Queen Mary
University of London

DR ADRIAN BEVAN

PRACTICAL MACHINE LEARNING

ETHICS

LECTURE PLAN

- Ethics
- Frameworks
- Examples
- Summary
- Suggested Reading

QMUL Summer School: https://www.qmul.ac.uk/summer-school/
Practical Machine Learning QMplus Page: https://qmplus.qmul.ac.uk/course/view.php?id=10006

ETHICS

- Knowing the difference between right and wrong enables you to behave ethically.
  - Relies on consideration of your moral compass.
- In our rapidly changing world technology presents many new opportunities.
  - It is natural to ask questions about capability of technology.
  - Deepens understanding in methods and possibilities.
  - However, just because you have the ability to do something with technology, does not mean you should do it.

FRAMEWORKS

- In 2016 the UK produced a Data Science Ethical Framework [1].

The guidance gives six principles which are based on existing law.
Fundamentally, the public benefit of doing the project needs to be balanced against the risks of doing so.

1. Start with clear user need and public benefit
2. Use data and tools which have the minimum intrusion necessary
3. Create robust data science models
4. Be alert to public perceptions
5. Be as open and accountable as possible
6. Keep data secure

Cabinet Office

Version 1.0
Published 19 May 2016

Data Science Ethical Framework

- The Data Protection and Intellectual Property Acts are intended to protect peoples rights in the UK; along with the new General Data Protection Regulation (GDPR).

[1] https://www.gov.uk/government/publications/data-science-ethical-framework

FRAMEWORKS 1. Start with clear user need and public benefit

- Identify a problem to solve.
- Ensure that the data is expected to be sufficient to solve the problem and to learn something useful.
- Ensure that the method is viable.
- Not ethical to collect data to do bad science; so need to understand that the method makes sense and that the data being collected will lead to some meaningful result.

[1] https://www.gov.uk/government/publications/data-science-ethical-framework

FRAMEWORKS 2. Use data and tools which have the minimum intrusion necessary

- Only use the features that are needed in the data in order to obtain results.
  - This is problematic as some features will lead to an intrinsic bias for some problems; requires care and attention.
    - *e.g. different ethnic groups are susceptible to different medical conditions. So including that data in a medical study related to one of those conditions could bias a model accordingly.*
      - Scientific/commercial problems have equivalent issues with features leading to bias.
- Data is ideally anonymised, or aggregated in such a way to retain anonymity. An exception to this would be identification of terrorists or criminals, where the aim is to identify individuals.
- Use of data from social media needs to be considered carefully.
- "The law states that you must take reasonable steps to ensure that individuals are will not be identifiable when you link data or combine it with other data in the public domain."[1]

[1] https://www.gov.uk/government/publications/data-science-ethical-framework

FRAMEWORKS 3. Create robust data science models

- Consider the data that you feed into an algorithm to make sure that you know how robust that is, and avoid potentially biasing the outcome.
- Consider the algorithm - there is little point trying to train a deep network with a few hundred or a few thousand training examples. The result will not be robust.
- Consider the cost of making an incorrect decision.
- Understand the output of the model that has been built - the question being answered may not be the one you want the model to address.
  - A study processing images of farmland into different types of usage while looking for Japaneese Knotweed can result in trees and buildings being classified as the same object… because they both cast shadows; not because they are the same thing.
  - As with Deep Thought in the Hitch Hikers Guide to the Galaxy; on asking the computer the answer to the ultimate question of the meaning of life the universe and everything, when you get the response 42 you need to understand what the computer thought your question was.

[1] https://www.gov.uk/government/publications/data-science-ethical-framework

FRAMEWORKS 4. Be alert to public perceptions

- The legal requirements of the data protection and IP acts, along with the GDPR lay down the boundaries of what is legal.
- Sometimes that is not sufficient and you need to consider that just because you can do something does not mean that you should do it.
- An example of unethical behaviour using algorithms is given by Volkswagen.
  - In 2015 the world learned that VW had been using software to trick emissions tests into categorising their cars as having lower emissions than when being driven normally. The first VW with this software installed was sold in the UK in 2008[2]. Up to 11 million cars were affected by this. The consequence for unethical behaviour was a drop in share price of 1/3 and VW set aside $18billion for associated costs (recalls, fines etc).

[1] https://www.gov.uk/government/publications/data-science-ethical-framework
[2] https://uk.reuters.com/article/volkswagen-emissions/vw-says-sold-first-uk-vehicle-with-emission-test-rigging-software-in-2008-idUKU8N10Z00520151012

FRAMEWORKS 5. Be as open and accountable as possible

- Where possible be open about a project to allow people to understand:
  - What data is collected;
  - How the data intends to be used;
  - Any social or other benefit expected from the work;
  - Ensure oversight and accountability;
- Provide a means for recourse should there be an incorrect decision made (this depends on the context of the problem).
- The aim of these points is to promote ethical practice.

[1] https://www.gov.uk/government/publications/data-science-ethical-framework

FRAMEWORKS 6. Keep data secure

- Data should be kept securely.
- General considerations:
  - Who has access to the data?
    - *Only people who need access to the data to address the problem under investigation should have access.*
  - How long will the data be stored?
    - *Typically we would expect the data to be stored for a finite time, after which it would be deleted. For example this is standard practice/constraint for a study performed that requires ethical approval from a University. After that fixed time period data would be deleted.*

[1] https://www.gov.uk/government/publications/data-science-ethical-framework

Quick checklist

Tick where you are on the scale

1. Start with clear user need and public benefit	**A.** How does the department and public benefit?	High public benefit (to society or to an individual)	Medium public benefit (to society or to an individual)	Low public benefit (to society or to an individual)
2. Use data and tools which have the minimum intrusion necessary	**B.** How intrusive and identifiable is the data you are working with?	Non-personal and therefore non-identifiable	Personal but non-sensitive	Personal, sensitive data which could be inferred or directly re-identified
	C. If identifying individuals, how widely are you searching personal data?	Querying against known individuals	Querying against a targeted group	Speculatively searching for needle in haystack
3. Create robust data science models	**D.** What is the quality of the data?	Representative and unbiased	Historical data which is biased and excludes certain groups	Inaccurate or missing data
	E. How automated are the decisions?	Human making decision based on analysis	Limited human oversight but regularly checked	No human oversight or method of checking
	F. What is the risk that someone will suffer a negative unintended consequence as a result of the project?	Low	Medium	High
4. Be alert to public perceptions	**G.** If personal data for operational purposes, how compatible was it with the reason collected?	Very compatible	Less compatible but fair	Not compatible
	H. Do the public agree with what you are doing?	Yes	Some would, some wouldn't or not sure what people think	No, or lots would have real concerns
5. Be as open and accountable as possible	**I.** How open can you be about the project?	Very open, and make open the tools and data for re-use	Open about project but not about data/tools	Cannot talk about project aim
	J. How much oversight and accountability is there throughout the project?	Throughout - including the decision made as a result of insight	Only at the beginning	None
6. Keep data secure	**K.** How secure is your data?	Very secure, with restricted access to a few named individuals	Secure and password protected	Openly available within the department

Some departments might find themselves at the left hand side of the scale, and others more on the right (blue), reflecting the nature of their department's work. This does not mean the project should not go ahead, but think carefully about it, and if possible, bring some elements to the green end of the scale.

*Not all may apply to your project

All fine? Go forward! | **Some issues?** Think carefully | **Tricky issues?** Extreme care & oversight

5

[1] https://www.gov.uk/government/publications/data-science-ethical-framework

EXAMPLES

- Examples of controversial use of technology can be found in the media.

Online streaming giant Netflix sent a festive tweet before Christmas "to the 53 people who've watched *A Christmas Prince* every day for the past 18 days: who hurt you?"

Thousands of users liked the tweet, but many others were concerned by its big brother overtones and declared it "creepy" that users' viewing habits were being tweeted for a cheap gag. Netflix responded, via technology website the Register, that "the privacy of our members' viewing is important to us".

Privacy has certainly been on the firm's agenda. In 2011 it successfully backed an amendment to water down the US Video Privacy Protection Act which meant social media sites could share rental information.

In 2012, Netflix changed its privacy rules in the wake of legal challenge by users who did not want their information retained once they had left the site. A few years earlier, Netflix had settled a lawsuit over an in-the-closet lesbian subscriber who feared that insufficiently anonymous information from her subscription history could out her.

Netflix's privacy statement openly says that it collects preferences and viewing history from its customers' accounts. It also states that it can't guarantee the security of the data it collects: "We use reasonable administrative, logical, physical and managerial measures to safeguard your personal information ... Unfortunately, no measures can be guaranteed to provide 100% security. Accordingly, we cannot guarantee the security of your information."

With privacy and security an ongoing issue for Netflix, perhaps tweeting jibes about its users' viewing habits wasn't the best Christmas present for customers.

• Spotify ran a curiously similar campaign over the festive season in 2016-17. The sample copy included "Dear person who played 'Sorry' 42 times on Valentine's Day, what did you do?"

Private Eye 1460 (2018)

Reproduced by kind permission of PRIVATE EYE magazine www.private-eye.co.uk

EXAMPLES

- A much higher-profile debacle can be found in the media: the use of facebook data of from users and their contacts as a result of taking personality tests.
  - The SLC group and its consultancy Cambridge Analytica became embroiled in a sensational media frenzy surrounding harvesting data from Facebook users.
  - Reportedly 50 million Americans an more than 1 million UK citizens had their data harvested [1].
  - This was used to target advertising related to the US election.
  - The story broke 17th March 2018, and on the 2nd May news broke that Cambridge Analytica was being closed down [2].
  - Facebook lost $120Bn off of its market price in July 2018 as a result of “slow growth rate in the past 2 years” [3].

[1] https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election
[2] https://www.theguardian.com/uk-news/2018/may/02/cambridge-analytica-closing-down-after-facebook-row-reports-say
[3] https://www.bbc.co.uk/news/world-us-canada-44978452

EXAMPLES

- A much higher-profile debacle can be found in the media: the use of facebook data of from users and their contacts as a result of taking personality tests.
  - The SLC group and its consultancy Cambridge Analytica became embroiled in a s cebook users
  - Repo s had their
  - This was used to target advertising related to the US election.
  - The story broke 17th March 2018, and on the 2nd May news broke that Cambridge Analytica was being closed down [2].
  - Facebook lost $120Bn off of its market price in July 2018 as a result of "slow growth rate in the past 2 years" [3].

Recent reports indicate that Facebook will be fined $5Bn over this unethical use of widespread data privacy violation.

e.g. see the BBC article on this story.

[1] https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election
[2] https://www.theguardian.com/uk-news/2018/may/02/cambridge-analytica-closing-down-after-facebook-row-reports-say
[3] https://www.bbc.co.uk/news/world-us-canada-44978452

EXAMPLES

Facebook terms and conditions state:

Sharing your content and information

You own all of the content and information that you post on Facebook, and you can control how it is shared through your privacy and application settings. In addition:

1. For content that is covered by intellectual property rights, such as photos and videos (IP content), you specifically give us the following permission, subject to your privacy and application settings: you grant us a non-exclusive, transferable, sub-licensable, royalty-free, worldwide licence to use any IP content that you post on or in connection with Facebook (IP Licence). This IP Licence ends when you delete your IP content or your account, unless your content has been shared with others and they have not deleted it.
2. When you delete IP content, it is deleted in a manner similar to emptying the recycle bin on a computer. However, you understand that removed content may persist in backup copies for a reasonable period of time (but will not be available to others).
3. When you use an application, the application may ask for your permission to access your content and information as well as content and information that others have shared with you. We require applications to respect your privacy, and your agreement with that application will control how the application can use, store and transfer that content and information. (To learn more about Platform, including how you can control what information other people may share with applications, read our Data Policy and Platform page.)
4. When you publish content or information using the Public setting, it means that you are allowing everyone, including people not on Facebook, to access and use that information, and to associate it with you (i.e. your name and profile picture).
5. We always appreciate your feedback or other suggestions about Facebook, but you understand that we may use your feedback or suggestions without any obligation to compensate you for them (just as you have no obligation to offer them).

[1] https://www.facebook.com/legal/terms (Version corresponding to Date of last revision: 31 January 2018)

EXAMPLES

Facebook terms and conditions state:

Sharing your content and information

You own all of the content and information that you post on Facebook, and you can control how it is shared through your privacy and application settings. In addition:

1. For content that is covered by intellectual property rights, such as photos and videos (IP content), you specifically give us the following permission, subject to your privacy and application settings: you grant us a non-exclusive, transferable, sub-licensable, royalty-free, worldwide licence to use any IP content that you post on or in connection with Facebook (IP Licence). This IP Licence ends when you delete your IP content or your account, unless your content has been shared with others and they have not deleted it.
2. When you delete IP content, it is deleted in a manner similar to emptying the recycle bin on a computer. However, you understand that removed content may persist in backup copies for a reasonable period of time (but will not be available to others).
3. When you use an application, the application may ask for your permission to access your content and information as well as content and information that others have shared with you. We require applications to respect your privacy, and your agreement with that application will control how the application can use, store and transfer that content and information. (To learn more about Platform, including how you can control what information other people may share with applications, read our Data Policy and Platform page.)
4. When you publish content or information using the Public setting, it means that you are allowing everyone, including people not on Facebook, to access and use that information, and to associate it with you (i.e. your name and profile picture).
5. We always appreciate your feedback or other suggestions about Facebook, but you understand that we may use your feedback or suggestions without any obligation to compensate you for them (just as you have no obligation to offer them).

[1] https://www.facebook.com/legal/terms (Version corresponding to Date of last revision: 31 January 2018)

EXAMPLES

- From a corporate perspective this is clearly a practical solution.
  - Users sign over rights for (free) reuse of IP or that of anything posted or shared.
    - *Complicated persistency conditions should users decide to delete accounts or IP related data.*
- Is it ethical?

SUMMARY

- The issue of ethics in machine learning and AI is currently not legislated, and users have agreed to give unprecedented concessions with regard to their data to social media companies such as Facebook the moment that they sign up for an account.
  - Some media attention has appeared related to actions made by some companies in response to processing user data for effect.
    - Higher profile backlash has been seen in the case of the involvement of Cambridge Analytical in mining data from Facebook users.
- These events have ignited a long-overdue debate about who has the right to own data of users.
- Data scientists should think about the ethics of solving a given problem, and can look to guidance coming from IEEE and government guidelines on ethical data science.

SUGGESTED READING

- This is a rapidly changing area and while suggested reading is given here, it is likely that the area of ethics in data science will continue to evolve and these references may become outdated on a faster timescale than those resources suggested for other areas of this course.
- ***Ethics in the Real World*, Peter Singer, Princeton (2016)**
  - Brief essays on modern ethical issues, including essays on science and technology. The write contributes to Project Syndicate.
- **IEEE Ethics in Action:**
  - Ethically Aligned Design (V2 complete, work continues on an updated version of this document).
  - https://ethicsinaction.ieee.org
- **Data Science Ethical Framework (UK Government framework document):**
  - Framework described in a 17 page document (2016).
  - https://www.gov.uk/government/publications/data-science-ethical-framework
- **House of Lords select committee on *AI in the UK: ready, willing and able?***
  - https://www.parliament.uk/business/committees/committees-a-z/lords-select/ai-committee/news-parliament-2017/ai-report-published/