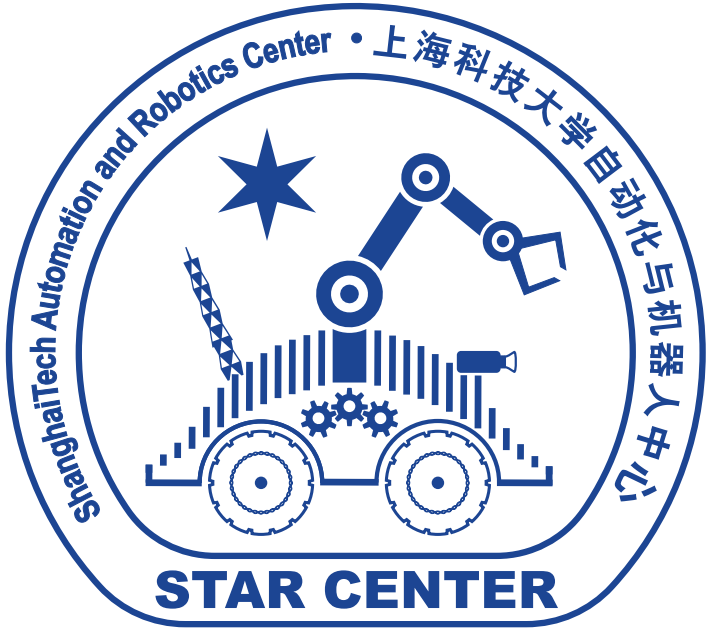
上海科技大学

ShanghaiTech University

CS283: **Robotics Spring 2025: Summary**

Sören Schwertfeger / **师泽仁**

ShanghaiTech University



HW4

- Is due today
- How is it going?

Project conclusion

- Due June 19:
  - Project Report:
    - Like intermediate report
    - With nice results and proper quantitative evaluation
    - Make look like a scientific paper
    - Use bibtex!
    - Put into git (folder: doc/final)
      - Include everything that is needed to generate the report in the git!
        - So don’t forget images/ the bib file
  - Project Demo:
    - Make an appointment with Prof. Schwertfeger and your TA to show the final demo of your project
    - Before: June 16th, 22:00 !

Project Webpage

- Write a text (word document) about your project for the general public – not too technical – not too many details
  - Some details can be written
  - Do not just copy the abstract/ intro from your final report – write a nice text for the general public!
- Provide a few images with captions (as images also extra files)
- Put into your group git (folder: doc/webpage)
- Prof. Schwertfeger will upload the data to the website – e.g. look at : https://robotics.shanghaitech.edu.cn/teaching all previous robotics and MoMa courses
- Also make a NICE video about your project. 4-8 minutes. Leave the video at good quality – size maybe 100 – 300 MB (MP4) – Prof. Schwertfeger will compress it to make a web version
  - Avoid showing other people; do not talk in the video; do not add music;
  - Add a title page: same info as on report
  - Add to your git folder

Final

- June 10
  - 16:00 – 18:00 in 1D-104

- Content:
  - All lectures
    - Take a look at facts, algorithms, concepts
  - Take a look at the homeworks again
  - Sample exam: https://robotics.shanghaitech.edu.cn/sites/default/files//files/final_Example.pdf

- You are allowed to bring **3** A4 sheets (so 6 pages) of info to the exams. You can write/ print anything on those sheets. On top of **every page** (so 6 times) there needs to be your **name (pinyin), student ID and ShanghaiTech email** address. We will check every cheat sheet before the exam and **confiscate** every sheet without name or with a name that is not yours.

- No electronics/ calculator/ smartwatch allowed

Midterm -> Final

- Questions in final can be:
  - Similar to the midterm
  - New questions (e.g. other algorithm)

- => learn for your Midterm mistakes!

- We may include questions from the guest lecture!

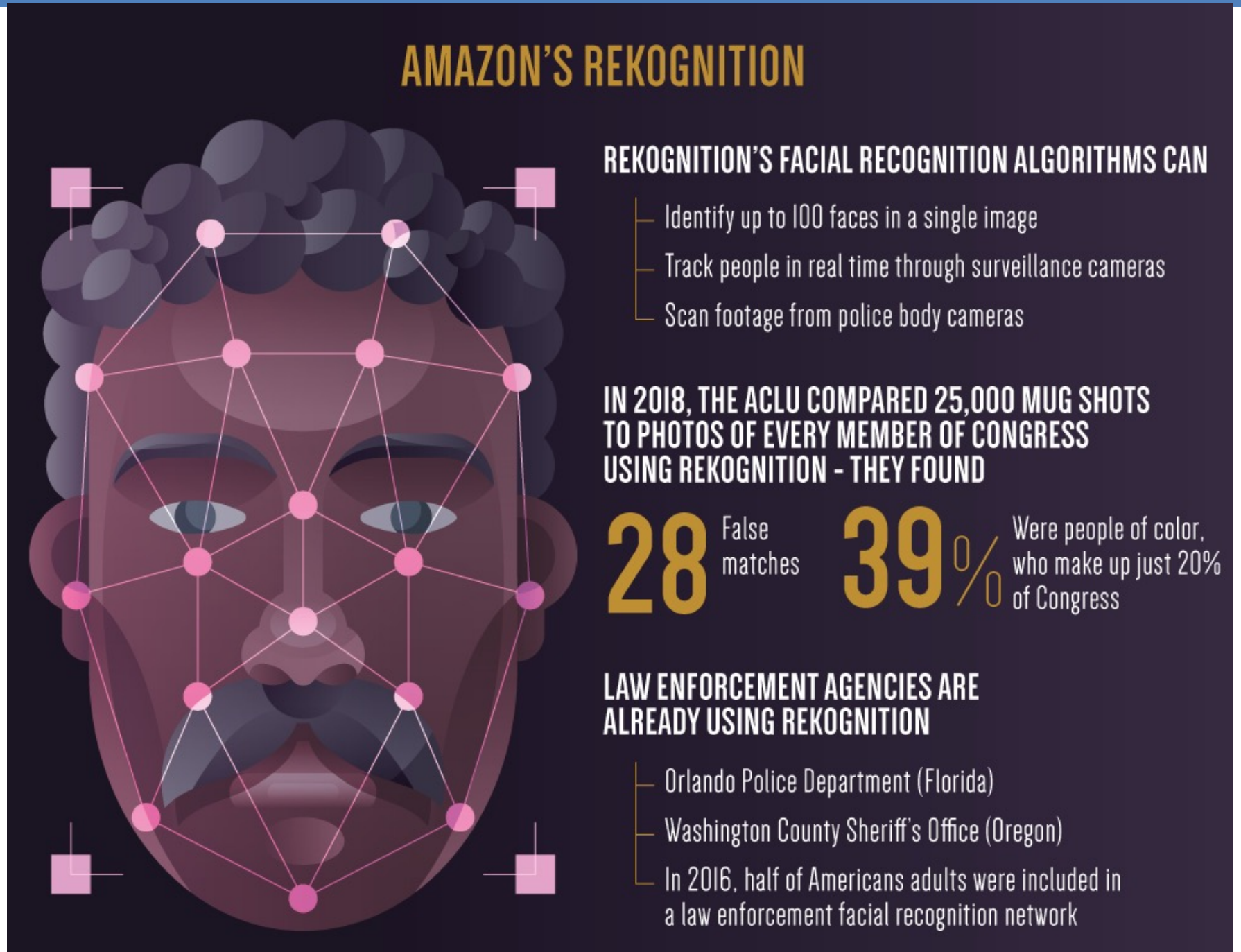
Project Meetings!

- Today is last class => spend lots of time on your project!
- We will make another project meeting in the week of May 26 – May 30!
- Also, other meetings afterwards!

ROBOTICS ETHICS

Continued…

When AI goes wrong

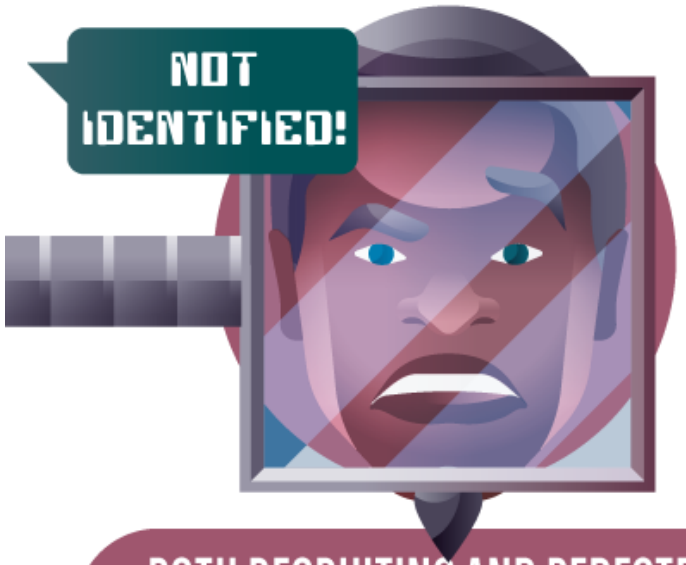


HOW A.I. BIAS HAPPENS

WOMEN NEED NOT APPLY

Starting in 2014, Amazon began training an A.I. to review job candidates

- The system was trained using resumes submitted over 10 years — MOST CAME FROM MEN
- The A.I. concluded that "male" was a preferred quality for new hires, and started FILTERING OUT FEMALE CANDIDATES



DETECTING DARKER SKIN TONES

In a 2019 study, researchers found that the object detection models used in driverless cars were better at identifying pedestrians with lighter skin

- The study used a standardized set of photos to train their A.I. — but found their DATASET CONTAINED 3X AS MANY LIGHT SKINNED PEOPLE
- The A.I. quickly learned to identify light skinned pedestrians, but STRUGGLED TO IDENTIFY DARKER SKIN TONES

BOTH RECRUITING AND PEDESTRIAN DETECTION ALGORITHMS FAILED BECAUSE THEY WERE TRAINED ON BAD DATA — A.I. LEARNED BIAS FROM HUMANS

MAKING ETHICAL A.I.

START WITH DATA

A.I. training data must reflect real diversity and control for existing bias

- Amazon's recruiting algorithm was trained to eliminate female candidates. INSTEAD, IT COULD HAVE BEEN PROGRAMMED TO IGNORE GENDER
- Pedestrian identification algorithms struggle to identify darker skin tones. Rather than monitoring success overall, the TRAINING DATA COULD HAVE WEIGHTED DARK SKIN DATA POINTS MORE HEAVILY

CONSIDER THE PROCESS

When training A.I., programmers typically split their dataset into 2 parts

- Half is used to TRAIN THE A.I.
- Half is used to VERIFY AND MEASURE SUCCESS

If the initial dataset is flawed, the test will have the same bias

MONITOR FOR UNKNOWNS

Programmers must monitor for unintentional bias appearing in their A.I.

- Subtle patterns can lead A.I. to PERPETUATE HUMAN BIAS
- Amazon's recruiting algorithm preferred VERBS LIKE "EXECUTED" AND "CAPTURED" — WHICH TEND TO BE MORE USED BY MALES

8 Ethical Questions in AI

Bias:
Is AI fair?

Liability:
Who is responsible for AI?

Security:
How do we protect access to AI from bad actors?

Human Interaction:
Will we stop talking to one another?

Employment:
Is AI getting rid of jobs?

Wealth Inequality:
Who benefits from AI?

Power & Control:
Who decides how to deploy AI?

Robot Rights:
Can AI suffer?

www.logikk.com LOGIKK

https://www.logikk.com/articles/8-ethical-questions-in-artificial-intelligence/

Ethical AI: Many open questions and topics:

- **Autonomy and liability**
- **Ethical principles in robotics**
- **Enhancement technologies**: ethical issues
- Defining ethical guidelines for the design, use, and operation of robots
- Privacy & management of personal data
- **Ethical frameworks: universal or region-specific?**
- The role of industry and society in the definition of safety standards
- AI technology to block unethical/mendacious social-media communication
- **Accountability** in autonomous systems
- **Embedding values** and norms into intelligent systems
- Ethics and **standardization**
- Raising ethical awareness among stakeholders
- **Transparency** in autonomous systems
- Political and **legal** frameworks
- **Formal and mathematical** frameworks for robot ethics
- Implementations and engineering studies
- User and HCI/HRI studies at the intersection of the above issues

https://clawar.org/icres2023/

Ethical AI: Scientific Discussion

- ICRES 2023 is the 8th edition of the International Conference on Robot Ethics and Standards series
  - https://clawar.org/icres2023/
- IEEE Robotics and Automation Society: Technical Committee on Robot Ethics
  - Framework for raising and addressing the urgent ethical questions prompted by and associated with robotics research and technology.
  - https://www.ieee-ras.org/robot-ethics
- Conference on Robotics, AI and Humanity, Science, Ethics and Policy organized jointly by the Pontifical Academy of Sciences (PAS) and the Pontifical Academy of Social Sciences (PASS)
  - http://www.pas.va/content/accademia/en/events/2019/robotics/statementrobotics.html

Autonomous Weapons are attractive to the military

- Don’t endanger the life of your own soldiers.
- Robots don’t get tired – they can operate 24/7.
- Superior situational awareness through 360-degree view with advanced sensors & instantaneous communication with various [illegible] the field
- Stay highly vigilant the whole day
- No fear: perform the most dangerous attacks without degraded performance
- Ease logistics: no need for food, shelter, medical attention or rotation of units.
- Superior performance in battle, e.g. speed of target acquisition, firing accuracy, battle coordination, overall strategy

THeMIS robot from the Estonia company MILREM Robotics.
The 2t semi-autonomous robot can drive with a speed of up to 20km/h.

Programs Experts CSIS | CENTER FOR STRATEGIC & INTERNATIONAL STUDIES Regions Topics

Ukraine's Future Vision and Current Capabilities for Waging AI-Enabled Autonomous Warfare

Photo: Anton Petrus/GETTY IMAGES

Table of Contents

- Executive Summary

Introduction

The Ukrainian Military's Strategic Vision and Technological Road Map

Report by **Kateryna Bondar**
Published March 6, 2025

Available Downloads

Kateryna Bondar
Fellow, Wadhwani AI Center

https://www.csis.org/analysis/ukraines-future-vision-and-current-capabilities-waging-ai-enabled-autonomous-warfare

Key Findings:

- The Ukrainian military's objective is to remove warfighters from direct combat and replace them with autonomous unmanned systems.
- Autonomy is not yet present on the battlefield in the war in Ukraine.
- The current deployment of AI is partial in scope, enhancing certain functions and addressing some operational challenges rather than enabling full system autonomy.
- Ukrainian forces have widely adopted small and medium first-person-view (FPV) drones as platforms that may be quickly adapted for diverse missions through modular design and interchangeable equipment.
- Ukraine's defense industry is developing standalone AI-driven software that can be integrated across various platforms to expand battlefield autonomy.
- Delegating target recognition to AI-enabled automatic target recognition (ATR) systems onboard unmanned platforms reduces human limitations and allows locking on to targets up to 2 km away.
- Autonomous navigation makes drones strikes three to four times more likely to succeed.
- Human oversight remains pivotal—particularly for engagement decisions—reflecting a human-in-the-loop approach that could shift toward higher-level supervision in the future while still maintaining human control of the system.

Don’t be evil?

A survey of the tech sector’s stance on lethal autonomous weapons

- Table ranks companies according to the level of concern regarding their potential (unintended) contribution to the development of lethal autonomous weapons.
  - https://www.paxforpeace.nl/publications/all-publications/dont-be-evil
- Autonomous weapons: Good? Or Bad?

HIGH CONCERN	Company working on military/security applications of relevant technologies + chose not to answer our survey’s questions in a meaningful way.
MEDIUM CONCERN	Company working on military/security applications of relevant technologies + answered that it was not working on lethal autonomous weapons;
	or Company not known as working on military/security applications of relevant technologies + chose not to answer our survey’s questions in a meaningful way.
BEST PRACTICE	Company answered to explain its policy on how it ensures its technology is not contributing to lethal autonomous weapons.
-	Unknown.

COMPANY	BEST PRACTICE	MEDIUM CONCERN	HIGH CONCERN	HQ	RELEVANT TECHNOLOGY	RELEVANT MILITARY/ SECURITY PROJECTS	COMMIT TO NOT DEVELOP
AerialX				Canada	Counter-drone systems	DroneBullet	
Airobotics				Israel	Autonomous drones	Border security patrol bots	
Airspace Systems				US	Counter-drone systems	Airspace interceptor	
Alibaba				China	AI chips, Facial recognition	-	
Amazon				US	Cloud, Drones, Facial and speech recognition	JEDI, Rekognition	
Anduril Industries				US	AI platforms	Project Maven, Lattice	
Animal Dynamics				UK	Autonomous drones	Skeeter	X
Apple				US	Computers, Facial and speech recognition	-	
Arbe robotics				Israel	Autonomous vehicles	-	X
ATOS				France	AI architecture, cyber security, data management	-	
Baidu				China	Deep learning, Pattern recognition	-	
Blue Bear Systems				UK	Unmanned maritime and aerial systems	Project Mosquito/LANCA	
Cambricon				China	AI chips	-	
Citadel Defense				US	Counter-drone systems	Titan	
Clarifai				US	Facial recognition	Project Maven	
Cloudwalk Technology				China	Facial recognition	-	
Corenova Technologies				US	Autonomous swarming systems	HiveDefense, OFFSET	
DeepGlint				China	Facial recognition	-	
Dibotics				France	Autonomous navigation, Drones	'Generate'	
EarthCube				France	Machine learning	'algorithmic warfare tools of the future'	
Facebook				US	Social media, Pattern recognition, Virtual Reality	-	
General Robotics				Israel	Ground robots	Dogo	X
Google				US	AI architecture, Social media, Facial recognition	-	X
Heron Systems				US	AI software, ML, Drone applications	'solutions to support tomorrow's military aircraft'	

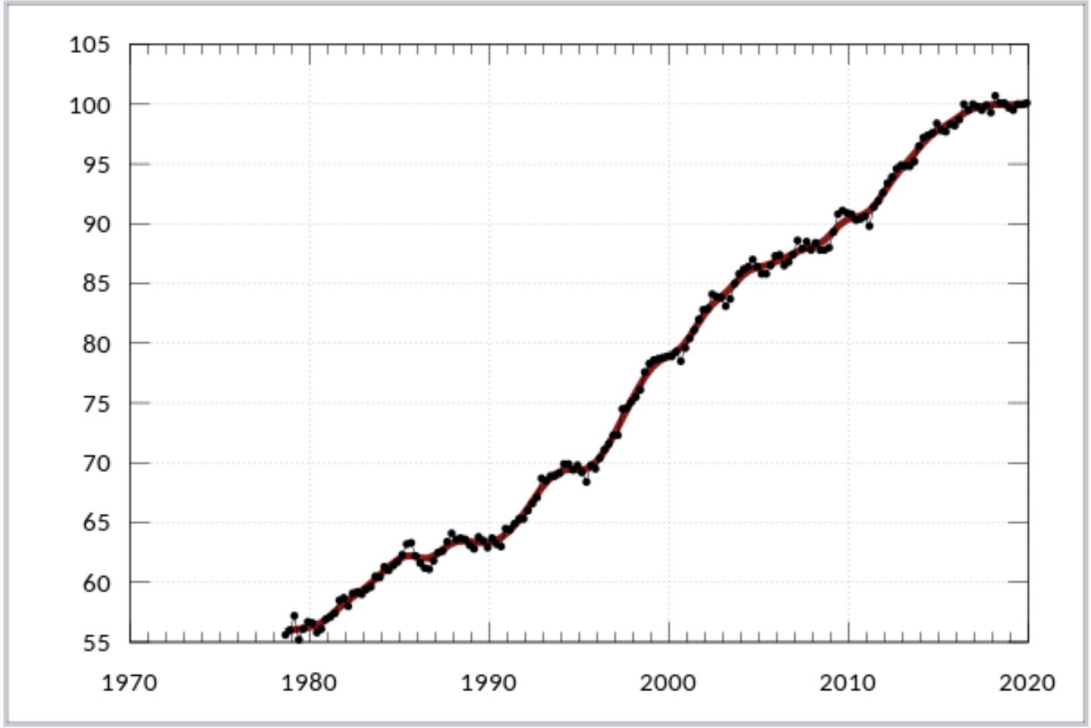
HiveMapper				US	Pattern recognition, Mapping	HiveMapper app	X
IBM				US	AI chips, Cloud, Super computers, Facial recognition	Nuclear testing super computers, ex-JEDI	
Innoviz				Israel	Autonomous vehicles	-	
Intel				US	AI chips, UAS	DARPA HIVE	
Megvii				China	Facial recognition	-	
Microsoft				US	Cloud, Facial recognition	HoloLens, JEDI	
Montvieux				UK	Data analysis, Deep learning	'Revolutionise human information relationship for defence'	
Naver				S. Korea	'Ambient Intelligence', Autonomous robots, Machine vision systems	-	
Neurala				US	Deep learning neural network software	Target identification software for military drones	
Oracle				US	Cloud, AI infrastructure, Big data	ex-JEDI	
Orbital Insight				US	Geospatial analytics	-	
Palantir				US	Data analytics	DCGS-A	
Percepto				Israel	Autonomous drones	-	
Roboteam				Israel	Unmanned systems; AI software	Semi-autonomous military UGVs	
Samsung				S. Korea	Computers and AI platforms	-	
SenseTime				China	Computer vision, Deep learning	SenseFace, SenseTotem for police use	
Shield AI				US	Autonomous (swarming) drones	Nova	
Siemens				Germany	AI, Automation	KRNS, TRADES	
Softbank				Japan	Telecom, Robotics	-	X
SparkCognition				US	AI systems, Swarm technology	'works across the defense and national security space in the U.S.'	
Synesis				Belarus	AI- and Cloud-based applications, Pattern recognition	Kipod	
Taiwan Semiconductor				Taiwan	AI chips	-	
Tencent				China	AI applications, Cloud, ML, Pattern recognition	-	
Tharsus				UK	Robotics	-	
VisionLabs				Russia	Visual recognition	-	X
Yitu				China	Facial recognition	Police use	

DISRUPTIVE AI & ROBOTICS

Productivity

- Crucial factor in production performance of firms and nations
- Increasing national productivity: raise living standards & help businesses to be more profitable
- Important measure: output per worker (e.g. GDP per worker)

- Robotics: potential to increase productivity manyfold!



Labour productivity growth in Australia since 1978, measured by GDP per hour worked (indexed)

Traditional, Contemporary & Future Industrial Robots

Mobile robot platform coupled with a dual-arm articulated robot works autonomously processing parts for lights-out manufacturing. (Courtesy of OTTO Motors)

Sparks of Artificial General Intelligence: Early experiments with GPT-4

Sébastien Bubeck Varun Chandrasekaran Ronen Eldan Johannes Gehrke
Eric Horvitz Ece Kamar Peter Lee Yin Tat Lee Yuanzhi Li Scott Lundberg
Harsha Nori Hamid Palangi Marco Tulio Ribeiro Yi Zhang

Microsoft Research

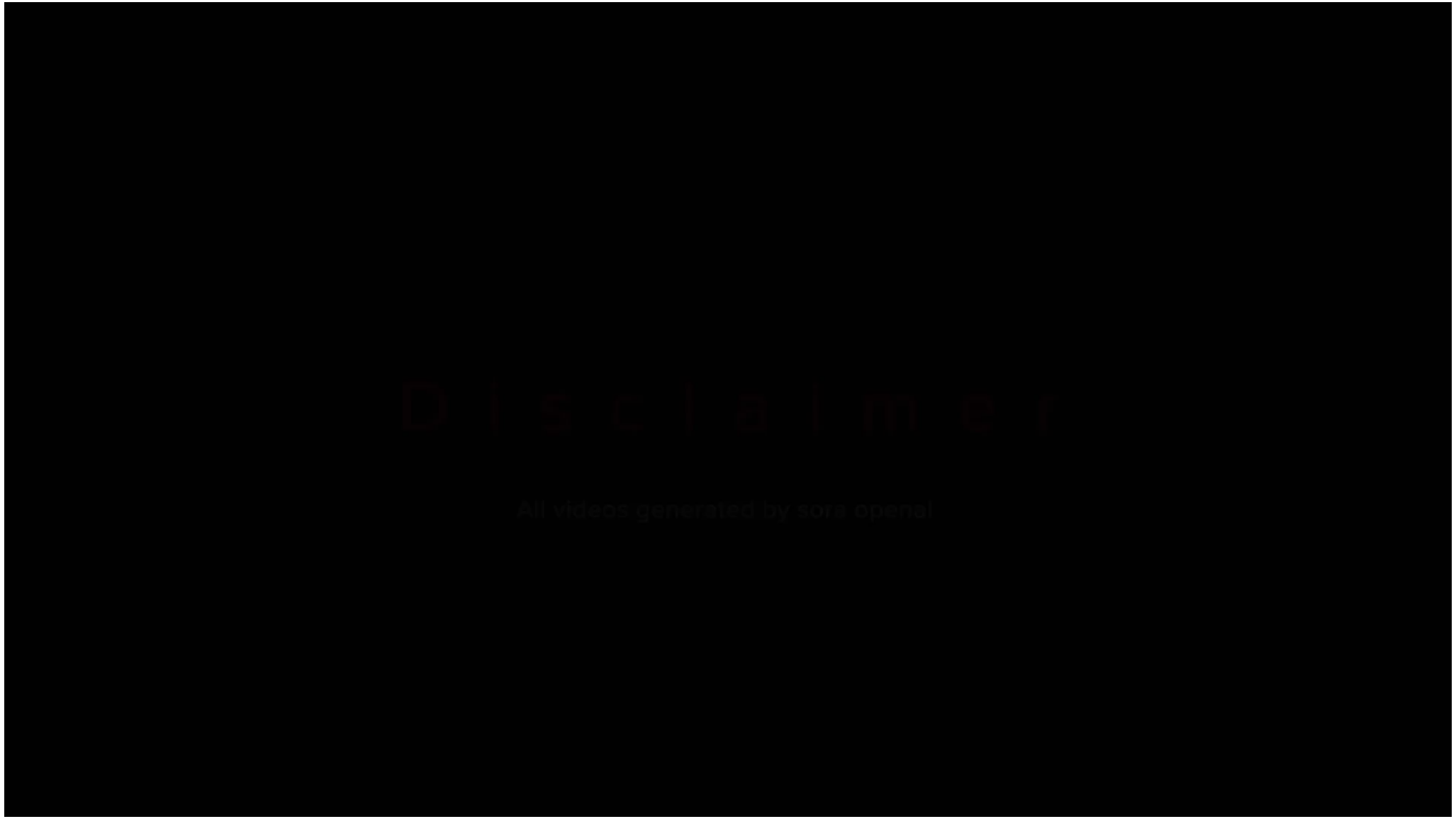
Abstract

Artificial intelligence (AI) researchers have been developing and refining large language models (LLMs) that exhibit remarkable capabilities across a variety of domains and tasks, challenging our understanding of learning and cognition. The latest model developed by OpenAI, GPT-4 [Ope23], was trained using an unprecedented scale of compute and data. In this paper, we report on our investigation of an early version of GPT-4, when it was still in active development by OpenAI. We contend that (this early version of) GPT-4 is part of a new cohort of LLMs (along with ChatGPT and Google's PaLM for example) that exhibit more general intelligence than previous AI models. We discuss the rising capabilities and implications of these models. We demonstrate that, beyond its mastery of language, GPT-4 can solve novel and difficult tasks that span mathematics, coding, vision, medicine, law, psychology and more, without needing any special prompting. Moreover, in all of these tasks, GPT-4's performance is strikingly close to human-level performance, and often vastly surpasses prior models such as ChatGPT. Given the breadth and depth of GPT-4's capabilities, we believe that it could reasonably be viewed as an early (yet still incomplete) version of an artificial general intelligence (AGI) system. In our exploration of GPT-4, we put special emphasis on discovering its limitations, and we discuss the challenges ahead for advancing towards deeper and more comprehensive versions of AGI, including the possible need for pursuing a new paradigm that moves beyond next-word prediction. We conclude with reflections on societal influences of the recent technological leap and future research directions.

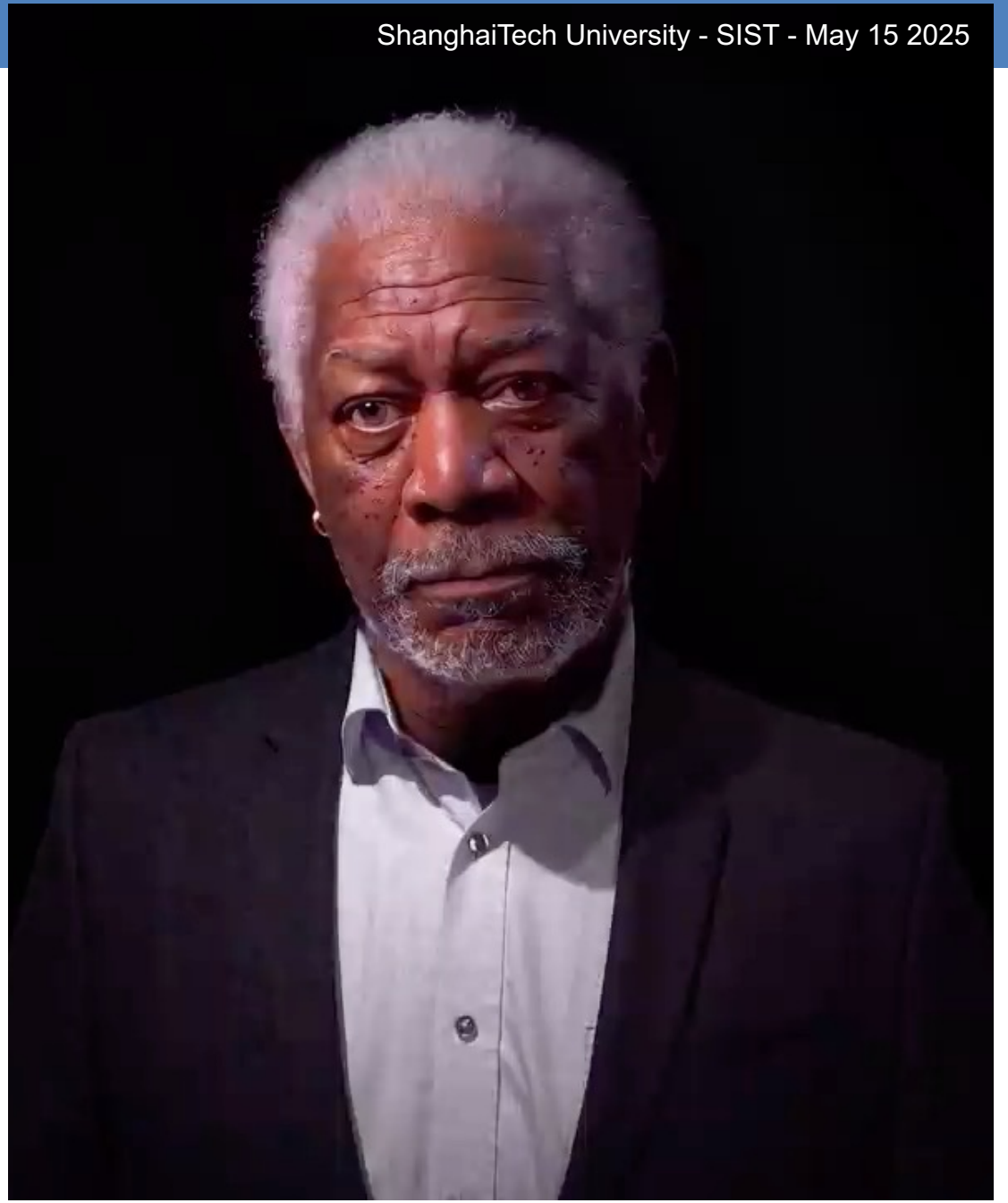
https://arxiv.org/pdf/2303.12712.pdf

April 13, 2023

Sora Text to Video AI introduced Feb 2024

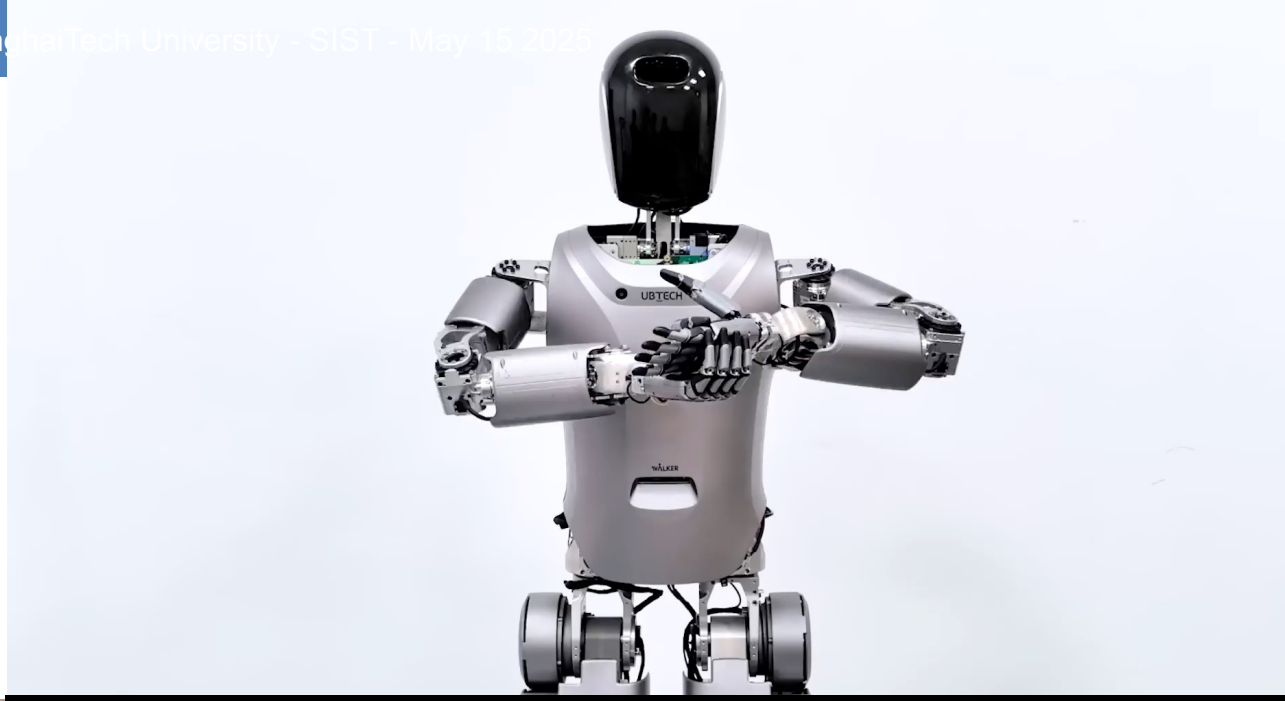


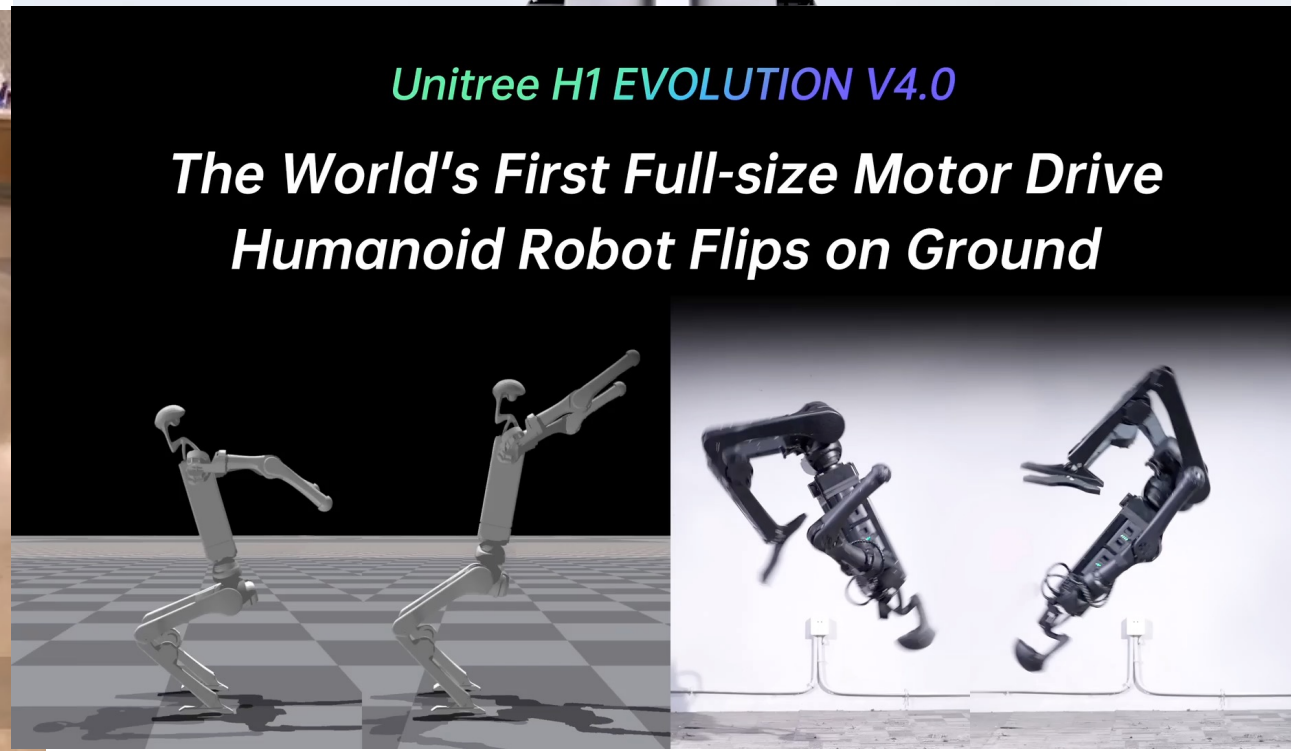
Deep Fakes



https://www.youtube.com/watch?v=oxXpB9pSETo

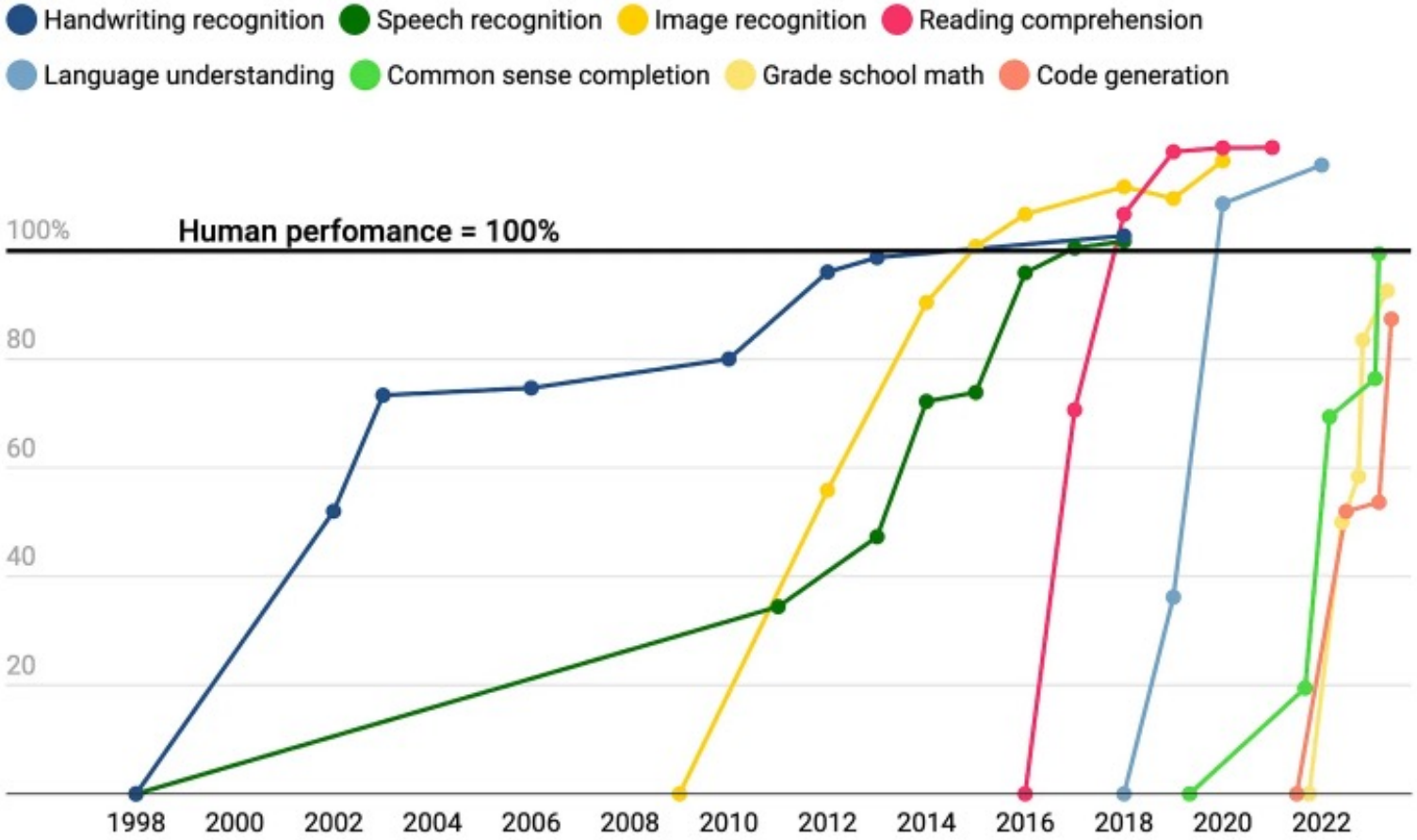
Huge trend: AI enabled Mobile Manipulation & Humanoid Robotics





AI has surpassed humans at a number of tasks and the rate at which humans are being surpassed at new tasks is increasing

State-of-the-art AI performance on benchmarks, relative to human performance



For each benchmark, the maximally performing baseline reported in the benchmark paper is taken as the "starting point", which is set at 0%. Human performance number is set at 100%. Handwriting recognition = MNIST, Language understanding = GLUE, Image recognition = ImageNet, Reading comprehension = SQuAD 1.1, Reading comprehension = SQuAD 2.0, Speech recognition = Switchboard, Grade school math = GSK8k, Common sense completion = HellaSwag, Code generation = HumanEval.

Chart: Will Henshall for TIME • Source: ContextualAI

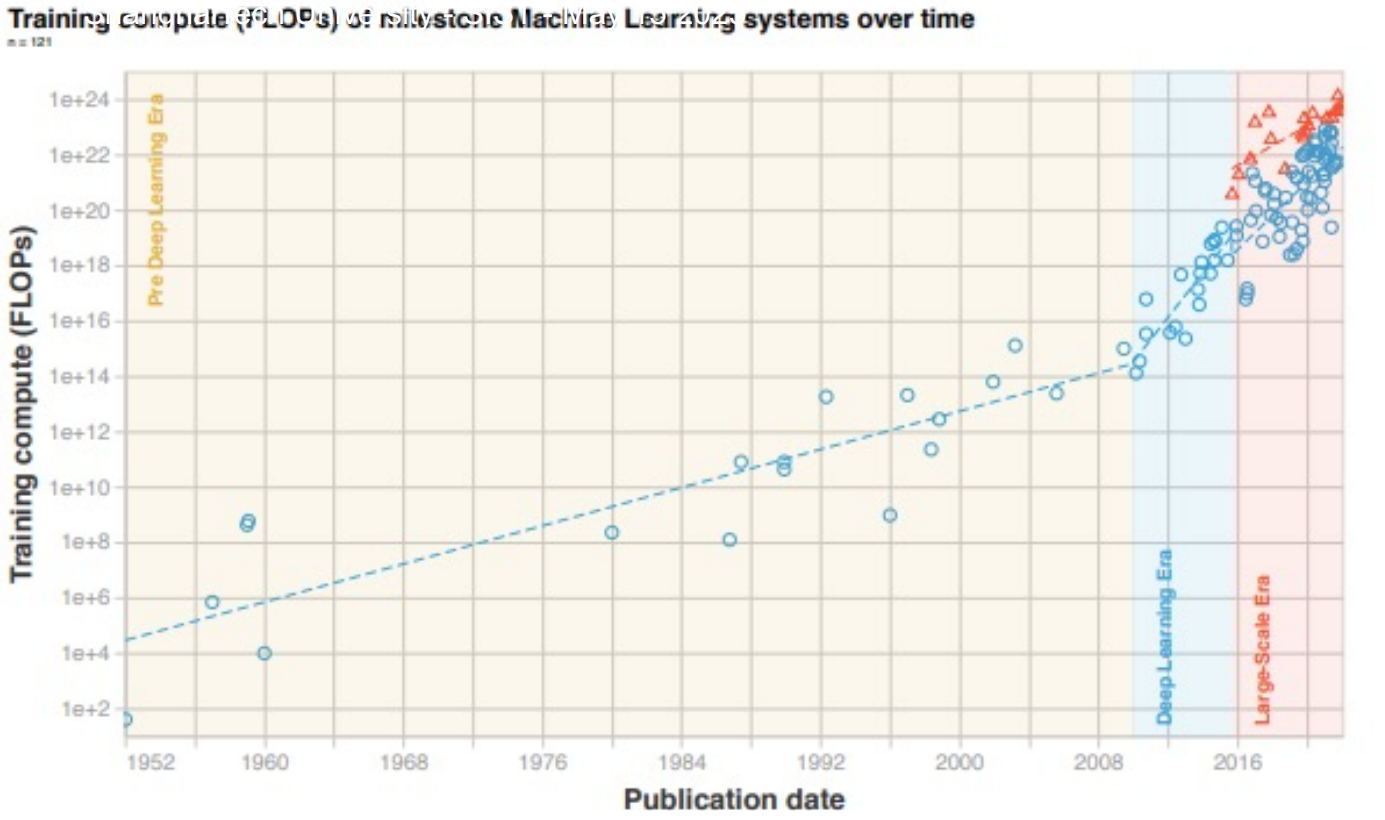
TIME

Exponential Growth of Compute in ML

- Large scale model training (such as LLM):

  Doubled every 9.9 month
  =>
  100x every 5.5 years!

  10000x after 11 years…



Period	Data	Scale (start to end)	Slope	Doubling time
1952 to 2010 Pre Deep Learning Trend	All models ($n = 19$)	3e+04 to 2e+14 FLOPs	0.2 OOMs/year [0.1; 0.2; 0.2]	21.3 months [17.0; 21.2; 29.3]
2010 to 2022 Deep Learning Trend	Regular-scale models ($n = 72$)	7e+14 to 2e+18 FLOPs	0.6 OOMs/year [0.4; 0.7; 0.9]	5.7 months [4.3; 5.6; 9.0]
September 2015 to 2022 Large-Scale Trend	Large-scale models ($n = 16$)	4e+21 to 8e+23 FLOPs	0.4 OOMs/year [0.2; 0.4; 0.5]	9.9 months [7.7; 10.1; 17.1]

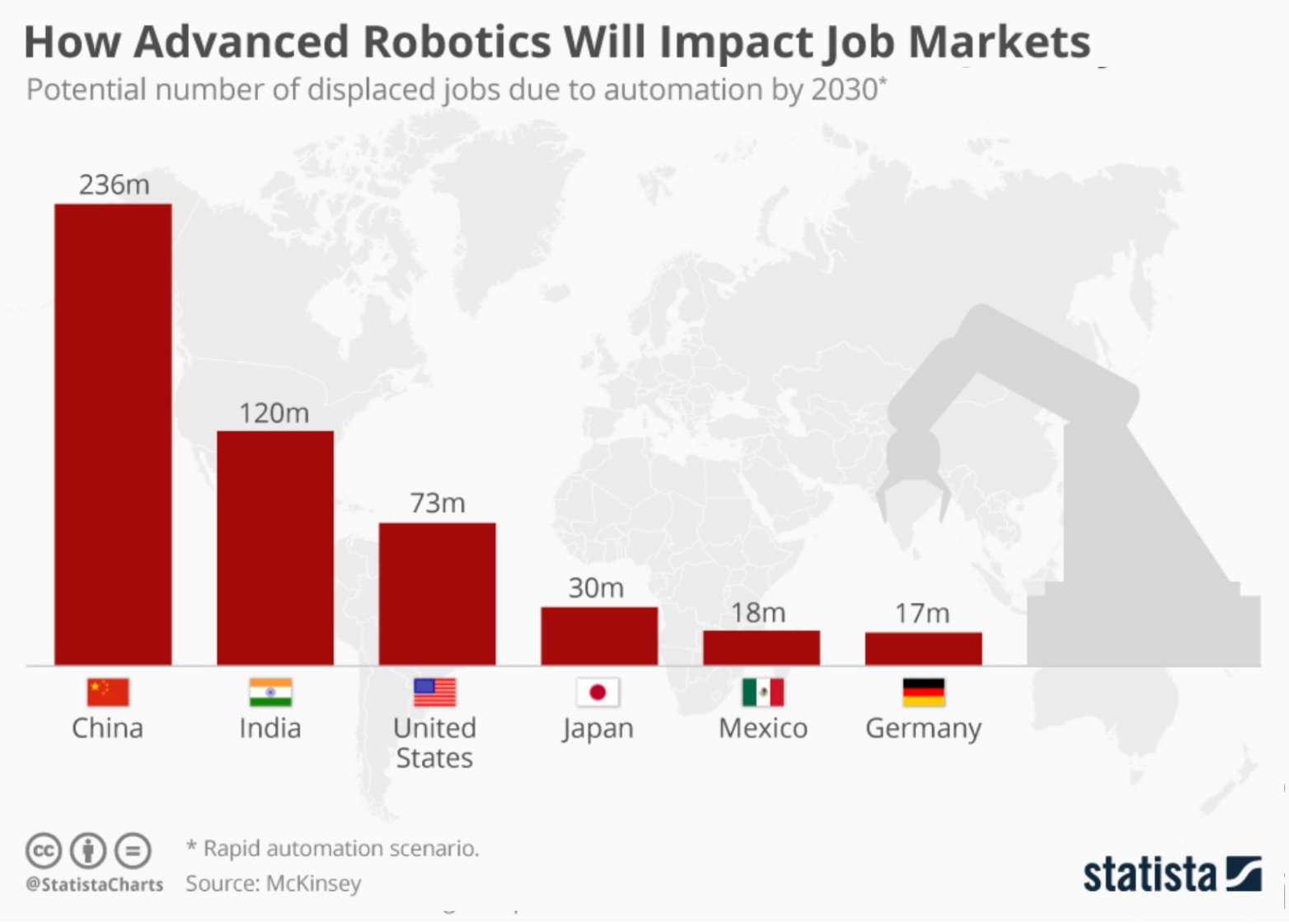
Table 2: Summary of our main results. In 2010 the trend accelerated along the with the popularity of Deep Learning, and in late 2015 a new trend of large-scale models emerged.

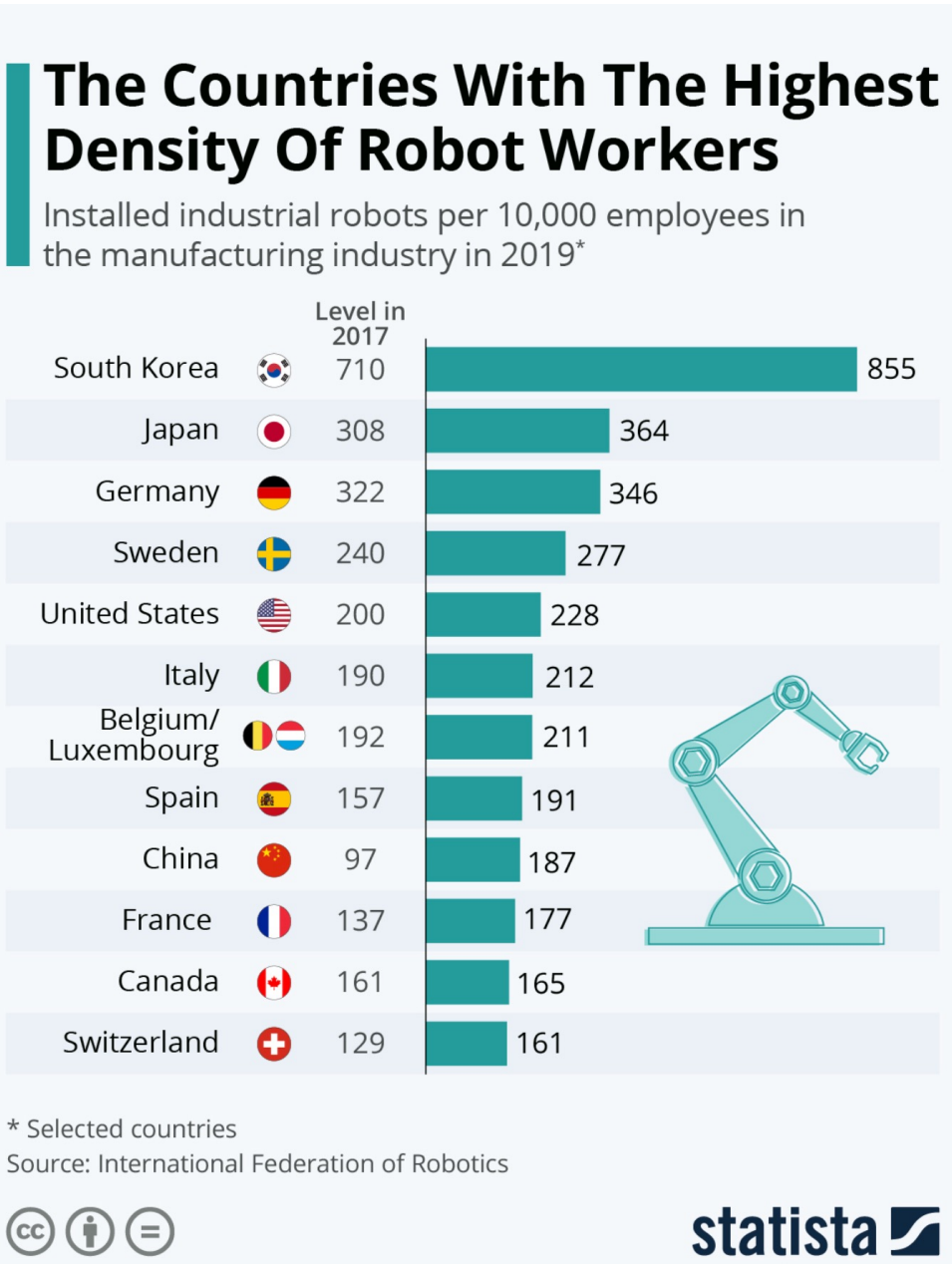
Sevilla, J., Heim, L., Ho, A., Besiroglu, T., Hobbhahn, M., & Villalobos, P. (2022, July). Compute trends across three eras of machine learning. In *2022 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.


UBTECH Humanoid Robot Walker S
Workstation Assistant in NIO Production Line

Immediate impact on jobs

https://www.statista.com/topics/1476/industrial-robots/



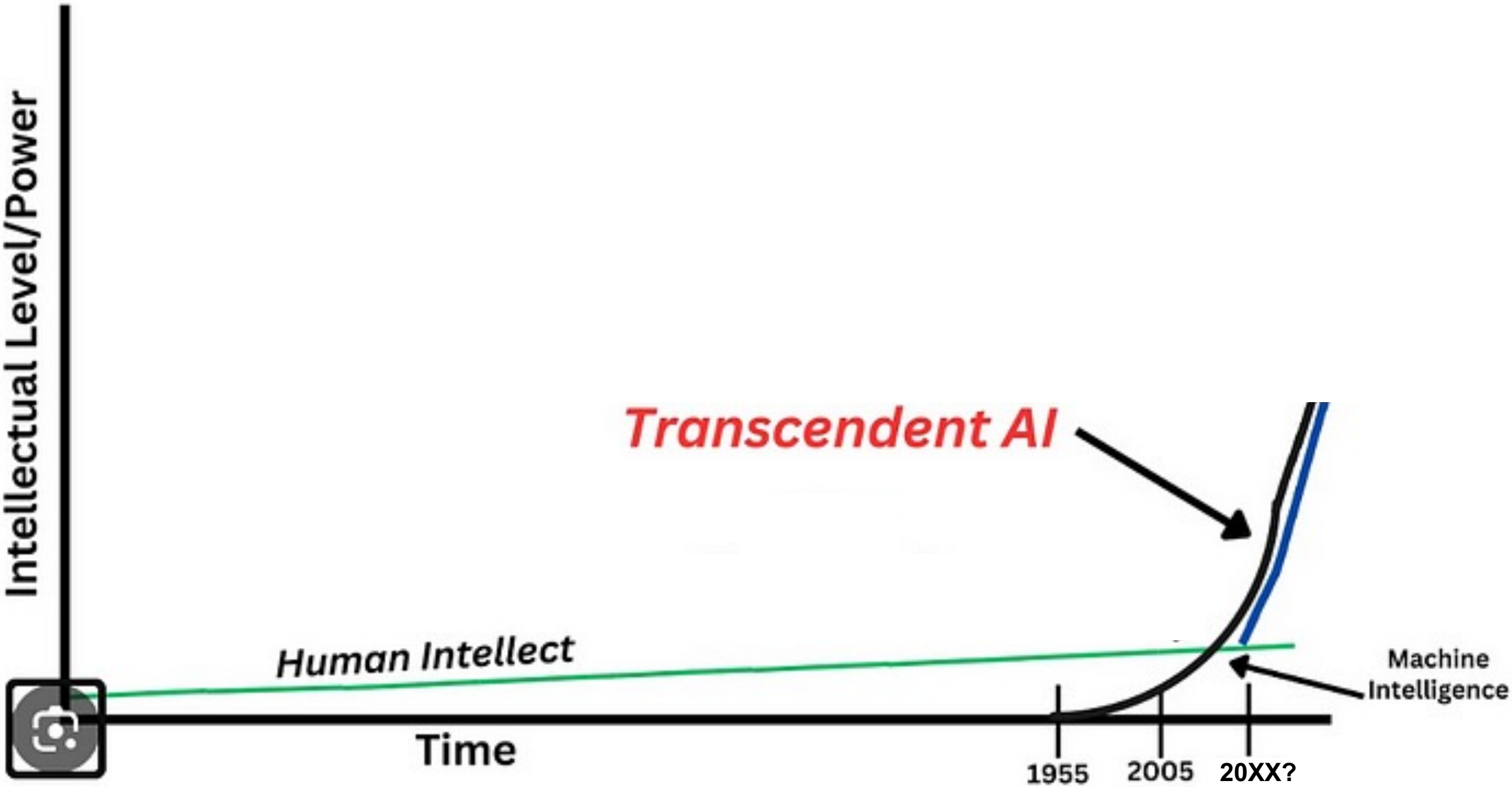


AI & Robotics may change Everything

- Industry Robotics & Service robotics & AI develop quickly and may have very big impact!
- AI & Robots may / will change our daily life:
  - At work (if we have work) we will interact with them daily
  - Autonomous transportation services (cars/ Didi, Uber) may replace privately owned cars (and public transport?) (China has about 1.5 million taxi drivers + Didi + Bus drivers + Delivery drivers)
  - In our home they (hopefully) will do work

- Most work may be done by AI & Robots =>
  - Great productivity: Products and services may become VERY cheap!
  - Unemployment: Many jobs will be lost.
    - Option 1: Lost jobs will be replaced by new jobs (e.g. robot engineer, prompt engineers, AI programmer, AI artists)
    - Option 2: Most jobs will no be replaced => big unemployment

- Potentially: new (updated) social systems might emerge
  - Unconditional basic income?
  - Communism 2.0?
  - …

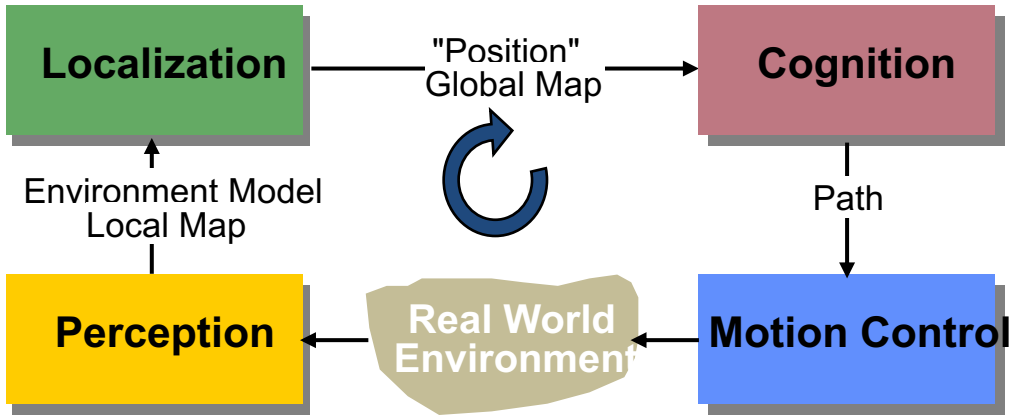

Intellectual Level/Power
Transcendent AI
Human Intellect
Machine Intelligence
Time
1955
2005
20XX?

BEHAVIOR BASED ROBOTICS

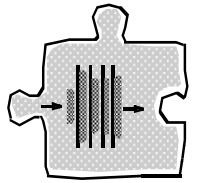
"Small Model" – "Little Brain"

Control Architectures / Strategies

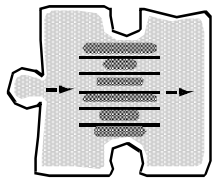
- Control Loop
  - **dynamically changing**
  - **no compact model available**
  - **many sources of uncertainty**



- Three Approaches
  - Classical AI (Big Model)
    - complete modeling
    - function based
    - horizontal decomposition



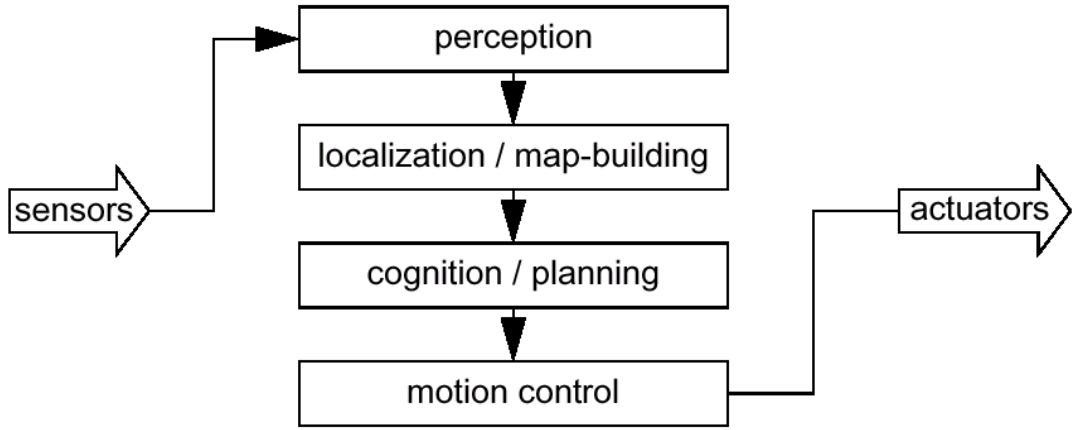
  - New AI (Nouvelle AI; Small Model; Behavior Based Robotics)
    - sparse or no modeling
    - behavior based
    - vertical decomposition
    - bottom up

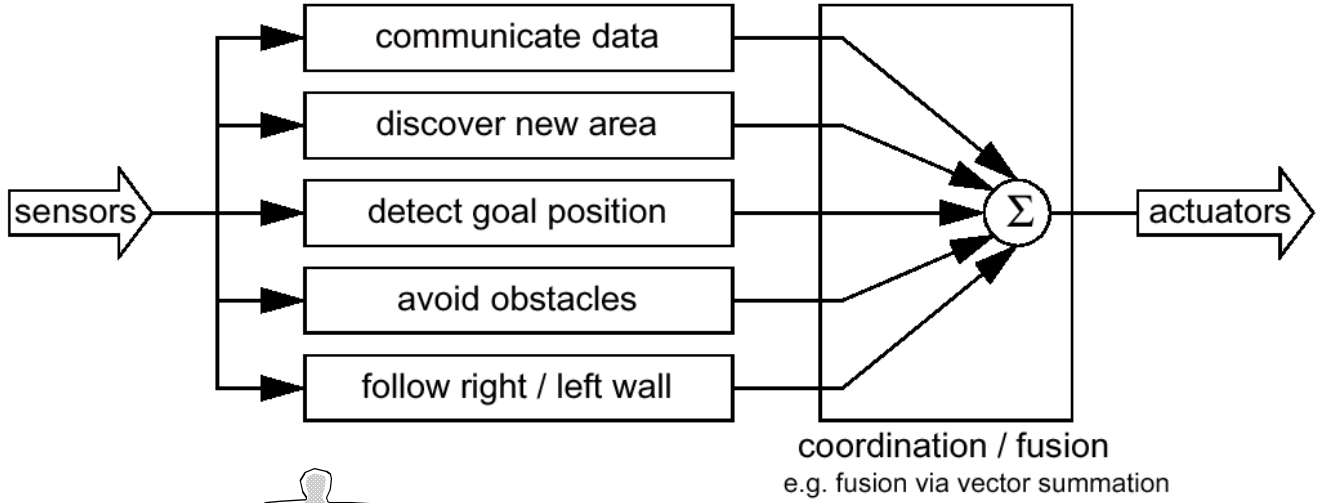


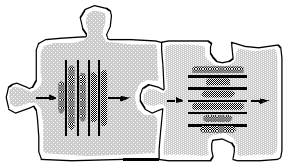
  - DL/ Reinforcement Learning

Two Approaches

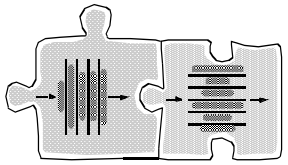
- Classical AI
  (model based navigation)
  - complete modeling
  - function based
  - horizontal decomposition
- New AI
  (behavior based navigation)
  - sparse or no modeling
  - behavior based
  - vertical decomposition
  - bottom up
- Possible Solution
  - Combine Approaches (= Hybrid Approach)

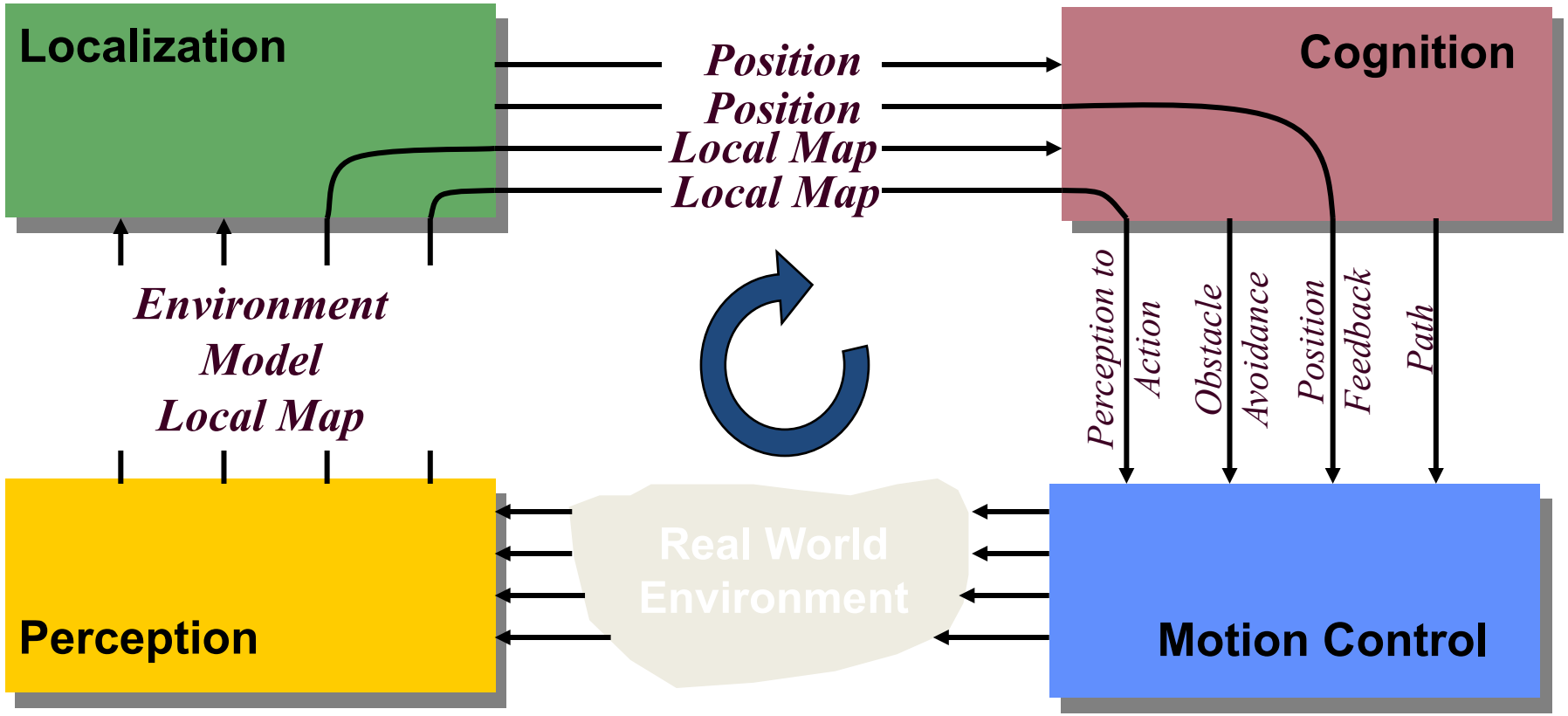






Mixed Approach Depicted into the General Control Scheme



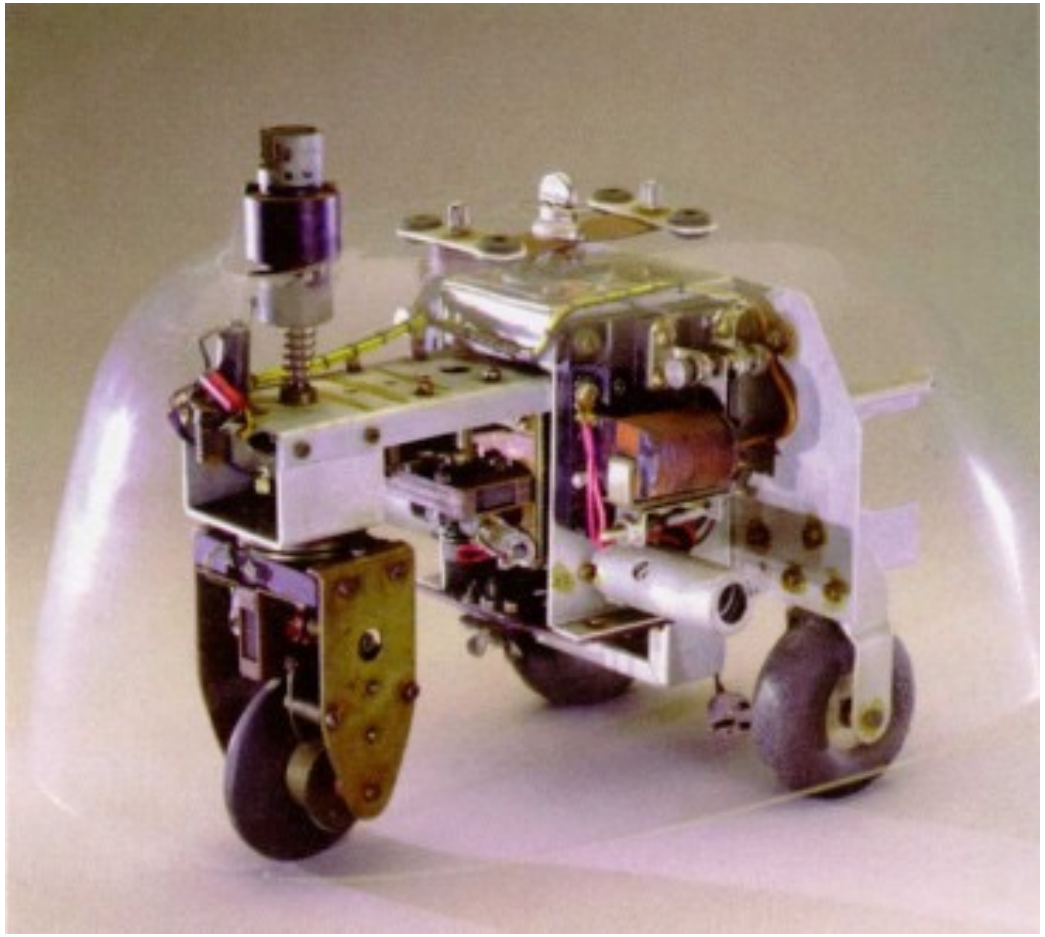


Emergence

- Adaptive behavior
  - emerges from complex interactions between body, world and brain
- Non-centrally controlled (or designed) behavior
  - results from the interactions of multiple simple components
- Meanings:
  - Surprising situations or behaviors
  - Property of system not contained in any of its parts
  - Behavior resulting from agent-environment interaction not explicitly programmed
- Ant colony:
  - self-organized; simple individuals; local interactions =>
  - emergent behavior - No global control

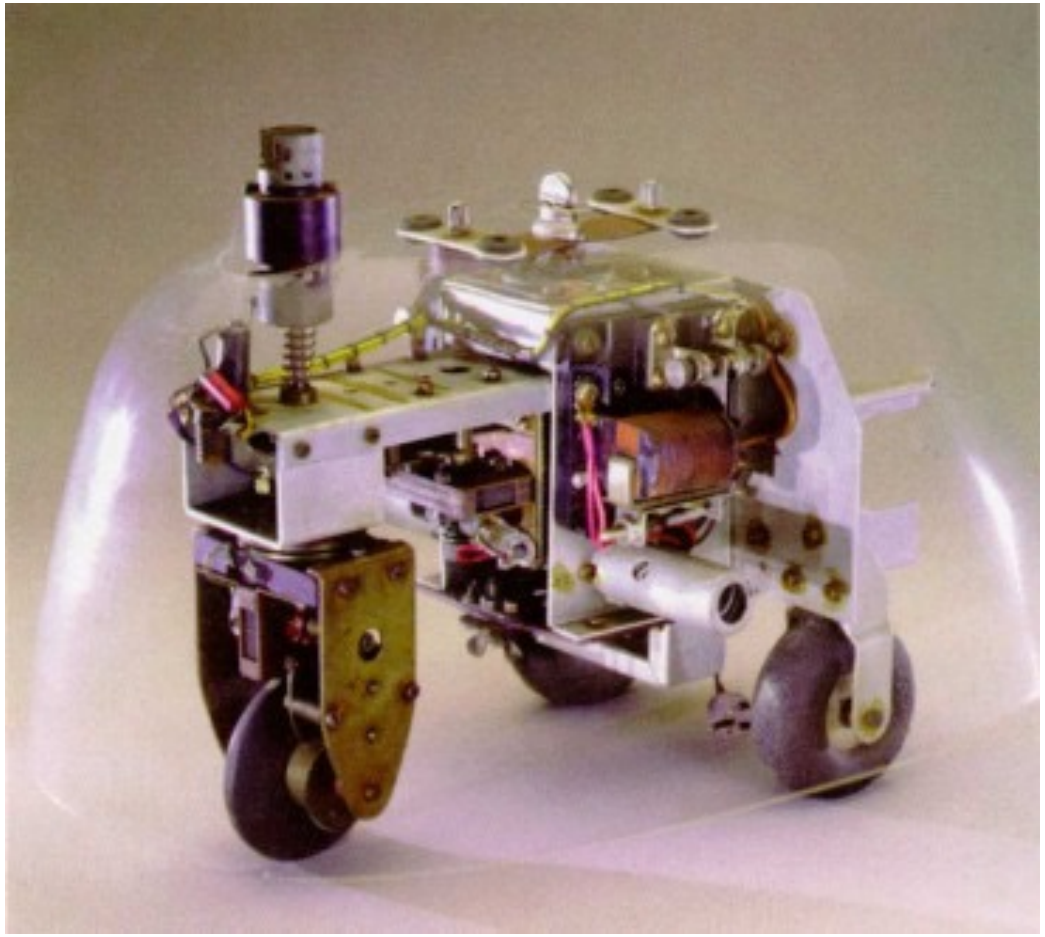
Grey Walter's Tortoise

- Turtle shape robots 1949
- Purely analogue electronics
- Phototaxis: go towards the light
- Sensors:
  - 1 photocell,
  - 1 bump sensor
- 2 motors
- Reactive control



Grey Walter's Tortoise

- Behaviors:
  - Seek light
  - Head toward weak light
  - Back away from bright light
  - Turn and push (obstacle avoidance)
  - Recharge battery

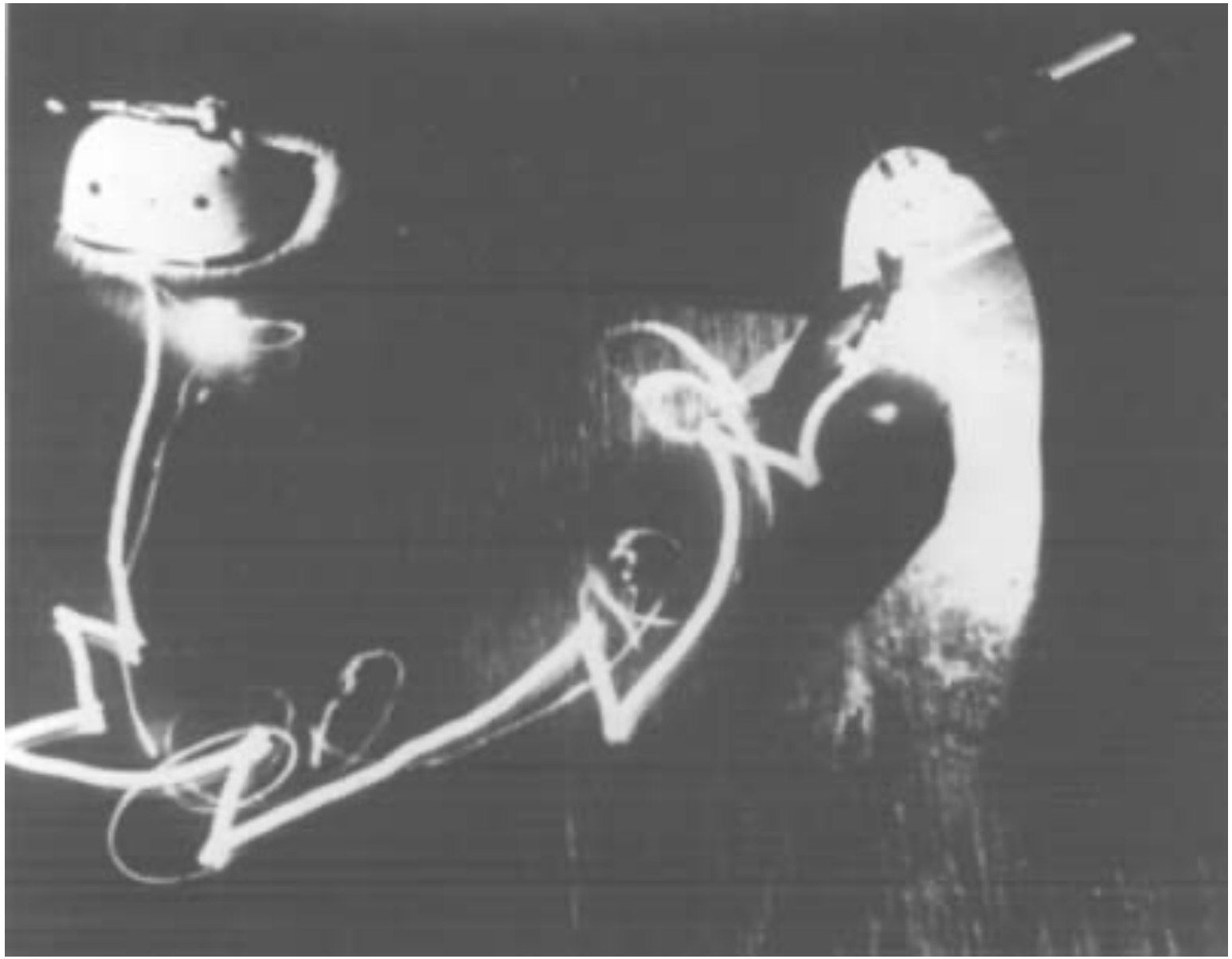


Turtle Principles

- Simple is better
  - e.g., clever recharging strategy
- Exploration/ speculation: keeps moving
  - except when charging
- Attraction:
  - motivation to approach light
- Aversion:
  - motivation to avoid obstacles, slopes

Tortoise behavior

- A path: a candle on top of the shell

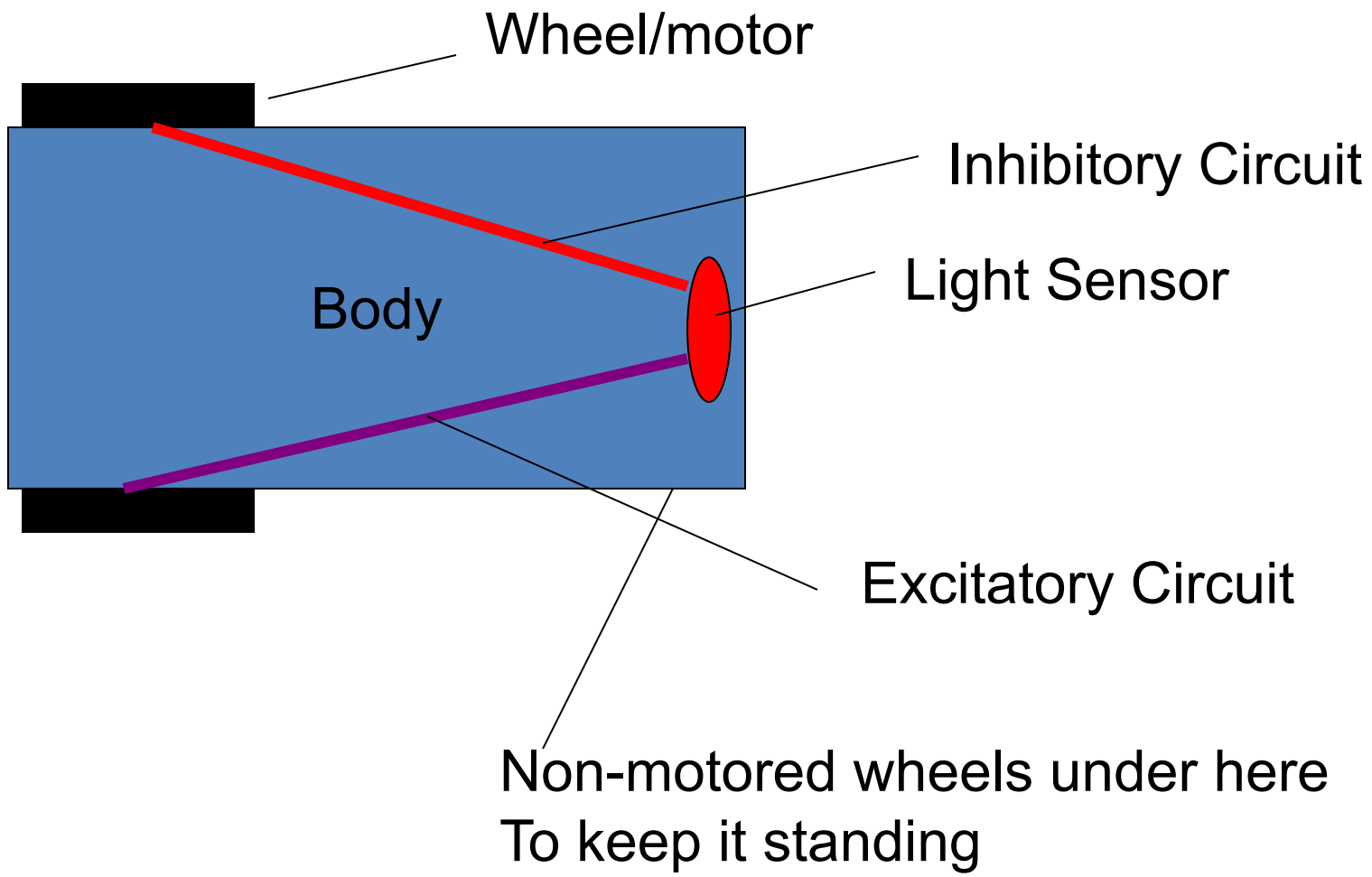


Video …


BRITISH
PATHÉ

Braitenberg's Vehicles

- Valentino Braitenberg (1926)
- 1984: "Vehicles: Experiments in Synthetic Psychology"

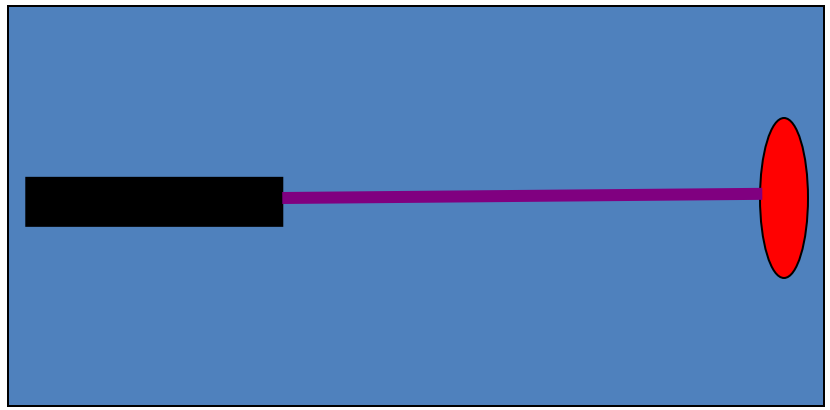


Material: Tommy Walsh http://my.fit.edu/~twalsh/Braitenberg%20Vehicles.ppt

Definitions

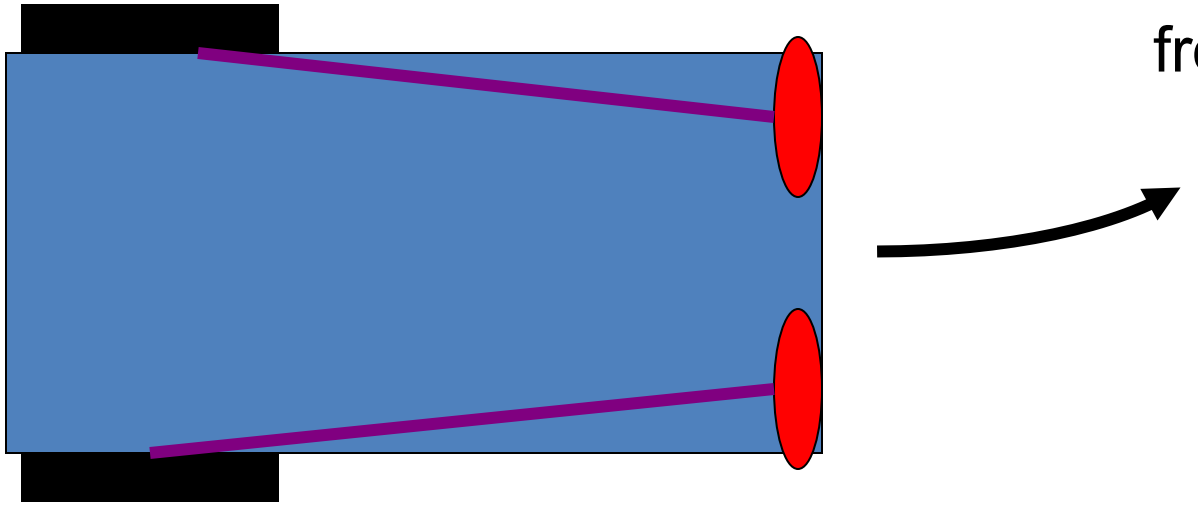
- Inhibitory circuit: when sensor gets activated, motor slows
- Excitatory circuit: when sensor gets activated, motor speeds
- Sensor is a light sensor, unless otherwise noted

Vehicle 1: Alive



Basic Braitenberg vehicle:
Goes towards light source

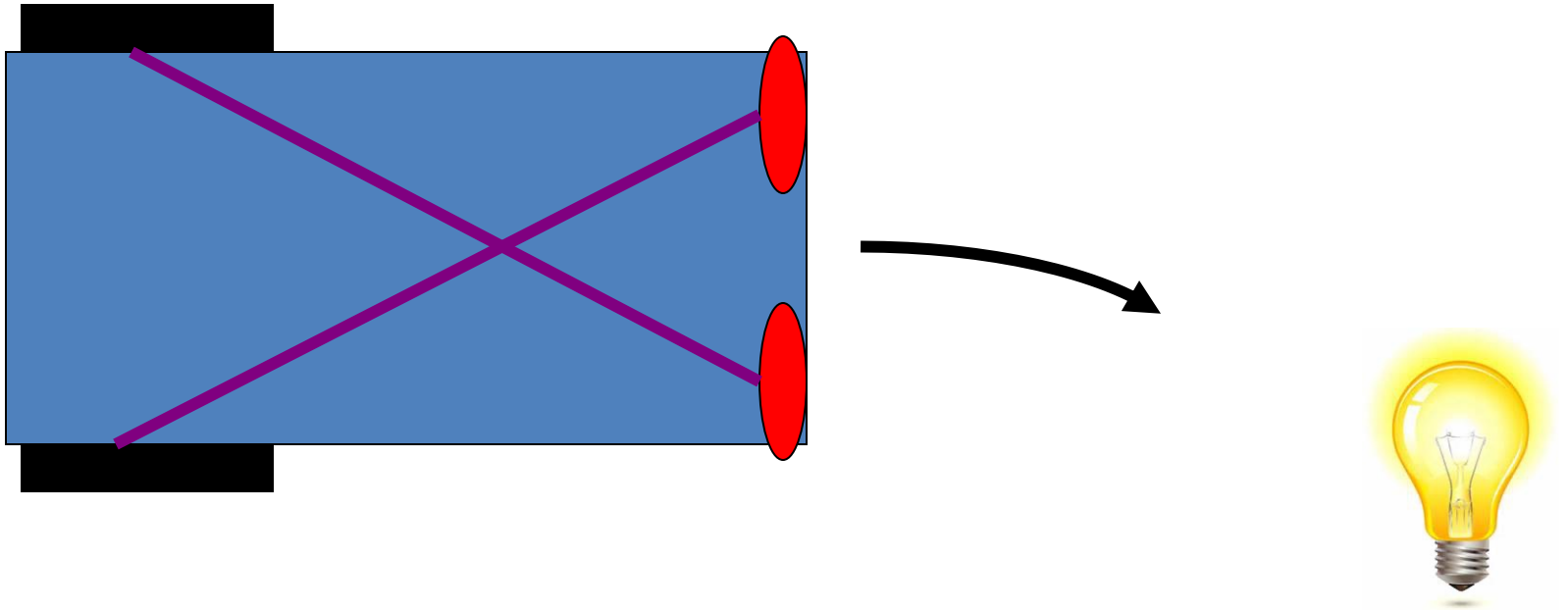
Vehicle 2: Cowardly



More light right →
right wheel turns faster →
turns towards the left, away from the light.

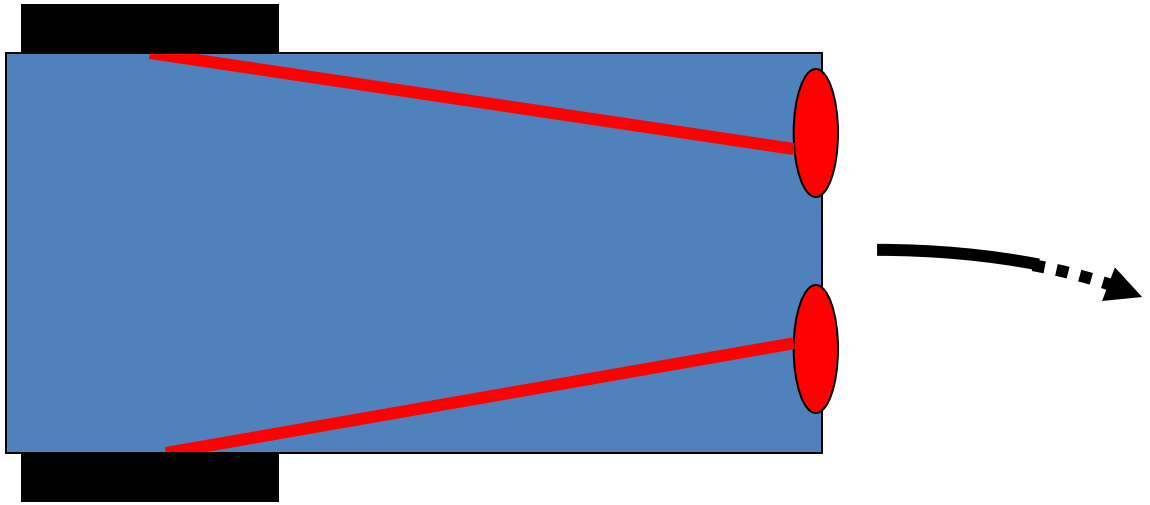
Demonstrates “fight or flight” instinct in animals
Turns away from light if one sensor is activated more than the other
If both are equal, light source is “attacked”

Vehicle 2b: Aggressive



Faces light source and drives toward it

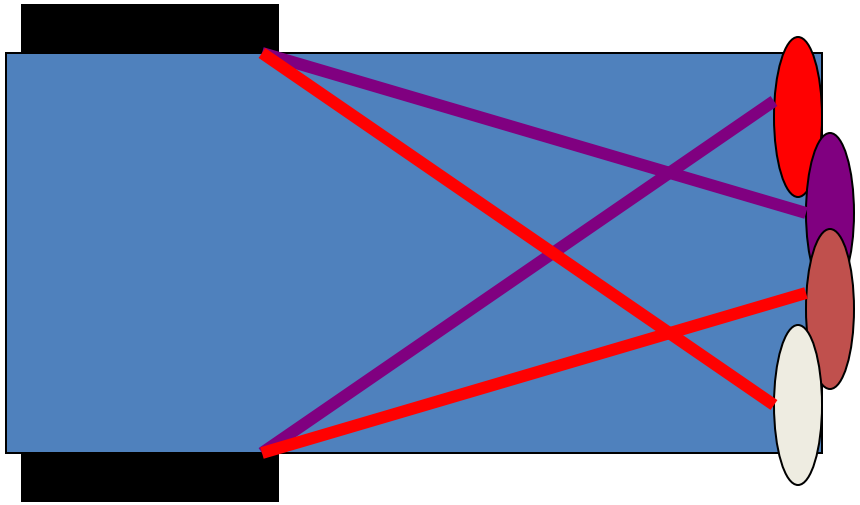
Vehicle 3: Loving



Drives forward
Faces the light source and slows down

Models love/adoration

A little more complicated: Vehicle 3c: Knowing

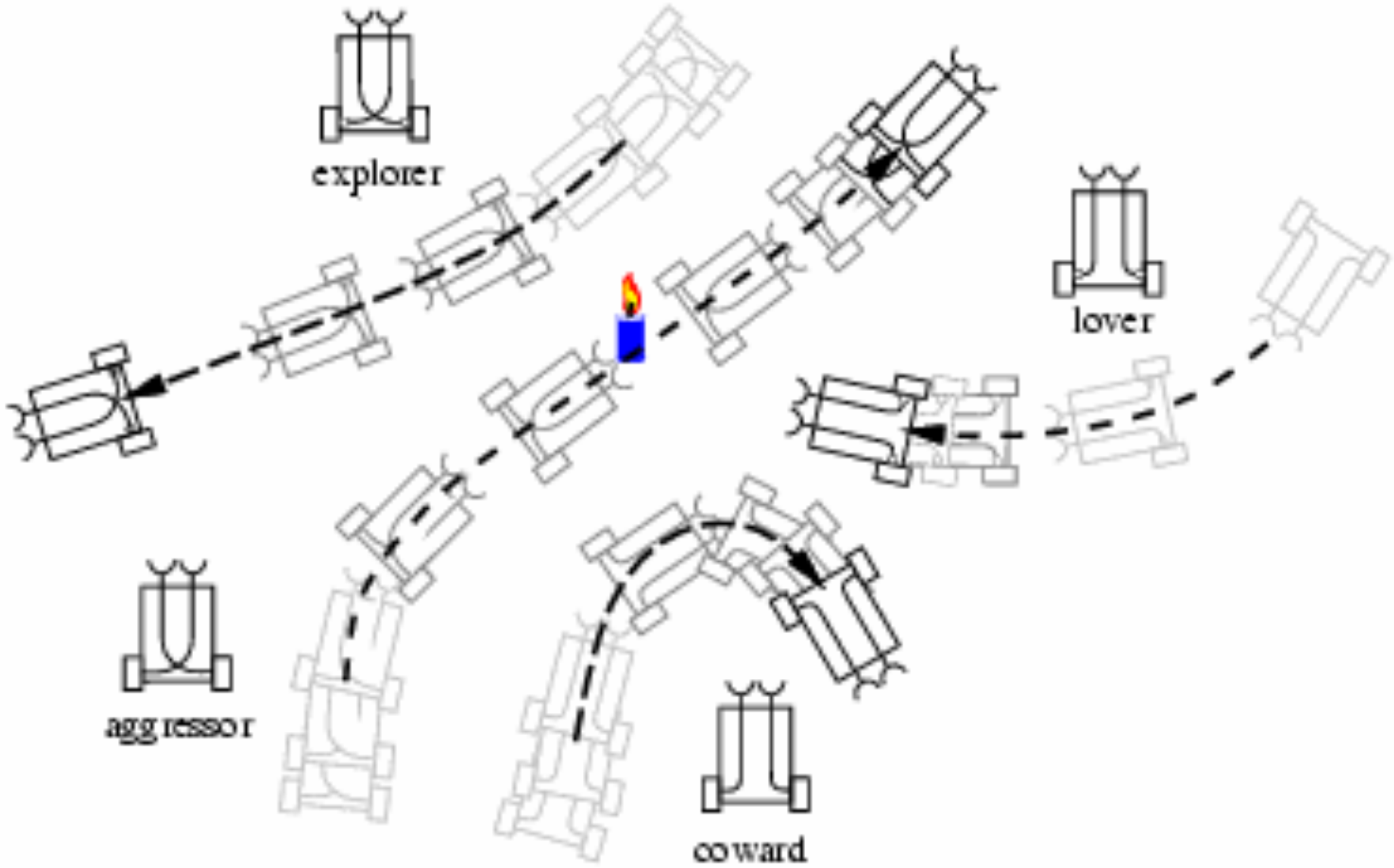


Light Sensor
Temperature Sensor
Organic Material Sensor
Oxygen Sensor

Different sensors:

Turns towards light,
doesn’t like heat,
loves organic material,
searches for best Oxygen

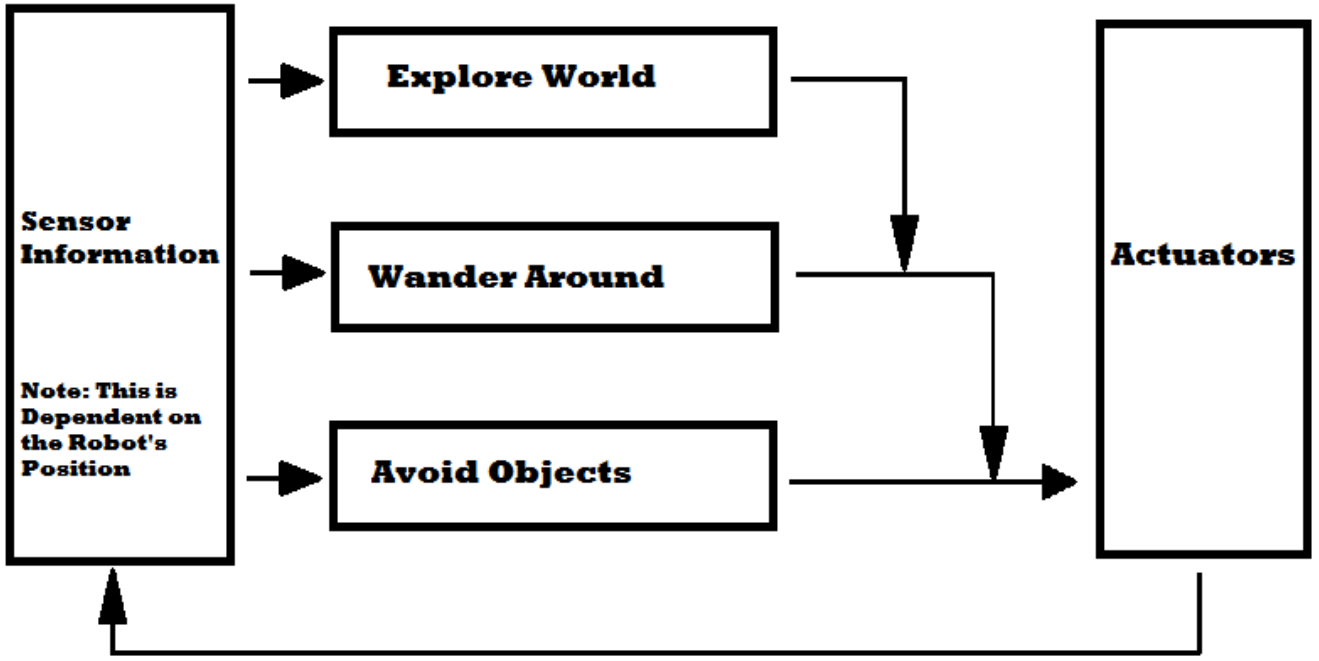
Emergent Behavior: Performs the brain function of simplest living beings


explorer
lover
aggressor
coward

Subsumption architecture

- Rodney Brooks (MIT; founder of iRobot and Rethink Robotics) 1991:
  - “The world is its own best model” =>
  - “Intelligence without representation”
- Emergent behaviors

- Conclusions:
  - Emergent behaviors quite interesting/ impressive.
  - More complex tasks often need more intelligence.
  - => Behaviors good for low level tasks.



GOOD LUCK WITH THE FINAL!