

How should society prepare for advances in Artificial Intelligence?

Alistair Knott

Dept of Computer Science, University of Otago
Soul Machines, Auckland



AI evangelists

AI evangelists



AI evangelists

- AI machines will achieve human-level intelligence in 2029, 'and will openly petition for recognition of this fact'.



AI evangelists

- AI machines will achieve human-level intelligence in 2029, 'and will openly petition for recognition of this fact'.
- Uploading human minds to machines will become possible in the 2030s, ushering in an era of human immortality.



AI prophets of doom

- The development of full artificial intelligence 'could be the greatest disaster in human history'...



AI prophets of doom

- The development of full artificial intelligence 'could be the greatest disaster in human history'...
- It could 'spell the end of the human race'.



AI prophets of doom

- 'Before the prospect of an intelligence explosion, we humans are like small children playing with a bomb. Such is the mismatch between the power of our plaything and the immaturity of our conduct.'



Evangelists

- 'I am optimistic. People who are naysayers and try to drum up these doomsday scenarios—I just don't understand it. It's really negative and in some ways I actually think it is pretty irresponsible.'



Prophets of doom

- 'I've talked to Mark about this. His understanding of the subject is limited.'



Outline of the talk

Outline of the talk

- 1 What can AI do right now?
- 2 What are the main impacts AI will have in the immediate future?
- 3 What aspects of near-future AI might need regulation?
- 4 How should our society prepare itself for the arrival of AI?

1. What can AI do right now?

Agile robots



Agile robots



Self-driving vehicles



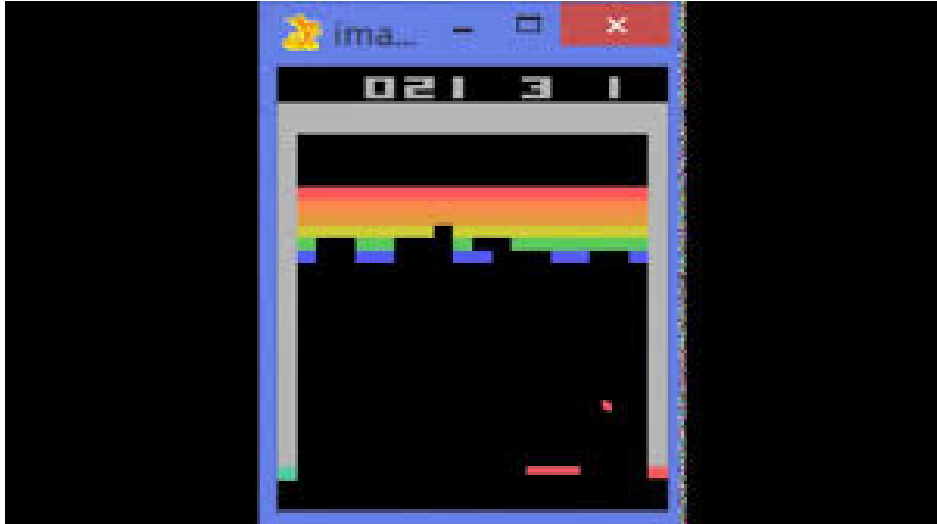
Self-driving vehicles



Open-domain question answering



Domain-general action learning



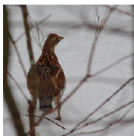
Visual object recognition



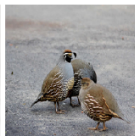
flamingo



cock



ruffed grouse



quail



partridge

...



Egyptian cat



Persian cat



Siamese cat



tabby



lynx

...



dalmatian



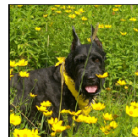
keeshond



miniature schnauzer



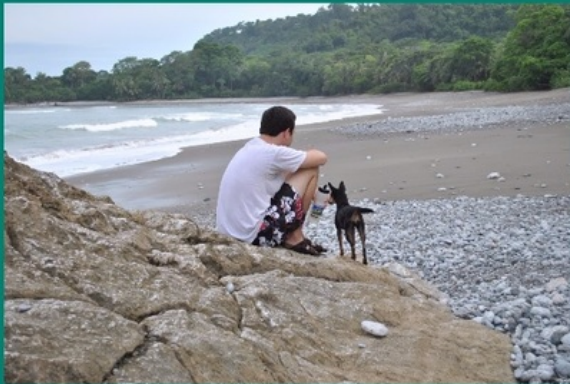
standard schnauzer



giant schnauzer

Scene description

I think it's a man and a dog on a rocky beach.



Machine translation



TranslateTurn off instant translation

EnglishSpanishFrenchEnglish - detected ▾

↔

EnglishSpanishArabic ▾

Translate

The quick brown fox jumped
over the lazy dog

44/5000

قفز الثعلب البني السريع فوق الكلب الكسول

qafaz althuelib albaniu alsarie fawq alkalb alkusul

2. What impacts will AI have in the immediate future?

2. What impacts will AI have in the immediate future?

I'll focus on **employment**.

An influential study: Frey and Osborne (2013)

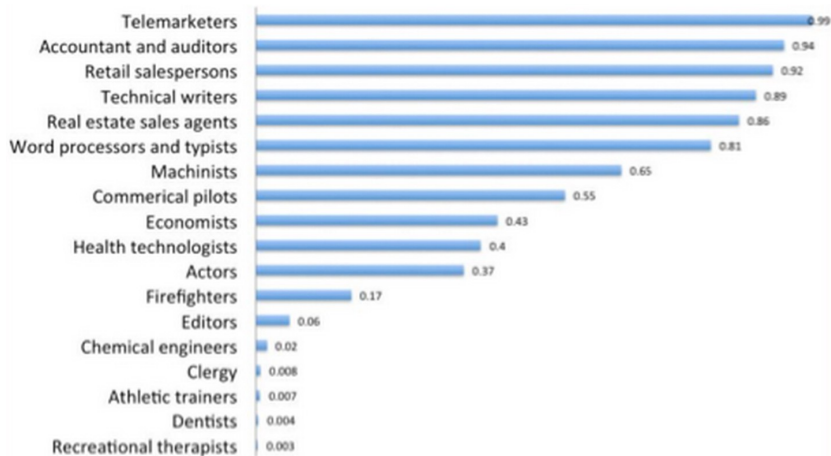
Frey & Osborne analysed a US database listing the 'key skills' needed for different job types. They estimated how 'automatable' each skill is.

An influential study: Frey and Osborne (2013)

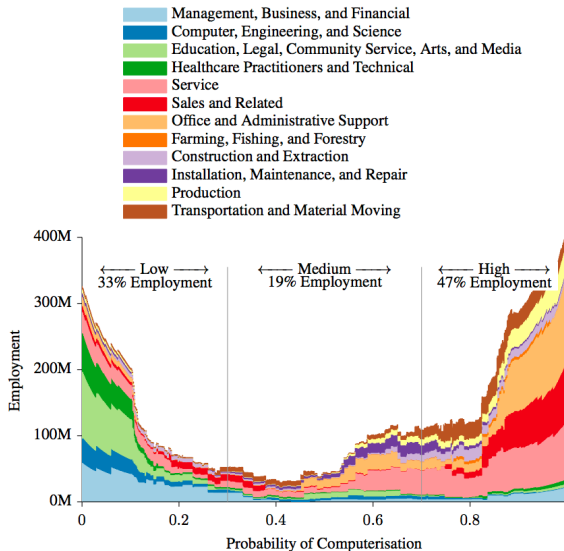
The main non-automatable skills are those involving perception/manipulation, creative intelligence, and social intelligence.

An influential study: Frey and Osborne (2013)

Then they estimated the probability of automation for each job type. Here are some examples:



A breakdown of F&O's estimates by job category



Some caveats

Many economists think that Frey and Osborne overestimate how automatable jobs are.

Some caveats

Many economists think that Frey and Osborne overestimate how automatable jobs are.

A common idea (see e.g. Arntz *et al.*, 2016):

- Most jobs have *some tasks* that are easy to automate, and *some tasks* that are hard.
- In this case, what we might see is a *restructuring* of jobs, that allocates the automatable tasks to computers.

But don't new technologies always create new jobs?

But don't new technologies always create new jobs?



But don't new technologies always create new jobs?



‘Don’t worry, there will always be jobs for horses.’

3. What aspects of AI might need regulation?

3. What aspects of AI might need regulation?

I'll touch on five topics:

- Employment
- Machine bias
- Transparency
- Accountability
- Ethics

Regulating **employment**

Regulating **employment**

Some key questions:

Regulating employment

Some key questions:

Are there certain **jobs we don't want computers to do?**

- E.g. teacher, carer, soldier. . .

Regulating employment

Some key questions:

Are there certain **jobs we don't want computers to do?**

- E.g. teacher, carer, soldier. . .

How do we revise **social security**, for human workers displaced by AI?

Regulating employment

Some key questions:

Are there certain **jobs we don't want computers to do?**

- E.g. teacher, carer, soldier. . .

How do we revise **social security**, for human workers displaced by AI?

- We could impose some kind of tax on AI systems. . .

Regulating employment

Some key questions:

Are there certain **jobs we don't want computers to do?**

- E.g. teacher, carer, soldier. . .

How do we revise **social security**, for human workers displaced by AI?

- We could impose some kind of tax on AI systems. . .
- We could institute some form of universal basic income. . .

Regulating employment

Some key questions:

Are there certain **jobs we don't want computers to do?**

- E.g. teacher, carer, soldier. . .

How do we revise **social security**, for human workers displaced by AI?

- We could impose some kind of tax on AI systems. . .
- We could institute some form of universal basic income. . .

Do we want a world where **people don't have to work?**

Regulating **machine bias**

Regulating **machine bias**

Some examples of bias in AI systems:

Regulating machine bias

Some examples of bias in AI systems:

- AI systems can learn word meanings from huge text corpora. These systems show all the biases exhibited by humans in word-association tests. (Caliskan *et al.*, 2017)

Regulating machine bias

Some examples of bias in AI systems:

- AI systems can learn word meanings from huge text corpora. These systems show all the biases exhibited by humans in word-association tests. (Caliskan *et al.*, 2017)

Phil - kitchen vs Sally - kitchen

Regulating machine bias

Some examples of bias in AI systems:

- AI systems are currently used in US courts to help judges make bail/parole decisions. These systems have been shown to have biases against black people. (Angwin *et al.*, 2016)

Regulating machine bias

Some examples of bias in AI systems:

- AI systems are currently used in US courts to help judges make bail/parole decisions. These systems have been shown to have biases against black people. (Angwin *et al.*, 2016)

Some open questions:

Regulating machine bias

Some examples of bias in AI systems:

- AI systems are currently used in US courts to help judges make bail/parole decisions. These systems have been shown to have biases against black people. (Angwin *et al.*, 2016)

Some open questions:

- How can we **test** AI systems for bias?

Regulating machine bias

Some examples of bias in AI systems:

- AI systems are currently used in US courts to help judges make bail/parole decisions. These systems have been shown to have biases against black people. (Angwin *et al.*, 2016)

Some open questions:

- How can we **test** AI systems for bias?
- Can we **legislate against bias** in AI systems?

Regulating machine bias

Some examples of bias in AI systems:

- AI systems are currently used in US courts to help judges make bail/parole decisions. These systems have been shown to have biases against black people. (Angwin *et al.*, 2016)

Some open questions:

- How can we **test** AI systems for bias?
- Can we **legislate against bias** in AI systems?
- Can we use AI systems to **reduce**, or **eliminate bias**?

Regulating **transparency**

Regulating transparency

Modern AI systems use **machine learning techniques** to make decisions:

- This involves consulting *huge databases*
- and performing *complex computations*.

Regulating transparency

Modern AI systems use **machine learning techniques** to make decisions:

- This involves consulting *huge databases*
- and performing *complex computations*.

Humans typically *can't understand* how these systems reach their conclusions.

Regulating transparency

Modern AI systems use **machine learning techniques** to make decisions:

- This involves consulting *huge databases*
- and performing *complex computations*.

Humans typically *can't understand* how these systems reach their conclusions.

Some questions:

Regulating transparency

Modern AI systems use **machine learning techniques** to make decisions:

- This involves consulting *huge databases*
- and performing *complex computations*.

Humans typically *can't understand* how these systems reach their conclusions.

Some questions:

- Should (some) AI systems be required to **explain their decisions**?

Regulating transparency

Modern AI systems use **machine learning techniques** to make decisions:

- This involves consulting *huge databases*
- and performing *complex computations*.

Humans typically *can't understand* how these systems reach their conclusions.

Some questions:

- Should (some) AI systems be required to **explain their decisions**?
- How can we **build** the 'explanation mechanism'?

Regulating **accountability**

Regulating **accountability**

Say a person is using a driverless car, and has an accident.

Who's to blame?

Regulating **accountability**

Say a person is using a driverless car, and has an accident.

Who's to blame?

- The person in the car?

Regulating **accountability**

Say a person is using a driverless car, and has an accident.

Who's to blame?

- The person in the car?
- The company who designed the AI system the car is using?

Regulating **accountability**

Say a person is using a driverless car, and has an accident.

Who's to blame?

- The person in the car?
- The company who designed the AI system the car is using?
- The AI system itself??

Regulating **accountability**

Say a person is using a driverless car, and has an accident.

Who's to blame?

- The person in the car?
- The company who designed the AI system the car is using?
- The AI system itself??

Some questions to consider:

Regulating **accountability**

Say a person is using a driverless car, and has an accident.

Who's to blame?

- The person in the car?
- The company who designed the AI system the car is using?
- The AI system itself??

Some questions to consider:

- AI systems are designed to behave very flexibly, in a wide range of circumstances. How can a company *guarantee the performance* of such systems?

Regulating **accountability**

Say a person is using a driverless car, and has an accident.

Who's to blame?

- The person in the car?
- The company who designed the AI system the car is using?
- The AI system itself??

Some questions to consider:

- AI systems are designed to behave very flexibly, in a wide range of circumstances. How can a company *guarantee the performance* of such systems?
- Say that driverless cars are more reliable than human drivers. How can companies be protected from crippling lawsuits?

Regulating the ethics of AI systems

Regulating the ethics of AI systems

Say we have an advanced AI system that can perform a wide range of tasks about the house. (Cooking, cleaning, tidying etc.)

Regulating the ethics of AI systems

Say we have an advanced AI system that can perform a wide range of tasks about the house. (Cooking, cleaning, tidying etc.)

- How can we ensure it doesn't do anything silly, or wrong?

Regulating the ethics of AI systems

Say we have an advanced AI system that can perform a wide range of tasks about the house. (Cooking, cleaning, tidying etc.)

- How can we ensure it doesn't do anything silly, or wrong?
- It may need **general principles** to regulate its behaviour.

Regulating the ethics of AI systems

Say we have an advanced AI system that can perform a wide range of tasks about the house. (Cooking, cleaning, tidying etc.)

- How can we ensure it doesn't do anything silly, or wrong?
- It may need **general principles** to regulate its behaviour.
- What principles do we give it?

Regulating the ethics of AI systems

Say we have an advanced AI system that can perform a wide range of tasks about the house. (Cooking, cleaning, tidying etc.)

- How can we ensure it doesn't do anything silly, or wrong?
- It may need **general principles** to regulate its behaviour.
- What principles do we give it?

If it's a system that *learns* how to behave, there are some interesting possibilities:

Regulating the ethics of AI systems

Say we have an advanced AI system that can perform a wide range of tasks about the house. (Cooking, cleaning, tidying etc.)

- How can we ensure it doesn't do anything silly, or wrong?
- It may need **general principles** to regulate its behaviour.
- What principles do we give it?

If it's a system that *learns* how to behave, there are some interesting possibilities:

- One possibility is that it should be taught in the same way that we **teach children**.

Regulating the ethics of AI systems

Say we have an advanced AI system that can perform a wide range of tasks about the house. (Cooking, cleaning, tidying etc.)

- How can we ensure it doesn't do anything silly, or wrong?
- It may need **general principles** to regulate its behaviour.
- What principles do we give it?

If it's a system that *learns* how to behave, there are some interesting possibilities:

- One possibility is that it should be taught in the same way that we **teach children**.
- However learning happens, it will probably involve AI systems **learning about human values**: both through explicit instruction, and through inference from human behaviour.

4. How should society prepare for advances in AI?

4. How should society prepare for advances in AI?

Let's think about New Zealand specifically. . .

(i) Interdisciplinary discussion groups

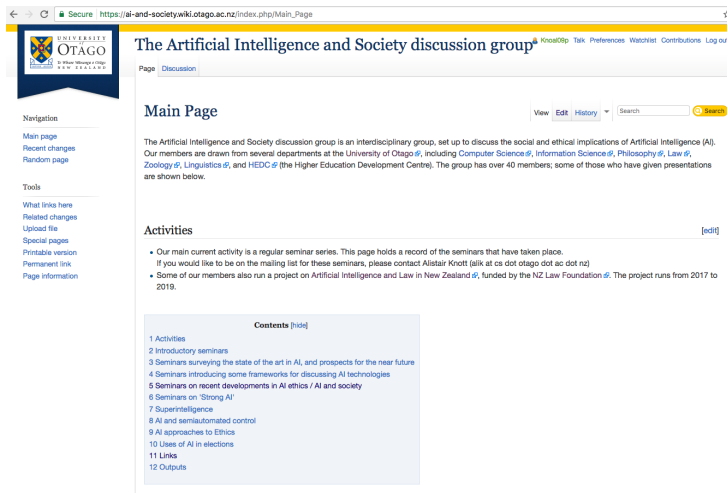
(i) Interdisciplinary discussion groups

The discussion about AI needs to cut across many disciplines.

- Technical AI people / computer scientists
- Business people
- Economists
- Lawyers
- Social scientists
- Ethicists
- ...

Some NZ groups I'm involved with

Otago's AI and Society Discussion Group



The screenshot shows a web browser displaying the 'Main Page' of 'The Artificial Intelligence and Society discussion group'. The browser's address bar shows the URL: https://ai-and-society.wiki.otago.ac.nz/index.php/Main_Page. The website features a navigation menu on the left with links like 'Main page', 'Recent changes', and 'Random page'. The main content area includes a description of the group as an interdisciplinary forum for discussing AI's social and ethical implications, listing member departments like Computer Science, Information Science, and Philosophy. It also has an 'Activities' section with a list of seminars and a 'Contents' table of contents.

The Artificial Intelligence and Society discussion group [Kno3DB](#) [Talk](#) [Preferences](#) [Watchlist](#) [Contributions](#) [Log out](#)

Page: [Discussion](#)

Main Page

[View](#) [Edit](#) [History](#) [Search](#)

The Artificial Intelligence and Society discussion group is an interdisciplinary group, set up to discuss the social and ethical implications of Artificial Intelligence (AI). Our members are drawn from several departments at the University of Otago [✎](#), including [Computer Science](#) [✎](#), [Information Science](#) [✎](#), [Philosophy](#) [✎](#), [Law](#) [✎](#), [Zoology](#) [✎](#), [Linguistics](#) [✎](#), and [HEDC](#) [✎](#) (the Higher Education Development Centre). The group has over 40 members; some of those who have given presentations are shown below.

Activities

[\[edit\]](#)

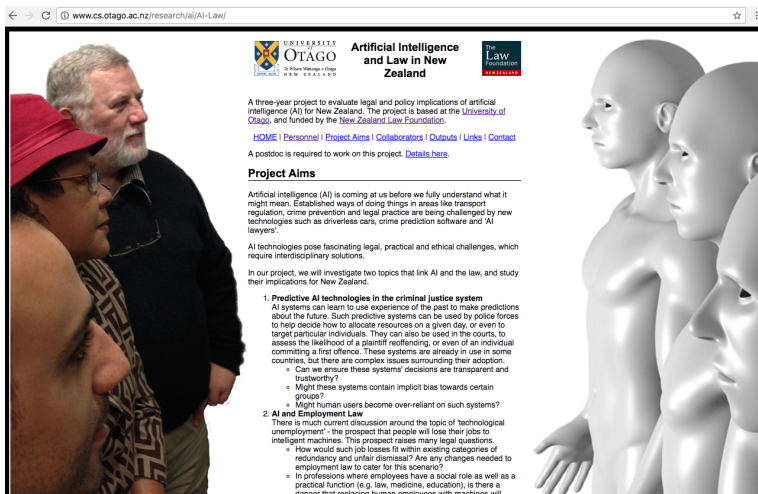
- Our main current activity is a regular seminar series. This page holds a record of the seminars that have taken place.
If you would like to be on the mailing list for these seminars, please contact Alistair Knott (alik at cs dot otago dot ac dot nz)
- Some of our members also run a project on Artificial Intelligence and Law in New Zealand [✎](#), funded by the NZ Law Foundation [✎](#). The project runs from 2017 to 2019.

Contents [\[hide\]](#)

- 1 Activities
- 2 Introductory seminars
- 3 Seminars surveying the state of the art in AI, and prospects for the near future
- 4 Seminars introducing some frameworks for discussing AI technologies
- 5 Seminars on recent developments in AI ethics / AI and society
- 6 Seminars on 'Strong AI'
- 7 Superintelligence
- 8 AI and semiautomated control
- 9 AI approaches to Ethics
- 10 Uses of AI in elections
- 11 Links
- 12 Outputs

Some NZ groups I'm involved with

An Otago project funded by the NZ Law foundation: **AI and Law In NZ**



← → ⓘ www.cs.otago.ac.nz/research/ai/[AI-Law] ☆ ⓘ

 **Artificial Intelligence and Law in New Zealand** 

A three-year project to evaluate legal and policy implications of artificial intelligence (AI) for New Zealand. The project is based at the [University of Otago](#), and funded by the [New Zealand Law Foundation](#).

[HOME](#) | [Personnel](#) | [Project Aims](#) | [Collaborators](#) | [Outputs](#) | [Links](#) | [Contact](#)

A postdoc is required to work on this project. [Details here.](#)

Project Aims

Artificial intelligence (AI) is coming at us before we fully understand what it might mean. Established ways of doing things in areas like transport regulation, crime prevention and legal practice are being challenged by new technologies such as driverless cars, crime prediction software and 'AI lawyers'.

AI technologies pose fascinating legal, practical and ethical challenges, which require interdisciplinary solutions.

In our project, we will investigate two topics that link AI and the law, and study their implications for New Zealand.

- Predictive AI technologies in the criminal justice system**

AI systems can learn to use experience of the past to make predictions about the future. Such predictive systems can be used by police forces to help decide how to allocate resources on a given day, or even to target particular individuals. They can also be used in the courts, to assess the likelihood of a plaintiff reoffending, or even of an individual committing a first offence. These systems are already in use in some countries, but there are complex issues surrounding their adoption.

 - Can we ensure these systems' decisions are transparent and trustworthy?
 - Might these systems contain implicit bias towards certain groups?
 - Might human users become over-reliant on such systems?
- AI and Employment Law**

There is much current discussion around the topic of 'technological unemployment' - the prospect that people will lose their jobs to intelligent machines. This prospect raises many legal questions.

 - How would such job losses fit within existing categories of redundancy and unfair dismissal? Are any changes needed to employment law to cater for this scenario?
 - In professions where employees have a social role as well as a practical function (e.g. law, medicine, education), is there a danger that replacing human employees with machines will

Some NZ groups I'm involved with

The AI Forum of New Zealand



(ii) Giving the next generation with the right skills

(ii) Giving the next generation with the right skills

Computer Science students need an understanding of the social consequences of AI. (And of IT more generally.)

(ii) Giving the next generation with the right skills

Computer Science students need an understanding of the social consequences of AI. (And of IT more generally.)

- Ethics should be an important (examinable) component of several CS courses.

(ii) Giving the next generation with the right skills

Computer Science students need an understanding of the social consequences of AI. (And of IT more generally.)

- Ethics should be an important (examinable) component of several CS courses.
- CS degrees should require social science papers.

(ii) Giving the next generation with the right skills

Computer Science students need an understanding of the social consequences of AI. (And of IT more generally.)

- Ethics should be an important (examinable) component of several CS courses.
- CS degrees should require social science papers.

Students planning on entering politics/law should be encouraged to become literate in IT (and/or other areas of science).

(ii) Giving the next generation with the right skills

Computer Science students need an understanding of the social consequences of AI. (And of IT more generally.)

- Ethics should be an important (examinable) component of several CS courses.
- CS degrees should require social science papers.

Students planning on entering politics/law should be encouraged to become literate in IT (and/or other areas of science).

Some recent Otago initiatives:

- LAWS 102 Introduction to Law and new Technologies
- LAWS 428 Law and Emerging Technologies
- Bachelor of Arts and Sciences

Bibliography

- J Angwin, J Larson, S Mattu, and L Kirchner. What algorithmic injustice looks like in real life. ProPublica, 2016.
- M Arntz, T Gregory, and U Zierahn. Automation for jobs in OECD countries: A comparative analysis. OECD Social, Employment and Migration Working Papers, No. 189, OECD Publishing, Paris, 2016.
- A Caliskan, J Bryson, and A Narayanan. Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334):183–186, 2017.
- C Frey and M Osborne. The future of employment: how susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114:254–280, 2013.