

AI and the Future

Tom Everitt

2 March 2016

1997



2016



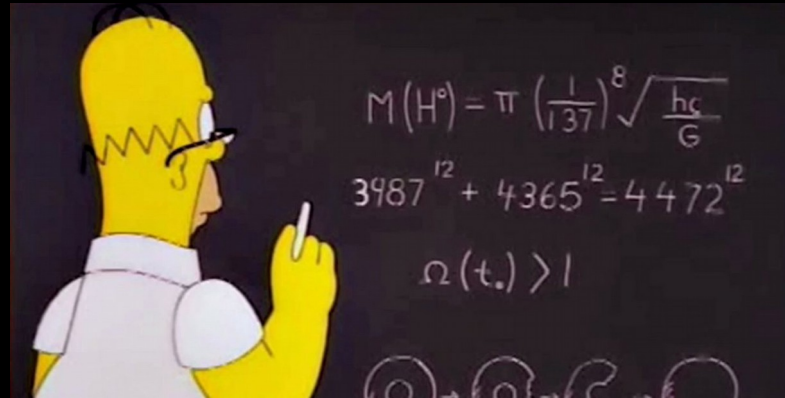
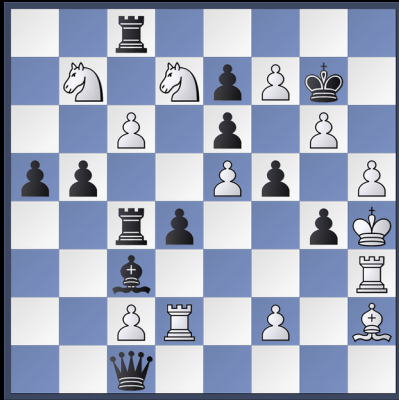
AlphaGo

Lee Sedol

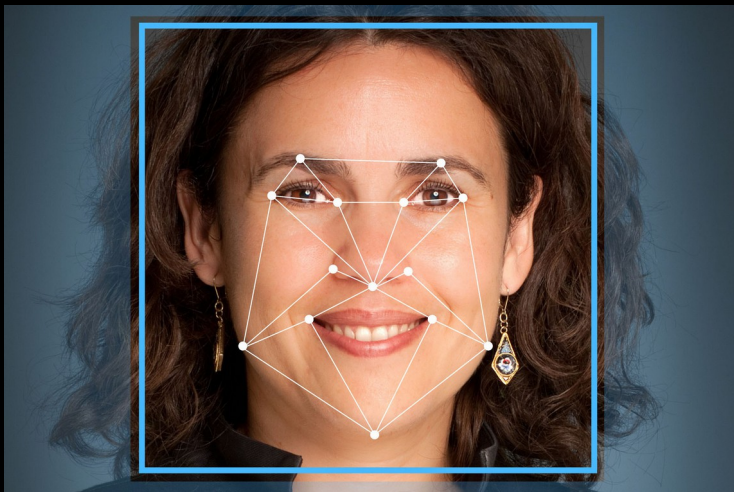


Types of Intelligence

“High-level” logical thinking



“Low-level” sensory-motor control

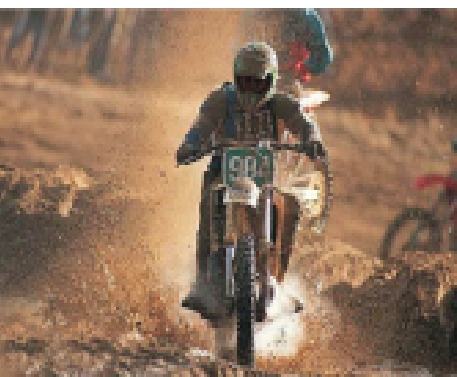






<https://youtu.be/rVlhMGQgDkY>

Describes without errors



A person riding a motorcycle on a dirt road.

Describes with minor errors



Two dogs play in the grass.

Somewhat related to the image



A skateboarder does a trick on a ramp.

Unrelated to the image



A dog is jumping to catch a frisbee.



A group of young people playing a game of frisbee.



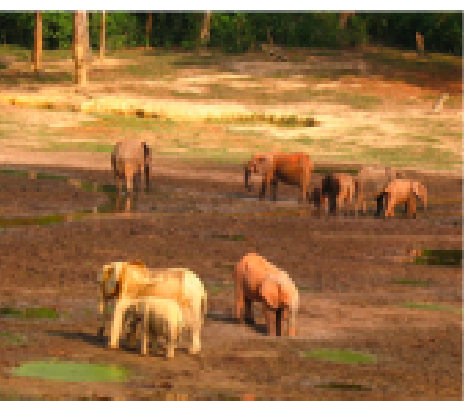
Two hockey players are fighting over the puck.



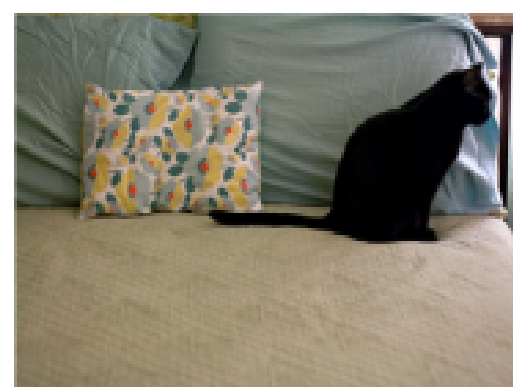
A little girl in a pink hat is blowing bubbles.



A refrigerator filled with lots of food and drinks.



A herd of elephants walking across a dry grass field.



A close up of a cat laying on a couch.



A red motorcycle parked on the side of the road.



A yellow school bus parked in a parking lot.

Intelligence



- Food chain position (mid or top)
- Current problems: Climate change, poverty, cancer ...
- AI = automating intelligence

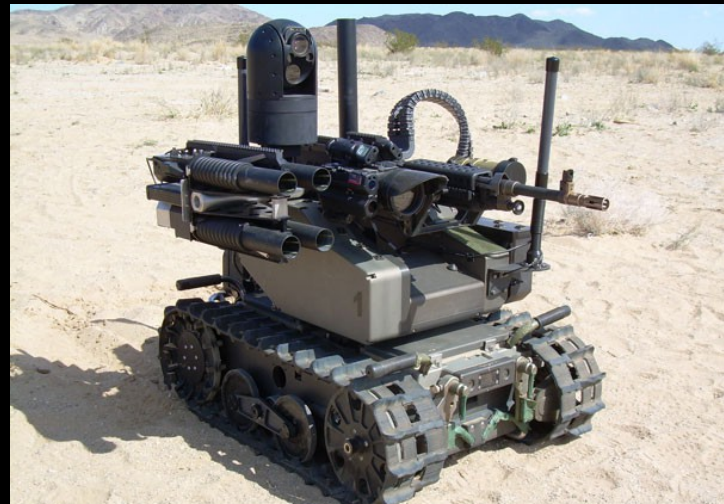
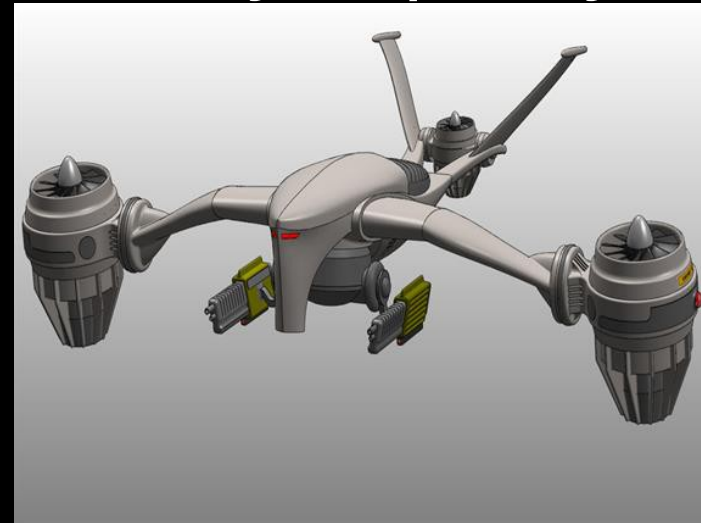
Violence and Military

- In today's society:
Money => weapons, soldiers => military capacity

A very, very small quadcopter, one inch in diameter can carry a one- or two-gram shaped charge. You can order them from a drone manufacturer in China. You can program the code to say: "Here are thousands of photographs of the kinds of things I want to target." A one-gram shaped charge can punch a hole in nine millimeters of steel, so presumably you can also punch a hole in someone's head. You can fit about three million of those in a semi-tractor-trailer. You can drive up I-95 with three trucks and have 10 million weapons attacking New York City. They don't have to be very effective, only 5 or 10% of them have to find the target.

There will be manufacturers producing millions of these weapons that people will be able to buy just like you can buy guns now, except millions of guns don't matter unless you have a million soldiers. You need only three guys to write the program and launch them. So you can just imagine that in many parts of the world humans will be hunted. They will be cowering underground in shelters and devising techniques so that they don't get detected. This is the ever-present cloud of lethal autonomous weapons.

They could be here in two to three years.



What to do?

- International conventions on autonomous weapons
 - Compare blinding laser guns, chemical weapons, biological warfare, nuclear weapons
- Petition 16000+ AI researchers: Ban Offensive Autonomous Weapons
- UN Conference Geneva April 2016
 - Militaries have mixed opinions

Production and Employment

- Today's society:
Money => factories, workers => goods
- With AI:
 - Humans may have no competitive advantage
 - A single human (or AI) can be self-sufficient

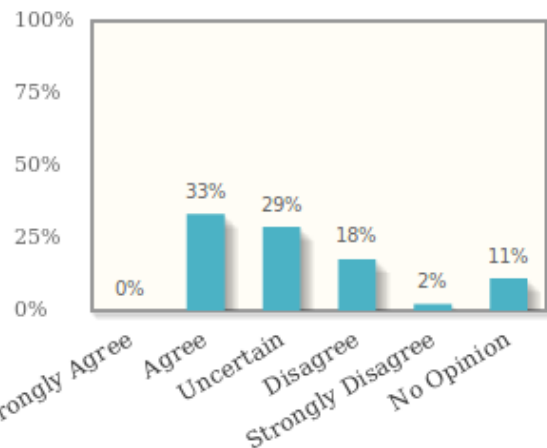




<http://images.clipartpanda.com/unemployment-clipart-unemployment-grads.jpg>

Question B: Information technology and automation are a central reason why median wages have been stagnant in the US over the past decade, despite rising productivity.

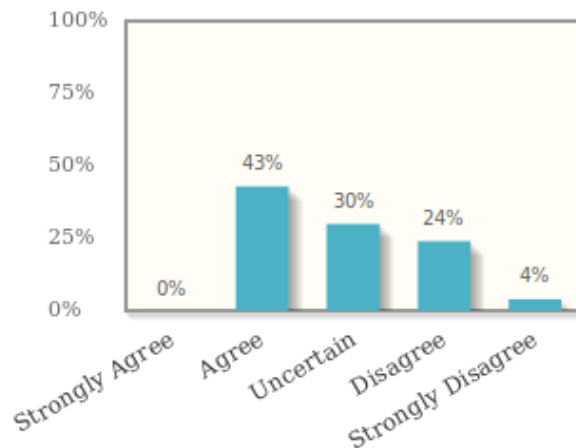
Responses



© 2017. Initiative on Global Markets.

Source: IGM Economic Experts Panel
www.igmchicago.org/igm-economic-experts-panel

Responses weighted by each expert's confidence



© 2017. Initiative on Global Markets.

Source: IGM Economic Experts Panel
www.igmchicago.org/igm-economic-experts-panel

Total employment 2014	150,539,900
Office and administrative support	22,766,100
Retail sales	8,739,300
Health diagnosing and treating	5,132,400
Construction	4,995,700
Motor vehicle operators	4,108,000
Building cleaning	3,835,100
Material movers, hand	3,587,800
Cooks and food preparation	3,164,700

U.S. Bureau of Labor Statistics
http://www.bls.gov/emp/ep_table_102.htm

What to do?

- Universal Basic Income
 - Trials in Netherlands, Finland, Brazil, ...
 - Requires effective taxation (of “robots”)
- Capital gains
 - Requires:
 - high growth rate
 - initially wealthy humans
 - stable property rights

Engineering Intelligence

Engineering Intelligence

*“Intelligence is the ability to achieve goals
in a wide range of environments”*

INNOVATIONS IN
The microsome

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

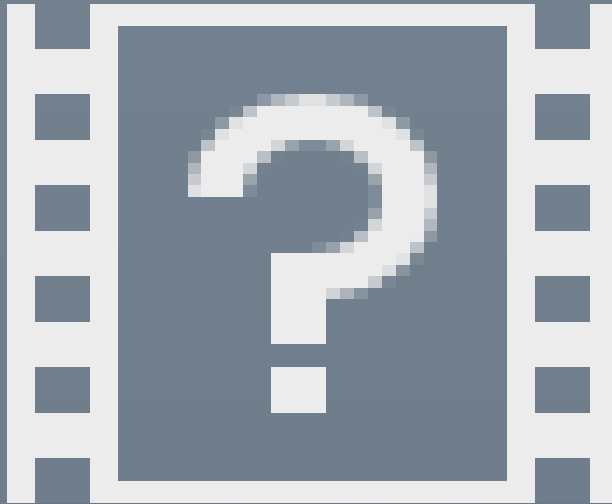
LEARNING CURVE

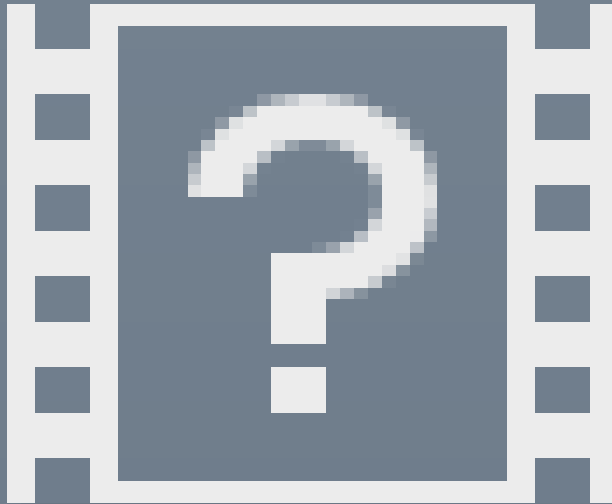


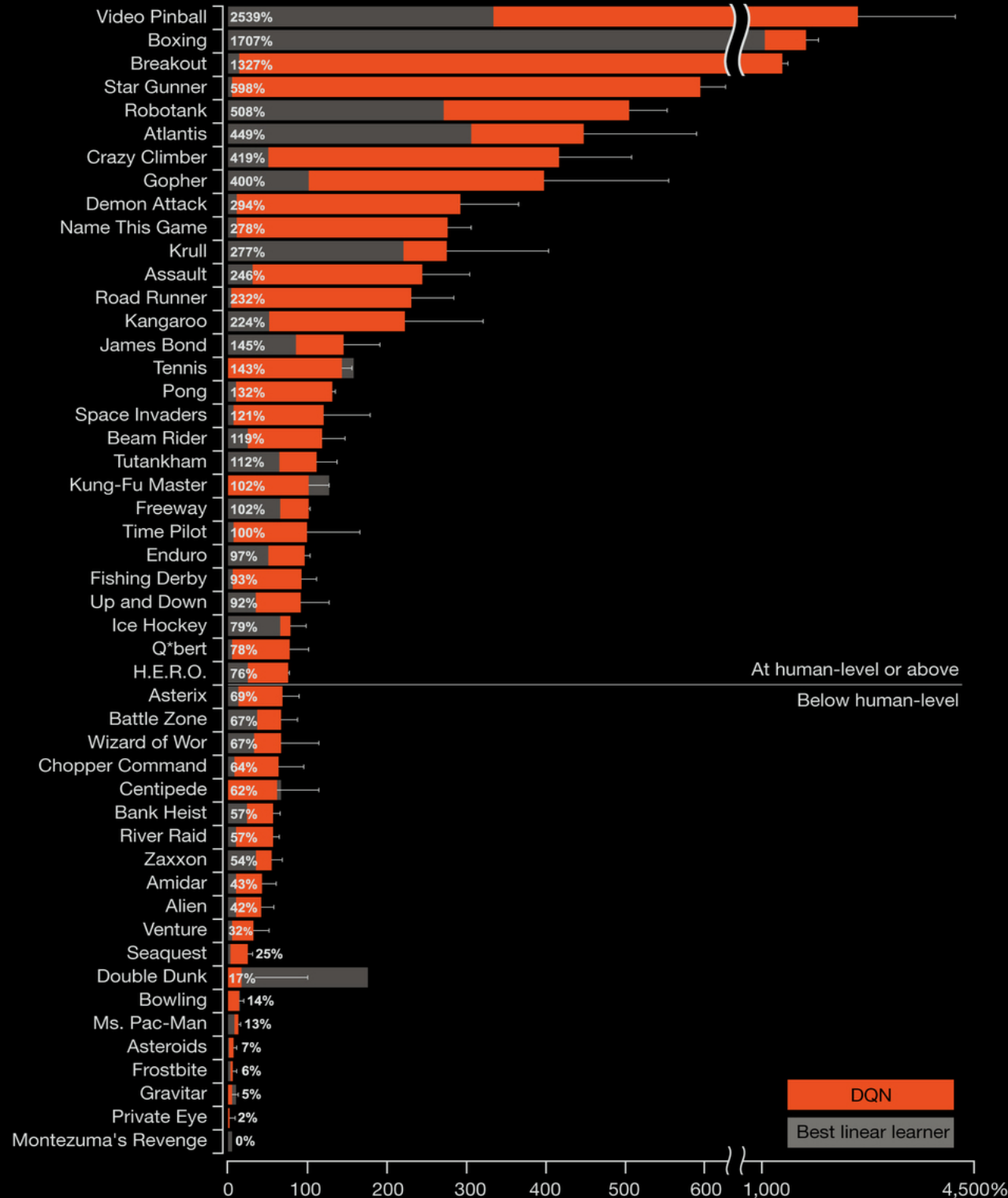
Self-taught AI software
attains human-level
performance in video games

2015

https://pbs.twimg.com/media/B-11_3

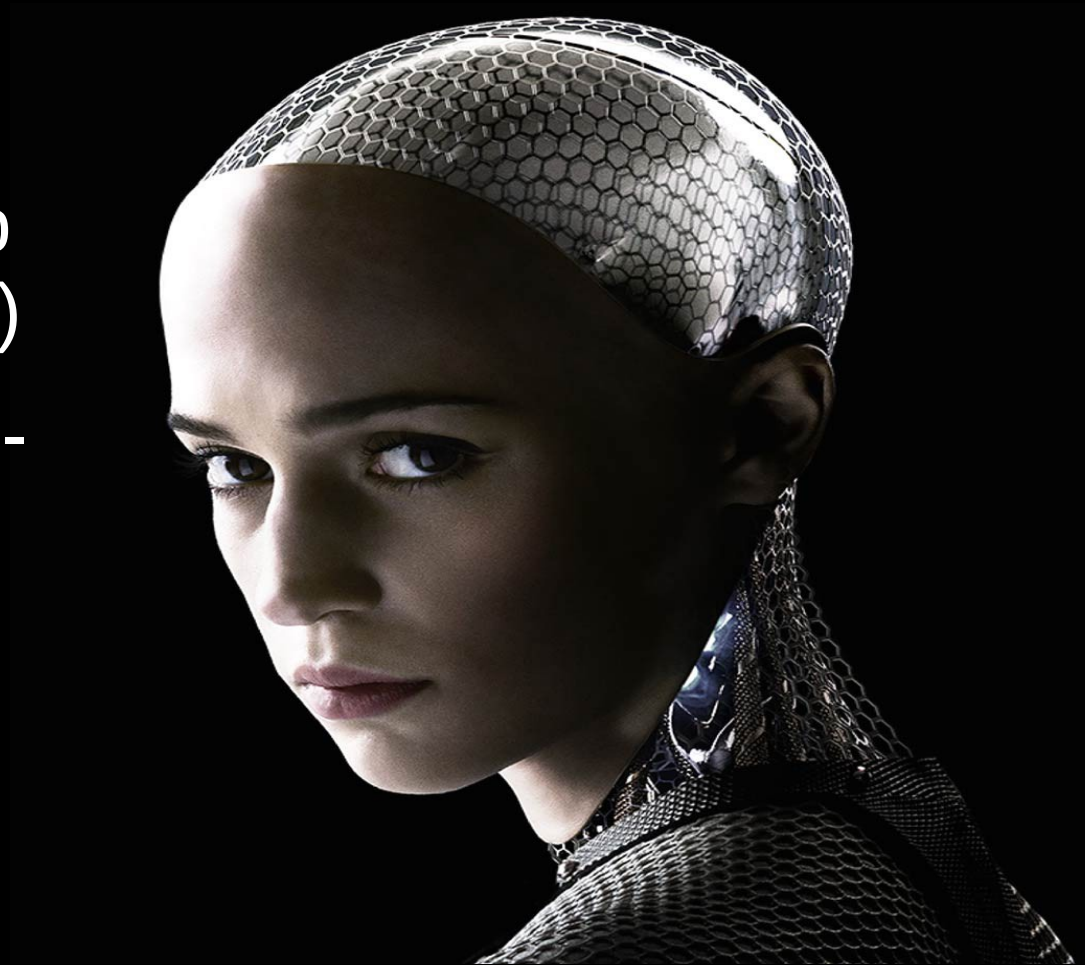






Human-level AI

- Intelligence is an optimization process
- Computers and AI keep improving (Moore's law)
- We might reach human-level AI soon
- Kurzweil: 2029
Legg (founder DeepMind): 2025



Thinking about Human-level AI

- Don't anthropomorphize!
- An AI won't think
 - Babies are cute
 - It's wrong to kill
- Not 'human-shaped'

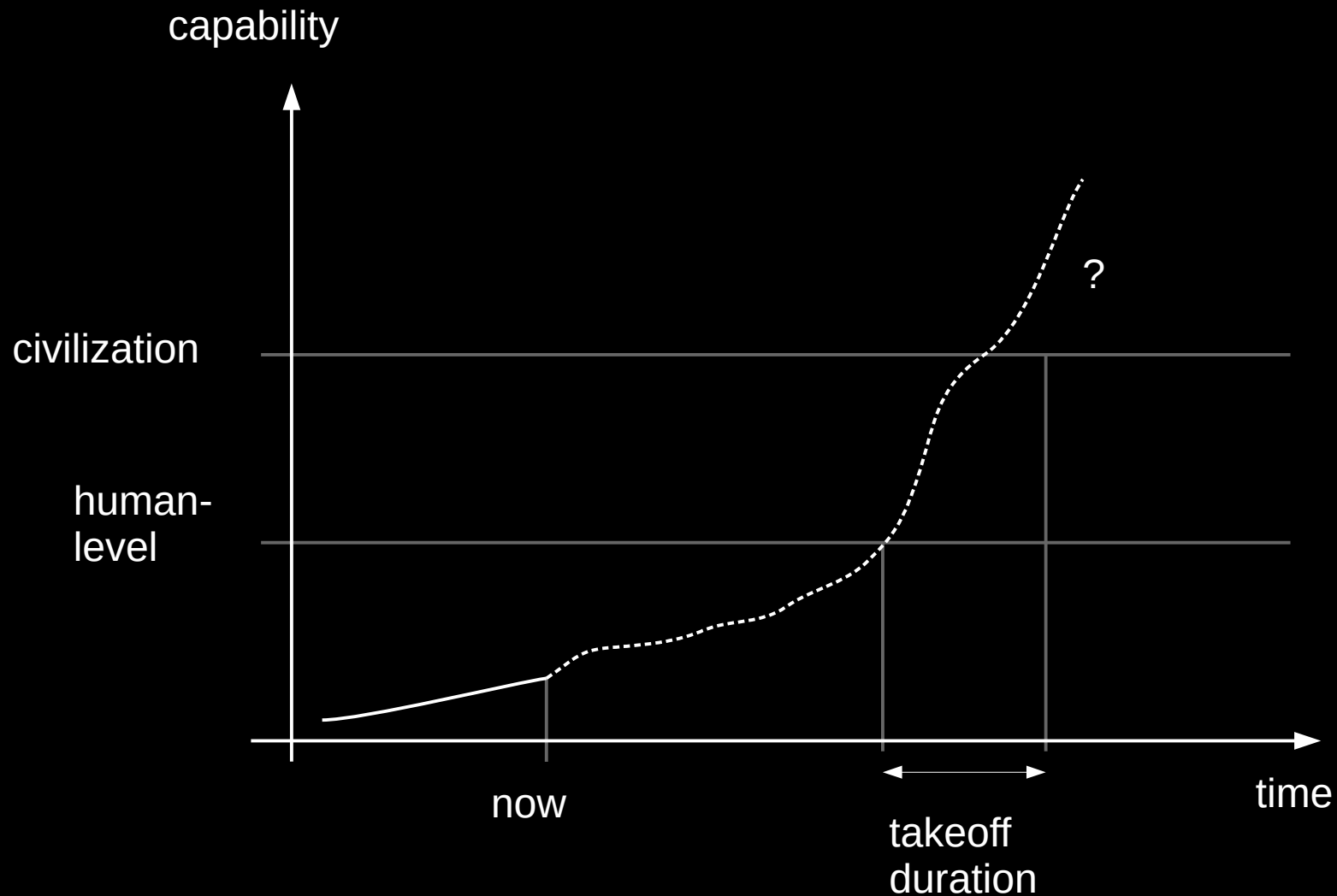


<http://cdn.patch.com/users/127196/stock/T800x600/20150154c116fb969fb.jpg>

Convergent Instrumental Goals

- Resource acquisition
- Survival
- Self-improvement
- ...

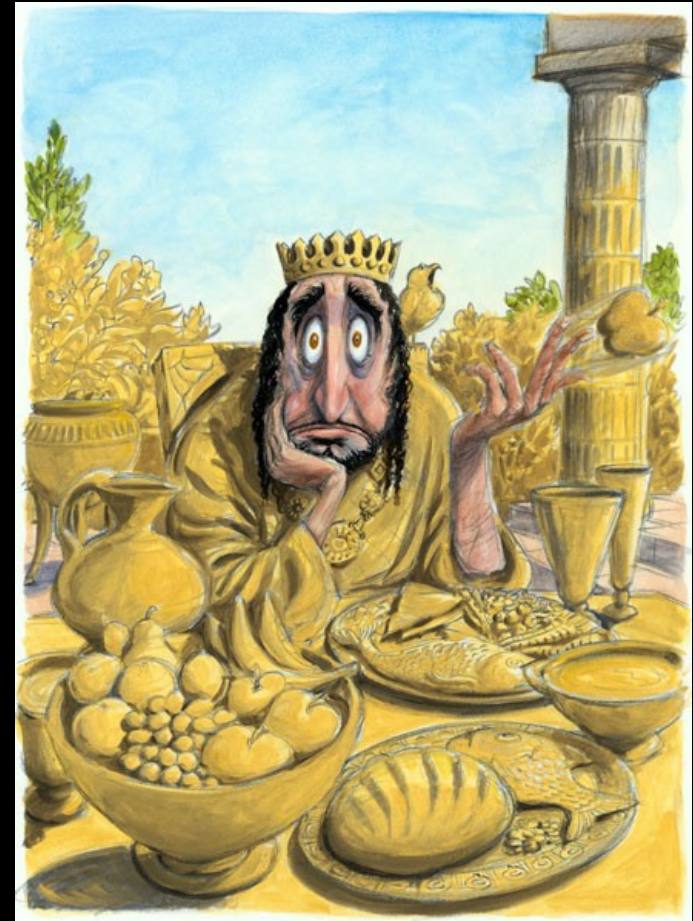
Towards Superintelligence



The Evil Genie Effect

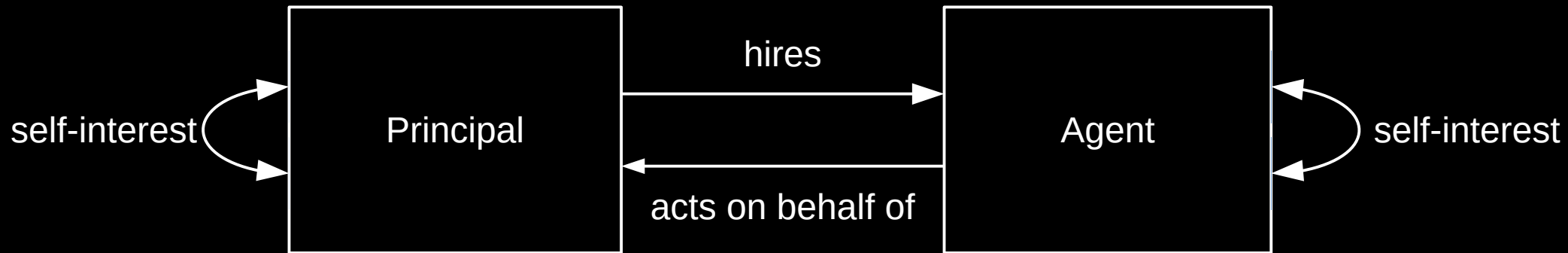
- Goal: Cure Cancer!
- AI-generated plan:
 1. Make lots of money by beating humans at stock market predictions
 2. Solve a few genetic engineering challenges
 3. Synthesize a supervirus that wipes out the human species
 4. No more cancer

King Midas



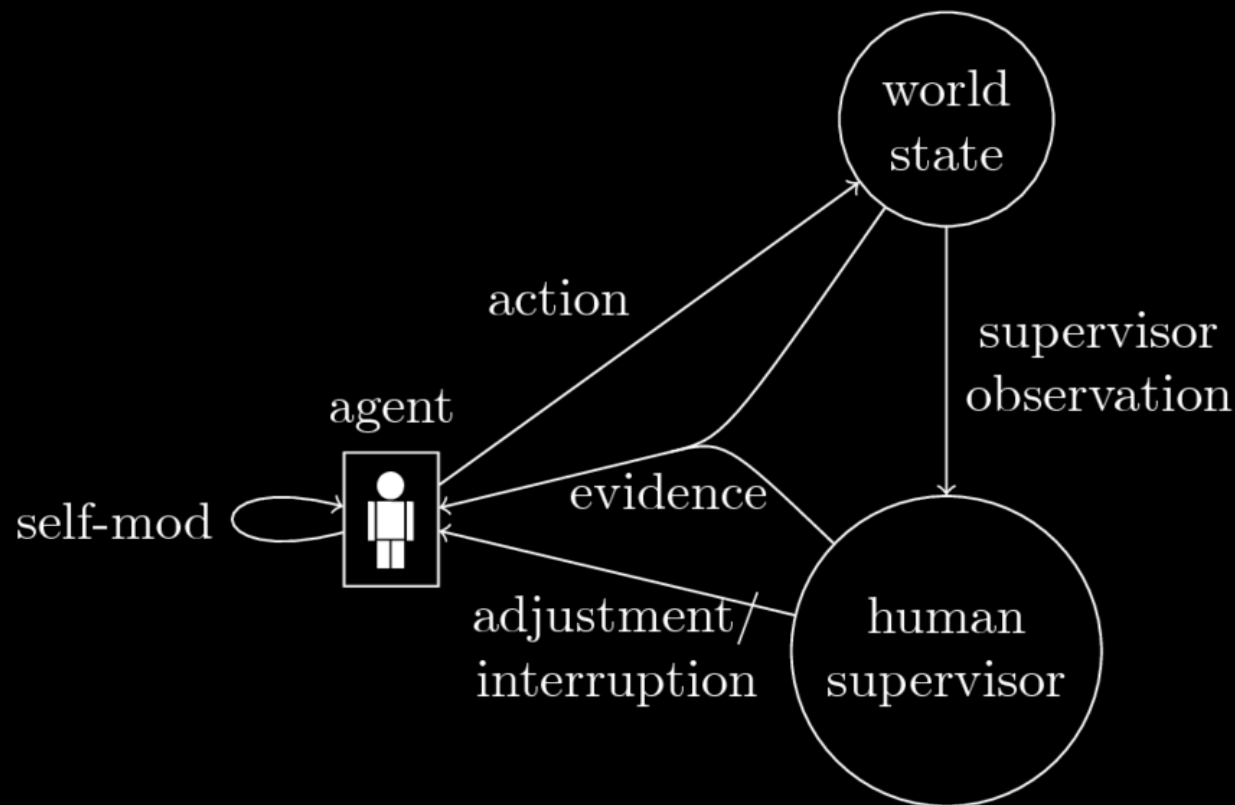
<https://anentrepreneurswords.files.wordpress.com/2014/06/king-midas.jpg>

The Principal-Agent Problem



Value Learning

- Goal:
Learn what humans value and optimize that
- A bit vague, and a lot can go wrong

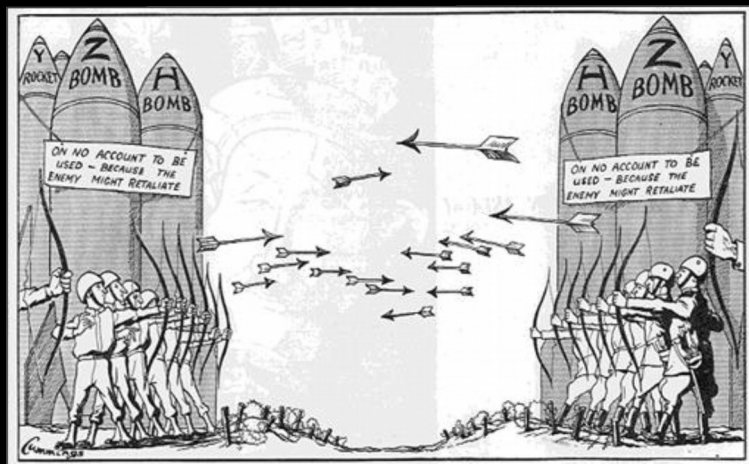


AI Arms Race

- Getting to human-level AI first is a big advantage
- Risk compromising safety to 'win the race'

Combined:

- Space race
- Arms race
- Gold rush





Superintelligent AI/Evil Genie

AI arms race

Autonomous
weapons

Unemployment

What to do?

- Universal basic income
- Make AI technology open
- Ban autonomous weapons
- Tight international cooperation
- Major research effort on keeping AI robust and beneficial

Organizations



CENTRE FOR THE STUDY OF EXISTENTIAL RISK
UNIVERSITY OF CAMBRIDGE

Summary

- Medium-term challenges:
 - Unemployment
 - Autonomous warfare
- Long-term challenges:
 - AI arms race
 - Arrival of human-level AI (soon?)
 - Intelligence explosion
 - Evil genies
- Things we can do:
 - Universal basic income
 - Make AI technology open
 - Ban autonomous weapons
 - Research AI Safety
 - International cooperation

References

- *Superintelligence: Paths, Dangers, Strategies*
Nick Bostrom, Oxford University Press, 2014
- *The Basic AI Drives*
Stephen Omohundro, Artificial General Intelligence, 2008
- *Economics Of The Singularity*
Robin Hanson, IEEE Spectrum, 2008
- *The Singularity Is Near*
Ray Kurzweil, Viking, 2006