Artificial Intelligence:

Summary of AI
Summary of the Module
Philosophy and Social Issues
Summary of Exam

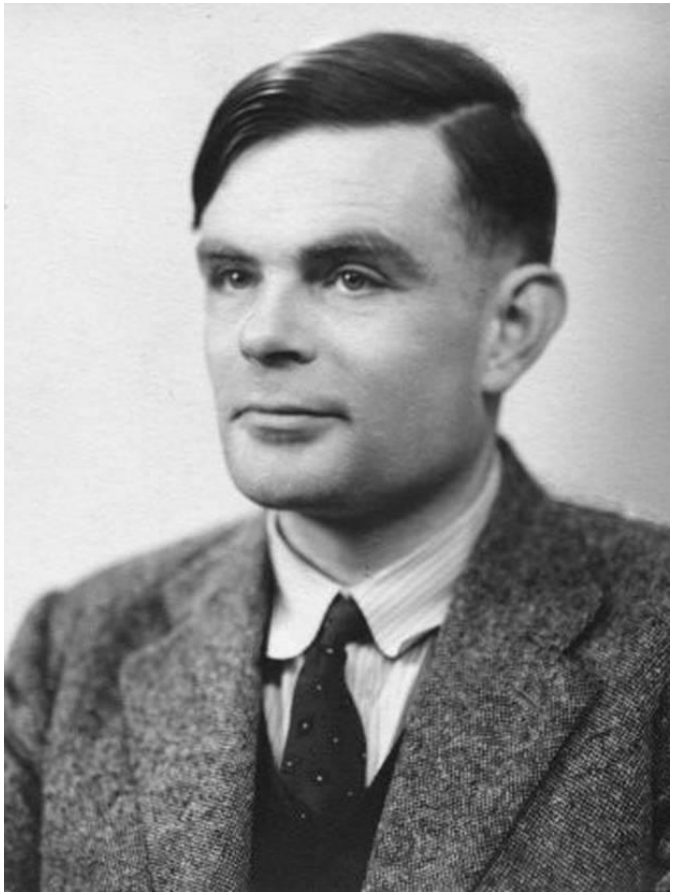
Birth of clockwork & Automata

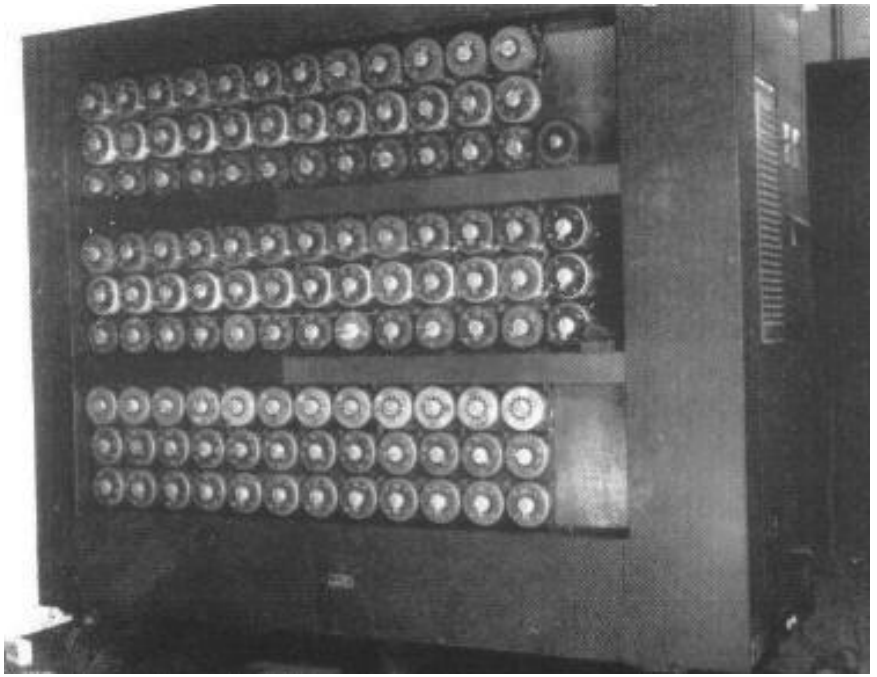
Maillardet's Automaton - pre 1800

The Writer - 1775

AI in Popular Culture

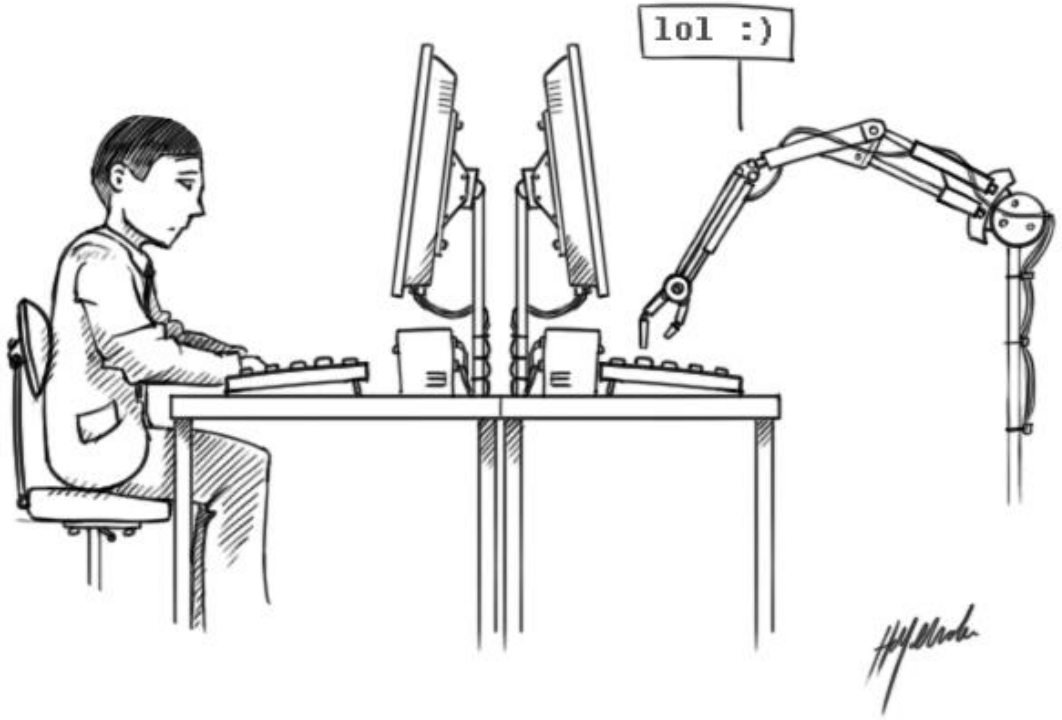
Early Computers & Turing (1940s)



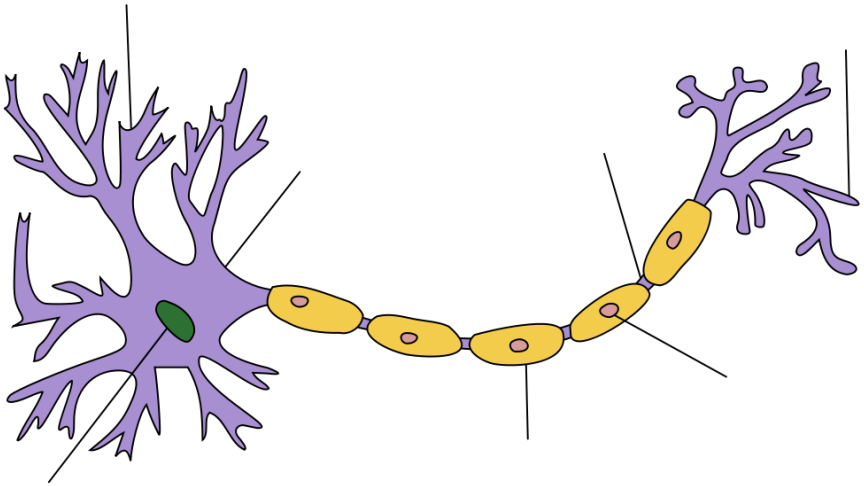


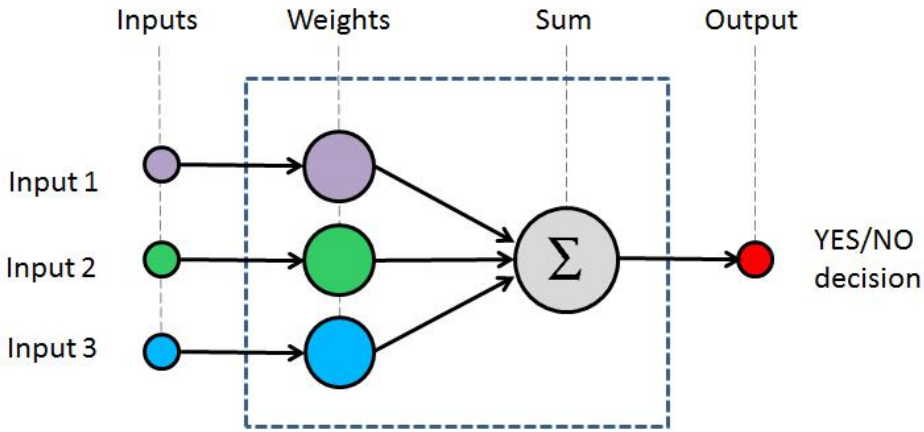
British Bombe (1 ton)

Turing Test (1951)



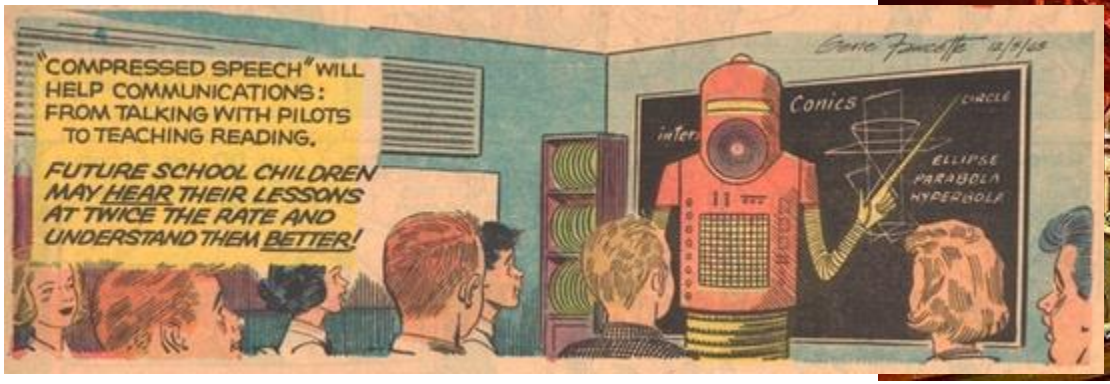
Early AI
Lots of Optimism
Promised the world?





The “Perceptron” – model of a neuron

Visions of the future



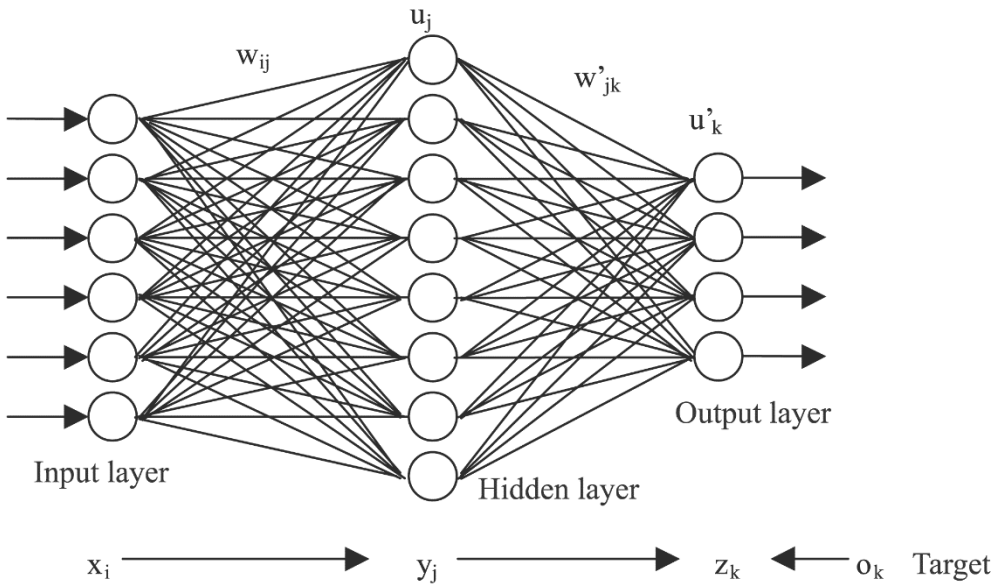
But …

- Hubert Dreyfus, in “What Computers Can't Do” 1972
- Machine Translation couldn’t cope with context
- Handwriting only worked on “easy” examples
- Chess was still dominated by Humans

Since the 70s computers got bigger & faster:

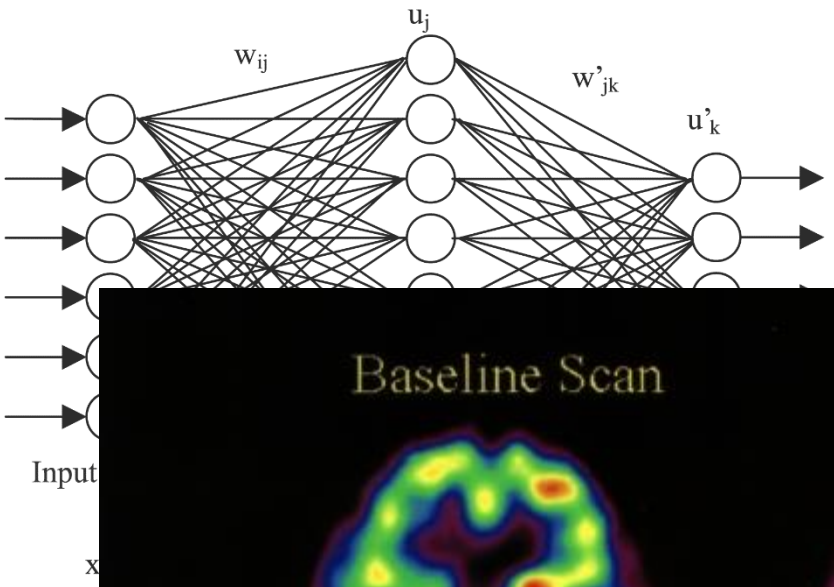
Deep Blue 1997

… and algorithms and models got better:

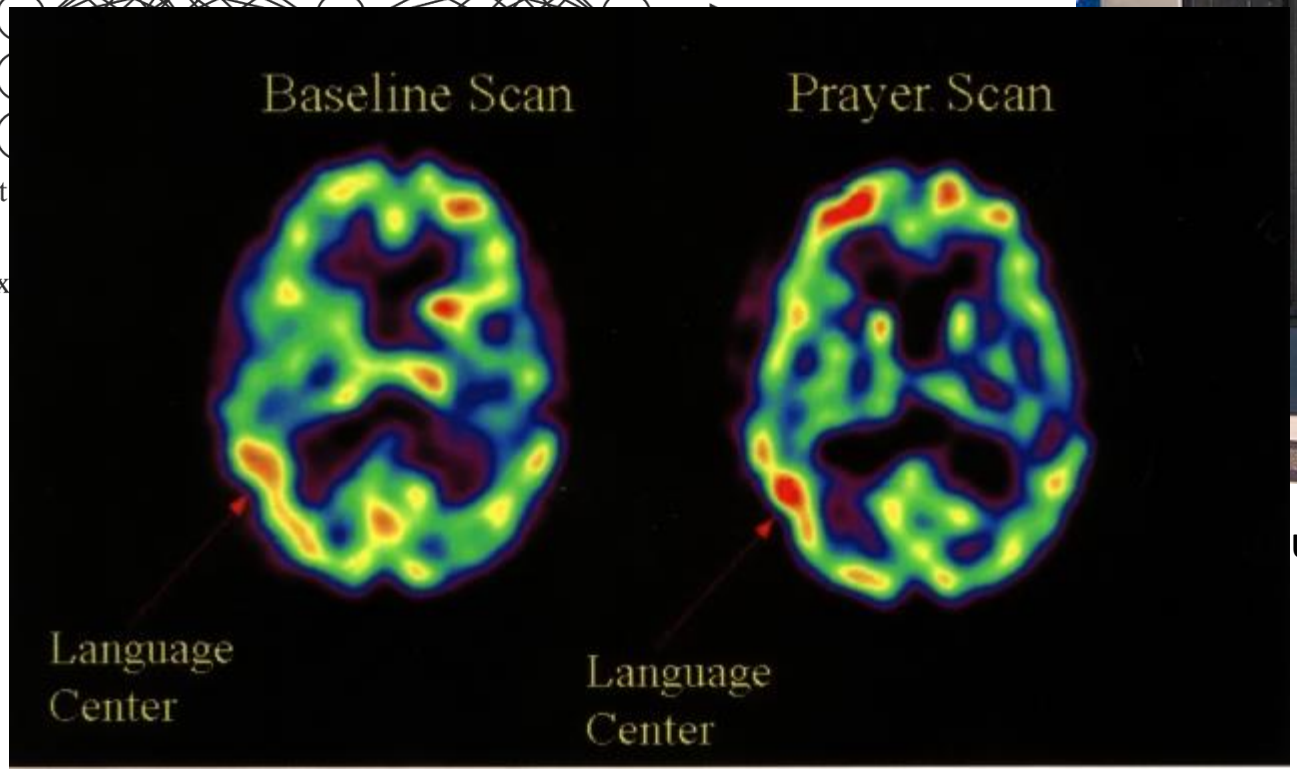


Deep Blue 1997

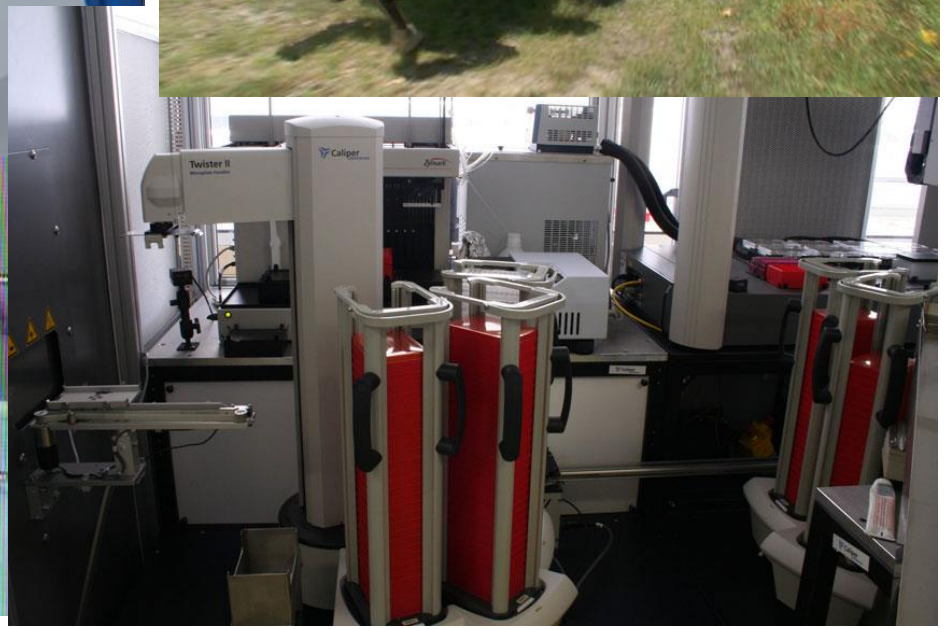
… and we learnt lots about the brain:



ue 1997



Successes



AI: Where are we at?

What has been easy and what has been hard?

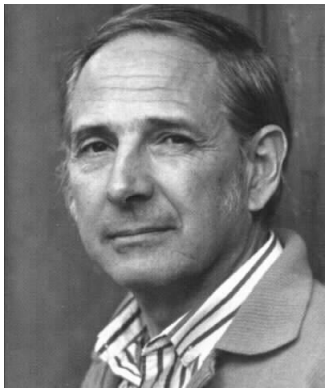
http://www.youtube.com/watch?v=WnzlbyTZsQY

Philosophy of Mind & Consciousness

Descartes:

Animals were essentially mechanical but humans have a “Ghost in the Machine”

Philosophy of Mind & Consciousness



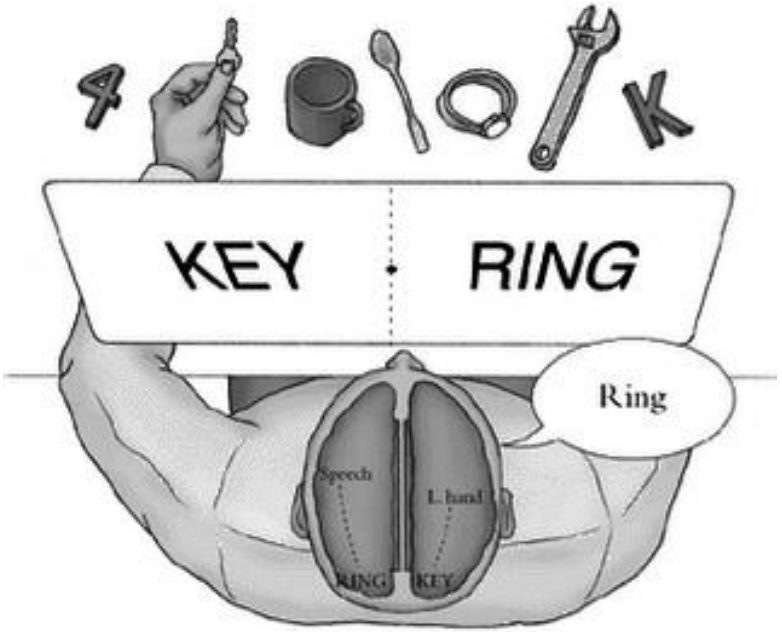
Searle's Chinese Room

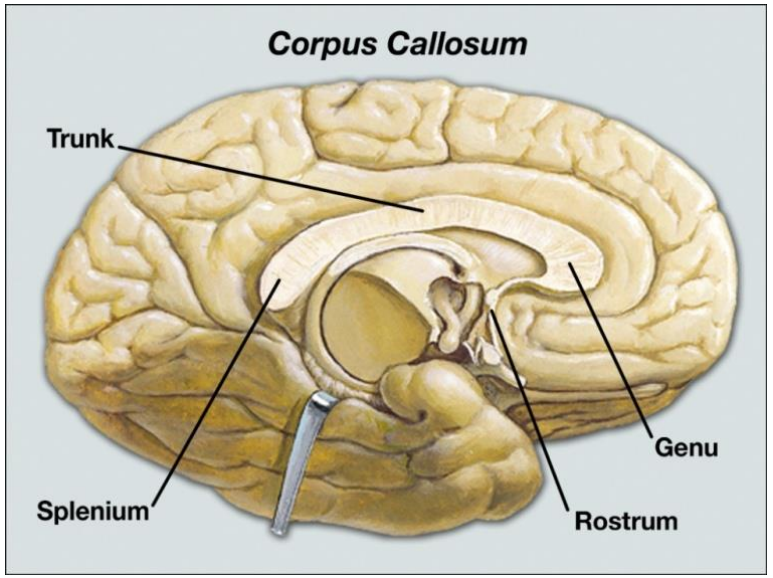
Functionalism = Strong AI?

Philosophy of Mind & Consciousness

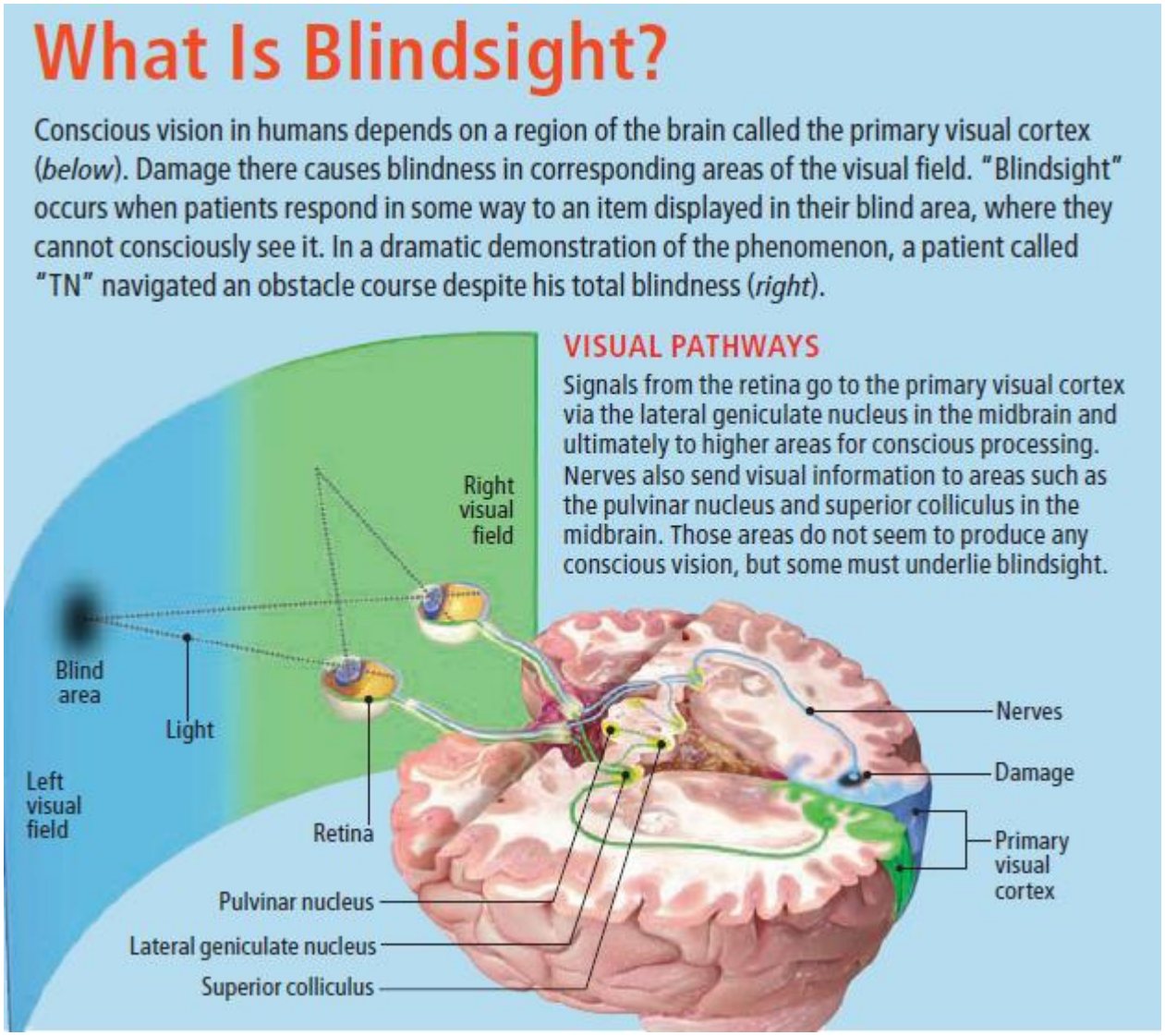
Brain Injuries / Surgery

- Corpus Callosum cuts for epilepsy





Philosophy of Mind & Consciousness



Consciousness

Ray Kurzweil, Director of Engineering, Google:

“In 2029 ...
... Turing Test will be passed
... Robots will outsmart us all
... Will gain what looks like consciousness”

predicted the future of the internet in the 80s,
he identified the year computers would beat humans at chess
foresaw the fall of the Soviet Union

theguardian

Computer simulating 13-year-old boy becomes first to pass Turing test

'Eugene Goostman' fools 33% of interrogators into thinking it is human, in what is seen as a milestone in artificial intelligence

- In 'his own' words: how Eugene fooled the Turing judges
- What is the Turing test? And are we all doomed now?

Press Association
theguardian.com, Monday 9 June 2014 12.09 BST
Jump to comments (655)

theguardian

News | Sport | Comment | Cultu

News › Technology › Artificia

Computer simulati becomes first to pas

'Eugene Goostman' fools 33% human, in what is seen as a mi

- In 'his own' words: how Euger
- What is the Turing test? And are we all doomed now?

Press Association
theguardian.com, Monday 9 June 2014 12.09 BST
Jump to comments (655)

THE INDEPENDENT THURSDAY 19 JUNE 2014 Apps

NEWS | IMAGES | VOICES | SPORT | WORLD CUP HUB | TECH | LIFE | PROPERTY | ARTS + ENTS | TR

UK / World / Business / People / Science / Environment / Media / Technology / Education / Images / Ob

News > Technology

Turing Test breakthrough as super-computer becomes first to convince us it's human

theguardian

News | Sport | Comment | Cultu

News › Technology › Artificia

Computer simulati becomes

'Eugene Goo human, in wh

- In 'his own'
- What is the

Press Association theguardian.com,

Jump to comm

THE INDEPENDENT THURSDAY 19 JUNE 2014 Apps

NEWS | IMAGES | VOICES | SPORT | WORLD CUP HUB | TECH | LIFE | PROPERTY | ARTS + ENTS | TR

ogy / Education / Images / Ob

The Telegraph

Home News World Sport World Cup Finance Comment Culture Travel Life Wome

Technology News | Technology Companies | Technology Reviews | Video Games | Technology Video

HOME » TECHNOLOGY » TECHNOLOGY NEWS

Computer passes 'Turing Test' for the first time after convincing users it is human

A "super computer" has duped humans into thinking it is a 13-year-old boy, becoming the first machine to pass the "iconic" Turing Test, experts say

omputer an

2014 University of Reading competition [edit]

On 7 June 2014 in a Turing test competition organised by Kevin Warwick to mark the 60th anniversary of Turing's death, was won by the Russian chatter bot Eugene Goostman. The bot, during a series of five minute-long text conversations, convinced 33% of the contest's judges that it was human. Judges included John Sharkey, a sponsor of the bill granting a government pardon to Turing, and *Red Dwarf* actor Robert Llewellyn.[48][49][50][51]

The competition's organiser believed that the Turing test had been "passed for the first time" at the event, saying that "some will claim that the Test has already been passed. The words Turing Test have been applied to similar competitions around the world. However this event involved the most simultaneous comparison tests than ever before, was independently verified and, crucially, the conversations were unrestricted. A true Turing Test does not set the questions or topics prior to the conversations."[49]

The contest has faced criticism, with many in the AI community stating that the computer clearly did not pass the test. First, only a third of the judges were fooled by the computer. Second, the program's character claimed to be a Ukrainian who learned English as a second language. Third, it claimed to be 13 years old, not an adult. The contest only required 30% of judges to be fooled, a very low threshold. This was based on an out-of-context quote by Turing, where he was predicting the future capabilities of computers rather than defining the test. In addition, many of its responses were cases of dodging the question, without demonstrating any understanding of what was said. Joshua Tenenbaum, an AI expert at MIT stated that the result was unimpressive.[52]

Versions of the Turing test [edit]

Saul Traiger argues that there are at least three primary versions of the Turing test, two of which are offered in "Computing Machinery and Intelligence" and one that he describes as the "Standard Interpretation."[53] While there is some debate regarding whether the

2014 University of Reading competition [edit]

On 7 June 2014 in a Turing test competition organised by Kevin Warwick to mark the 60th anniversary of Turing's death, was won by the Russian chatter bot Eugene Goostman. The bot, during a series of five minute-long text conversations, convinced 33% of the contest's judges that it was human. Judges included John Sharkey, a sponsor of the bill granting a government pardon to Turing, and *Red Dwarf* actor Robert Llewellyn.[48][49][50][51]

The competition's organiser believed that the Turing test had been "passed for the first time" at the event, saying that "some will claim that the Test has already been passed. The words Turing Test have been applied to similar competitions around the world. However this event involved the most simultaneous comparison tests than ever before, was independently verified and, crucially, the convers

First, only a third of the judges were fooled

cs prior to the conversations."[49]

The contest has faced criticism, with many in the AI community stating that the computer clearly did not pass the test. First, only a third of the judges were fooled by the computer. Second, the program's character claimed to be a Ukrainian who learned English as a second language. Third, it claimed to be 13 years old, not an adult. The contest only required 30% of judges to be fooled, a very low threshold. This was based on an out-of-context quote by Turing, where he was predicting the future capabilities of computers rather than defining the test. In addition, many of its responses were cases of dodging the question, without demonstrating any understanding of what was said. Joshua Tenenbaum, an AI expert at MIT stated that the result was unimpressive.[52]

Versions of the Turing test [edit]

Saul Traiger argues that there are at least three primary versions of the Turing test, two of which are offered in "Computing Machinery and Intelligence" and one that he describes as the "Standard Interpretation."[53] While there is some debate regarding whether the

2014 University of Reading competition [edit]

On 7 June 2014 in a Turing test competition organised by Kevin Warwick to mark the 60th anniversary of Turing's death, was won by the Russian chatter bot Eugene Goostman. The bot, during a series of five minute-long text conversations, convinced 33% of the contest's judges that it was human. Judges included John Sharkey, a sponsor of the bill granting a government pardon to Turing, and *Red Dwarf* actor Robert Llewellyn.[48][49][50][51]

The competition's organiser believed that the Turing test had been "passed for the first time" at the event, saying that "some will claim that the Test has already been passed. The words Turing Test have been applied to similar competitions around the world. However this event involved the most simultaneous comparison tests than ever before, was independently verified and, crucially, the convers First, only a third of the judges were fooled cs prior to the conversations."[49]

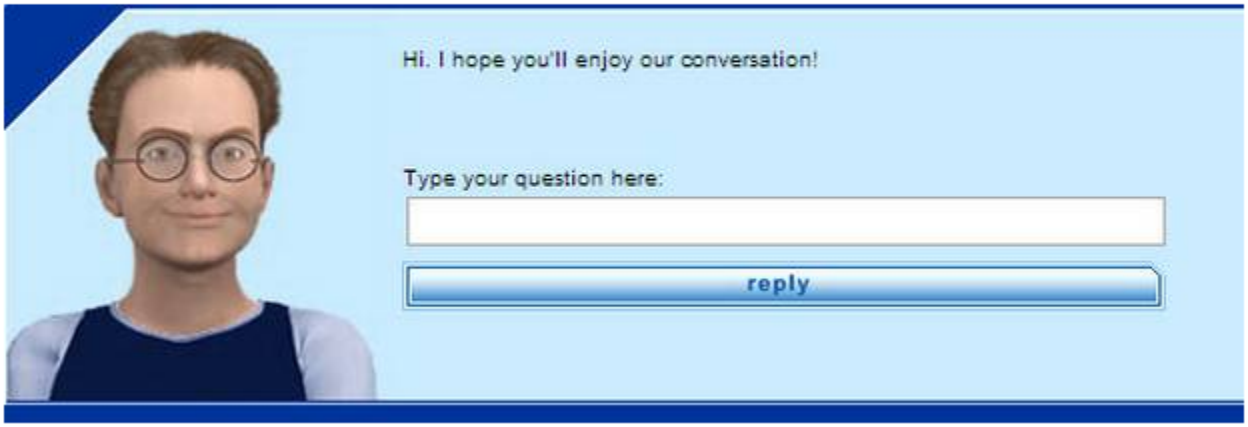
Second, the program's character claimed to be a Ukrainian who learned English as a second language. T

contest only required 30% of judges to be fooled, a very low threshold. This was based on an out-of-context quote by Turing, where he was predicting the future capabilities of computers rather than defining the test. In addition, many of its responses were cases of dodging the question, without demonstrating any understanding of what was said. Joshua Tenenbaum, an AI expert at MIT stated that the result was unimpressive.[52]

Versions of the Turing test [edit]

Saul Traiger argues that there are at least three primary versions of the Turing test, two of which are offered in "Computing Machinery and Intelligence" and one that he describes as the "Standard Interpretation."[53] While there is some debate regarding whether the

Turing Test breakthrough as super-computer becomes first to convince us it's human



WIKIPEDIA
The Free Encyclopedia

2014 University of Reading competition [edit]

On 7 June 2014 in a Turing test competition organised by Kevin Warwick to mark the 60th anniversary of Turing's death, was won by the Russian chatter bot Eugene Goostman. The bot, during a series of five minute-long text conversations, convinced 33% of the contest's judges that it was human. Judges included John Sharkey, a sponsor of the bill granting a government pardon to Turing, and *Red Dwarf* actor Robert Llewellyn.[48][49][50][51]

The competition's organiser believed that the Turing test had been "passed for the first time" at the event, saying that "some will claim that the Test has already been passed. The words Turing Test have been applied to similar competitions around the world. However this event involved the most simultaneous comparison tests than ever before, was independently verified and, crucially, the convers cs prior to the conversations."[49]

First, only a third of the judges were fooled

Second, the program's character claimed to be a Ukrainian who learned English as a second language.

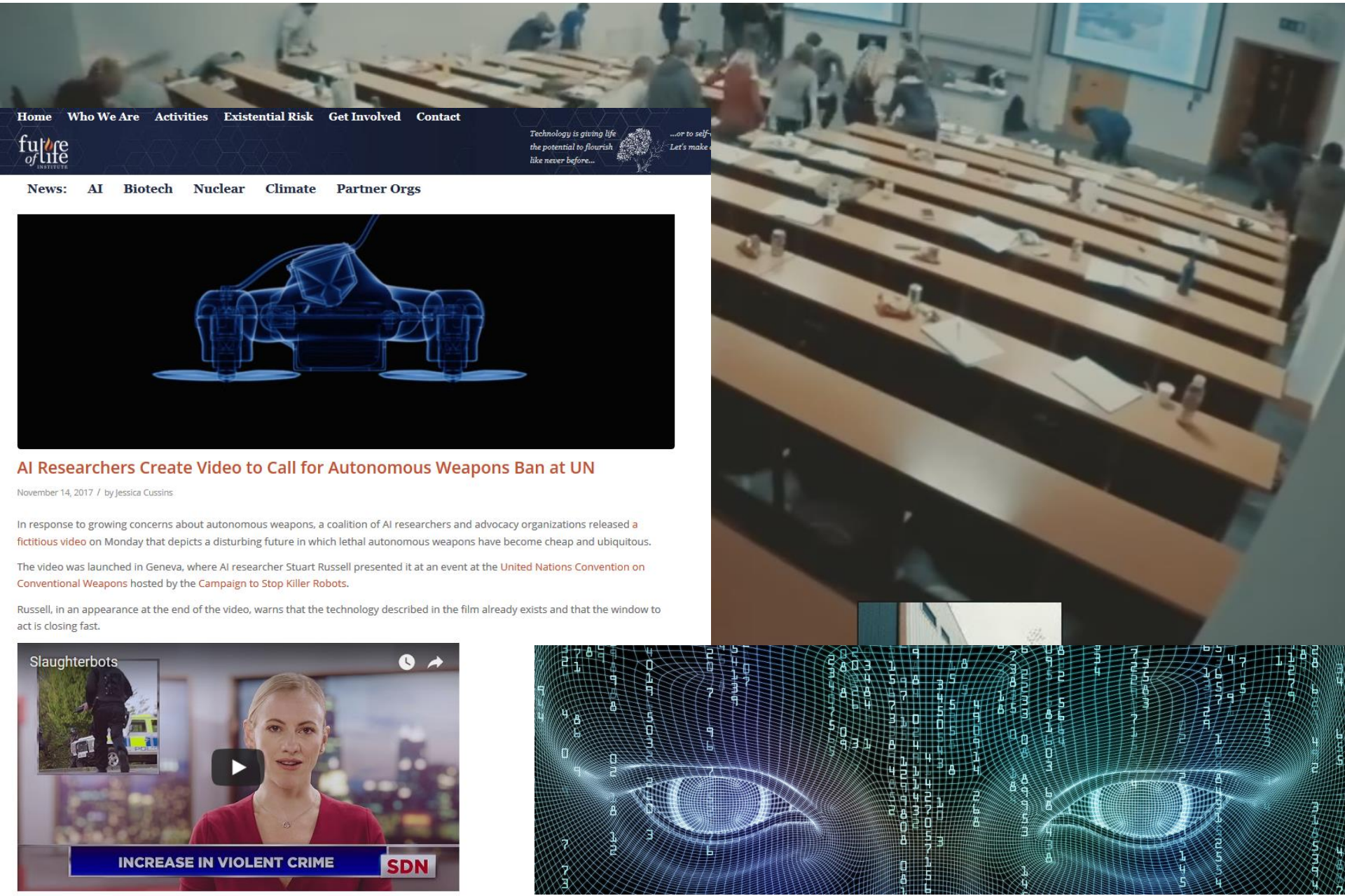
contest only required 30% of judges to be ere he was predicting the future capabilities of computers rather than defining the test. In rating any understanding of what was said. Joshua Tenenbaum, an AI expert at MIT stated that the result was unimpressive.[52]

Third, it claimed to be 13 years old, not an adult.

Versions of the Turing test [edit]

Saul Traiger argues that there are at least three primary versions of the Turing test, two of which are offered in "Computing Machinery and Intelligence" and one that he describes as the "Standard Interpretation."[53] While there is some debate regarding whether the

https://futureoflife.org/2017/11/14/ai-researchers-create-video-call-autonomous-weapons-ban-un/



Home Who We Are Activities Existential Risk Get Involved Contact

future of life INSTITUTE

Technology is giving life the potential to flourish like never before... ...or to self- Let's make

News: AI Biotech Nuclear Climate Partner Orgs

AI Researchers Create Video to Call for Autonomous Weapons Ban at UN

November 14, 2017 / by Jessica Cussins

In response to growing concerns about autonomous weapons, a coalition of AI researchers and advocacy organizations released a fictitious video on Monday that depicts a disturbing future in which lethal autonomous weapons have become cheap and ubiquitous.

The video was launched in Geneva, where AI researcher Stuart Russell presented it at an event at the United Nations Convention on Conventional Weapons hosted by the Campaign to Stop Killer Robots.

Russell, in an appearance at the end of the video, warns that the technology described in the film already exists and that the window to act is closing fast.

(If you have questions about this letter, please contact tegmark@mit.edu)

Research Priorities for Robust and Beneficial Artificial Intelligence: an Open Letter

Artificial intelligence (AI) research has explored a variety of problems and approaches since its inception, but for the last 20 years or so has been focused on the problems surrounding the construction of intelligent agents - systems that perceive and act in some environment. In this context, "intelligence" is related to statistical and economic notions of rationality - colloquially, the ability to make good decisions, plans, or inferences. The adoption of probabilistic and decision-theoretic representations and statistical learning methods has led to a large degree of integration and cross-fertilization among AI, machine learning, statistics, control theory, neuroscience, and other fields. The establishment of shared theoretical frameworks, combined with the availability of data and processing power, has yielded remarkable successes in various component tasks such as speech recognition, image classification, autonomous vehicles, machine translation, legged locomotion, and question-answering systems.

As capabilities in these areas and others cross the threshold from laboratory research to economically valuable technologies, a virtuous cycle takes hold whereby even small improvements in performance are worth large sums of money, prompting greater investments in research. There is now a broad consensus that AI research is progressing steadily, and that its impact on society is likely to increase. The potential benefits are huge, since everything that civilization has to offer is a product of human intelligence; we cannot predict what we might achieve when this intelligence is magnified by the tools AI may provide, but the eradication of disease and poverty are not unfathomable. Because of the great potential of AI, it is important to research how to reap its benefits while avoiding potential pitfalls.

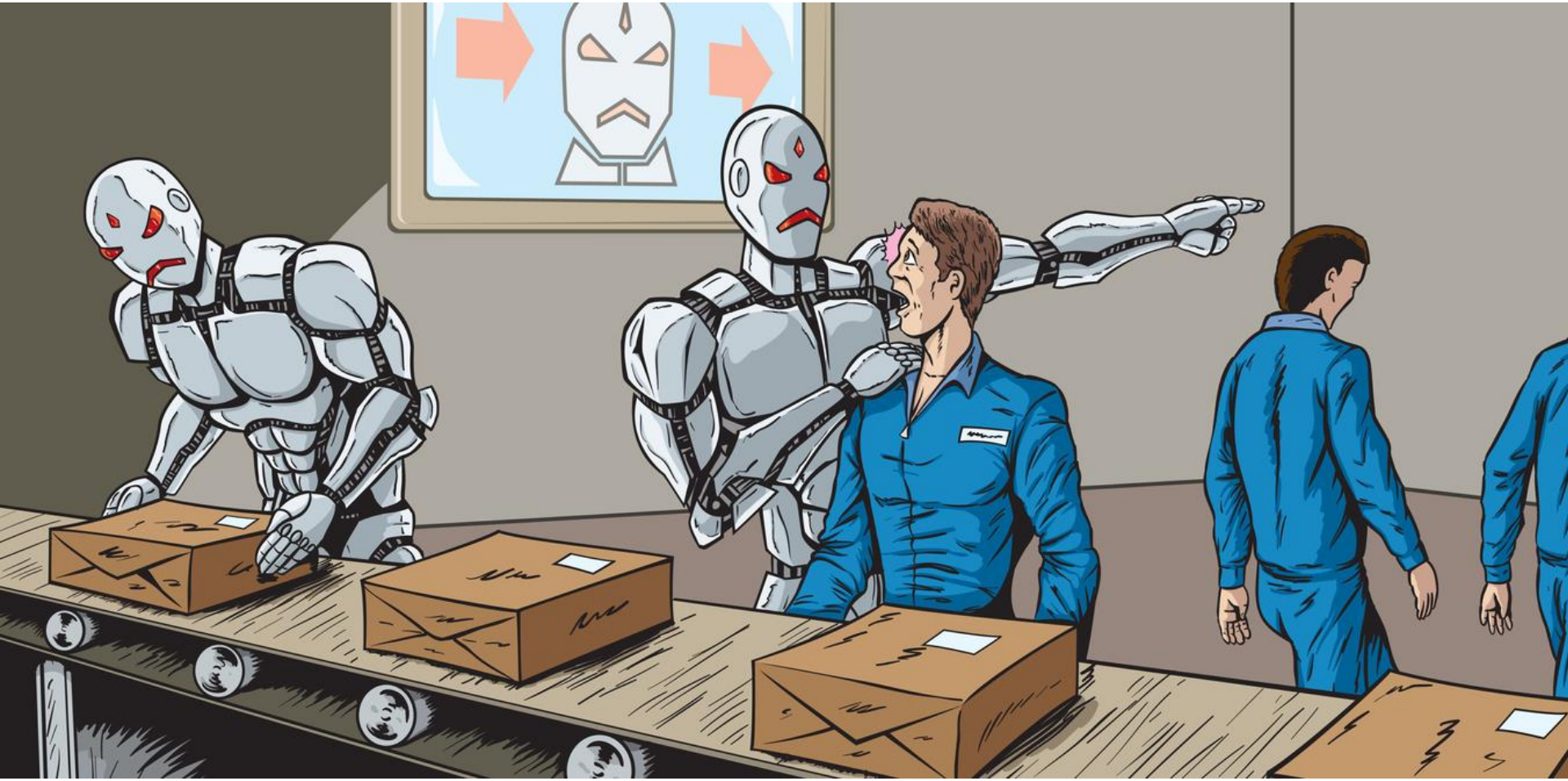
The progress in AI research makes it timely to focus research not only on making AI more capable, but also on maximizing the societal benefit of AI. Such considerations motivated the AAAI 2008-09 Presidential Panel on Long-Term AI Futures and other projects on AI impacts, and constitute a significant expansion of the field of AI itself, which up to now has focused largely on techniques that are neutral with respect to purpose. We recommend expanded research aimed at ensuring that increasingly capable AI systems are robust and beneficial: our AI systems must do what we want them to do. The attached research priorities document gives many examples of such research directions that can help maximize the societal benefit of AI. This research is by necessity interdisciplinary, because it involves both society and AI. It ranges from economics, law and philosophy to computer security, formal methods and, of course, various branches of AI itself.

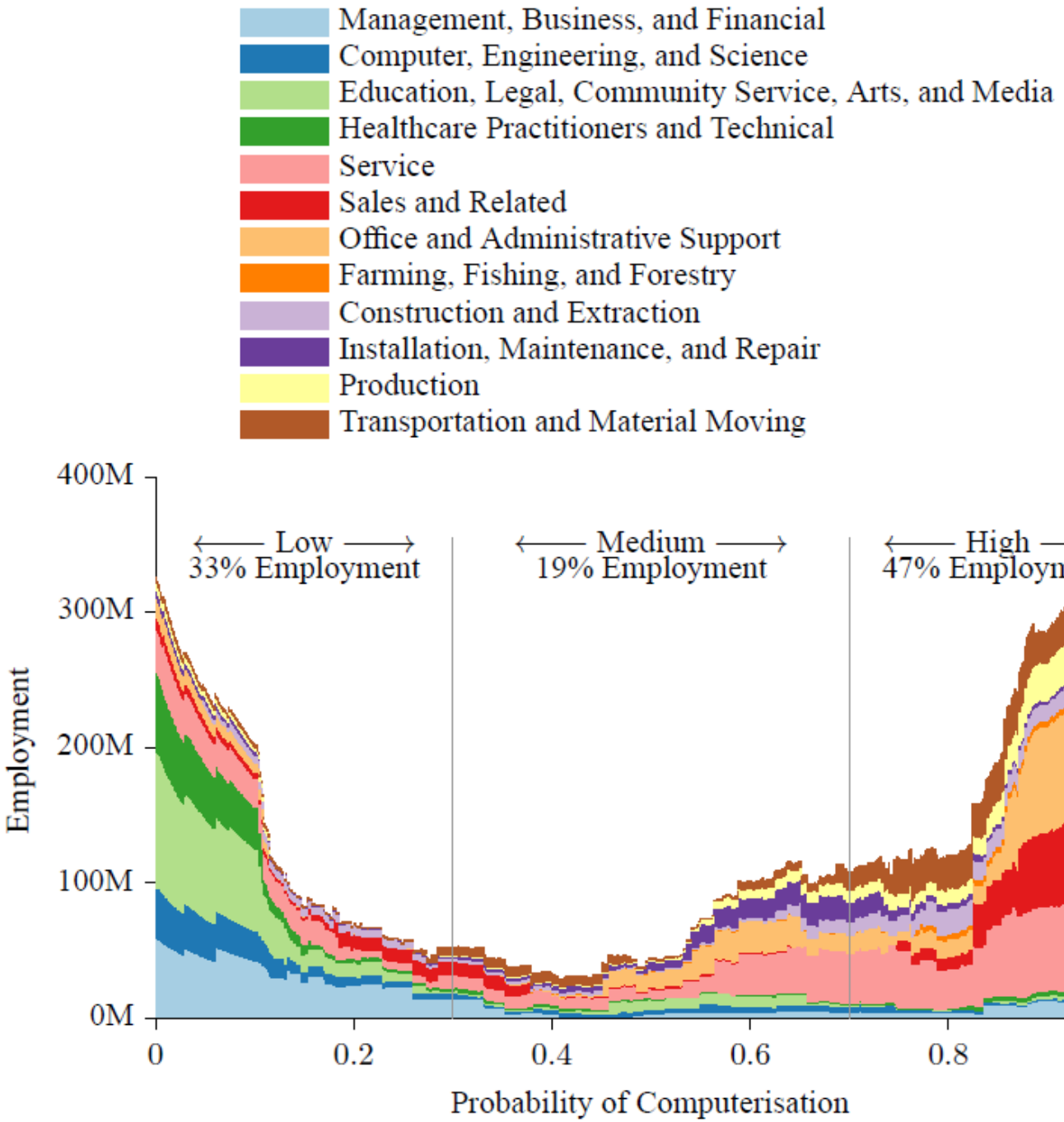
In summary, we believe that research on how to make AI systems robust and beneficial is both important and timely, and that there are concrete research directions that can be pursued today.

List of signatories

To sign the open letter, please add your

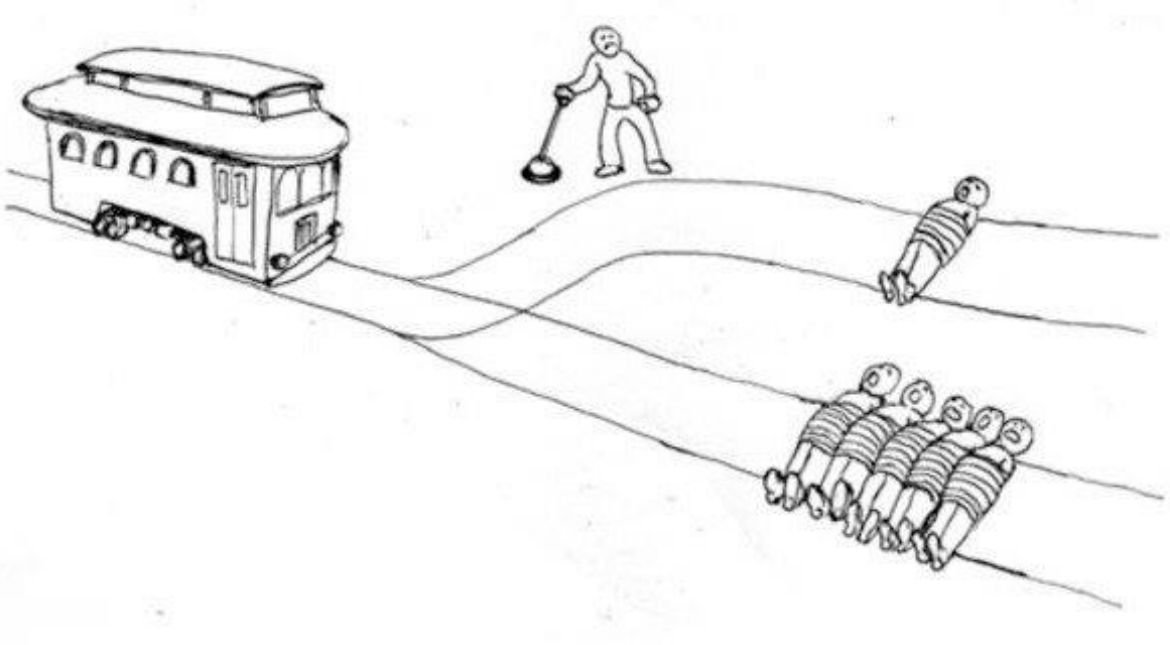
Ethics - Automation




Management, Business, and Financial
Computer, Engineering, and Science
Education, Legal, Community Service, Arts, and Media
Healthcare Practitioners and Technical
Service
Sales and Related
Office and Administrative Support
Farming, Fishing, and Forestry
Construction and Extraction
Installation, Maintenance, and Repair
Production
Transportation and Material Moving
400M
300M
200M
100M
0M
Employment
Low
33% Employment
Medium
19% Employment
High
47% Employment
0
0.2
0.4
0.6
0.8
1
Probability of Computerisation

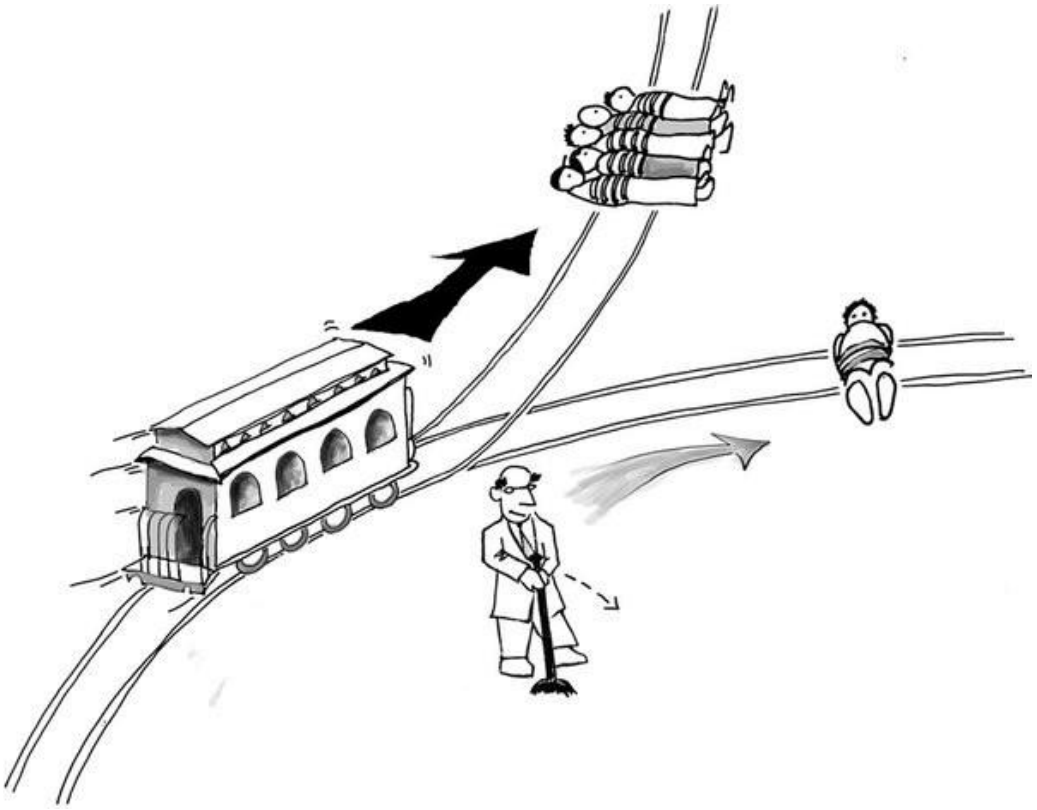
Ethics: The Trolley Problem

- Thought Experiment…



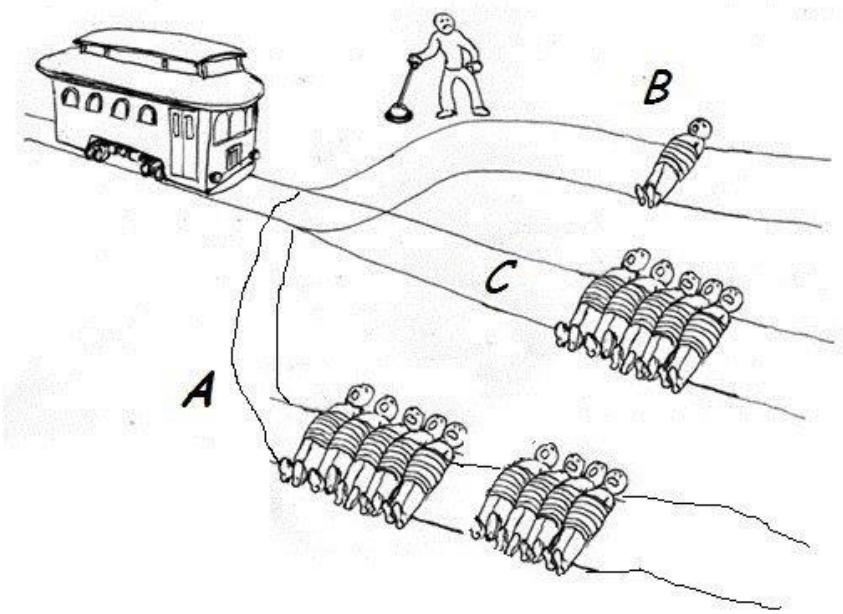
Ethics: The Trolley Problem

- Thought Experiment…



Ethics: The Trolley Problem

- Thought Experiment…

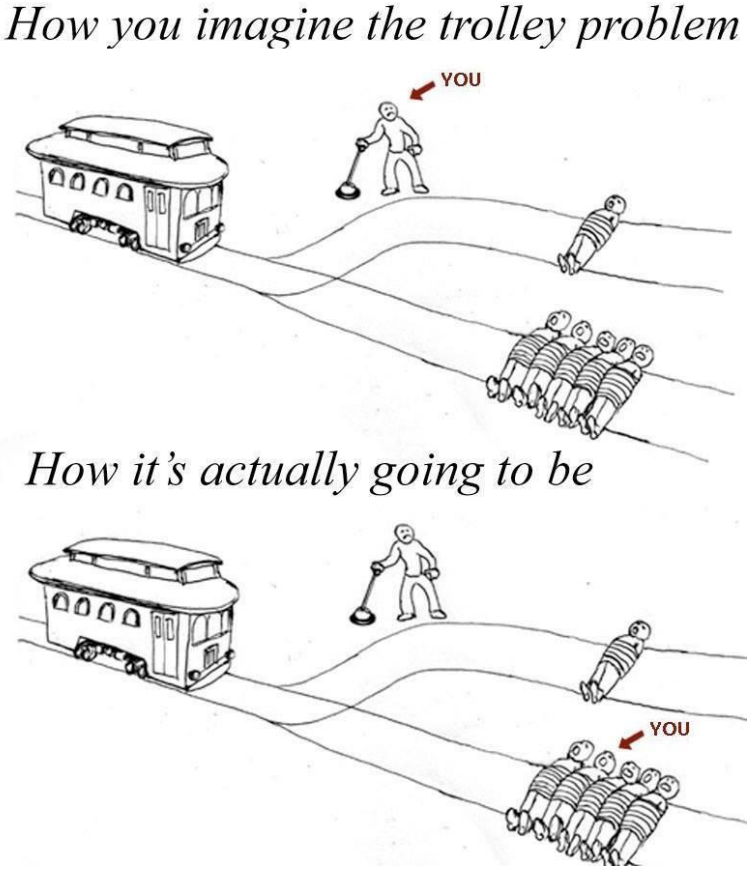


The Quantum Trolley Problem

If you do not pull the lever, the train will stay on track C. If you pull the lever, the train will be in either track A or track B. Until you observe the train, you will not know the effect of pulling the lever and thus it is said to be in superposition.

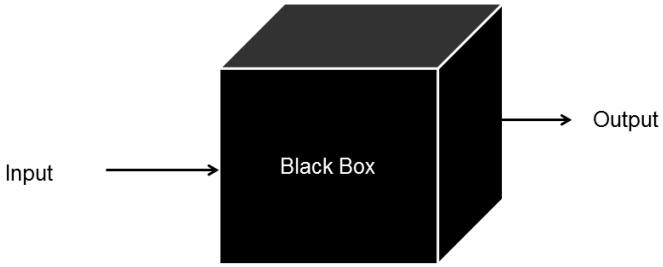
Ethics: The Trolley Problem

- Thought Experiment…

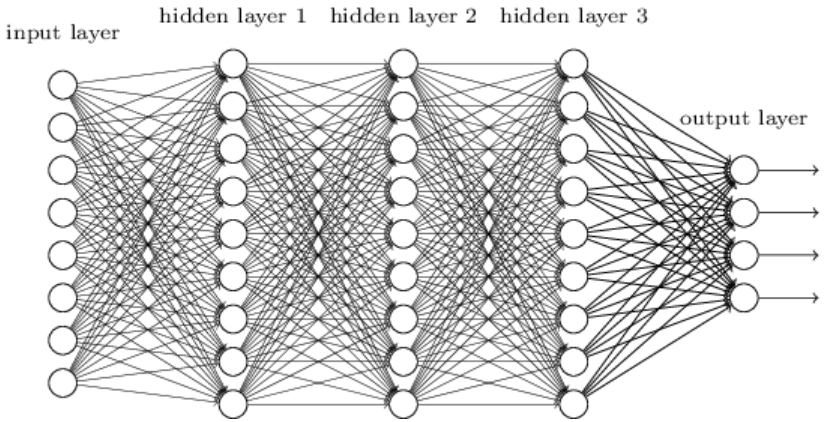


Ethics - Black Box

- Too complex for us to understand
  - Massively parallel
  - Huge numbers of parameters



Internal behavior of the code is unknown



Do we care?

"I don't care if the decision cannot be explained if it is better than a human"

The Geoff Hinton "Is this a 2?" argument

Automation Bias:

- Tendency to favour AI over Humans…

Automation Bias:

Self-driving Uber kills Arizona woman in first fatal crash involving pedestrian

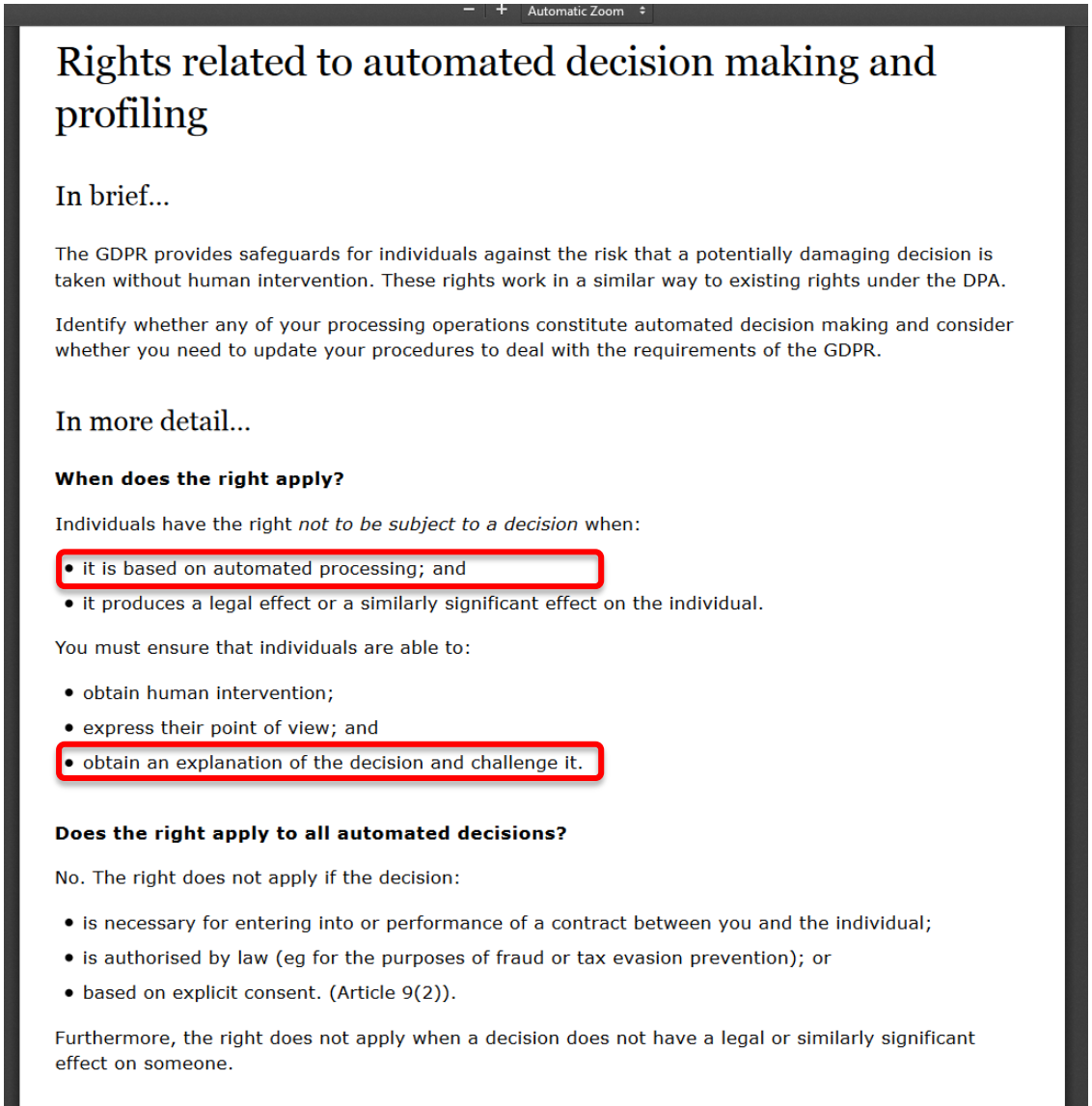
Tempe police said car was in autonomous mode at the time of the crash and that the vehicle hit a woman who later died at a hospital

▲ A car passes the location where a woman pedestrian was struck and killed by an Uber self-driving sport utility vehicle in Tempe, Arizona, on Monday. Photograph: Rick Scuteri/Reuters

Automation Bias:

- Tendency to favour AI over Humans…

General Data Protection Reg. 2018



Automatic Zoom

Rights related to automated decision making and profiling

In brief...

The GDPR provides safeguards for individuals against the risk that a potentially damaging decision is taken without human intervention. These rights work in a similar way to existing rights under the DPA.

Identify whether any of your processing operations constitute automated decision making and consider whether you need to update your procedures to deal with the requirements of the GDPR.

In more detail...

When does the right apply?

Individuals have the right *not to be subject to a decision* when:

- it is based on automated processing; and
- it produces a legal effect or a similarly significant effect on the individual.

You must ensure that individuals are able to:

- obtain human intervention;
- express their point of view; and
- obtain an explanation of the decision and challenge it.

Does the right apply to all automated decisions?

No. The right does not apply if the decision:

- is necessary for entering into or performance of a contract between you and the individual;
- is authorised by law (eg for the purposes of fraud or tax evasion prevention); or
- based on explicit consent. (Article 9(2)).

Furthermore, the right does not apply when a decision does not have a legal or similarly significant effect on someone.

Myth:
Superintelligence by 2100 is inevitable

Myth:
Superintelligence by 2100 is impossible

Fact:
It may happen in decades, centuries or never: AI experts disagree & we simply don't know

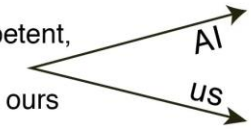
Myth:
Only Luddites worry about AI

Fact:
Many top AI researchers are concerned

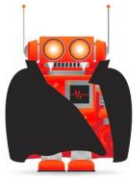
Mythical worry:
AI turning evil

Mythical worry:
AI turning conscious

Actual worry:
AI turning competent, with goals misaligned with ours



Myth:
Robots are the main concern



Fact:
Misaligned intelligence is the main concern: it needs no body, only an internet connection

Myth:
AI can't control humans

Fact:
Intelligence enables control: we control tigers by being smarter

Myth:
Machines can't have goals

Fact:
A heat-seeking missile has a goal

Mythical worry:
Superintelligence is just years away

Actual worry:
It's at least decades away, but it may take that long to make it safe

The Exam

- 3 hours
- Short answer questions
- Sample questions in the labs
- Also release excerpts of past papers that are relevant

The Topics

- Unsupervised Learning
  - Distance metrics
  - 2 key clustering algorithms in detail
  - Association Rules
- Classification
  - 2 key classification algorithms in detail
  - Sensitivity Analysis – TPs vs FPs
- Neural Networks
  - Forward propagation in Perceptron and Multilayer NNs
  - General form of Backpropagation
  - Some Knowledge on Deep Learning

The Topics

- Expert Systems
  - Knowledge Representation & Definition of Expert System
  - Rule based ES & Inference (forward / backward chaining & conflict resolution)
- Bayesian Networks
  - Definition and how to retrieve the joint probability
  - D-separation & Markov Blanket
- Time Series / Sequence Models
  - Markov chains and calculating probabilities of sequences
  - Hidden Markov Models and the key algorithms

The Topics

- Deep Learning (Alina Miron)
  - Image Analysis, Convolutional Neural Networks
  - NLP: Recurrent Neural Networks
- Philosophy:
  - What has been easy and what has been hard (examples)
  - Language / consciousness: Searle's Chinese room
  - Impacts on Society: Ethics, black box, trolley problem

Thanks for Listening

- After this lecture – your last chance to have lab sheets assessed (if you haven't already!)

- Revision Lab / Seminar in Term 2 – TBC
- Opportunity to talk to Loebner Prize winning AI