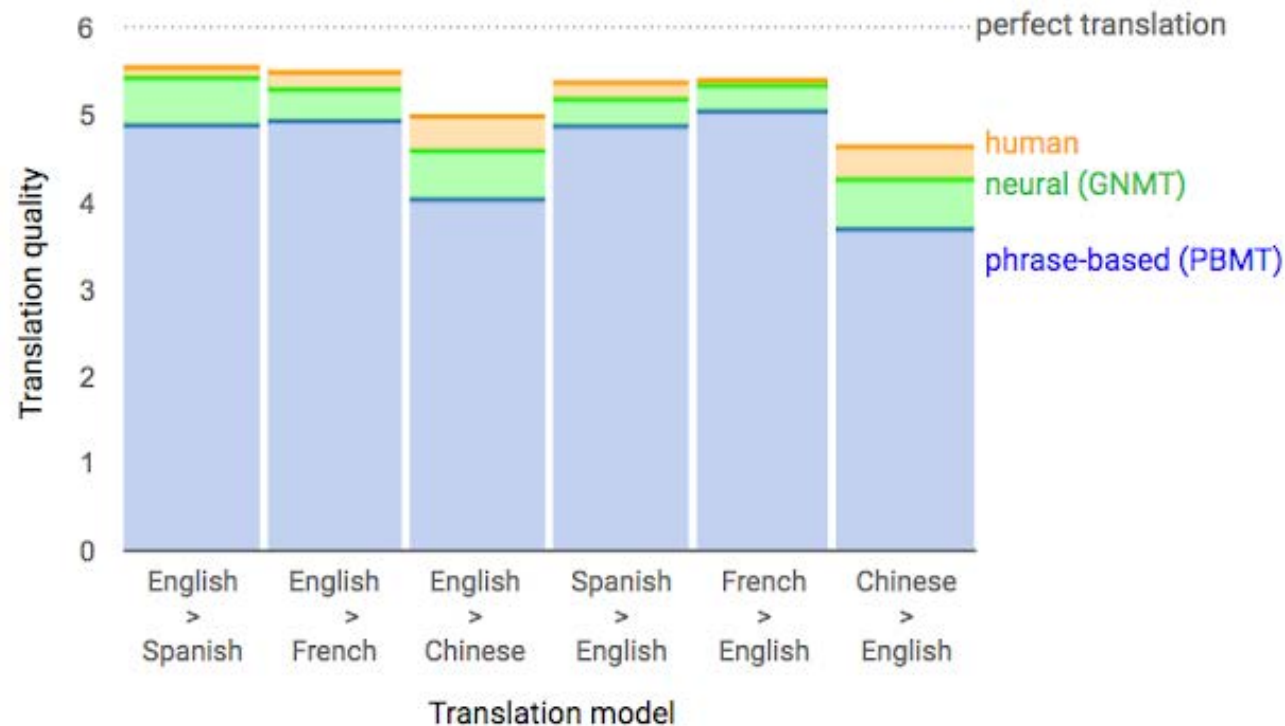


LLM performance and timeline comparison (Source: information is beautiful)

DEEP LEARNING: SOME (NOT SO) RECENT APPLICATIONS

Language Translation (2016)

- Google Neural Machine Translation (GNMT) System



Source: <https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html>

Language to Action (SNU, 2018)

Supplementary Material for ICRA 2018

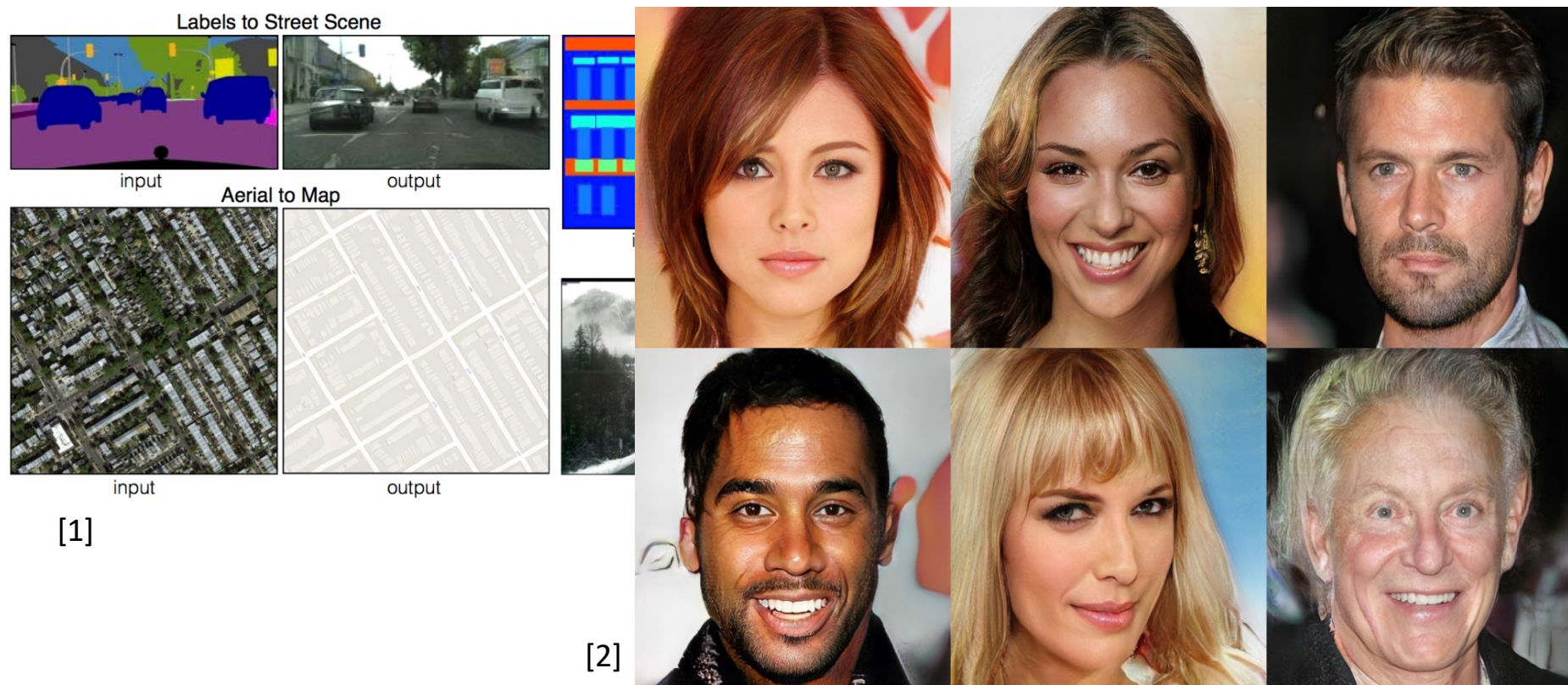
Text2Action: Generative Adversarial Synthesis from Language to Action

Hyemin Ahn, Timothy Ha, Yunho Choi, Hwiyeon Yoo, and Songhwai Oh

CPSLAB, Department of Electrical and Computer Engineering, Seoul National University

Synthesis

- Speech (WaveNet, Google DeepMind)
- Images



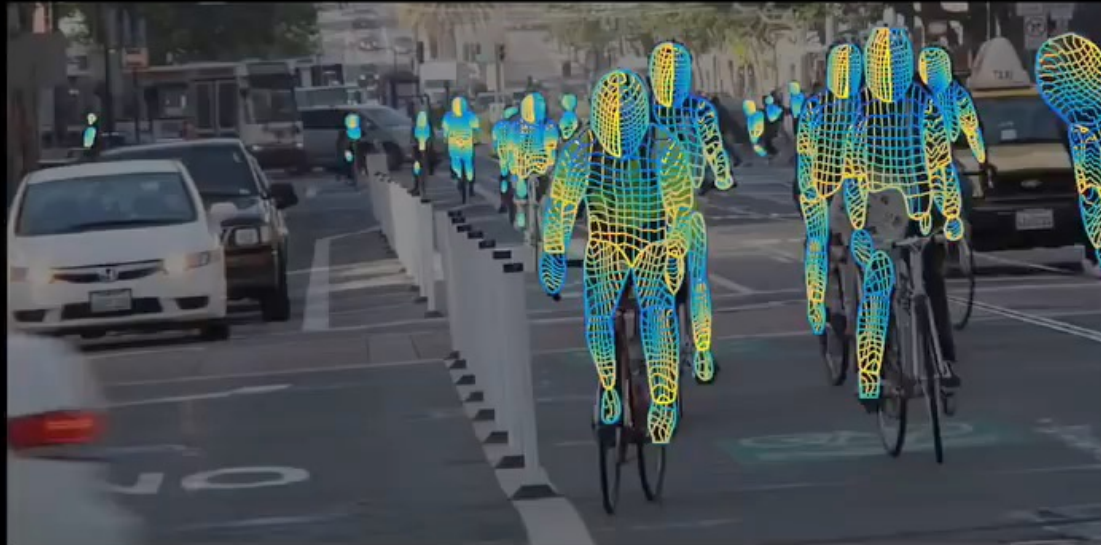
[1] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros , “Image-to-Image Translation with Conditional Adversarial Nets,” CVPR 2017.
[2] Karras, T., Aila, T., Laine, S., & Lehtinen, J. **Progressive growing of GANs for improved quality, stability, and variation.** ICLR 2018.

Video Synthesis (University of Washington, 2018)



Pose Estimation (DensePose, Facebook, 2018)

DensePose: Dense Human Pose Estimation In The Wild



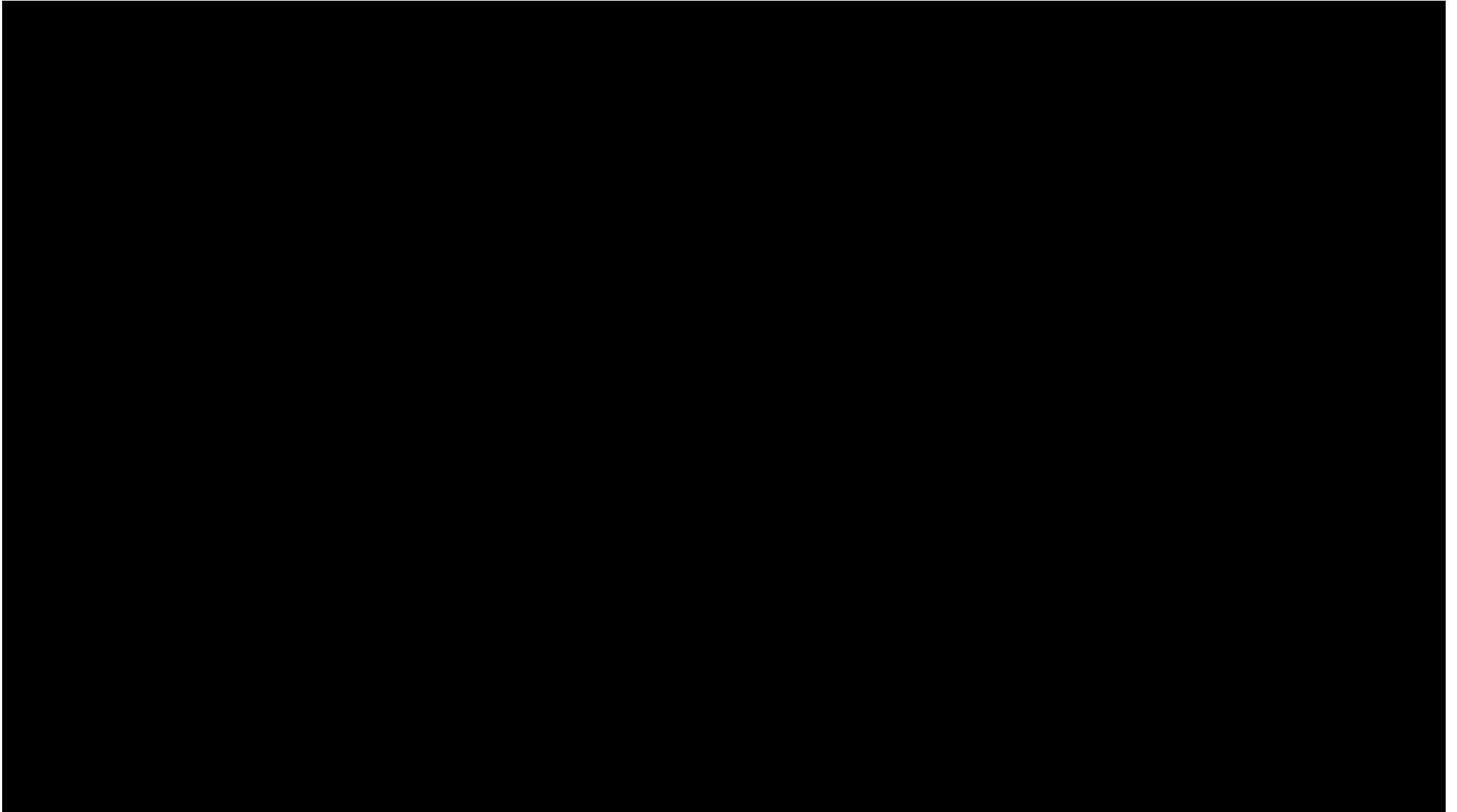
Rıza Alp Güler *
INRIA, CentraleSupélec

Natalia Neverova
Facebook AI Research

Iasonas Kokkinos
Facebook AI Research

* Rıza Alp Güler was with Facebook AI Research during this work.

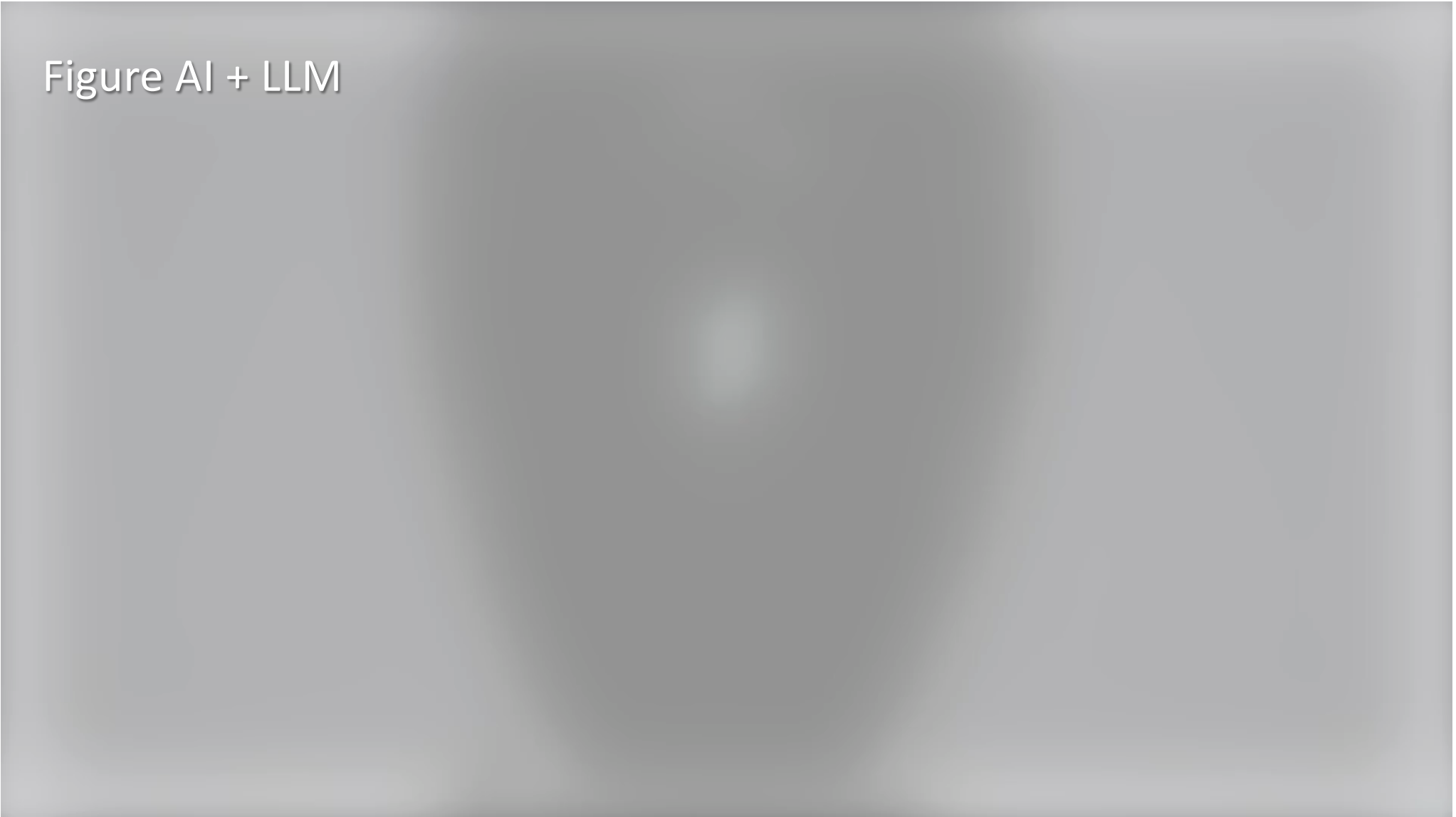
Autonomous Driving (Wayve, 2018)



CONVERGENCE OF ROBOTICS AND AI

Figure AI + LLM

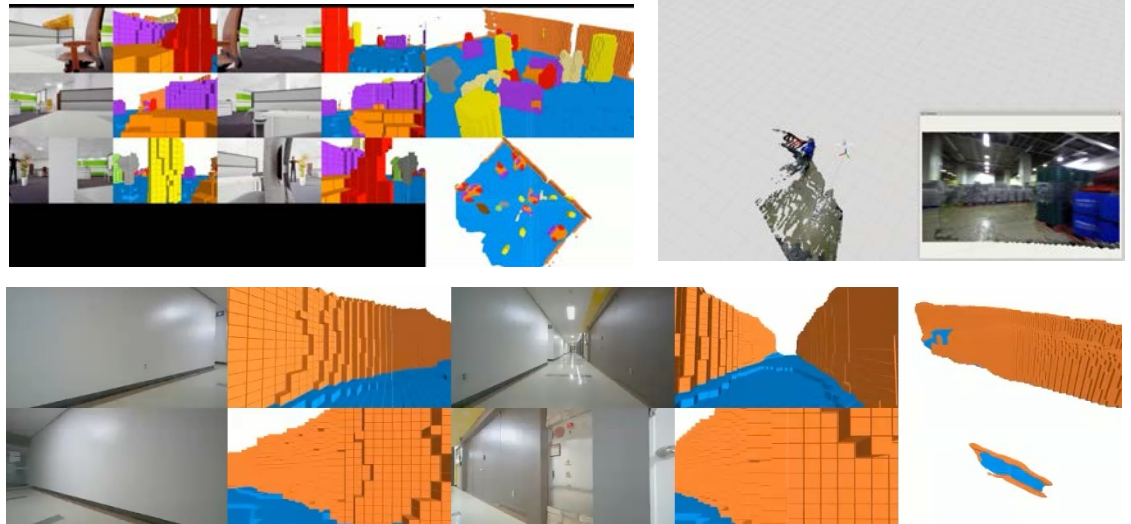
Figure AI + LLM



Robot Learning

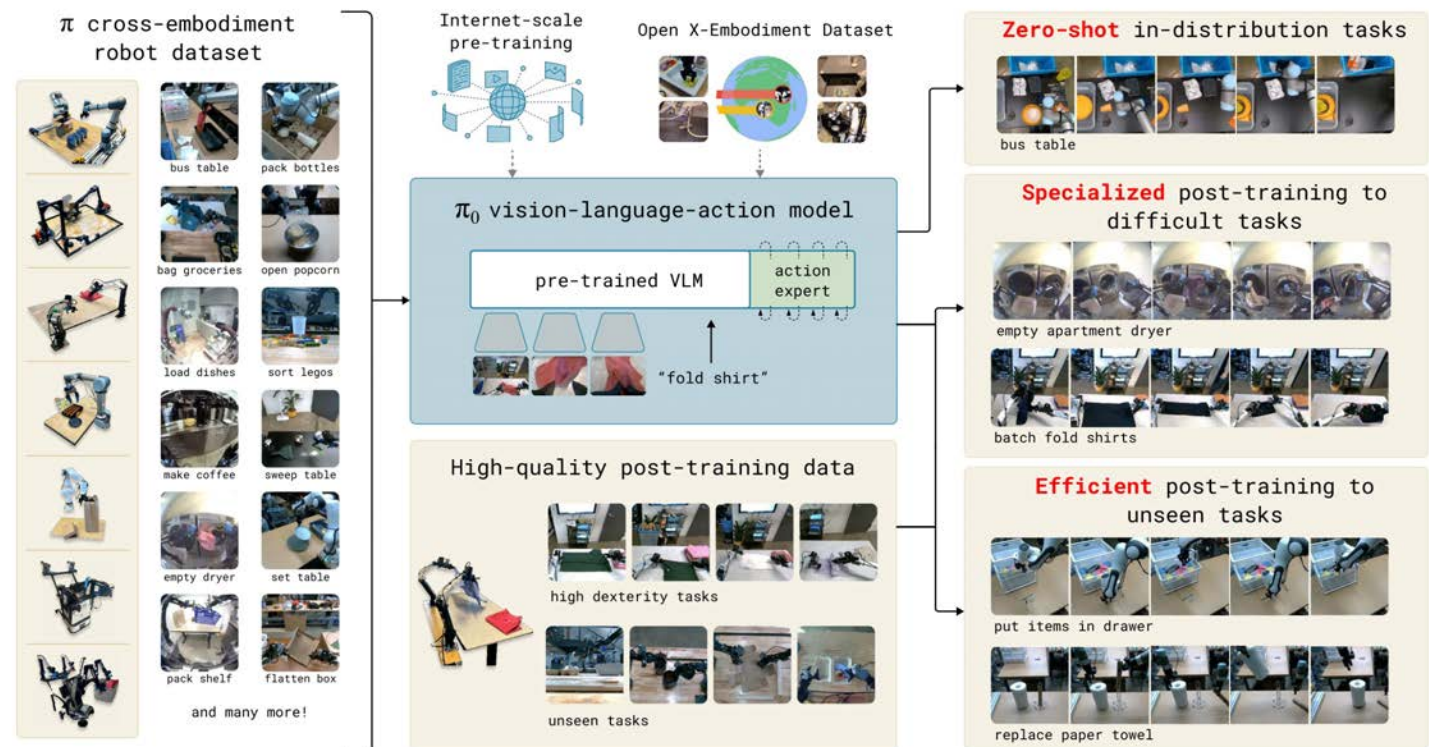
- AI technology for robots
 - Enables robots to
 - Perform complex but everyday tasks in unseen environments
 - Interact with unfamiliar objects
 - Learn common sense
 - Solve open-world problems

Robot Foundation Model

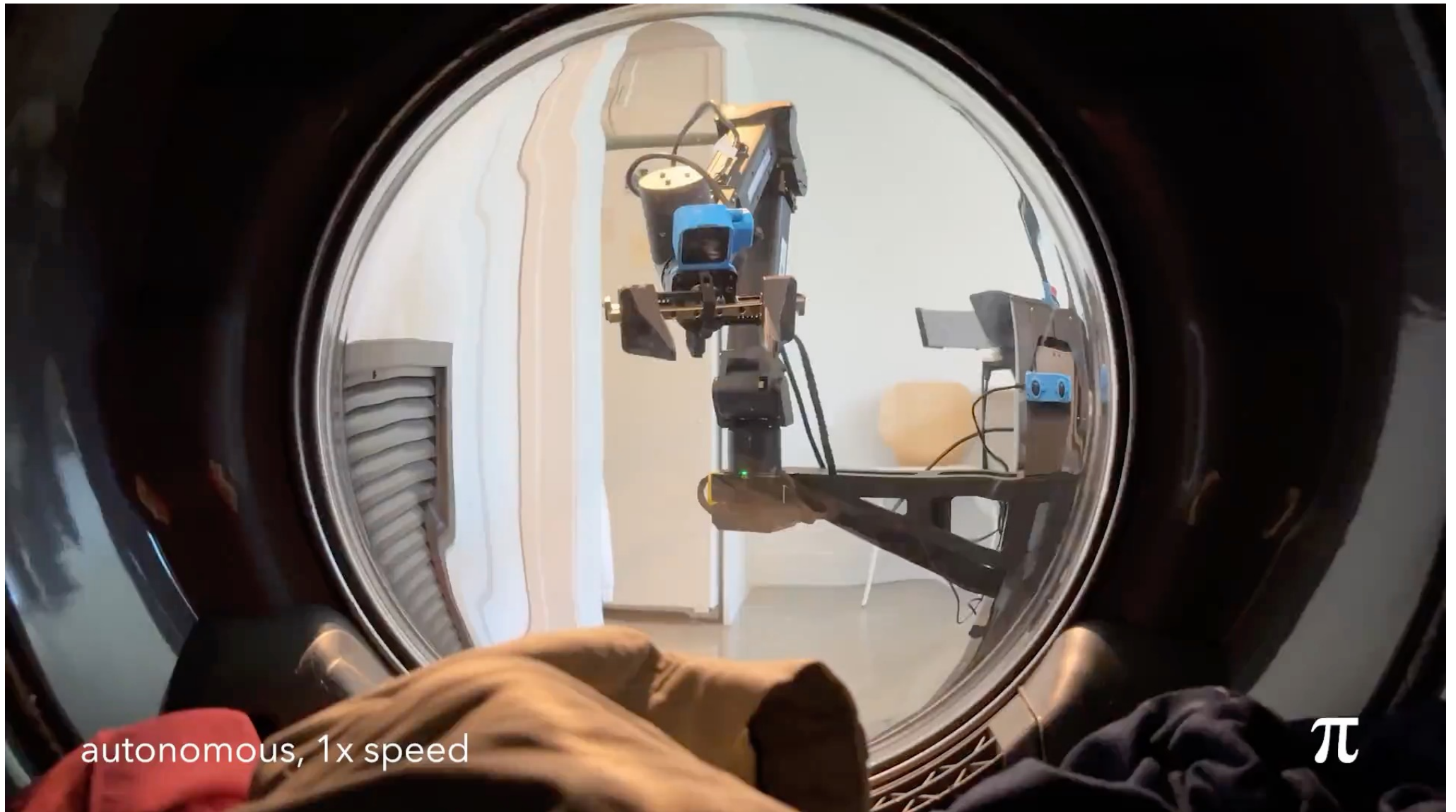


Physical Intelligence: $\pi 0$

- Startup (founded in 2023)
- Funding: over \$1B
- Valuation: \$11B
- Vision language action (VLA) model



source: <https://www.physicalintelligence.company/blog/pi0>



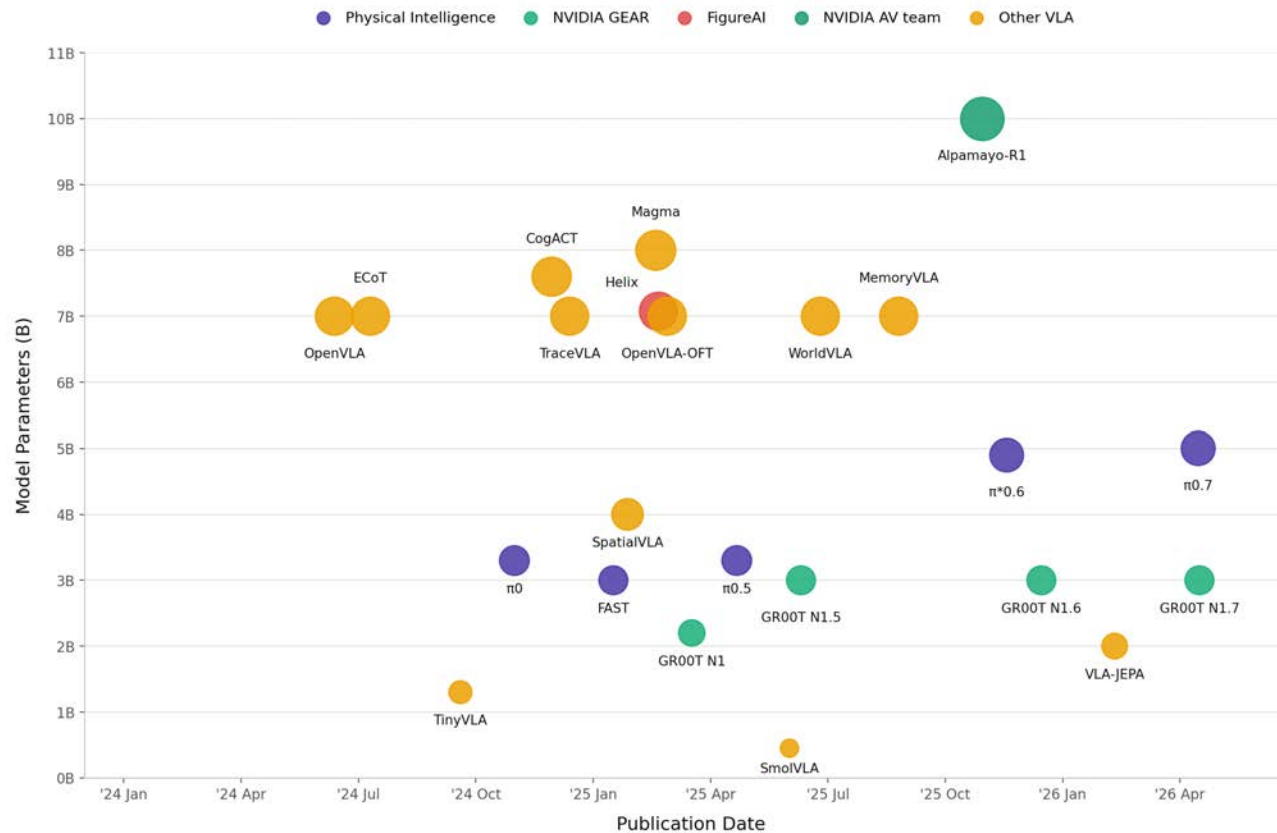
Application Domains

- Manufacturing
- Warehouses and logistics
- Agriculture
- Healthcare and medicine
- Service robots for home and offices
- and more ...

Tesla Optimus

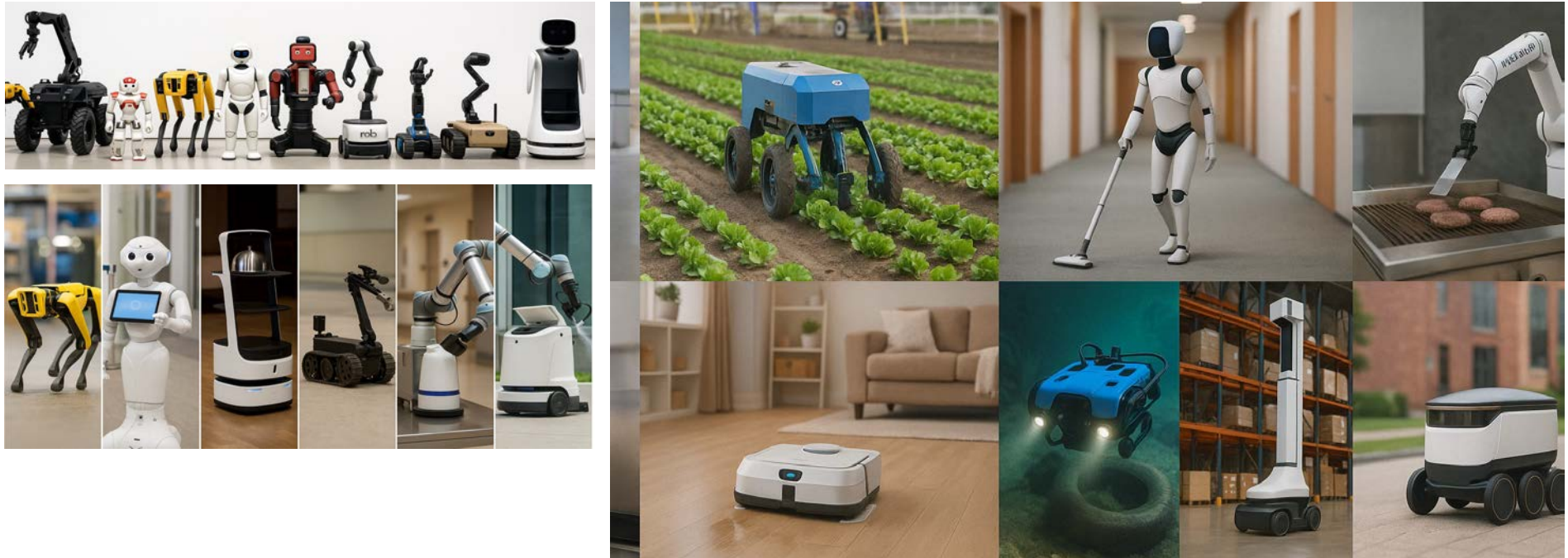


Robot Foundation Models



Challenges in Building a Robot Foundation Model

- Diversity in robots and tasks



- Lack of data
- Real-world is just hard