上海科技大学
ShanghaiTech University

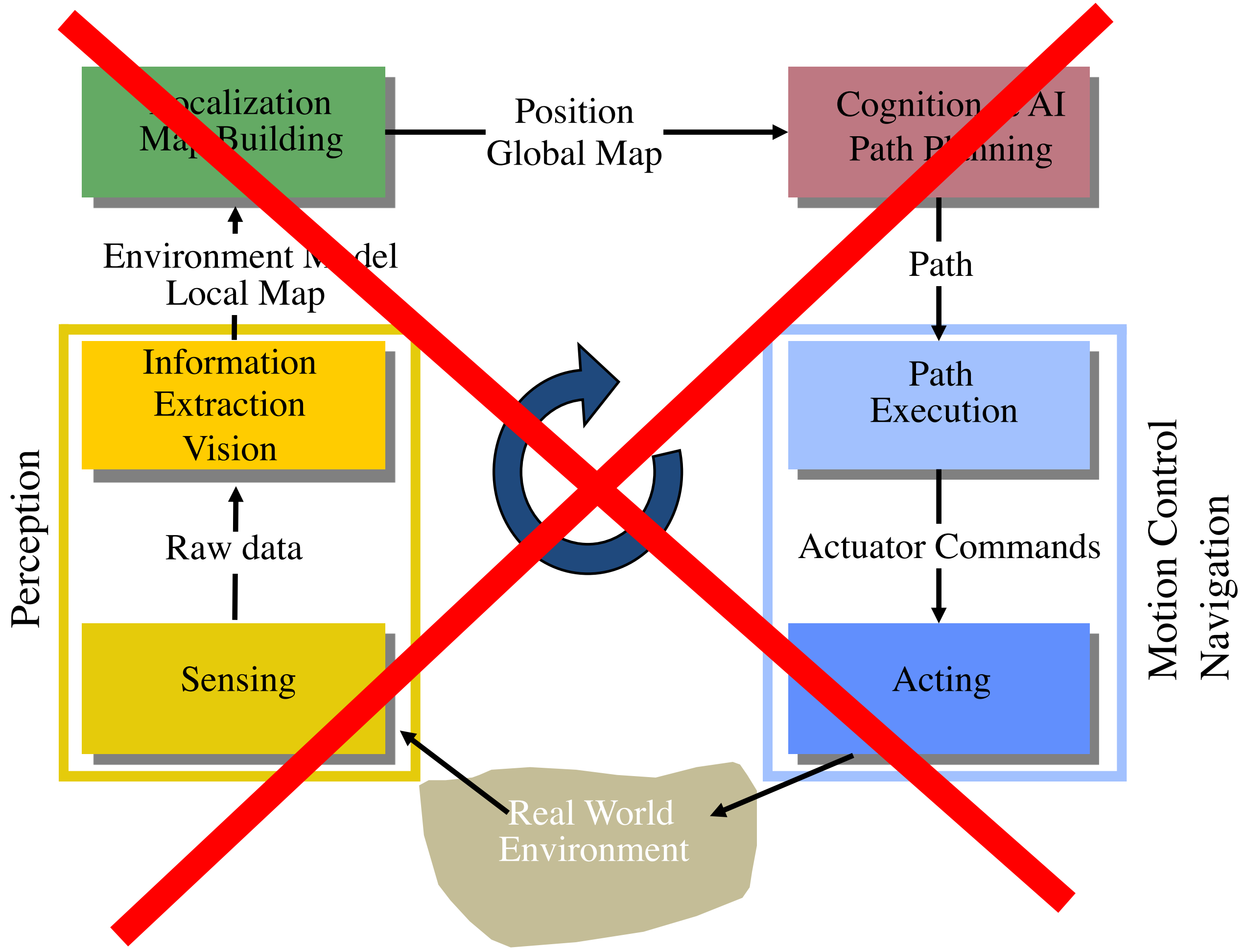
CS283: **Robotics Fall 2019: DL & Ethics**

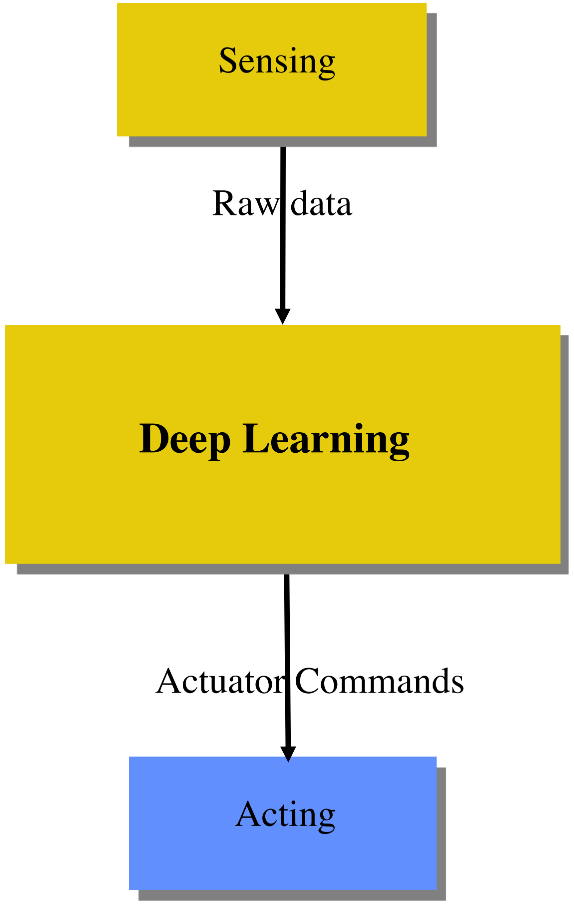
Sören Schwertfeger / **师泽仁**

ShanghaiTech University

DEEP LEARNING IN ROBOTICS

End-to-End Deep Learning





Deep Learning

- Very hot trend in Machine Learning / AI
- Traditional approaches: Hand tuned features, e.g.:
  - Feature extraction (Computer Vision);
    Hidden Markov Models & Statistics (Natural Language Processing)
- Deep Learning approach:
  - Learning of parameters for Artificial Neural Network (ANN) based on training data
- Recent success based on:
  - Huge amounts of (training) data
  - Lots of computing power
  - Better DL architectures

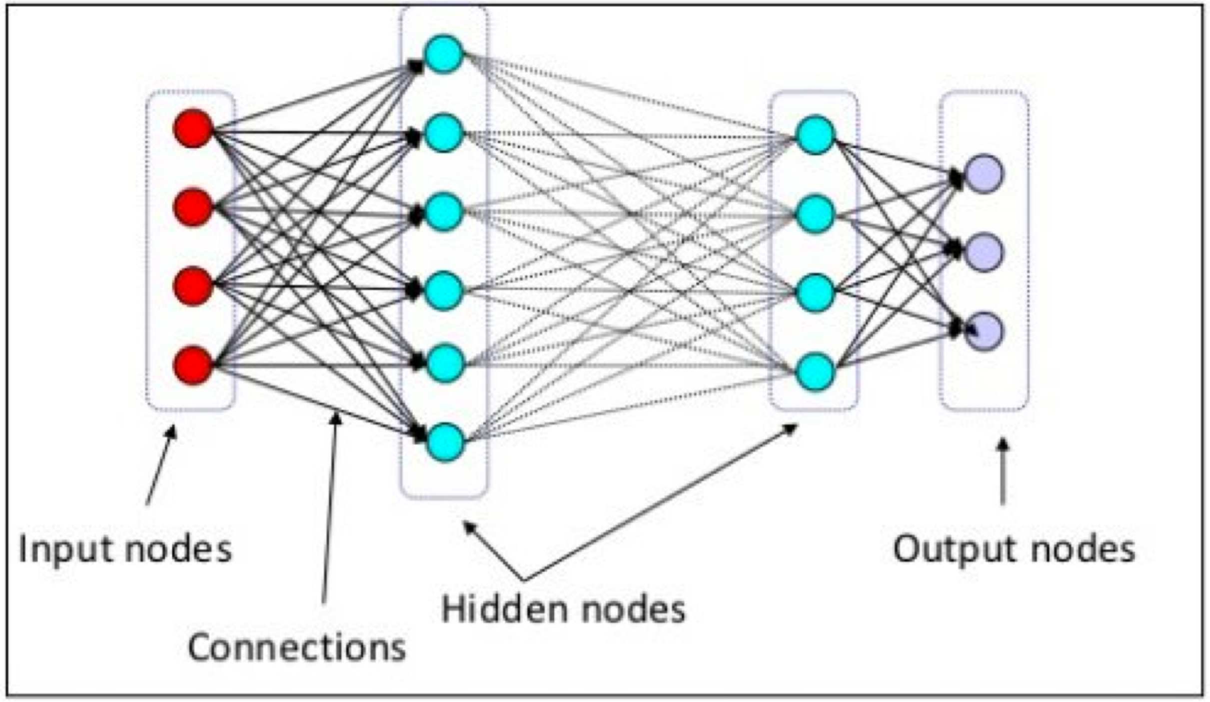


Image Source: Matthew Mayo

- **Traditional Pattern Recognition: Fixed/Handcrafted Feature Extractor**

Feature Extractor → Trainable Classifier →

- **Mainstream Pattern Recognition 9until recently)**

Feature Extractor → Mid-Level Features → Trainable Classifier →

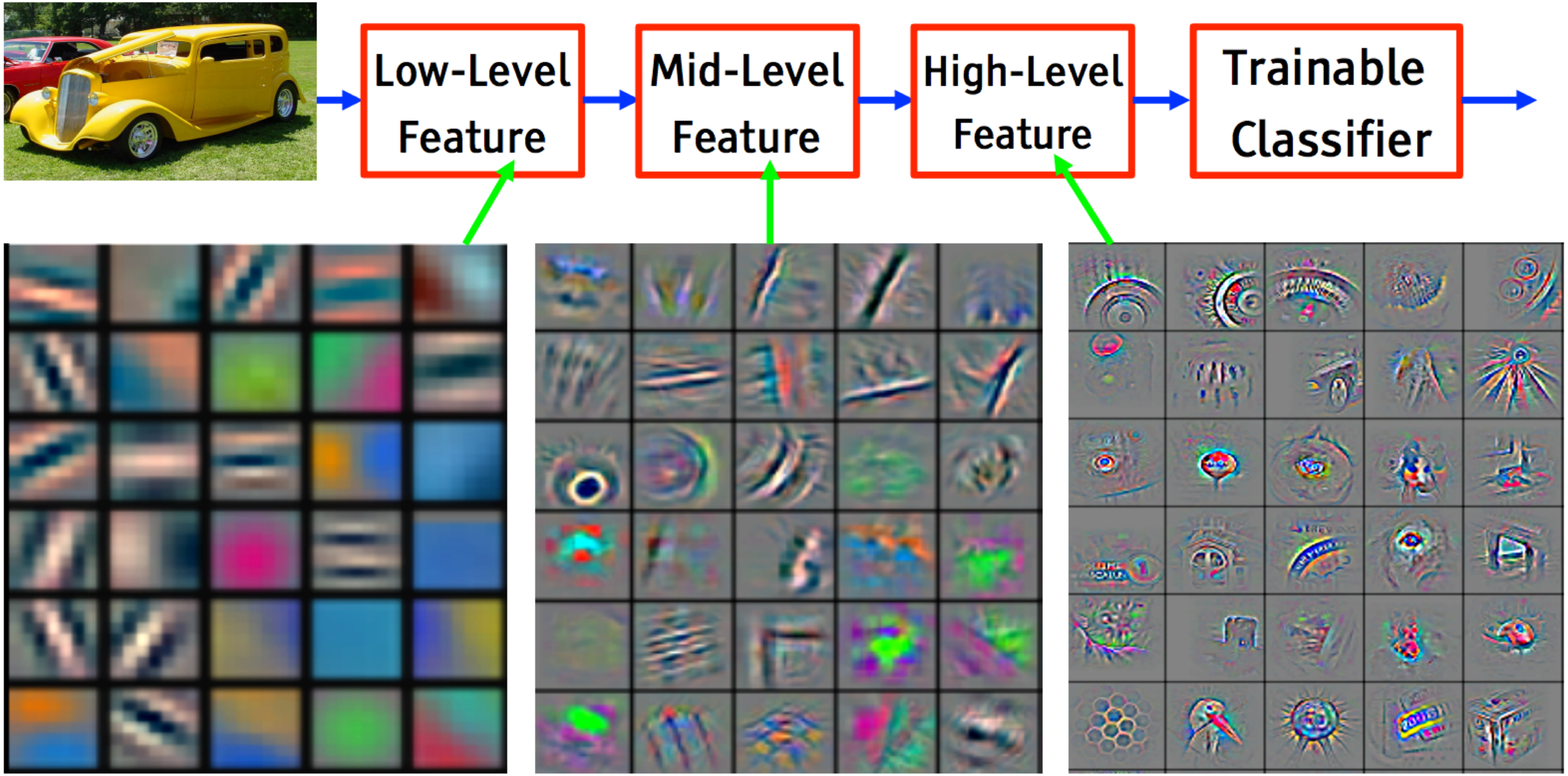
- **Deep Learning: Multiple stages/layers trained end to end**

Low-Level Features → Mid-Level Features → High-Level Features → Trainable Classifier →

NIPS’2015 DL Tutorial http://www.iro.umontreal.ca/~bengioy/talks/DL-Tutorial-NIPS2015.pdf

Source:
NIPS'2015 Tutorial
Geoff Hinton, Yoshua Bengio & Yann LeCun

Multiple Layers: Levels of Abstraction



Amazon Picking Challenge

- Robot stow and take (pick) products from shelves
- Challenges in: Computer vision (find object and its pose [position & orientation]) and Manipulation (planning, grasping)
- Since 2015 – now co-located with RoboCup
- Winner 2016: Team Delft
  - Suction gripper (easy!)
  - 100 items per hour (human: 400)
  - Failure rate: 16.7 %
  - Deep learning to find objects:
    - Stereo camera for 3D
  - ROS Industrial

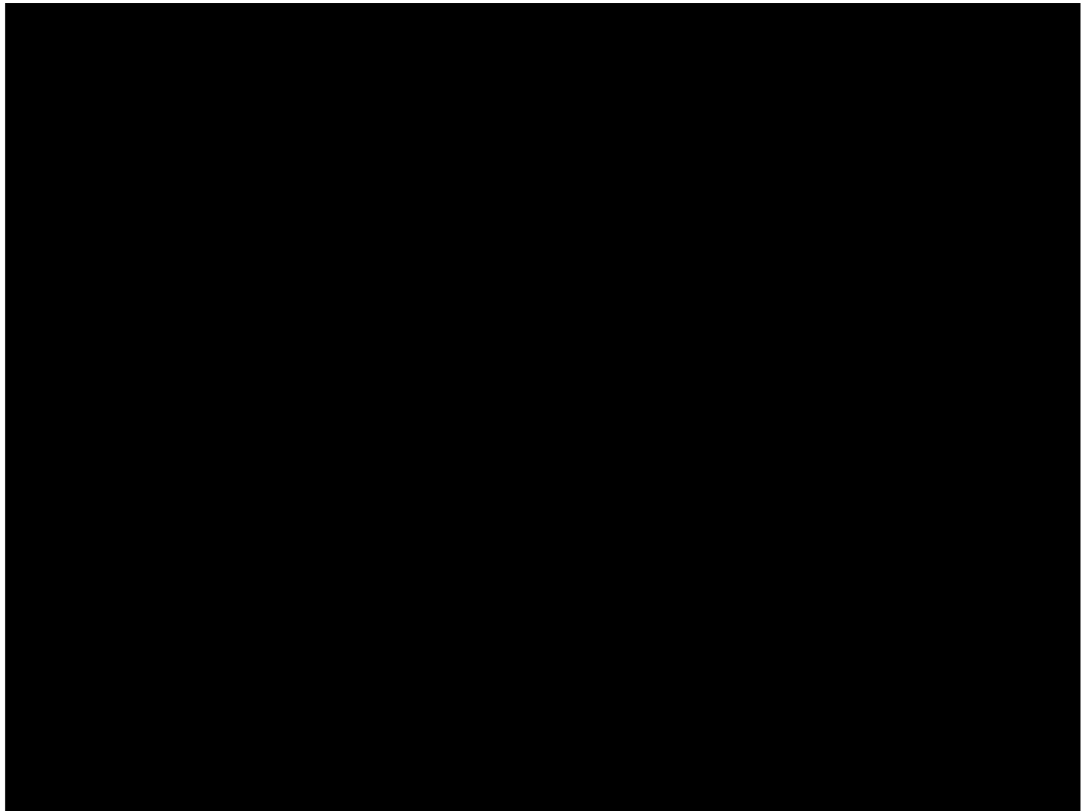
https://www.theverge.com/2016/7/5/12095788/amazon-picking-robot-challenge-2016

End-to-End Deep Reinforcement Learning

- From sensors to actuation: one layered or recurrent neural network! =>
  - NOT classical general control scheme (Perception, SLAM, Cognition & Planning, Navigation)

- Needs reward signal: sparse, noisy, delayed!
- Take time into account: input frames are related!

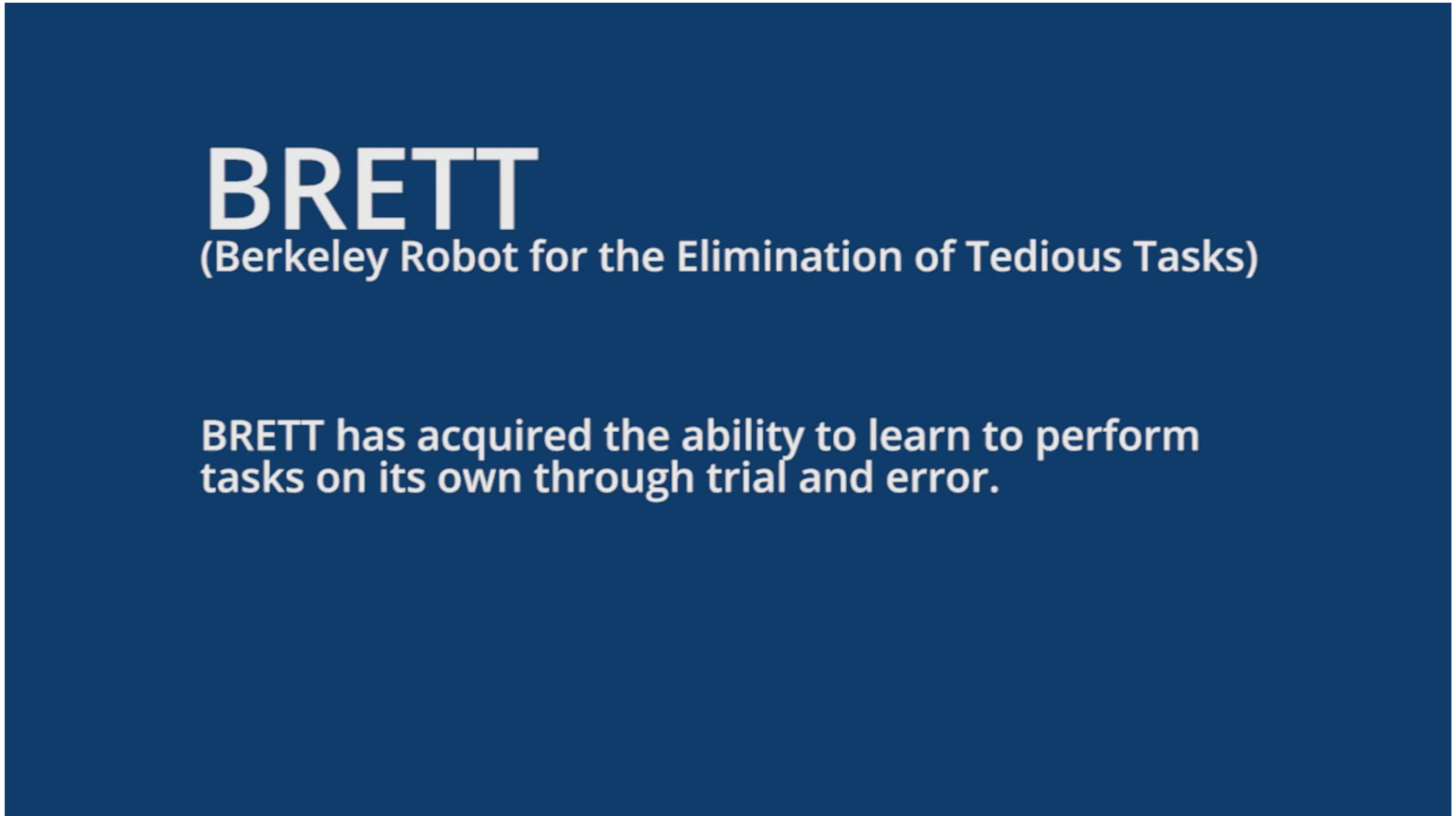
- Gained interest 2013 again with:
  - Deep Mind (google) playing ATARI 2600 games
  - Video: Breakout
  - Learned 7 games
  - Surpasses human expert in 3



https://www.cs.toronto.edu/~vmnih/docs/dqn.pdf

BRETT: Berkeley Robot for Learning Tedious Tasks: Deep Reinforcement Learning

- ”There are no labeled directions, no examples of how to solve the problem in advance. There are no examples of the correct solution like one would have in speech and vision recognition programs”
- Learn simple tasks in 10 minutes; learn vision and control together in 3 hours
- Pieter Abbeel of UC Berkeley;
- 2015



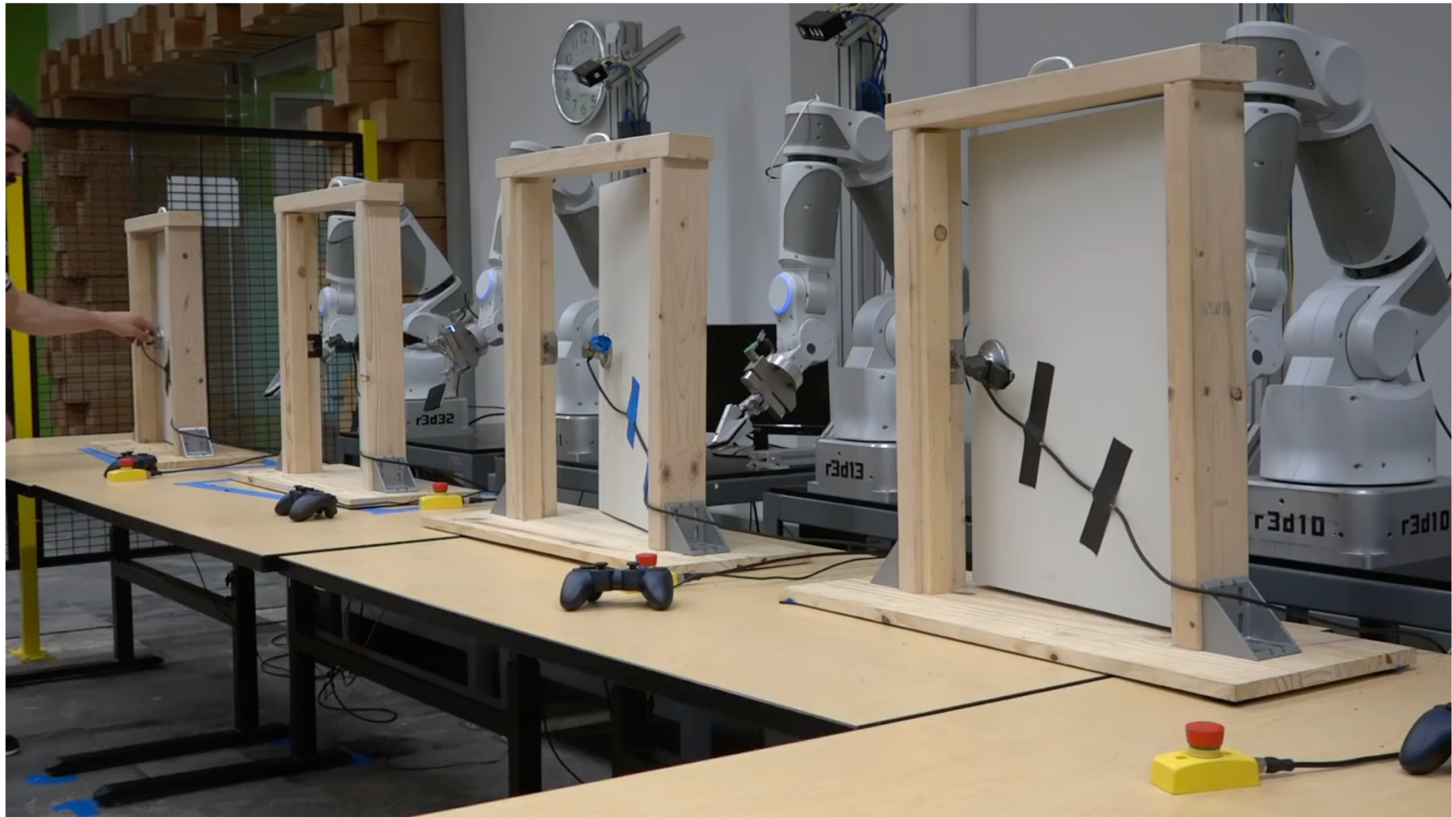
http://news.berkeley.edu/2015/05/21/deep-learning-robot-masters-skills-via-trial-and-error/

Google Door Opening Project

- Learn to open doors using Reinforcement learning
  - Learning reward: opening the door
  - Much harder than purely digital learning: very slow iterations!
  - Simulation only helps a bit:
    real world much more complex

- Google and
  UC Berkeley Sergey Levine

- Google very secretive …



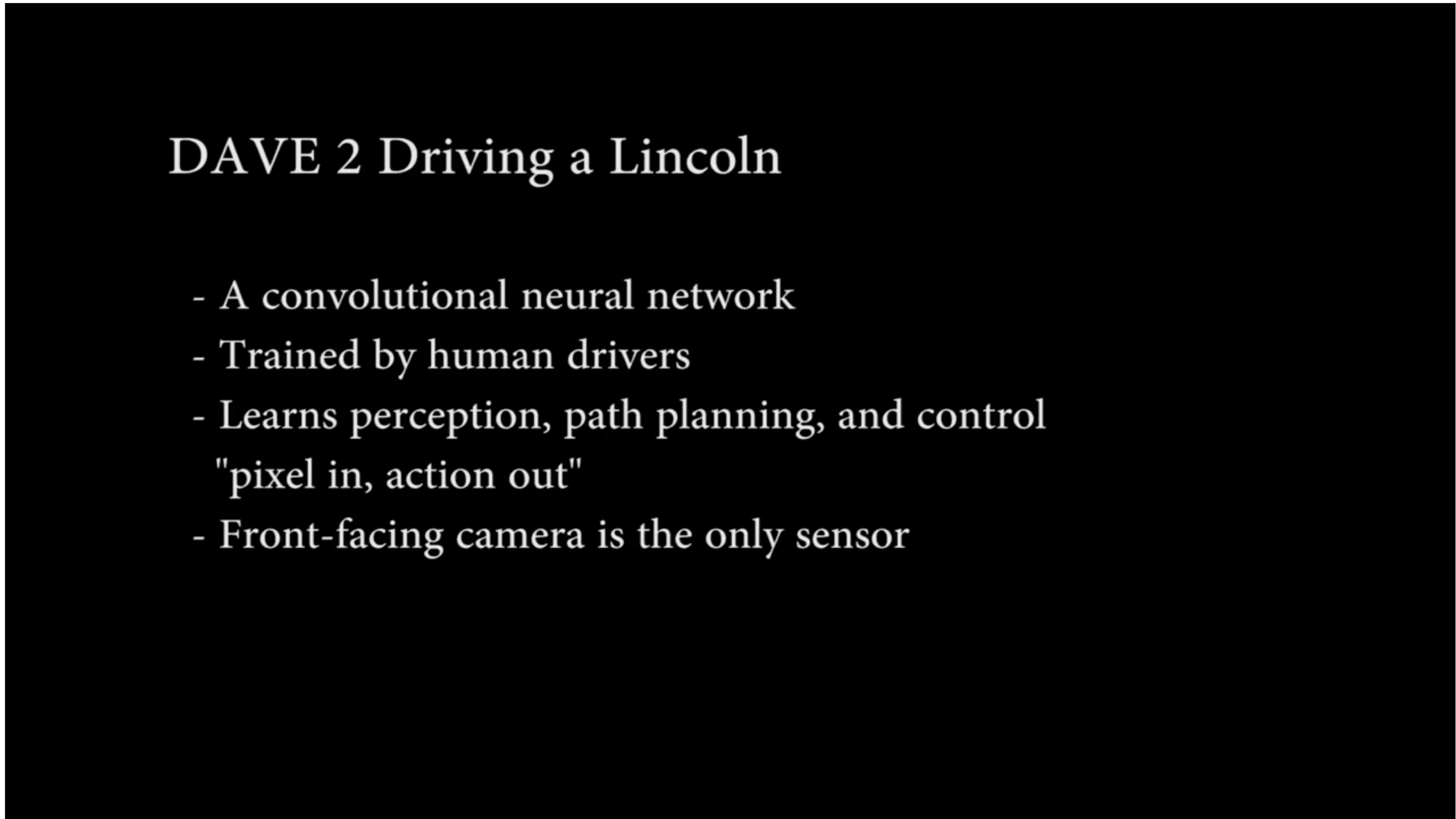
https://www.wired.com/2017/01/googles-go-playing-machine-opens-door-robots-learn/

Nvidia end-to-end deep learning self driving car

- Raw pixel of single camera => steering command of car
  - 30 FPS; single-image control (no history)
  - Nvidia Drive PX car computer (ARM cores and Nvidia GPU)
  - Human steering angle for training

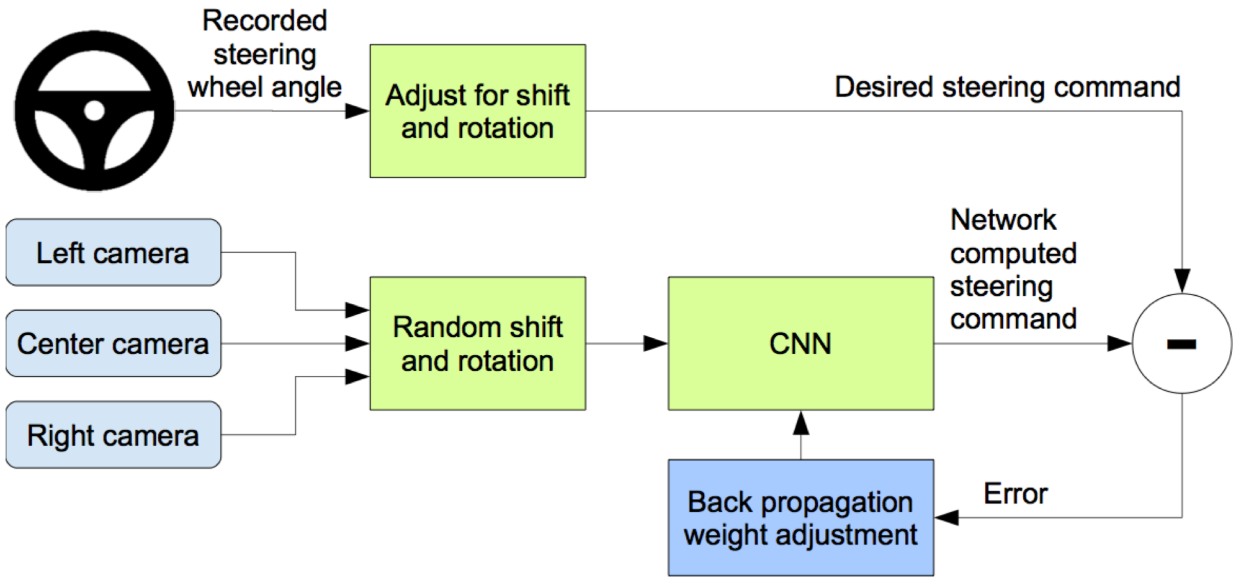
- End-to-end: NO explicit:
  - Lane detection
  - Road detection
  - Obstacle detection

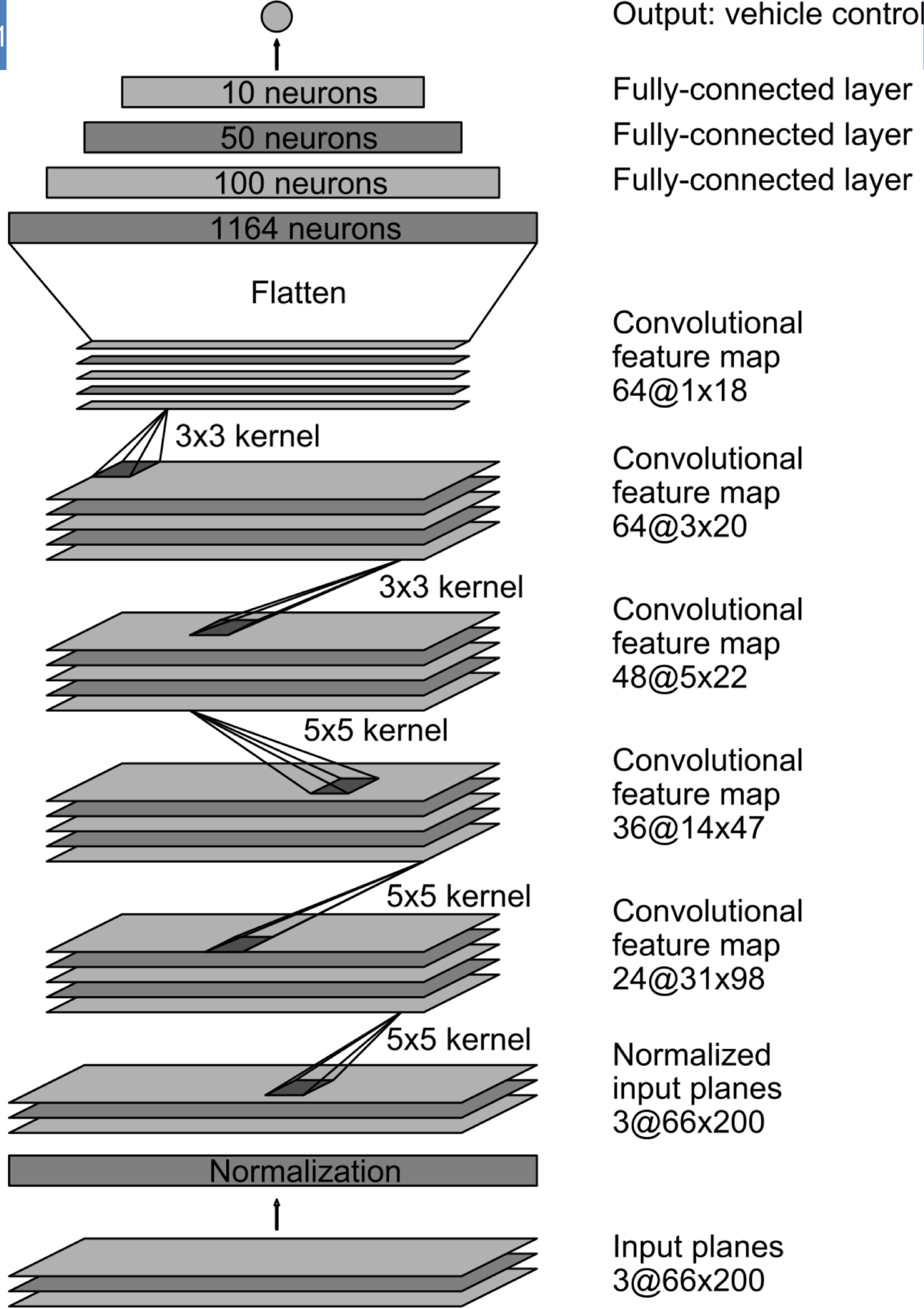
https://devblogs.nvidia.com/parallelforall/deep-learning-self-driving-cars/



- Training:
  - Additionally use left and right camera: negative examples!
  - Highway, residential roads, unpaved roads, car parks
  - Different weather conditions

- Result: Autonomous 98% of the time => 2% driver intervention





Problems with Deep Learning

- 99% success rate sounds good, but 1% failure is often unacceptable (e.g. autonomous car)
  - Failures are unavoidable =>
  - need quality estimate/ uncertainty of the result!
  - Often not available for DL ☹
- Lack of theory regarding deep learning
  - Acts like a black box…
- No introspection of how or why a DL system is behaving like it is =>
  - No safety guarantees possible
- Deep learning only part of an overall AI system
  - Hand-crafted methods can still be very powerful
  - Modelling useful (with input from DL)
  - Statistical methods
  - Reasoning
  - Planning

ADMIN

- Final Exam: Dec 5 during lecture hours
- HW 4: will be published soon…

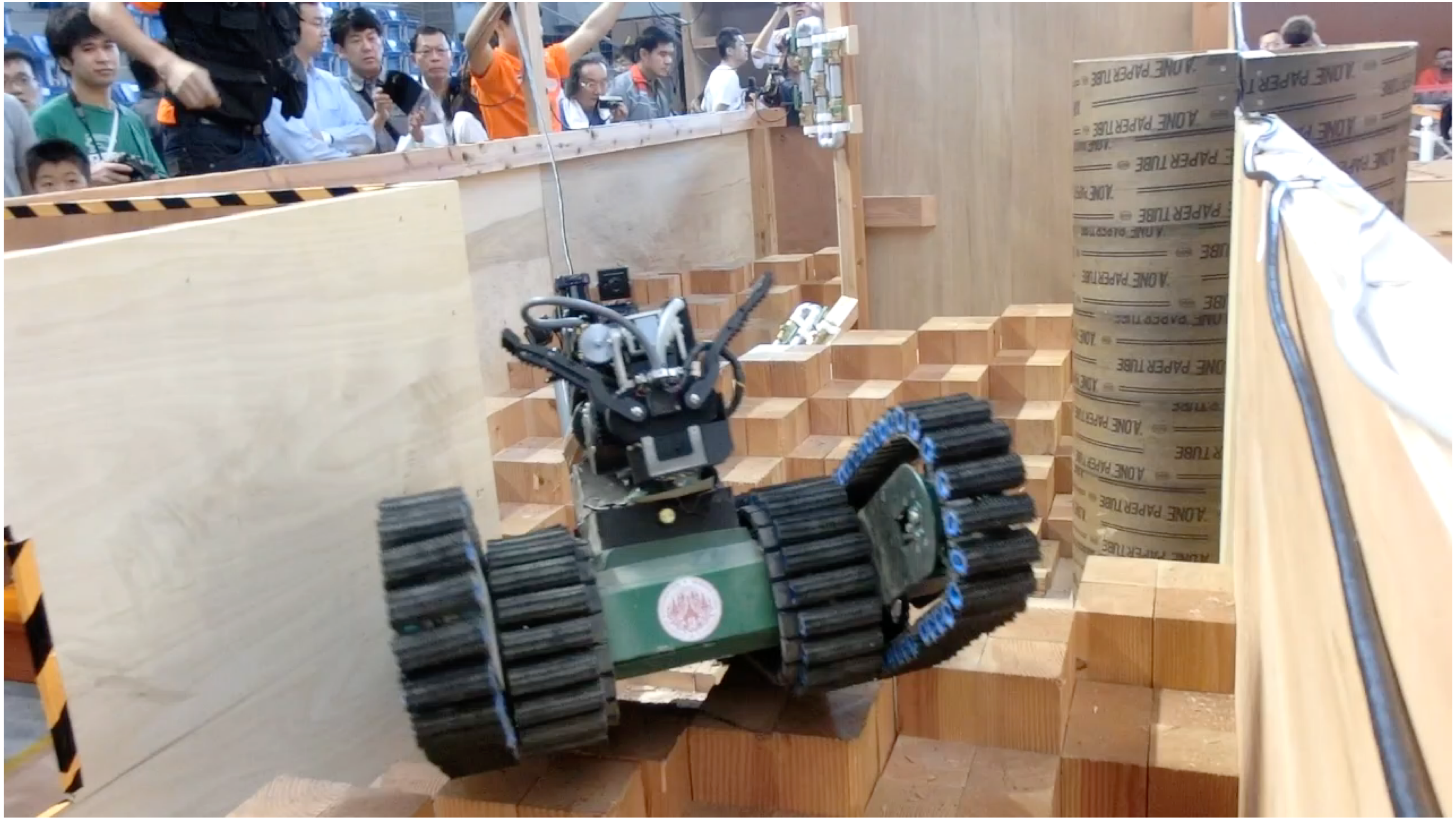
Paper Presentation

- Presentation Schedule:
  - During lecture hours
  - 4 Presentation slots:
    - Presentation 1: Nov 12
    - Presentation 2: Nov 19
    - Presentation 3: Nov 21
    - Presentation 4: Nov 28
- After Oct 21st: loose 33%; After Nov 2nd: loose 50%
- 10 minute presentation plus 1 minute project relevance plus 3 minutes questions =>
  - 15 minute per student x 8 students => 2 hours
- Every student has to attend the 2 presentations slots defined in the schedule!
  - Attendance will be checked!
- You will be using my Laptop
- A laser pointer will be provided

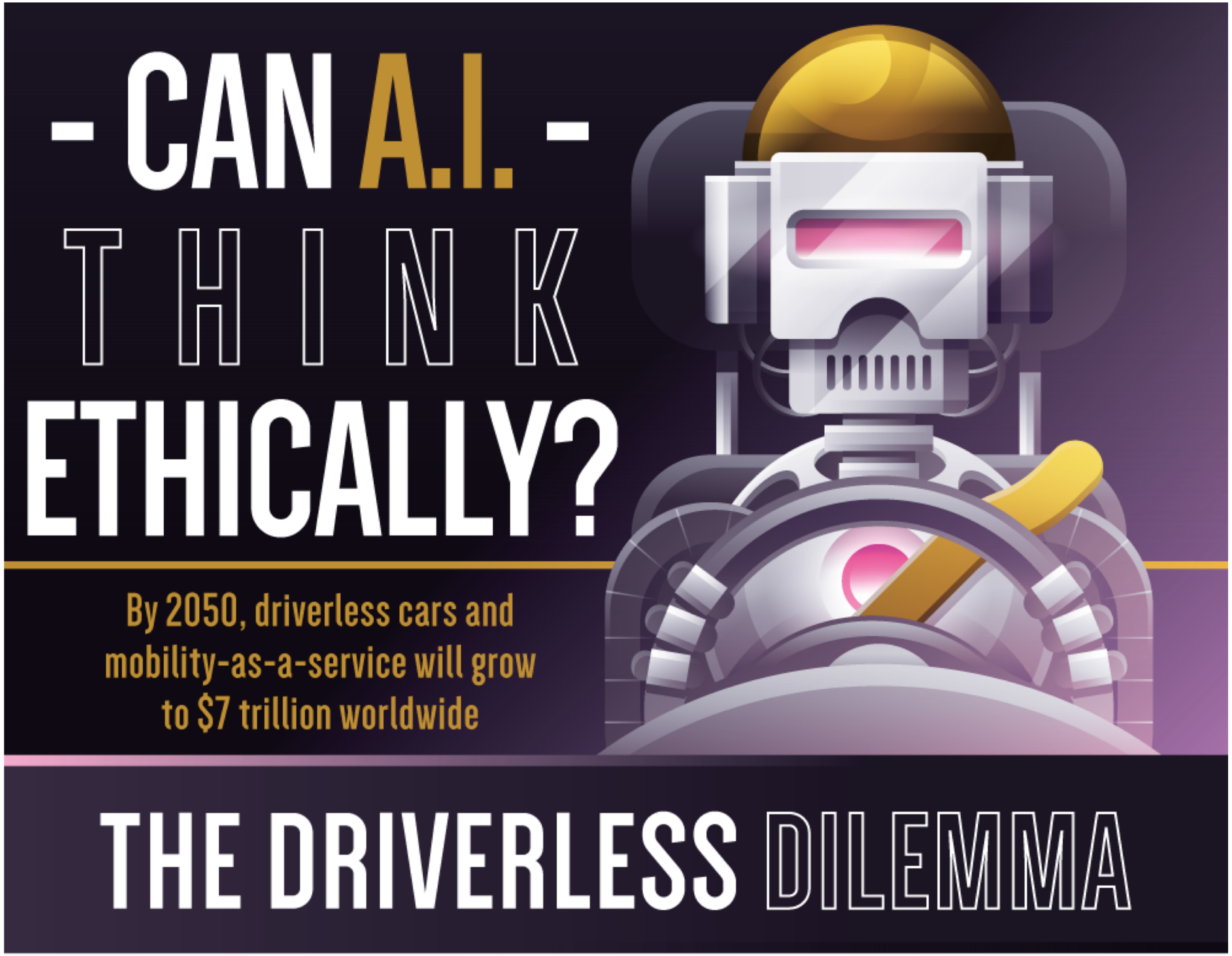
Robotics Project Teams

Project	姓名	Name	Presentation	Attendence	Student Advisor
Life science Fetch robot	张亦正	Yizheng Zhang	1	1 & 2	Xiaoling
	朱佳会	Jiahui Zhu	1	1 & 2	
Elevator	余泽浩	Zehao Yu	1	1 & 2	Xiaoling
	钱深瀚	Shenhan Qian	1	1 & 2	
	陈夏宁	Xianing Chen	1	1 & 2	
Rover Manipulation	李阳	Yang Li	1	1 & 2	Xiaoling
	陈睿卿	Ruiqing Chen	1	1 & 2	
	陈亮	Liang Chen	1	1 & 2	
Rover SLAM	王志伟	Zhiwei Wang	2	2 &3	Hongyu
	严亦晖	Yihui Yan	2	2 &3	
Car project	唐伟鸿	Weihong Tang	2	2 &3	Hongyu
	唐韧之	Renzhi Tang	2	2 &3	
Car project	侯镜阳	Jingyang Hou	2	2 &3	Hongyu
	王翌舟	Yizhou Wang	2	2 &3	
Furniture Free Mapping	何振鹏	Zhenpeng He	2	2 &3	Sören
	孙豪	Hao Sun	2	2 &3	
Mapping Robot	赵希亭	Xiting Zhao	3	3 & 4	Hongyu
	杨之杰	Zhijie Yang	3	3 & 4	
	万浩川	Haochuan Wan	3	3 & 4	
Wifi Localization	祝宁之	Ningzhi Zhu	3	3 & 4	Xiaoling
	周宇	Yu Zhou	3	3 & 4	
Event Camera	李润东	Rundong Li	3	3 & 4	Haofei
	黄坤	Kun Huang	3	3 & 4	
	高翎	Ling Gao	3	3 & 4	
Underwater stereo SLAM	张苁卉	Conghui Zhang	4	1 & 4	Haofei
	吴德明	Deming Wu	4	1 & 4	
	牛天星	Tianxing Niu	4	1 & 4	
Omni Camera	崔佳迪	Jiadi Cui	4	1 & 4	Qingwen
	金磊	Lei Jin	4	1 & 4	
Factorization	林洲洋	Zhouyang Lin	4	1 & 4	Qingwen
	高国亮	Guoliang Gao	4	1 & 4	

Projects: Safety is important!



AI ETHICS



https://cybersecuritydegrees.com/ethical-ai/

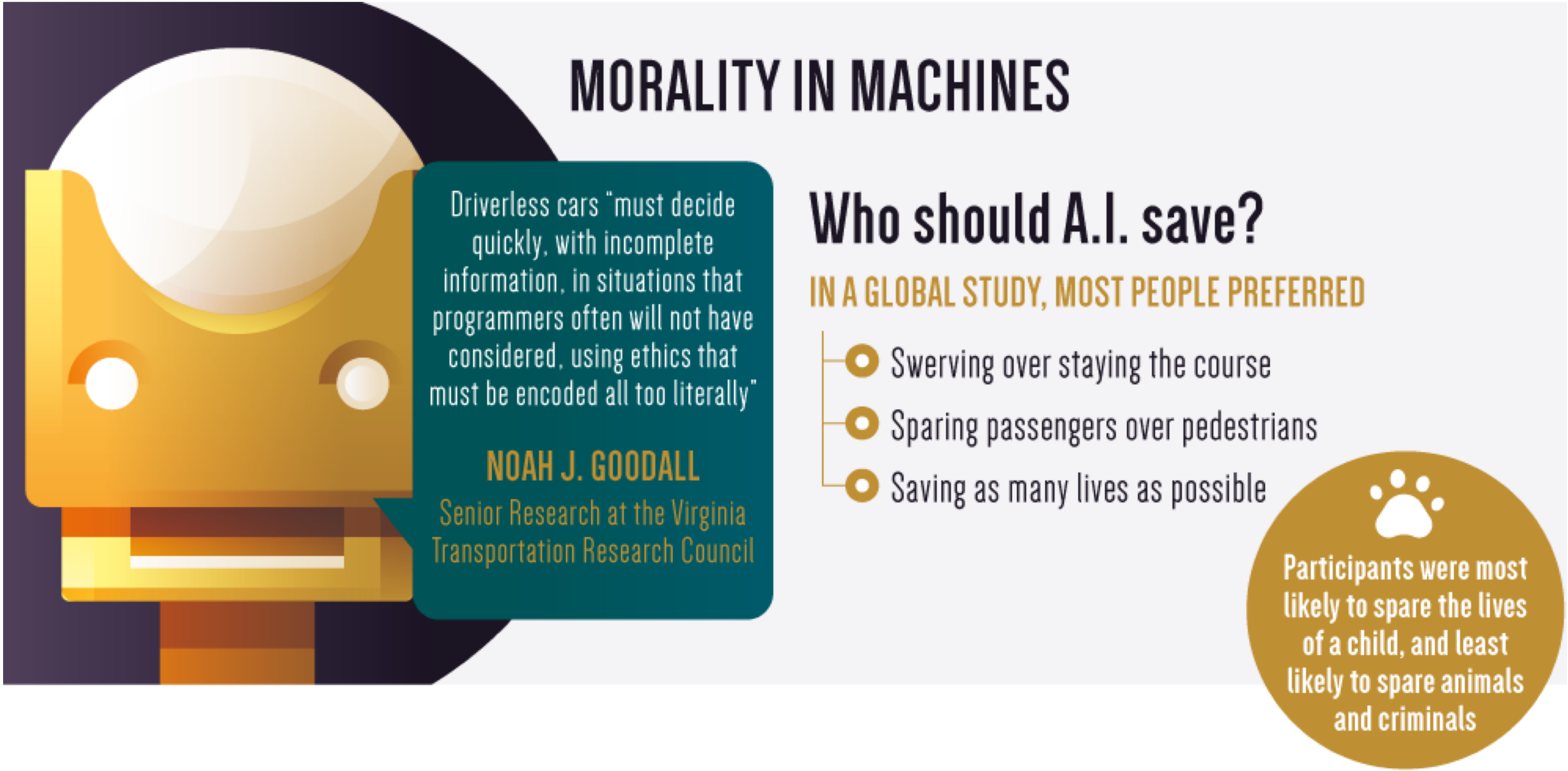
FROM 2035 TO 2045

Consumers will regain up to **250 MILLION HOURS OF FREE TIME** from behind the wheel

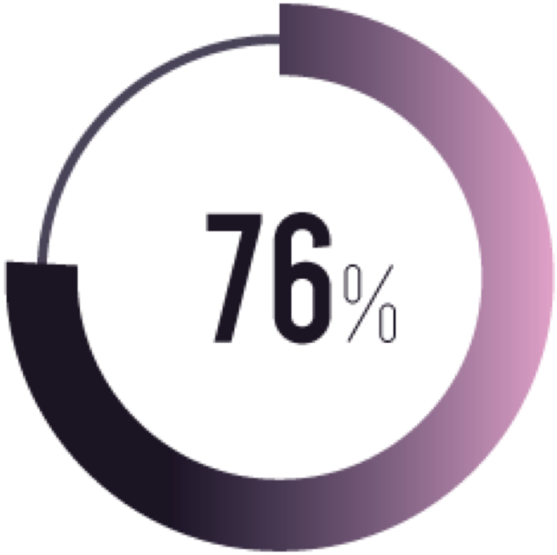
$234 BILLION IN PUBLIC COSTS will be saved by reducing accidents from human error

Driverless cars can **ELIMINATE 90% OF TRAFFIC FATALITIES** — Saving 1 million lives every year

HOW WILL DRIVERLESS CARS DETERMINE WHOSE LIFE SHOULD BE SPARED?


MORALITY IN MACHINES
Driverless cars "must decide quickly, with incomplete information, in situations that programmers often will not have considered, using ethics that must be encoded all too literally"
NOAH J. GOODALL
Senior Research at the Virginia Transportation Research Council
Who should A.I. save?
IN A GLOBAL STUDY, MOST PEOPLE PREFERRED
Swerving over staying the course
Sparing passengers over pedestrians
Saving as many lives as possible
Participants were most likely to spare the lives of a child, and least likely to spare animals and criminals

MINIMIZED HARM VS. PASSENGER PROTECTION



of people felt driverless cars should SAVE AS MANY LIVES AS POSSIBLE

<

BUT, very few were willing to buy a vehicle programmed to minimize harm

They prefer cars programmed to PROTECT PASSENGERS AT ALL COSTS

DRIVERLESS CARS WILL SAVE LIVES, BUT PROGRAMMING THEM TO DO SO COULD SLOW THEIR ADOPTION AND COST MANY MORE LIVES

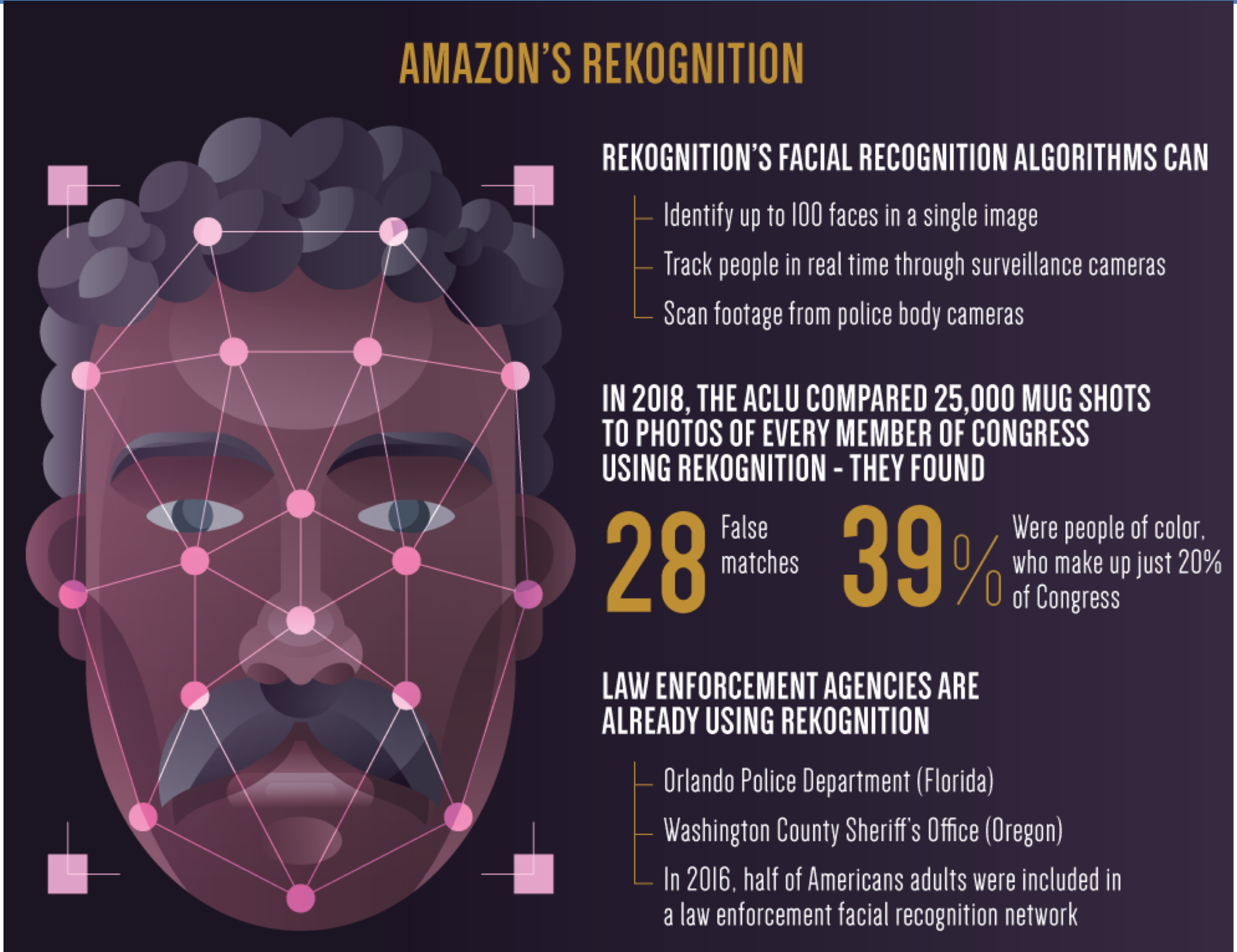
REAL LIFE APPLICATIONS GROW EVEN MORE COMPLEX

In an accident causing injuries but not fatalities, should A.I.

- **DISTRIBUTE INJURIES EVENLY**, harming more people less severely?
- **CONSIDER THE LIKELIHOOD** and severity of potential injuries?
- Take into account the **QUALITY OF LIFE EFFECT** of resulting injuries?
- Is it better to **HOSPITALIZE 5 PEOPLE, OR KILL 1?**

AS A.I. ADVANCES, IT BECOMES RESPONSIBLE FOR MORE MORAL AND ETHICAL DECISION MAKING

When AI goes wrong

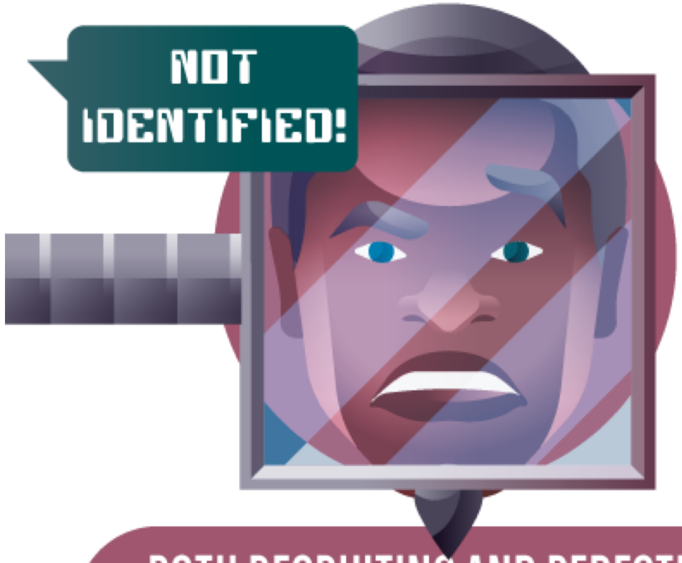


HOW A.I. BIAS HAPPENS

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
- Defining ethical guidelines for the design, use and operation of robots
- **Enhancement technologies: ethical issues**
- Privacy and the 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
- **Transparency** in autonomous systems
- Embedding values and norms into intelligent systems
- Ethics and standardization
- Raising ethical awareness among stakeholders
- Political and legal frameworks
- Formal and mathematical frameworks for robot ethics
- Implementations and engineering studies

https://www.icres2019.org

Ethical AI: Scientific Discussion

- 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

- ICRES 2019 is the fourth edition of the International Conference on Robot Ethics and Standards series
  - https://www.icres2019.org/

- 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

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'	

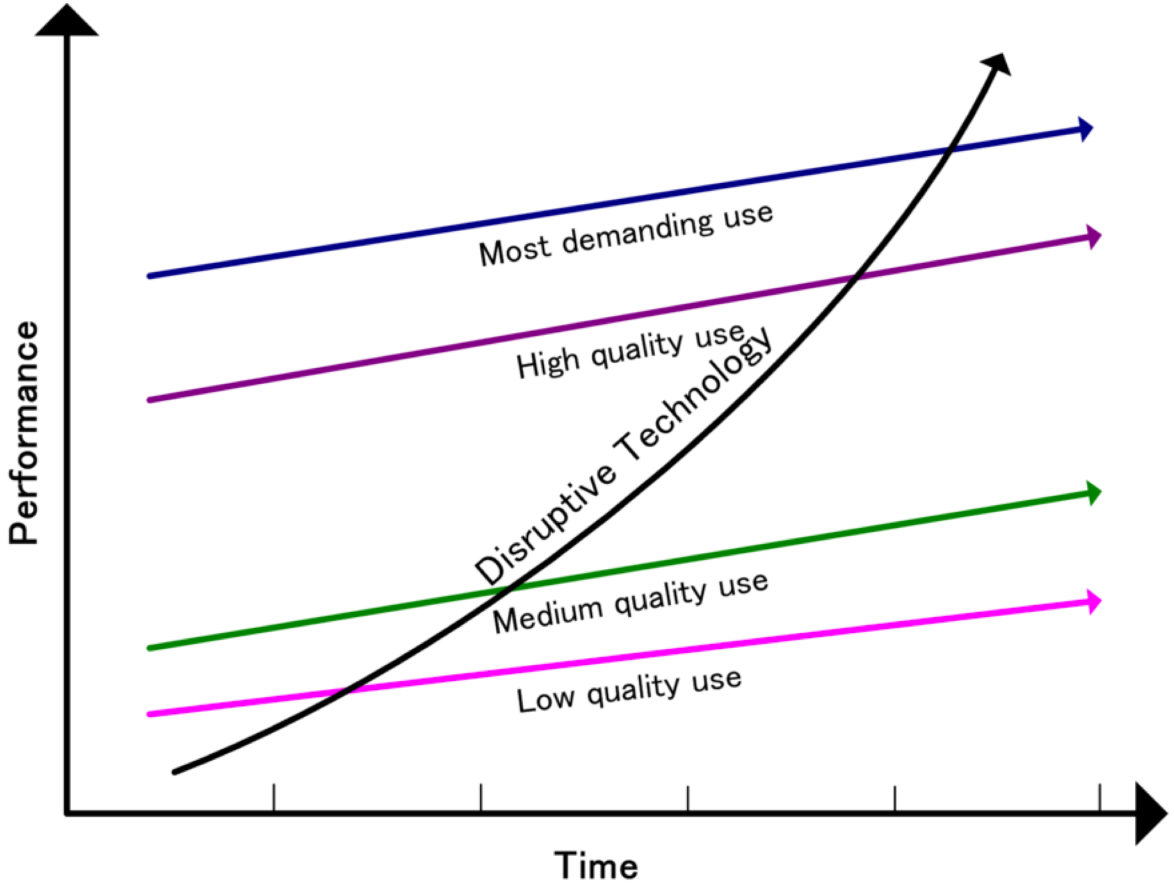
HiveMapper				US	Pattern recognition, Mapping	HiveMapper app	X
IBM				US	AI chips, Cloud, Super computers, Facial recognition	Nuclear testing super computers, ex-JEDI	
Innoviz				Israel	Autonomous vehicles	-	
Intel				US	AI chips, UAS	DARPA HIVE	
Megvii				China	Facial recognition	-	
Microsoft				US	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	

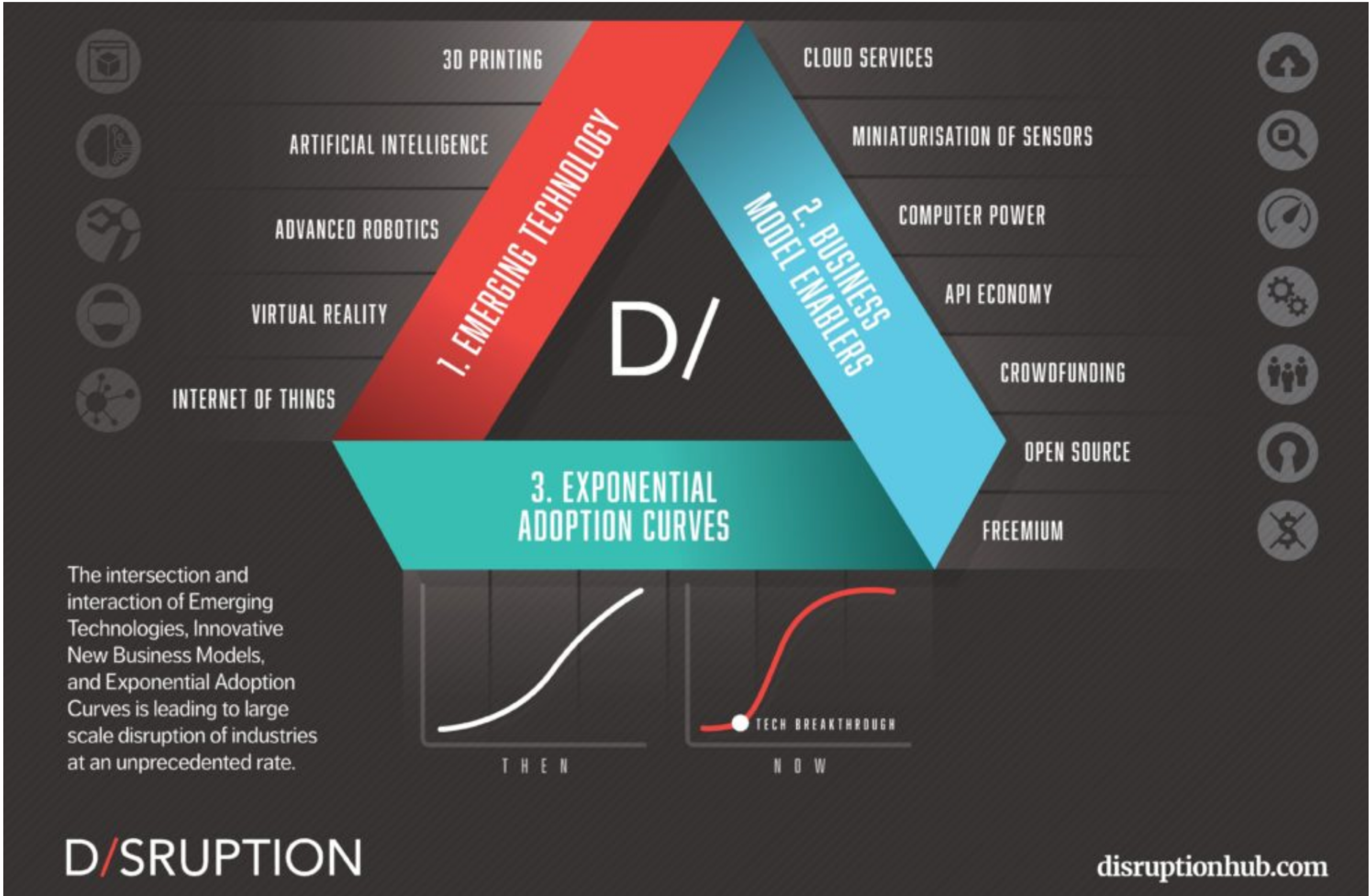
DISRUPTIVE AI & ROBOTICS

Disruptive Innovation

- **Disruptive Innovation** is an innovation that creates a **new market** and eventually **disrupts** an existing market, **displacing** established market-leading firms, products, and alliances.

- Examples:
  - Automobiles (Disrupts Railway, Horse carriage)
  - Smartphones -> Compact Digital Cameras -> Analog (Chemical) Cameras
  - Smartphones -> Laptops, Computers, PDAs
  - Streaming Videos -> Video Rentals
  - Internet -> Mail, Telephone, TV, Books, basically everything
  - …

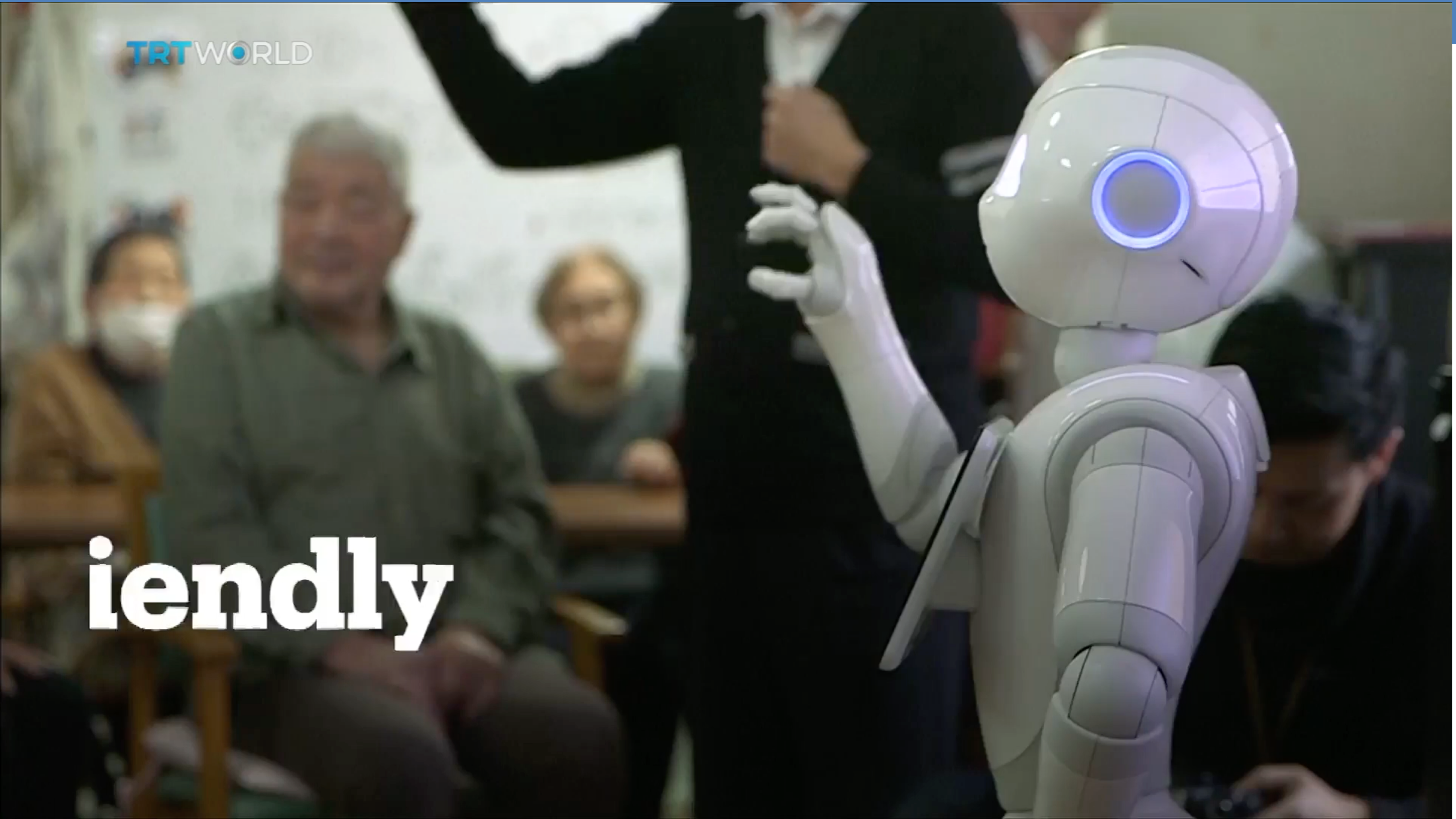



3D PRINTING
ARTIFICIAL INTELLIGENCE
ADVANCED ROBOTICS
VIRTUAL REALITY
INTERNET OF THINGS
1. EMERGING TECHNOLOGY
2. BUSINESS MODEL ENABLERS
3. EXPONENTIAL ADOPTION CURVES
D/
CLOUD SERVICES
MINIATURISATION OF SENSORS
COMPUTER POWER
API ECONOMY
CROWDFUNDING
OPEN SOURCE
FREEMIUM
The intersection and interaction of Emerging Technologies, Innovative New Business Models, and Exponential Adoption Curves is leading to large scale disruption of industries at an unprecedented rate.
TECH BREAKTHROUGH
THEN
NOW
D/SRUPTION
disruptionhub.com

AI & Robotics may change Everything

- AI & Robots 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?)
  - 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, AI programmer)
    - Option 2: Most jobs will no be replaced => big unemployment
- Potentially: new (updated) social systems might emerge
  - Unconditional basic income?
  - Communism 2.0?
  - …

DISCUSSION

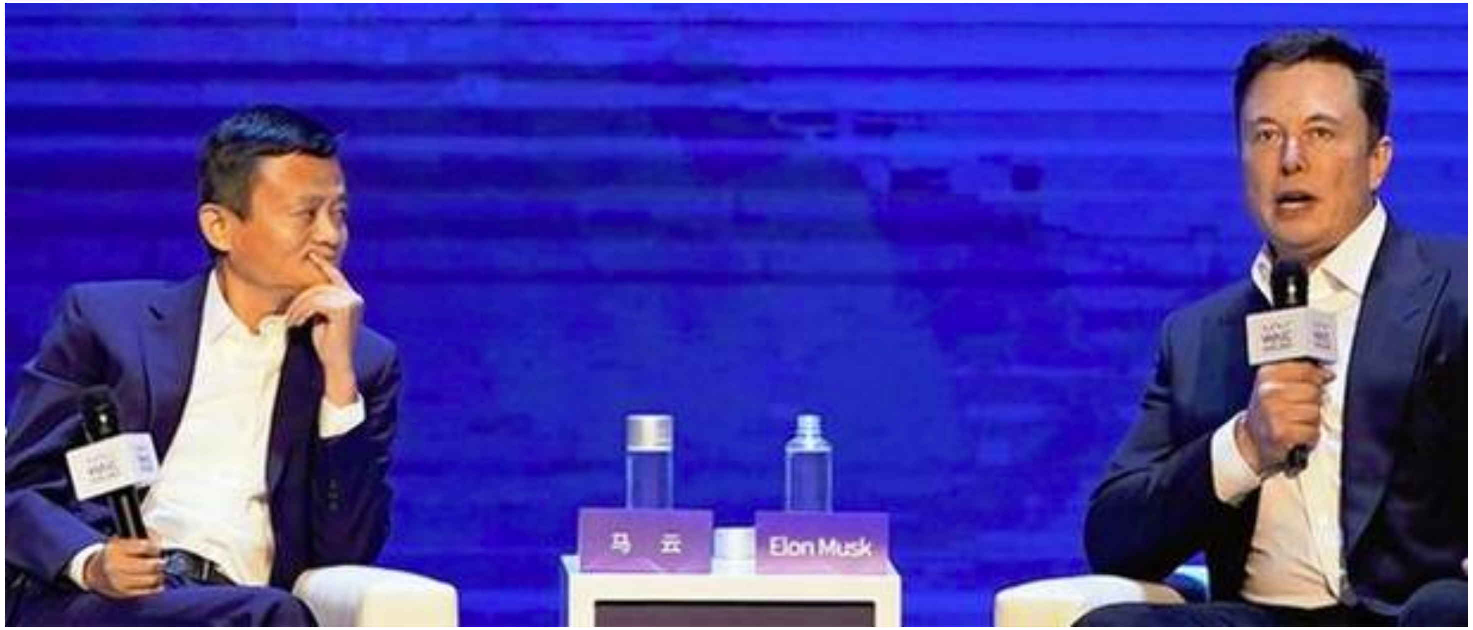

TRT WORLD
iendly

Robots for Elderly care…

- Over 80% of Japanese positive about robotic nursing care
  - Respondents who said they are ready to or want to receive nursing care from robots: 84.3%. Of the respondents who prefer not to use robotic nursing care, 46.9 percent — the largest group — said the reason for their choice was that they would prefer to be taken care of by humans.
    - https://www.japantimes.co.jp/news/2018/11/15/national/80-japanese-positive-robotic-nursing-care/#.XV-r_ZMza9Y

- Paper: Granny and the robots: Ethical issues in robot care for the elderly. Issues:
  1. the potential reduction in the amount of human contact;
  2. an increase in the feelings of objectification and loss of control;
  3. a loss of privacy
  4. a loss of personal liberty;
  5. deception and infantilisation;
  6. the circumstances in which elderly people should be allowed to control robots
     - http://staffwww.dcs.shef.ac.uk/people/a.sharkey/sharkey-granny.pdf

This August: Ma vs Musk: Tech tycoons spar on future of artificial intelligence



Jack Ma believes artificial intelligence poses no threat to humanity, but Elon Musk called that “famous last words” as the billionaire tech tycoons faced off on Aug 29 in an occasionally animated debate on futurism in Shanghai.

https://www.thehindubusinessline.com/info-tech/ma-vs-musk-tech-tycoons-spar-on-future-of-artificial-intelligence/article29291335.ece