


FACULTY OF MATHEMATICS,
PHYSICS AND INFORMATICS
Comenius University
Bratislava

Neural Networks for Computer Vision

Lecture 13: Science and Ethics of Deep Learning

Ing. Viktor Kocur, PhD., RNDr. Zuzana Černeková, PhD.

15.12.2021

Contents

Ensuring Validity of Deep Learning Results

ML algorithms heavily rely on training data. This brings about several issues:

- Overfitting
- Reproducibility
- Identification of meaningful components
- Reliability/Explainability
- Different distribution of data in real-life

Overfitting

We can broadly overfit the data in two main ways:

- Direct overfitting of parameters

Overfitting

We can broadly overfit the data in two main ways:

- Direct overfitting of parameters
  - We can check for this using the val set

Overfitting

We can broadly overfit the data in two main ways:

- Direct overfitting of parameters
  - We can check for this using the val set
- Overfitting of hyperparameters

Overfitting

We can broadly overfit the data in two main ways:

- Direct overfitting of parameters
  - We can check for this using the val set
- Overfitting of hyperparameters
  - We use the test set to make sure we did not optimize hyper-parameters specifically to optimize on the val set

Evaluation on test set

Evaluation on test set is the gold standard for most areas of deep learning research. However care must be taken to ensure the validity of the test set evaluation! This mostly means that the **test set should not be used to fine-tune the model hyperparameters**.

Evaluation on test set

What should you do when you perform some research and at the last stage you perform evaluation and your model does not perform well on the test set?

Evaluation on test set

What should you do when you perform some research and at the last stage you perform evaluation and your model does not perform well on the test set?

In such case you might consider changing some elements of the model including the training process. This is directly in contradiction with the principle of not using the test set to guide the training process in any way!

Evaluation on test set

What should you do when you perform some research and at the last stage you perform evaluation and your model does not perform well on the test set?

In such case you might consider changing some elements of the model including the training process. This is directly in contradiction with the principle of not using the test set to guide the training process in any way!

Such instances happen all the time and there is no best way to avoid them. If you do make any changes they should still be mostly guided by the validation data. From the scientific perspective it is usually better to choose common sense values for some parameters even if the resulting test evaluation would be worse.

Reproducibility

Reproducibility issues can be mitigated to a large extent by

- Publishing your code and making it easy to run
- Making sure that all the datasets are publicly available
- Provide pre-trained models to easily download and use

Reproducibility - stochastic methods

Another issue regarding reproducibility comes from the fact that some methods are highly stochastic. In deep learning the training data is usually shuffled and weights get initialized randomly.

Reproducibility - stochastic methods

Another issue regarding reproducibility comes from the fact that some methods are highly stochastic. In deep learning the training data is usually shuffled and weights get initialized randomly.

Ideally we would train the network multiple times and report the means and standard deviations of the resulting metrics. In practice this gets done only on smaller experiments which do not require so much time.

Identification of meaningful components

Many modern deep learning papers present a solution to a problem while utilizing multiple previously published methods and techniques.

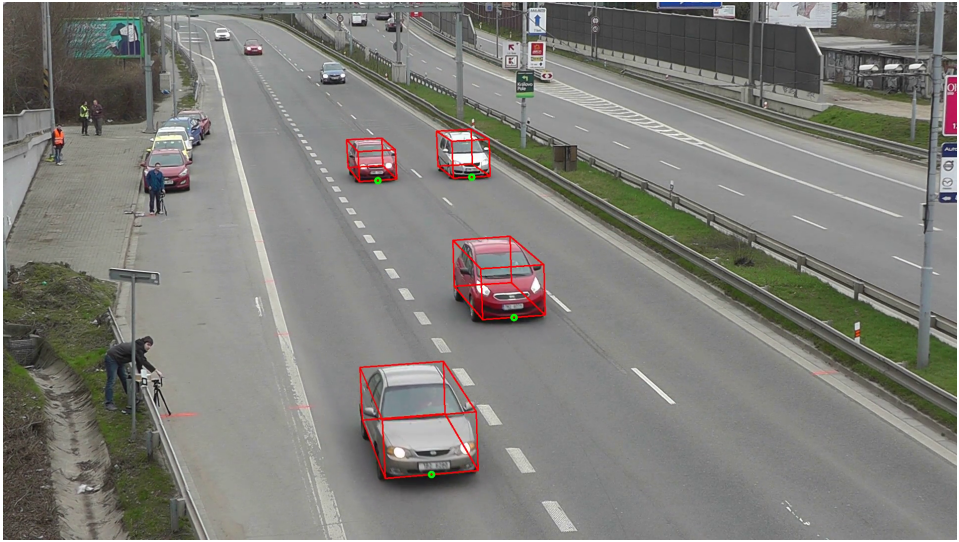
It is possible that some components of the proposed solution are not necessary or they do not bring any benefit. How can we tell if this is the case?

Ablation studies

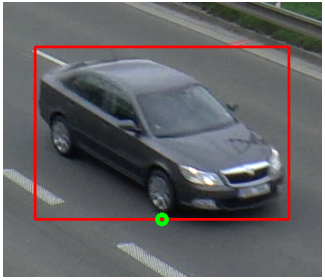
To gauge the benefits of individual elements of a given method we can perform so-called ablation studies. These typically include various experiments and evaluations with:

- One or more elements of the proposed method removed
- Smaller/Larger models to enable better comparison with other approaches
- Previously published methods with new techniques added

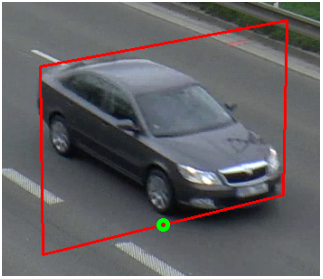
Ablation study - example - speed estimation



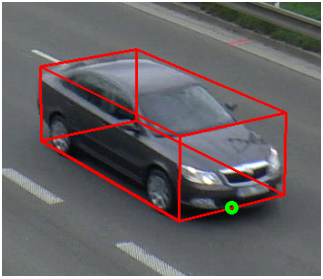
Ablation study - example



Orig2D



Trans2D



Trans3D

Ablation study - example

Method	Mean error (km/h)	Median error (km/h)	Recall (%)	Precision (%)	FPS
Trans3D (960 × 540)	**0.77**	**0.58**	83.02	86.90	43
Trans3D (640 × 360)	0.79	0.60	83.10	83.53	62
Trans3D (480 × 270)	0.92	0.72	79.88	88.39	**70**
Trans2D (640 × 360)	0.93	0.69	77.58	84.06	62
Orig2D (640 × 360)	0.94	0.73	85.03	88.00	62
MaskRCNN (1024 × 768)	0.88	0.64	81.89	88.44	5
SochorAuto [1]	1.10	0.97	83.34	**90.72**	-
SochorManual [1]	1.32	0.95	83.34	**90.72**	-
DubskaAuto[2]	8.22	7.87	**90.08**	73.48	-

[1] Jakub Sochor, Roman Juránek, and Adam Herout. “Traffic surveillance camera calibration by 3d model bounding box alignment for accurate vehicle speed measurement.” In: *Computer Vision and Image Understanding* 161 (2017), pp. 87–98

[2] Markéta Dubská, Adam Herout, and Jakub Sochor. “Automatic Camera Calibration for Traffic Understanding.” In: *Proceedings of the British Machine Vision Conference*. BMVA Press, 2014

Bad or insufficient data

In practice we may find out that the data a system was trained and evaluated on is not representative of the actual distribution of data the system uses!

Bad or insufficient data

In practice we may find out that the data a system was trained and evaluated on is not representative of the actual distribution of data the system uses!

Usual fixes:

- Get better/more data
- Change loss functions/training process to better reflect the actual distribution.

Distribution shift

Distribution shift may occur with time as the data that is input into the system changes over time. This can be addressed in following ways:

- Collecting new data as the system is deployed
- Reporting/checking the system performance after it has been deployed especially on new data
- Re-training or fine-tuning the model with new data

Reliability

When deploying machine learning systems care must be taken to ensure their reliability. Ensuring reliability is generally in the domain of Reliability Engineering.

The objectives of reliability engineering, in decreasing order of priority, are:[3]

- To apply engineering knowledge and specialist techniques to prevent or to reduce the likelihood or frequency of failures.

[3] Patrick O'Connor and Andre Kleyner. *Practical reliability engineering*. John Wiley & Sons, 2012

Reliability

When deploying machine learning systems care must be taken to ensure their reliability. Ensuring reliability is generally in the domain of Reliability Engineering.

The objectives of reliability engineering, in decreasing order of priority, are:[3]

- To apply engineering knowledge and specialist techniques to prevent or to reduce the likelihood or frequency of failures.
- To identify and correct the causes of failures that do occur despite the efforts to prevent them.

[3] Patrick O'Connor and Andre Kleyner. *Practical reliability engineering*. John Wiley & Sons, 2012

Reliability

When deploying machine learning systems care must be taken to ensure their reliability. Ensuring reliability is generally in the domain of Reliability Engineering.

The objectives of reliability engineering, in decreasing order of priority, are:[3]

- To apply engineering knowledge and specialist techniques to prevent or to reduce the likelihood or frequency of failures.
- To identify and correct the causes of failures that do occur despite the efforts to prevent them.
- To determine ways of coping with failures that do occur, if their causes have not been corrected.

[3] Patrick O'Connor and Andre Kleyner. *Practical reliability engineering*. John Wiley & Sons, 2012

Reliability

When deploying machine learning systems care must be taken to ensure their reliability. Ensuring reliability is generally in the domain of Reliability Engineering.

The objectives of reliability engineering, in decreasing order of priority, are:[3]

- To apply engineering knowledge and specialist techniques to prevent or to reduce the likelihood or frequency of failures.
- To identify and correct the causes of failures that do occur despite the efforts to prevent them.
- To determine ways of coping with failures that do occur, if their causes have not been corrected.
- To apply methods for estimating the likely reliability of new designs, and for analysing reliability data.

[3] Patrick O'Connor and Andre Kleyner. *Practical reliability engineering*. John Wiley & Sons, 2012

Reliability - preventing failure

There are various methods of failure prevention. Some examples:

- Good definition of bounds within which the system will work. Eg. constraints on the input data.

Reliability - preventing failure

There are various methods of failure prevention. Some examples:

- Good definition of bounds within which the system will work. Eg. constraints on the input data.
- Careful consideration of failure modes (e.g. adversarial examples) and preventative measures (e.g. training models to be robust w.r.t. adversarial examples, noise etc.).

Reliability - preventing failure

There are various methods of failure prevention. Some examples:

- Good definition of bounds within which the system will work. Eg. constraints on the input data.
- Careful consideration of failure modes (e.g. adversarial examples) and preventative measures (e.g. training models to be robust w.r.t. adversarial examples, noise etc.).
- External auditing of systems (e.g. institutional/governmental certification).

Reliability - preventing failure

There are various methods of failure prevention. Some examples:

- Good definition of bounds within which the system will work. Eg. constraints on the input data.
- Careful consideration of failure modes (e.g. adversarial examples) and preventative measures (e.g. training models to be robust w.r.t. adversarial examples, noise etc.).
- External auditing of systems (e.g. institutional/governmental certification).
- Verification that the algorithm works as intended (e.g. using explainability techniques)

Reliability - monitoring and auditing

It is important to continually monitor the performance of the system to catch failure modes. This is especially important due to effects such as distribution shift.

- Deploying the system in parts (e.g. at start system only works as a guide for an expert who makes the final decisions)

Reliability - monitoring and auditing

It is important to continually monitor the performance of the system to catch failure modes. This is especially important due to effects such as distribution shift.

- Deploying the system in parts (e.g. at start system only works as a guide for an expert who makes the final decisions)
- Selecting appropriate metrics and monitoring them continually

Reliability - monitoring and auditing

It is important to continually monitor the performance of the system to catch failure modes. This is especially important due to effects such as distribution shift.

- Deploying the system in parts (e.g. at start system only works as a guide for an expert who makes the final decisions)
- Selecting appropriate metrics and monitoring them continually
- Setting up mechanisms to detect failures and address them

Reliability - monitoring and auditing

It is important to continually monitor the performance of the system to catch failure modes. This is especially important due to effects such as distribution shift.

- Deploying the system in parts (e.g. at start system only works as a guide for an expert who makes the final decisions)
- Selecting appropriate metrics and monitoring them continually
- Setting up mechanisms to detect failures and address them
- Make sure that there is a clear organizational responsibility when it comes to deployed ML models

Explainability

ML systems can be deployed in areas with significant social impact. There may therefore be demand from society for the right to explanation. The right to explanation is still an evolving legal topic.

Explainable AI (XAI) is still an open problem in AI research a probably will stay become increasingly relevant in the future.

Ethics of AI

- Algorithmic bias
- Data and compute in the hands of private corporations
- Environmental impact of AI
- Inequality and unemployment
- Responsibility for actions of AI systems
- Misaligned AI
- Misuse of AI
- Copyright issues related to deep learning
- Robotic rights

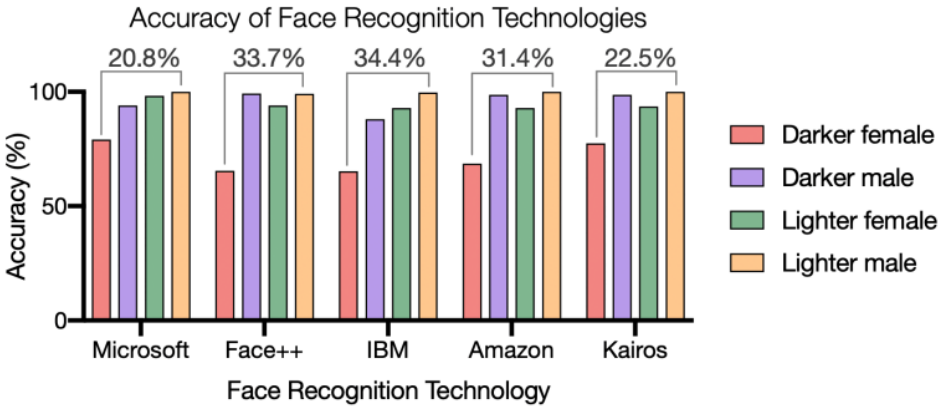
Algorithmic bias

Humans are inherently biased in many ways. Social sciences show us that these biases can have very significant negative effects on society. The problems can persist even after biases were remedied e.g. systemic racism.

These biases can propagate to ML systems in many ways:

- Bad data collection
- Good data collection, but the existing distribution contains undesirable biases
- Biased developers

Algorithmic bias



Joy Buolamwini and Timnit Gebru. “Gender shades: Intersectional accuracy disparities in commercial gender classification.” In: *Conference on fairness, accountability and transparency*. PMLR. 2018, pp. 77–91

Algorithmic bias

Amazon ditched AI recruiting tool that favored men for technical jobs

Specialists had been building computer programs since 2014 to review résumés in an effort to automate the search process

Image adopted from: Reuters. “Amazon ditched AI recruiting tool that favored men for technical jobs.” In: *The Guardian* (2018). ISSN: 0261-3077. URL: https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine

Algorithmic bias

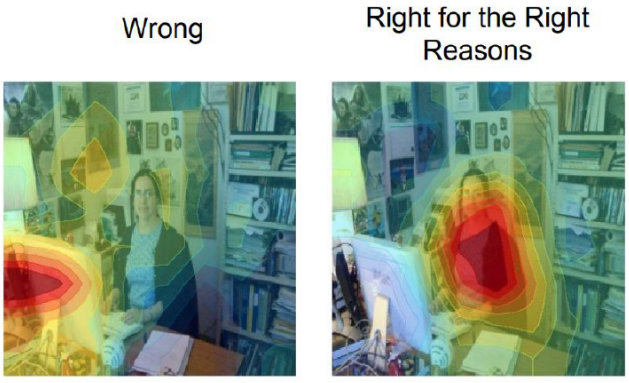


Image adopted from: Lisa Anne Hendricks et al. "Women also snowboard: Overcoming bias in captioning models." In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018, pp. 771–787

Algorithmic bias - Algocracy

Compass is a software which rates the recidivism risk of recidivism and is intended to be used in the criminal justice system.

In forecasting who would re-offend, the algorithm made mistakes with black and white defendants at roughly the same rate but in very different ways.[7]

- The formula was particularly likely to falsely flag black defendants as future criminals, wrongly labeling them this way at almost twice the rate as white defendants.
- White defendants were mislabeled as low risk more often than black defendants.

[7] Julia Mattu et al. *Machine Bias*. URL: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=8JGdLgdwlU0izBRjVJUOv_PQFUET-J7_

How to deal with algorithmic bias

- Evaluate your models for algorithmic biases
- Be skeptical of collected data
- Understand the limitations of your own experiences when considering existing biases
- Aim for diversity in the development team
- Aim to understand the related power structures

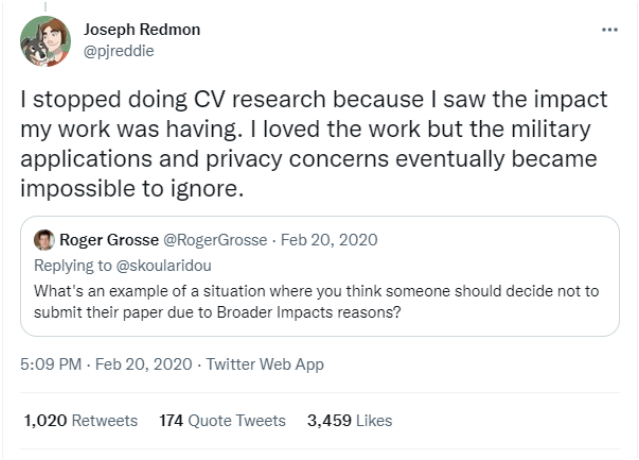
Misaligned AI

Misuse of AI

AI is powerful technology and as any types of technologies these can be misused. Some general areas of concern are:

- Military applications
- Surveillance applications (e.g. face recognition)
- Deepfakes/impersonation
- Disruptive tech with unintended consequences

A not-so-fun fact



Joseph Redmon
@pjreddie

I stopped doing CV research because I saw the impact my work was having. I loved the work but the military applications and privacy concerns eventually became impossible to ignore.

> Roger Grosse @RogerGrosse · Feb 20, 2020
> Replying to @skoularidou
> What's an example of a situation where you think someone should decide not to submit their paper due to Broader Impacts reasons?

5:09 PM · Feb 20, 2020 · Twitter Web App

1,020 Retweets **174** Quote Tweets **3,459** Likes

Mass-scale data collection

AI and the related concept of Big Data make personal data increasingly more valuable to those with resources to leverage such data. This issue has several elements which complicate the solutions to these problems:

- A single data-point has very low value
- Data without resources to process it is not as valuable
- The value of individual data-points increases as more data is available

Mass-scale data collection

AI and the related concept of Big Data make personal data increasingly more valuable to those with resources to leverage such data. This issue has several elements which complicate the solutions to these problems:

- A single data-point has very low value
- Data without resources to process it is not as valuable
- The value of individual data-points increases as more data is available

This results in a situation where users are willing to exchange their data for free services as this is rationally a good transaction.

Mass-scale data collection

There are several attempts to remedy this such as GDPR. However time has shown that this is not a sufficient solution to the problem.

- GDPR theoretically allows users more control of their data, but this is limited in practice
- GDPR does not deal with selling of data.
- It might be possible to infer information about non-consenting parties from consenting ones.

Computational and data supremacy

Large corporations are able to invest huge resources to collect data and train models. This results in a huge potential for these corporations to develop significantly more powerful models which may push them ahead in developing AI technology without public oversight.

Example: JFT-3B is an internal Google dataset and a larger version of the JFT-300M dataset. It consists of nearly **3 billion images, annotated with a class-hierarchy of around 30k labels** via a semi-automatic pipeline. In other words, the data and associated labels are noisy.

Environmental impacts of AI

Computer Science > Computers and Society

[Submitted on 22 Jul 2019 (v1), last revised 13 Aug 2019 (this version, v3)]

Green AI

Roy Schwartz, Jesse Dodge, Noah A. Smith, Oren Etzioni

The computations required for deep learning research have been doubling every few months, resulting in an estimated 300,000x increase from 2012 to 2018 [2]. These computations have a surprisingly large carbon footprint [38]. Ironically, deep learning was inspired by the human brain, which is remarkably energy efficient. Moreover, the financial cost of the computations can make it difficult for academics, students, and researchers, in particular those from emerging economies, to engage in deep learning research.

This position paper advocates a practical solution by making efficiency an evaluation criterion for research alongside accuracy and related measures. In addition, we propose reporting the financial cost or "price tag" of developing, training, and running models to provide baselines for the investigation of increasingly efficient methods. Our goal is to make AI both greener and more inclusive---enabling any inspired undergraduate with a laptop to write high-quality research papers. Green AI is an emerging focus at the Allen Institute for AI.

Comments: 12 pages

Subjects: **Computers and Society (cs.CY)**; Computation and Language (cs.CL); Computer Vision and Pattern Recognition (cs.CV); Machine Learning (cs.LG); Methodology (stat.ME)

Cite as: arXiv:1907.10597 **[cs.CY]**

(or arXiv:1907.10597v3 **[cs.CY]** for this version)

Job automation

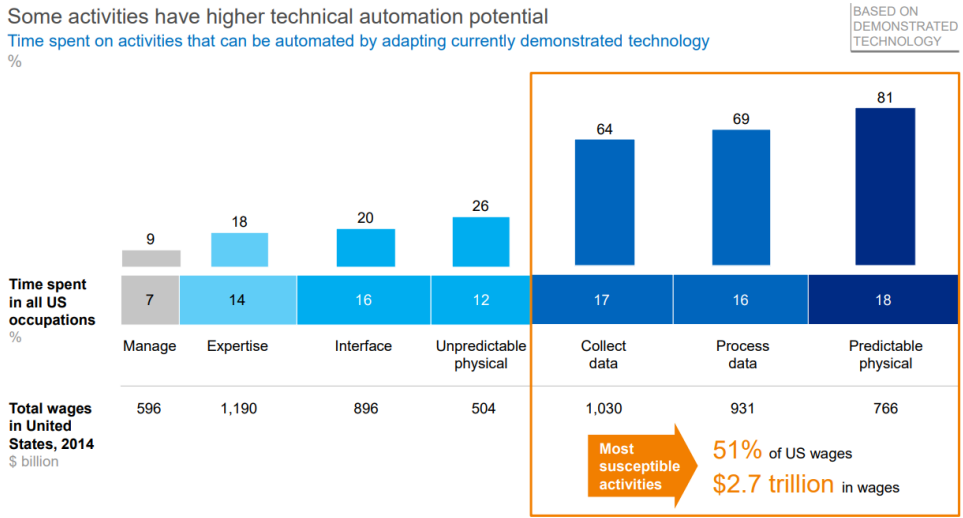


Image adopted from: James Manyika et al. “A future that works: AI, automation, employment, and productivity.” In: *McKinsey Global Institute Research, Tech. Rep* 60 (2017), pp. 1–135

Big Tech

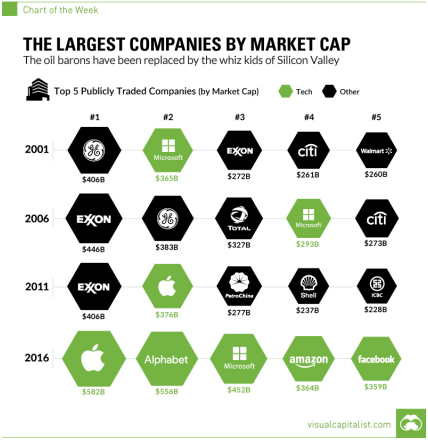


Image adopted from: Jeff Desjardins. *Chart: The Largest Companies by Market Cap Over 15 Years.* 2016. URL: https://www.visualcapitalist.com/chart-largest-companies-market-cap-15-years/

Big Tech

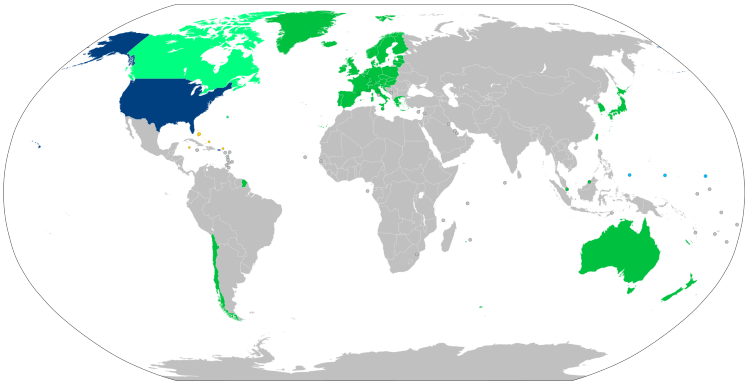
Big Tech has massive impact and power on a global scale. This is problematic as the company has an explicit legal obligation (fiduciary duty) to use its resources specifically to financially benefit its shareholders. This may result in misalignment with public good!

Big Tech

Big Tech has massive impact and power on a global scale. This is problematic as the company has an explicit legal obligation (fiduciary duty) to use its resources specifically to financially benefit its shareholders. This may result in misalignment with public good!

- Stakeholder capitalism
- Regulation from public institutions - might be tricky
- Strong ethics-conscious unions
- Solutions beyond capitalism

Conference access inequality



The most prestigious conference in CV is held in USA. Obtaining a visa is a significant barrier for researchers from developing countries.

Image adopted from: Page Version ID: 1058682332. 2021. URL: https://en.wikipedia.org/w/index.php?title=Visa_policy_of_the_United_States&oldid=1058682332

Responsibility for AI actions

Consider the scenario of a self-driving car making decisions. If a decision results in an accident who is responsible?

Responsibility for AI actions

Consider the scenario of a self-driving car making decisions. If a decision results in an accident who is responsible?

How should a self-driving car react to a situation where the only viable options are harm to the passengers or harm to some pedestrian? Who should decide?

Copyright issues

Suppose we train a GAN on a dataset of images created by many artists. Who can claim ownership rights of the resulting images? What if the network also outputs near-copies of the training data?

Patent issues

Another not-so-fun-fact: BatchNorm is patented by Google![11]

[11] Sergey Ioffe and Corinna Cortes. *Batch normalization layers*. US Patent 10,417,562. 2019

Robot rights

Can AI systems become self-aware? Are they capable of suffering?

If yes should we afford them rights?

Key takeaways

Many of the issues arising with modern technology cannot be solved with STEM approaches. We must therefore not dismiss the work of our colleagues in humanities such as philosophy, political science, sociology, gender/race studies and others to understand the related power structures, ethical issues and social impacts of the technology we may help develop and maintain.

Key takeaways