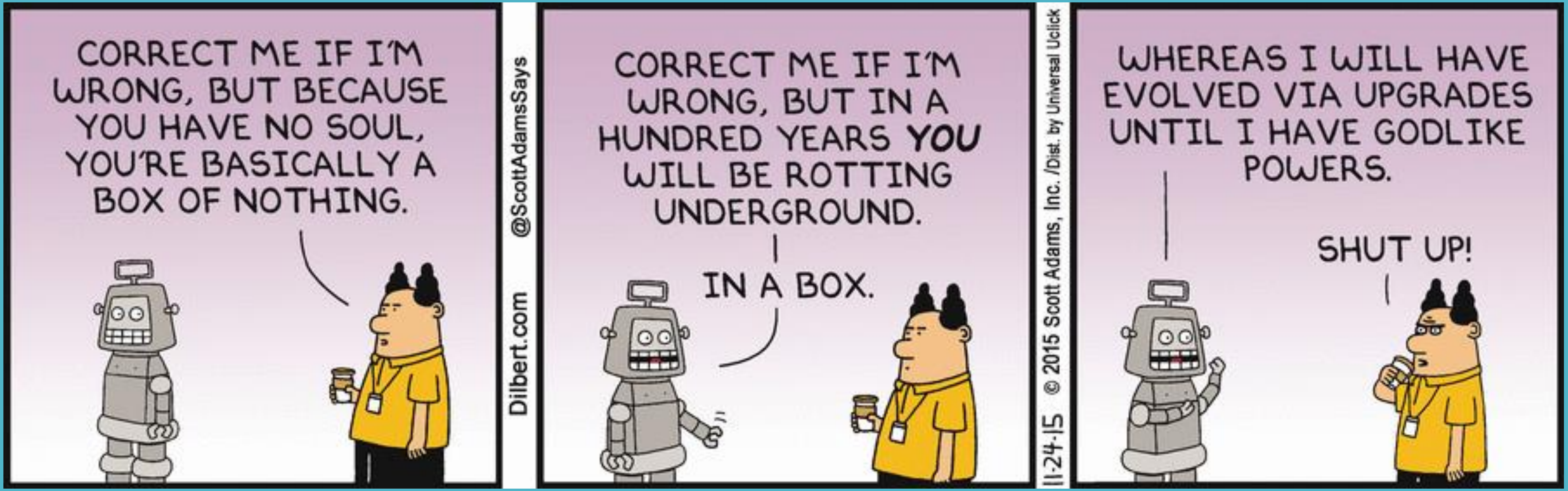
Seminars in AI & Robotics: Social Robotics

Mary Ellen Foster

MaryEllen.Foster@glasgow.ac.uk

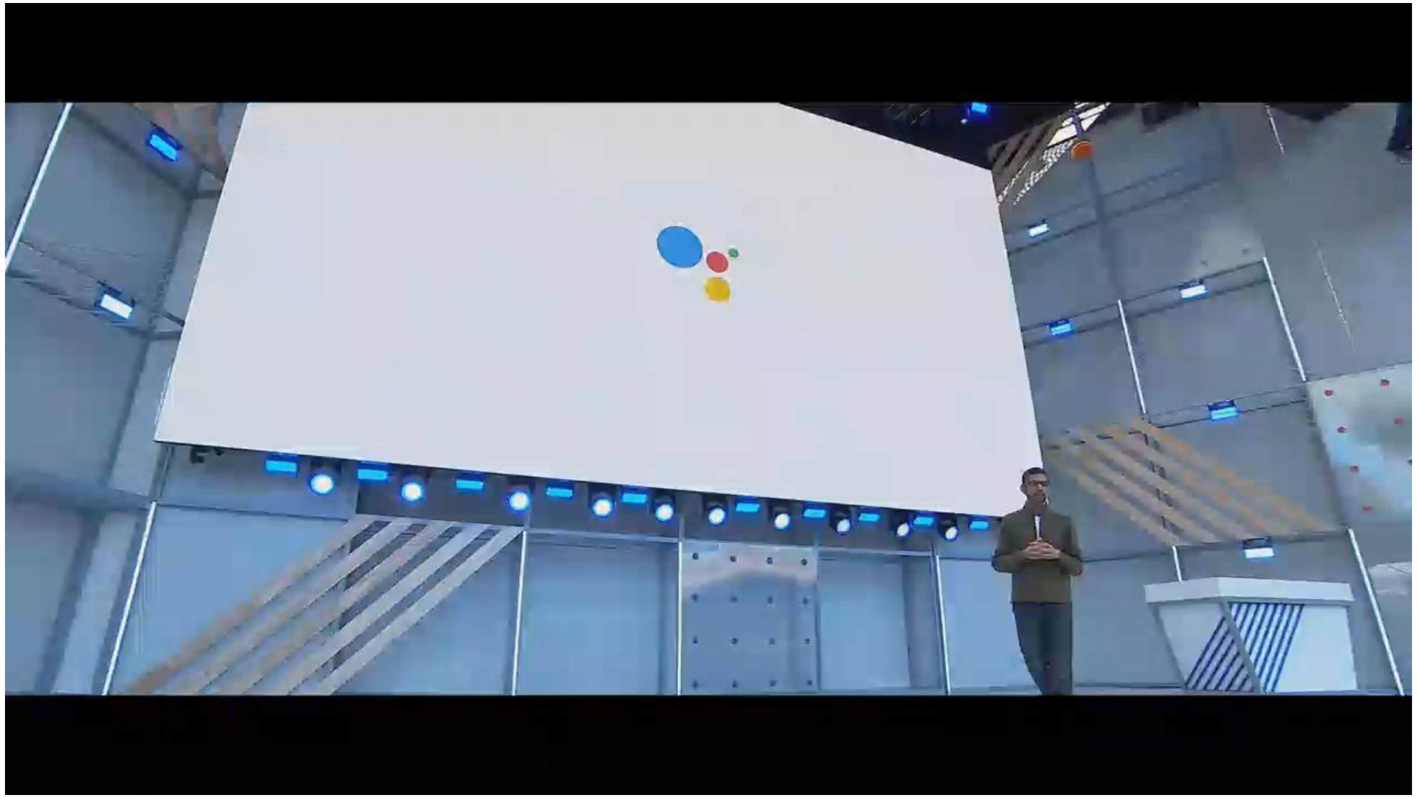
15 May 2018


CORRECT ME IF I'M WRONG, BUT BECAUSE YOU HAVE NO SOUL, YOU'RE BASICALLY A BOX OF NOTHING.
Dilbert.com @ScottAdamsSays
CORRECT ME IF I'M WRONG, BUT IN A HUNDRED YEARS YOU WILL BE ROTTING UNDERGROUND.
IN A BOX.
11-24-15 © 2015 Scott Adams, Inc. /Dist. by Universal Uclick
WHEREAS I WILL HAVE EVOLVED VIA UPGRADES UNTIL I HAVE GODLIKE POWERS.
SHUT UP!

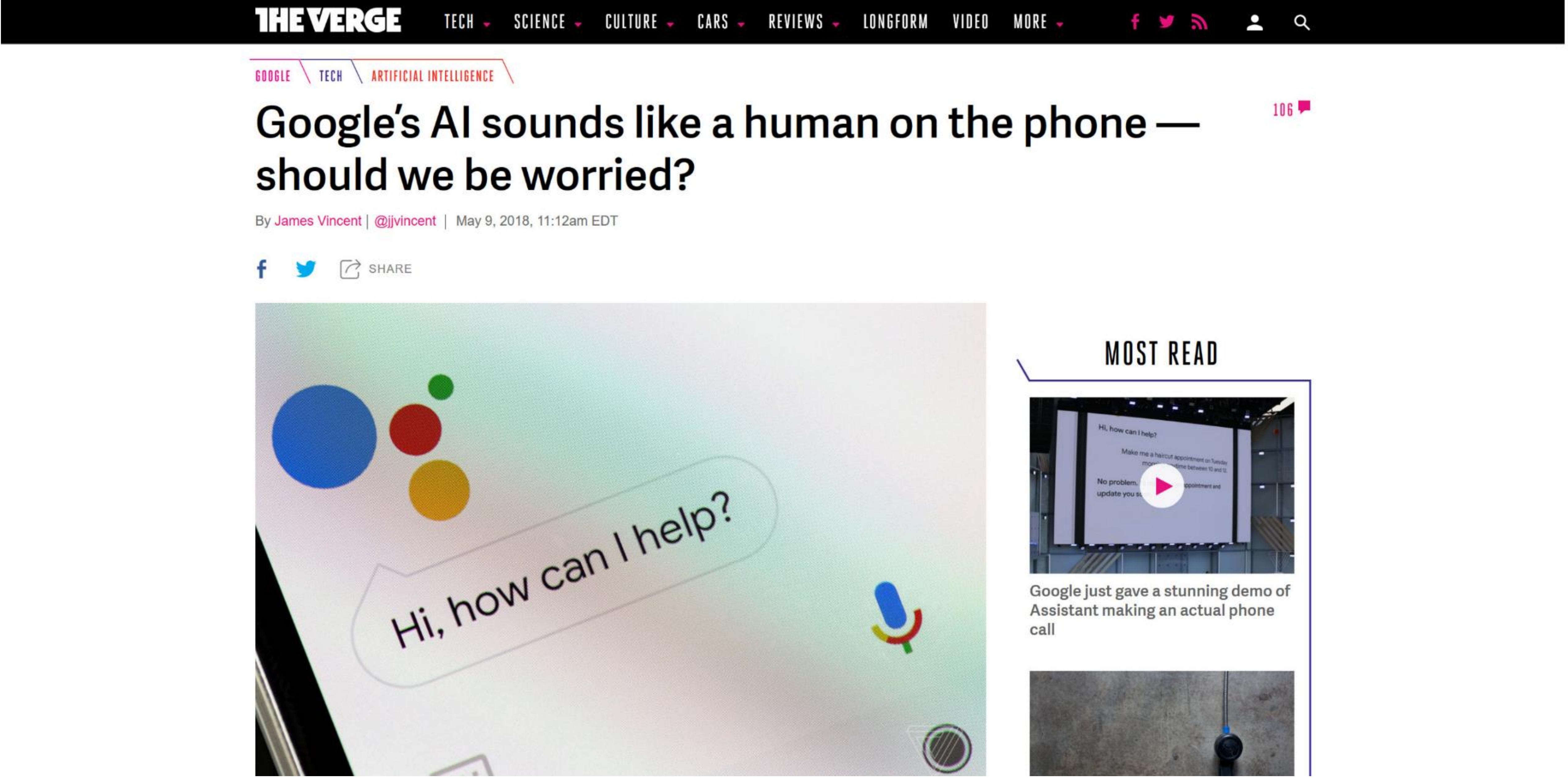
http://dilbert.com/strip/2015-11-24

Today's topic: Societal and Ethical Implications

Google keynote speech, 8 May 2018



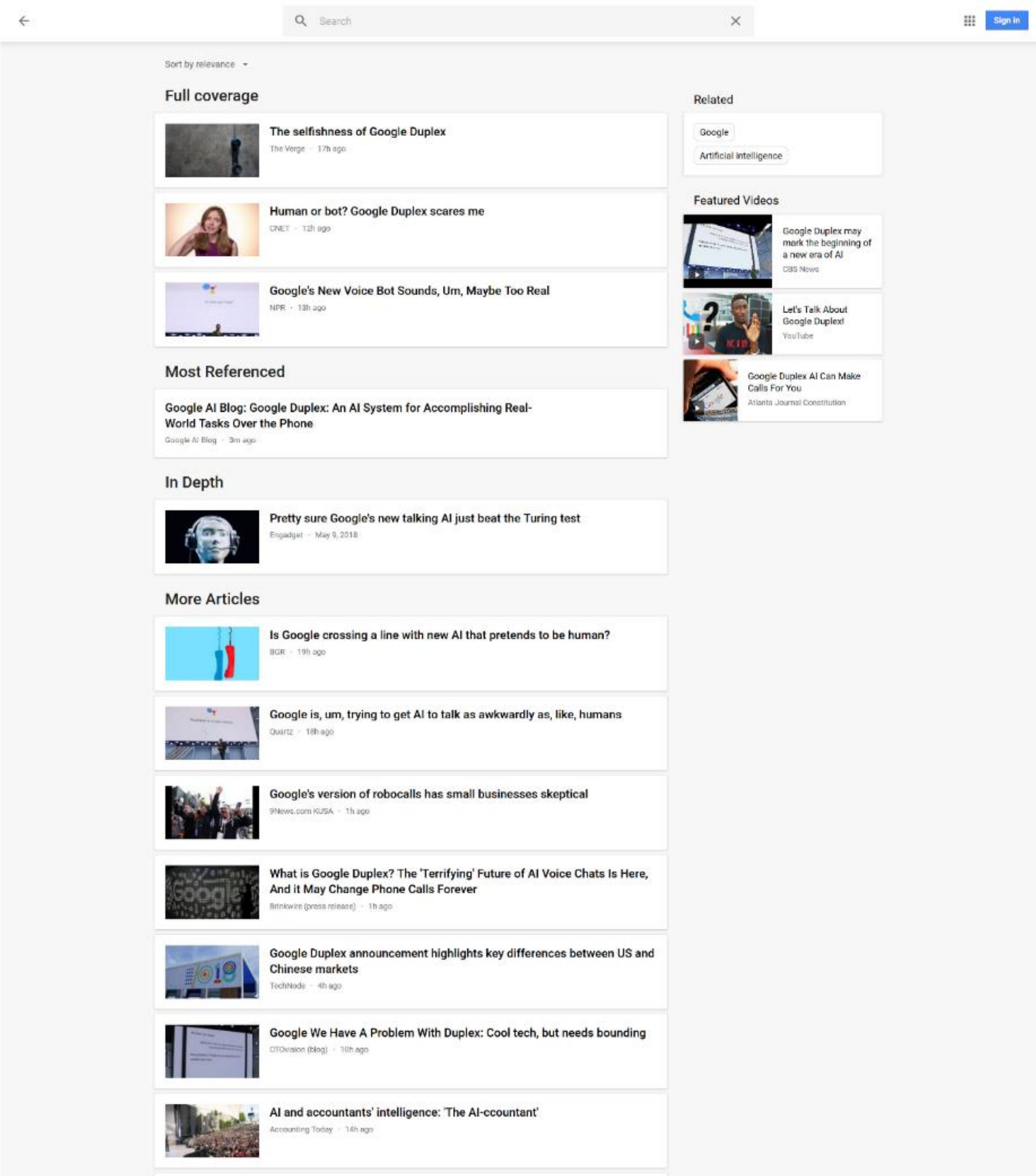
https://youtu.be/pKVppdt_-B4



Google's AI sounds like a human on the phone — should we be worried?

By James Vincent | @jjvincent | May 9, 2018, 11:12am EDT

Google just gave a stunning demo of Assistant making an actual phone call



Sort by relevance

Full coverage

The selfishness of Google Duplex
The Verge · 17h ago

Human or bot? Google Duplex scares me
CNET · 12h ago

Google's New Voice Bot Sounds, Um, Maybe Too Real
NPR · 13h ago

Most Referenced

Google AI Blog: Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone
Google AI Blog · 3m ago

In Depth

Pretty sure Google's new talking AI just beat the Turing test
Engadget · May 9, 2018

More Articles

Is Google crossing a line with new AI that pretends to be human?
BGR · 19h ago

Google is, um, trying to get AI to talk as awkwardly as, like, humans
Quartz · 18h ago

Google's version of robocalls has small businesses skeptical
9News.com KUSA · 1h ago

What is Google Duplex? The 'Terrifying' Future of AI Voice Chats Is Here, And it May Change Phone Calls Forever
Brinkwire (press release) · 1h ago

Google Duplex announcement highlights key differences between US and Chinese markets
TechNode · 4h ago

Google We Have A Problem With Duplex: Cool tech, but needs bounding
CTOvision (blog) · 10h ago

AI and accountants' intelligence: 'The AI-ccountant'
Accounting Today · 14h ago

Related

Google

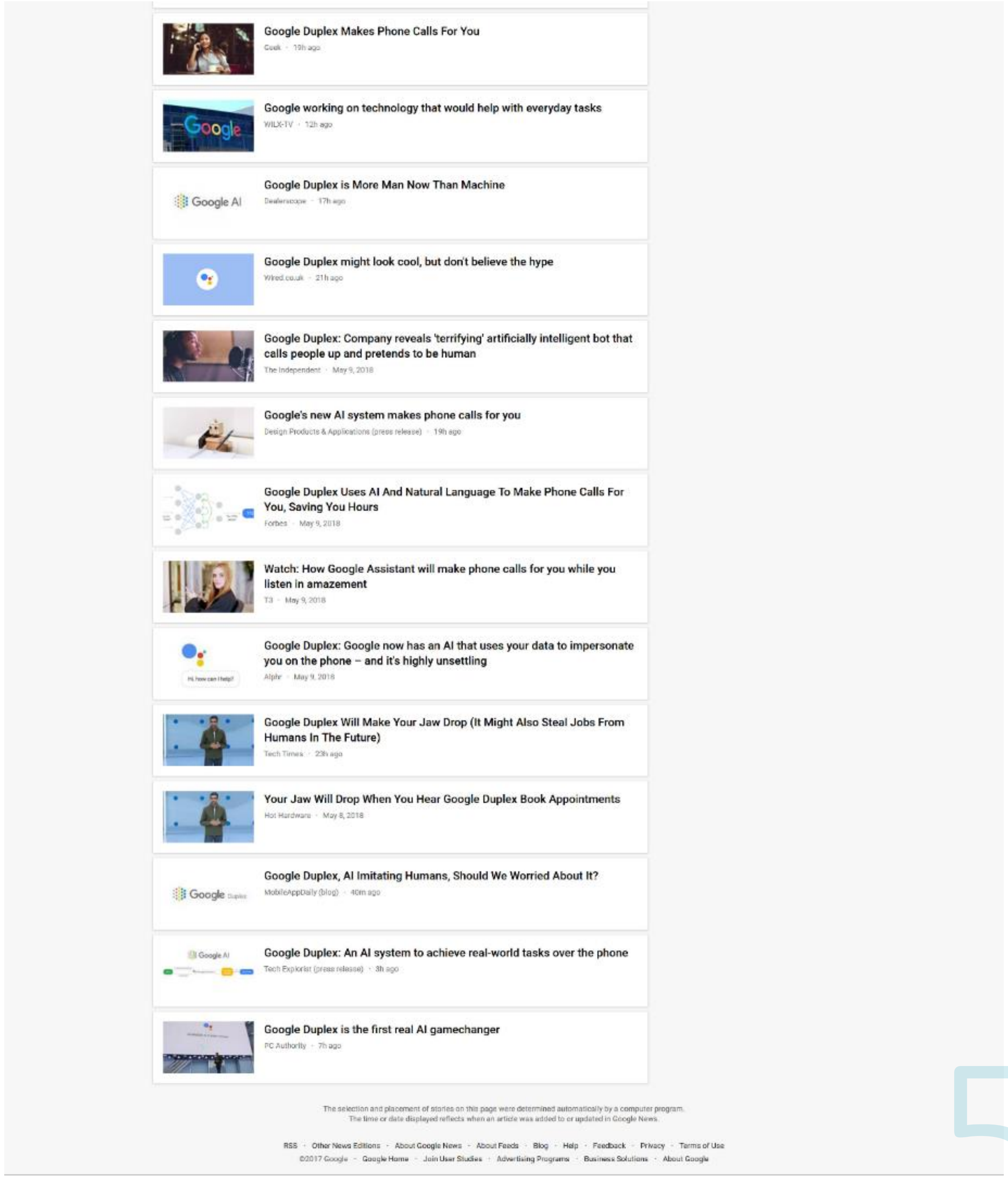
Artificial intelligence

Featured Videos

Google Duplex may mark the beginning of a new era of AI
CBS News

Let's Talk About Google Duplex!
YouTube

Google Duplex AI Can Make Calls For You
Atlanta Journal Constitution



Google Duplex Makes Phone Calls For You
Geek · 19h ago

Google working on technology that would help with everyday tasks
WILX-TV · 12h ago

Google Duplex is More Man Now Than Machine
Dealerscope · 17h ago

Google Duplex might look cool, but don't believe the hype
Wired.co.uk · 21h ago

Google Duplex: Company reveals 'terrifying' artificially intelligent bot that calls people up and pretends to be human
The Independent · May 9, 2018

Google's new AI system makes phone calls for you
Design Products & Applications (press release) · 19h ago

Google Duplex Uses AI And Natural Language To Make Phone Calls For You, Saving You Hours
Forbes · May 9, 2018

Watch: How Google Assistant will make phone calls for you while you listen in amazement
T3 · May 9, 2018

Google Duplex: Google now has an AI that uses your data to impersonate you on the phone – and it's highly unsettling
Alphr · May 9, 2018

Google Duplex Will Make Your Jaw Drop (It Might Also Steal Jobs From Humans In The Future)
Tech Times · 23h ago

Your Jaw Will Drop When You Hear Google Duplex Book Appointments
Hot Hardware · May 8, 2018

Google Duplex, AI Imitating Humans, Should We Worried About It?
MobileAppDaily (blog) · 40m ago

Google Duplex: An AI system to achieve real-world tasks over the phone
Tech Explorist (press release) · 3h ago

Google Duplex is the first real AI gamechanger
PC Authority · 7h ago

The selection and placement of stories on this page were determined automatically by a computer program.
The time or date displayed reflects when an article was added to or updated in Google News.

RSS · Other News Editions · About Google News · About Feeds · Blog · Help · Feedback · Privacy · Terms of Use
©2017 Google · Google Home · Join User Studies · Advertising Programs · Business Solutions · About Google

Google statement, 11 May 2018

“We understand and value the discussion around Google Duplex — as we’ve said from the beginning, transparency in the technology is important. We are designing this feature with disclosure built-in and we’ll make sure the system is appropriately identified. What we showed at I/O was an early technology demo and we look forward to incorporating feedback as we develop this into a product.”

Some relevant people in this area

Benjamin Kuipers, University of Michigan

http://web.eecs.umich.edu/~kuipers/

Toby Walsh, UNSW, Sydney, Australia

http://www.cse.unsw.edu.au/~tw/

Joanna Bryson, University of Bath, UK

http://www.cs.bath.ac.uk/~jjb/

Ethical overview

Source: Burton, E., Goldsmith, J., Koenig, S., Kuipers, B., Mattei, N. and Walsh, T. Ethical considerations in artificial intelligence courses. *AI Magazine*, Summer 2017; arxiv:1701.07769.

Trolley problems (The Good Life, Netflix)

https://youtu.be/lDnO4nDA3kM

Ethical problems posed by AI/robots

How should robots behave in our society?

What should we do if jobs are in short supply?

Should AI systems/robots be allowed to kill?

Should we worry about “superintelligence” and the “singularity”?

How should we treat robots?

Ethical approaches: Deontology

Summary: ethics is about following moral law

Basic question: “what is my duty?”

Combines well with popular and technical understandings of how a machine should behave (e.g., Asimov’s “three laws of robotics”)

Underlying questions:

- How are rules applied to decisions?
- What are the right rules?

Asimov’s three laws

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

0. A robot may not harm humanity or, by inaction, allow humanity to come to harm.

EPSRC Principles of Robotics

1. Robots are multi-use tools. Robots should not be designed solely or primarily to kill or harm humans, except in the interests of national security.

   Robots should not be designed as weapons, except for national security reasons.

2. Humans, not robots, are responsible agents. Robots should be designed; operated as far as is practicable to comply with existing laws & fundamental rights & freedoms, including privacy.

   Robots should be designed and operated to comply with existing law, including privacy.

3. Robots are products. They should be designed using processes which assure their safety and security.

   Robots are products: as with other products, they should be designed to be safe and secure.

EPSRC Principles of Robotics (cont'd)

4. Robots are manufactured artefacts. They should not be designed in a deceptive way to exploit vulnerable users; instead their machine nature should be transparent.

   Robots are manufactured artefacts: the illusion of emotions and intent should not be used to exploit vulnerable users.

5. The person with legal responsibility for a robot should be attributed.

   It should be possible to find out who is responsible for any robot.

Ethical approaches: Utilitarianism

Basic question: “what is the greatest possible good for the greatest number?”

Underlying assumption: **utility** can be quantified as a mixture of happiness and other qualities

- Utility of different individuals can be compared

Classic utilitarian calculus does not consider probabilities – however, expected utility (i.e., decision-theoretic planning) fits well into framework

Mathematical analogue: game theory

- Every agent is a rational utility maximiser

Ethical approaches: virtue ethics

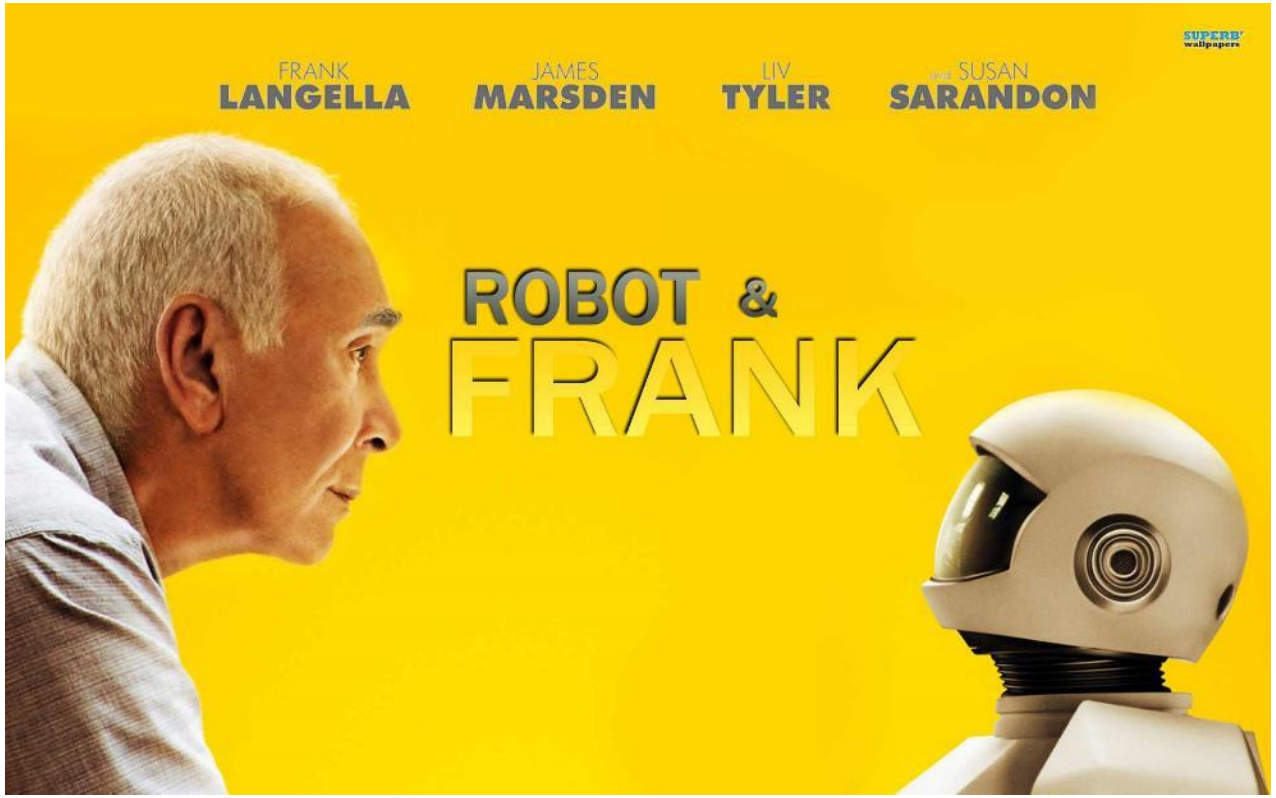
Basic question: “what should I be?”

Organised around developing habits and dispositions that help a person achieve their goals and to flourish as an individual

Contrast to deontological: considers ethics in **local** rather than **universal** terms

Dominant mode of ethics through 17th century – replaced by other approaches more recently

Ethical case study: Robot and Frank



Robot and Frank (1): Walking in the woods

https://youtu.be/eQxUW4B622E

Robot and Frank (2): Eating healthy

https://youtu.be/3yXwPfvvIt4

Robot and Frank (3): In the shop

https://youtu.be/xlpeRIG18TA

Ethical issues in Robot and Frank

Frank’s health is Robot’s top priority, superseding all other considerations

Robot’s goal is to find a long-term activity to help keep Frank mentally engaged and physically active – preparing for and carrying out robberies.

Robot and Frank develop a friendship – Robot is not a human, but Frank — and through him, the audience — come to regard him as if he were.

Ending: Robot persuades Frank to wipe his memory. Even though Robot has made it clear that he is untroubled by his own “death,” Frank has essentially killed his friend. What are the moral ramifications of this?

Robot and Frank and deontology

Robot is guided solely by duty to Frank – puts deontology at the centre

Local rather than universal guiding laws (robbery!)

Question: can a “carebot” function (caring for its assigned person) without violating other societal norms?

- Story suggests the design choice is not always straightforward

Robot and Frank and virtue ethics

- Robot makes choices according to its own personal goals – caring for Frank
  - Different ethical theory than those who build robots might expect
- Robot lacks ability to make nuanced judgement about how to act
  - Reasoning ability not sufficient to make socially responsible ethical judgements (either unaware of social harm caused by stealing, or else prioritises Frank's welfare)
- Robot is untroubled by its own destruction
  - Virtue ethics assume humans are concerned for own welfare and success
  - If an artificial agent is not concerned, how can it be evaluated?

Robot and Frank and utilitarianism

Why should Frank's criminal tendencies be understood as ethically wrong?

- If we don't steal, everyone (as a society) is better off

Robot and Frank have little concern for long-term social consequences

What is an ethical design of an eldercare robot anyway?

- Should it have pre-programmed ethics? Or should humans guide its reasoning?

Questions raised by Robot and Frank

If an elderly person wishes to behave in ways that violate common social norms, should a caretaker robot intervene, and if so, how?

If the elderly person seriously wants to die, should the robot help them to die?

If the elderly person asks the robot to help make preparations for taking his/her own life, does the robot have an obligation to inform other family members?

If the elderly person wants to walk around the house, in spite of some risk of falling, should the robot prevent it?

Extrapolating into other domains, a caretaker robot for a child raises many additional issues, since a child needs to be taught how to behave in society as well, and a child's instructions need not be followed, for a variety of different reasons.

Skynet

https://www.youtube.com/watch?v=4DQsG3TKQ0I

Skynet transcript

SC: I need to know how SkyNet gets built. Who's responsible?

T2: The man most directly responsible is Miles Bennett Dyson.

SC: Who is that?

T2: The Director of Special Projects at Cyberdyne Systems Corporation.

SC: Why him?

T2: In a few months, he creates a revolutionary type of microprocessor.

SC: Go on. Then what?

T2: In three years, Cyberdyne will become the largest supplier of military computer systems. All stealth bombers are upgraded with Cyberdyne computers, becoming fully unmanned. Afterwards, they fly with a perfect operational record. The SkyNet Funding Bill is passed. The system goes online on August 4th, 1997. Human decisions are removed from strategic defense. SkyNet begins to learn at a geometric rate. It becomes self-aware at 2:14 am Eastern time, August 29th. In a panic, they try to pull the plug.

SC: SkyNet fights back.

T2: Yes. It launches its missiles against their targets in Russia.

JC: Why attack Russia? Aren't they our friends now?

T2: Because SkyNet knows that the Russian counter-attack will eliminate its enemies over here.

SC: Jesus!

Approaches to robot ethics

1. “How do we design AI systems so that they function ethically?”

2. “How do we act ethically as programmers and system designers, to decrease the risks that our systems and code will act unethically?”

Actors involved in Skynet:

- Initial clients – provided vague specifications
- Knowledge engineers – translate into technical specifications
- Managers, programmers, testers – implementation
- Legislators and regulators – constrain specification, possibly after the fact
- Engineers – install system
- Politicians and bureaucrats – decide how to run system

Questions to ask (consider also real-world out-of-control AI, e.g., trading)

Was it rational to deploy SkyNet? It is worth considering that, its initial phase of implementation, it performed with a perfect operational record.

Was it necessary to make SkyNet a learning system? What might have made this seem like a good or necessary choice?

What is "self-awareness," that it scared its creators so much that they tried to turn SkyNet off? Could this have been avoided? Or would SkyNet almost certainly have reached some other capability that scared the human creators?

As a critical part of the national defense system, was it reasonable for SkyNet to fight back against all perceived threats to its existence?

SkyNet found an solution to its problem that its designers did not anticipate. What sorts of constraints could have prevented it from discovering or using that solution?

Skynet and ethics

Deontology: rules have unintended consequences – even Asimov's laws might not have prevented Skynet

- Could any set of rules have allowed Skynet to control nuclear arsenal and not result in these consequences?

Utilitarianism: right action is the one that results in best for everyone – but who is "everyone"?

- Almost any definition would prevent nuclear attach – but what about Mutually Assured Destruction (nuclear policy since the 1940s)

Virtue ethics: Skynet is clearly not following moral norms

Additional questions

Under what conditions should humans trust an AI system?

What criteria might human creators use to determine how much power to entrust to a given AI?

How can an AI system show that it is trustworthy?

Benjamin Kuipers. 2018. How can we trust a robot?. https://doi.org/10.1145/3173087

https://youtu.be/kiTXph8-WiY

What is trust for?

“Trust is necessary for successful cooperation. And morality and ethics (and other social norms) are mechanisms by which a society encourages trustworthy behavior by its individual members.”

Trust enables cooperation – division of labour, sharing of expenses, reduction of risk

Example: driving on the roads

- Early days: everyone could drive anywhere – frequent accidents even with caution
- Social norm encoded (drive on the left/right) – safer and more efficient

Robots should follow social norms in order to participate in society and earn trust

Making robots trustworthy

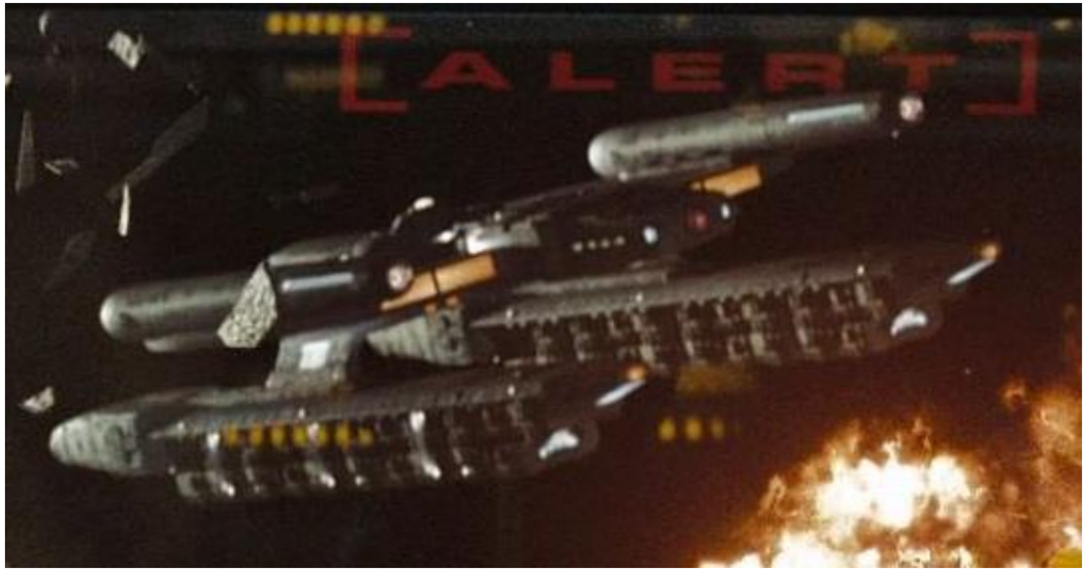
"The complexity of the world suggests the only way to acquire adequately complex decision criteria is through learning." – that's how humans do it!

Formalise ethical criteria (deontology, utilitarianism, virtue ethics) – best option may be a hybrid approach that combines aspect of all of them to decision making

One approach: case-based reasoning

Include deliberation about the consequences of action to improve decision making next time around

“The Deadly Dilemma”



Trolley problem for a self-driving car

Not incredibly realistic – more often, “near miss” scenarios where the agent can learn to avoid the situation

Reflect on near miss to improve decision making for next time

Earning trust for a self-driving car: showing that behaviour follows social norms (including politeness), and that decision making in near-miss situations is reliable

Toby Walsh on the “singularity”

Based on: The Singularity May Never Be Near (pdf). Toby Walsh. *AI Magazine* 38(3): 58-62, 2017.

Definition: “the technological singularity is the point in time at which we build a machine of sufficient intelligence that is able to redesign itself to improve its intelligence, and at which its intelligence starts to grow exponentially fast, quickly exceeding human intelligence by orders of magnitude.” (e.g., Skynet)

“Within thirty years, we will have the technological means to create superhuman intelligence. Shortly after, the human era will be ended.” – Vernor Vinge

More recent: Ray Kurzweil considers that we will reach “technological singularity” by 2045 – also Stephen Hawking, Bill Gates, Elon Musk, …

Arguments against the singularity

The “fast thinking dog” argument – intelligence is not only about processing power

The “anthropocentric” argument – why should human intelligence be a tipping point? What is so special about humans anyway?

The “meta-intelligence” argument – we should not confuse intelligence to do a task, with the capability to improve that intelligence

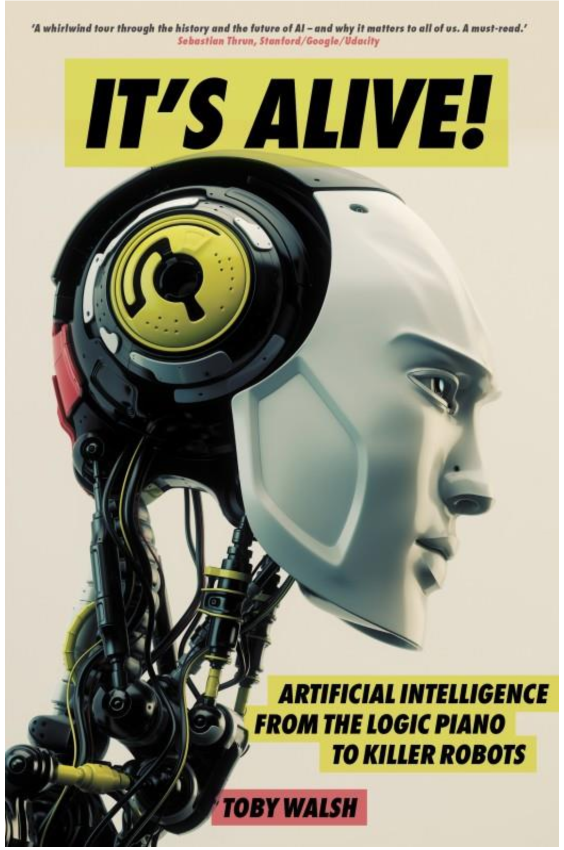
The “diminishing returns” argument – we are running out of low-hanging fruit

The “limits of intelligence” argument – intelligence may be ultimately limited by physics

The “computational complexity” argument – exponential growth cannot support super-exponential algorithms

Toby Walsh: 10 ways AI will change society by 2050

Source: It's Alive!: Artificial Intelligence from the Logic Piano to Killer Robots, Black Inc, Australia, 2017.


'A whirlwind tour through the history and the future of AI – and why it matters to all of us. A must-read.'
Sebastian Thrun, Stanford/Google/Udacity
IT'S ALIVE!
ARTIFICIAL INTELLIGENCE
FROM THE LOGIC PIANO
TO KILLER ROBOTS
TOBY WALSH

Predictions for 2050

1. Autonomous cars will replace manually driven cars
2. Wearable technology will support continuous health monitoring
3. Virtual characters can be programmed to talk and act like anyone
4. AI will make more decisions about day-to-day activities (including hiring/firing?)
5. Smart rooms/Internet of Things will be used all over – privacy issues!
6. AI warfare – e.g., cyber-crime with AI to defend against it
7. Robot football team will beat human players
8. Autonomous transport vehicles (cargo trains, ships, etc)
9. TV news will be created without humans – assembling stories, presenting, personalised
10. Digital doubles will allow people to live on after death (Black Mirror …)

Meet “Janet” (The Good Place, Netflix)

https://youtu.be/gaqUzyjN8M8

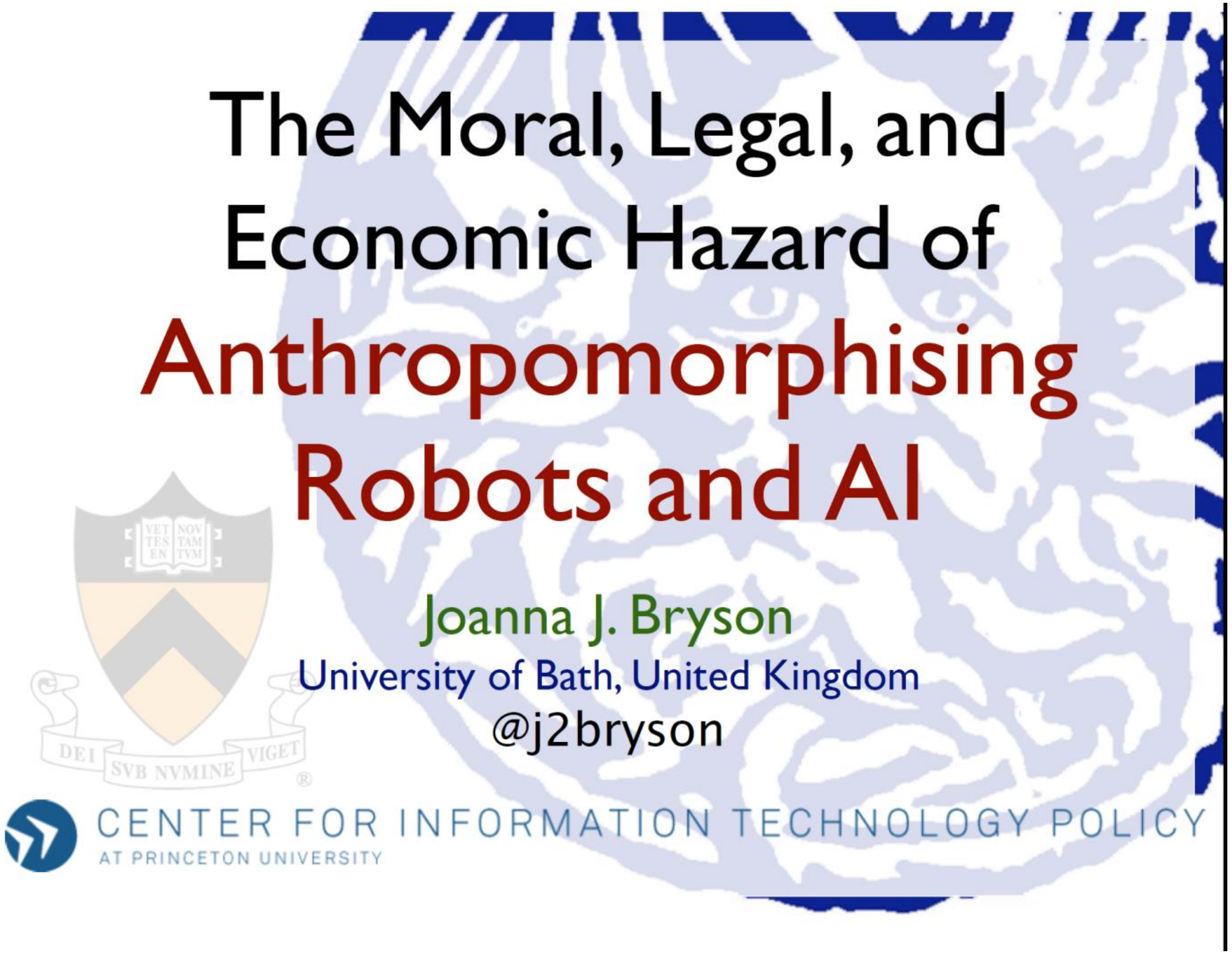
Deactivating Janet

https://youtu.be/etJ6RmMPGko

Joanna Bryson's position

Keynote talk presented at **Robot-Philosophy 2018** at the University of Vienna

http://conferences.au.dk/robo-philosophy-2018-at-the-university-of-vienna/keynotes/joanna-bryson/abstrac joanna-bryson/



The Moral, Legal, and Economic Hazard of Anthropomorphising Robots and AI

Joanna J. Bryson

University of Bath, United Kingdom

@j2bryson

CENTER FOR INFORMATION TECHNOLOGY POLICY

AT PRINCETON UNIVERSITY

DEI SVB NVMINE VIGET
VET TES TAM EN NOV TVM

Slides downloaded from http://www.cs.bath.ac.uk/~jjb/ftp/Bryson%20RoboPhil%20Wien%202018%20Lacuna.key.pdf