Fairness, Accountability, Transparency and Ethics

Ifeoma Adaji

Learning objectives

- At the end of the class, students should be able to:
  - Explain fairness and accountability in AI
  - Explain AI bias with examples
  - List common ways bias is introduced in AI
  - Describe steps governments are taking towards accountability in AI
  - Describe transparency in AI and its implications
  - Describe AI ethics and its implications

Artificial Intelligence

- Systems with the intellectual characteristic of humans, such as the ability to reason, discover meaning, generalize, or learn from past experience.
- A computer system that is able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision making, and translation between languages.
- AI is used for email spam filtering, image recognition, recommender systems, understanding human speech (Siri, Alexa), self driving cars, filters in social media, targeted ads in social media, determine whose posts you see in social media, etc.

Problems with AI - Bias

- No technology is free of its creators
- AI systems are not truly separate and autonomous, they start with us
- Technology always comes from and is designed by people; it is no more objective than we are
- AI bias is the underlying prejudice in data that is used to create AI algorithms, which can ultimately result in discrimination and other social consequences
- AI bias is lack of fairness
- AI is bias; bias can be based on
  - Who builds the technology
  - Which assumptions are programmed into them
  - How they're trained - data
  - How they're ultimately deployed
- The data you create for your system to learn from will be biased by how you see the world
  - If an algorithm is trained using data that doesn't represent a particular demographic group, the algorithm will likely be inaccurate when applied to people that are part of that group.

3 common ways bias is introduced

- Algorithmic bias. Algorithms are limited by the assumptions or views of its developers
- Sample bias. Algorithms are trained using incomplete data or data that does not represent a complete picture
- Measurement bias. Algorithms based on faulty sensors or faulty measurement devices, or when measuring devices are nor read/recorded properly

Examples of algorithmic bias

- Algorithm used to predict which patients would need extra medical care was in favour of one race over another.
  - https://www.scientificamerican.com/article/racial-bias-found-in-a-major-health-care-risk-algorithm/
- Facebook’s algorithm that made people turn against each other (2020)
  - https://www.independent.co.uk/life-style/gadgets-and-tech/news/facebook-algorithm-bias-right-wing-feed-a9536396.html
- Facebook’s news feed algorithmic bias (2016)
  - https://www.nytimes.com/2016/05/19/opinion/the-real-bias-built-in-at-facebook.html
  - https://gizmodo.com/former-facebook-workers-we-routinely-suppressed-conser-1775461006
  - https://tinyurl.com/2d5zvyc3 (how they’ve changed)
- Amazon’s hiring algorithm
  - https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine

Impact of AI bias

- Real world consequences
  - Discrimination (eg. Age discrimination)
  - Gender bias,
  - Racial prejudice.
  - Living in a bubble (news feed)

Improving algorithmic fairness

- Algorithms can have real world consequences so they should be fair
- AI should be designed to minimize bias
  - Eg. randomized block design, blinding, random sampling, bias –variance tradeoff.
- AI should be designed to promote inclusive representation
- Be aware of how your choices affect the rest of humanity
- Be aware of what data to use for AI
  - If you train your model with only specific type of data, what happens when data that is out of scope is introduced
  - all subjects should have an equal chance of being represented in the data; obtain data from traditionally underrepresented groups.
- Models should be retrained periodically with new data to start to remove historical biases, even though this will be more expensive
- People should be made aware when algorithms are being used in ways that impact their lives to ensure fairness
  - Knowing which of our data is being used as input to a model and access to any output generated

Improving algorithmic fairness

- Quick intervention when bias is identified
- Thorough investigation to identify source of bias and how it can be mitigated
- Frequent reviews of algorithms/systems
- Collect data on user identified bias (from users)
- Diversity of teams
- Standardizing processes including
  - unconscious bias training
  - rigorous peer review of algorithms before deployment to check for bias
  - independent post-implementation auditing of the fairness of algorithms to understand its impact on the most vulnerable people affected by it

Accountability

- Who should be held accountable for AI bias?
- Accountability is important
  - foundation of trust
  - acknowledgement and assumption of responsibility and “answerability” for actions, decisions, products and policies
- AI designers and developers are responsible for considering AI design, development, decision processes, and outcomes.
- US Government’s Accountability office recently developed first framework to help assure accountability and responsible use of AI
- “Goal is to help organizations and leaders move from theories and principles to practices that can actually be used to manage and evaluate AI in the real world”
- Framework
  - Defines conditions for accountability throughout AI lifecycle
  - Details specific questions to ask
  - Specifies audit procedures to use when assessing AI systems
  - Cover 4 dimensions: governance, data, performance, and monitoring

Four dimensions of AI accountability

- Assess governance structures
  - Understand the governance structure, have well-defined roles, responsibilities, and lines of authority
  - Document technical specifications of the particular AI system
- Understand the data.
  - Documentation of how data is being used for build the AI model and when it is in operation
  - Reliability and representativeness of data
  - Examine data for potential bias, inequity, or other societal concerns
- Define performance goals and metrics
  - After deployment of AI system, evaluate to ensure it meets its intended goals
- Review monitoring plans
  - Ongoing performance monitoring

Source:
https://hbr.org/2021/08/how-to-build-accountability-into-your-ai

In Canada….

A paper was developed at the request of the Government of Canada to support the G7 Multi-stakeholder Conference on Artificial Intelligence: Enabling the Responsible Adoption of AI on December 6, 2018.

- Many stakeholders could be engaged in creating a robust AI accountability framework
  - Policymakers in National Governments
  - Intergovernmental Organizations
  - Policymakers in Sub-National Governments
  - Corporations and Other Data Owners
  - Universities and Colleges
  - Advocacy Groups and Public Interest Organizations
  - Foundations
  - Professional Regulatory Bodies and Organizations

Source: https://www.ic.gc.ca/eic/site/133.nsf/eng/00005.html

In Canada…

- How govt. of Canada is ensuring responsible use of AI

https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai.html#toc1

Transparency of AI

- The issues of Fairness and Accountability have led to a call for *transparency* and *responsibility* in AI: AI that is explainable to employees and customers and aligns with a company's values
  - AI now has a huge impact on lives; it is used for critical decisions (bank loans, employment)
- Transparency means being able to assess the inner workings of a system
- Involves explaining how an AI-based decision was made, what the decision was based on
- Explanation should be sufficiently understandable to those affected by the model
- Allows humans see if models have been thoroughly tested, if models make sense, and why particular decisions are made

Questions on transparency

- What of Intellectual property issues? Competitors?
- Can people make sense of AI models?
- Will seeing the code help especially without the data that was used?
- Will companies be willing to share data (privacy issues)?

How do you create transparent AI?

- Technical correctness should be checked
- Developer of the model should explain how they addressed the problem, what technology was used and why, and what data sets where used.
- Others should be able to replicate the process or audit it
- Check that no groups are under-represented in the outcomes; make changes if necessary (can detect hidden bias)
- Use open source AI with care; if you understand how the model works
- Companies need to understand the technologies they use for decision making
- Follow the latest research

General Data Protection Regulation - GDPR

- Privacy & security law of the EU
- Imposes obligations on companies anywhere as so long as they target or collect data of people in the EU
- **Lawfulness, fairness and transparency** — Processing must be lawful, fair, and transparent to the data subject.
- **Accountability** — The data controller is responsible for being able to demonstrate GDPR compliance with all of these principles.

Challenges of openness

- What of Intellectual property issues? Competitors?
- Can people make sense of AI models?
- Will seeing the code help especially without the data that was used?
- Will companies be willing to share data (privacy issues)?
- Others?

AI Ethics

- What is good or bad? What is right or wrong?
- How developers, manufacturers, authorities and operators should behave to minimize ethical risks that results from using AI in society
  - Design, inappropriate application, intentional misuse of the technology
- Impact on decision making, equality and wellbeing
- Impact several societal sectors: employment, labour, social interaction, healthcare etc.
- Change human experience, raise concerns over the reliability of information sources, human dignity
- Who decides what is right/wrong, good/bad?

AI ethics

- Is it ethical to secretly collect personal data about someone without their knowledge and then sell that data to multiple third parties for use in targeted marketing?
- Is it ethical to ask a person to consent to tracking for the sole purpose of analyzing and improving their experience on a website, and then to honor their consent (or lack thereof)
- Is it ethical to give a person the option to opt out of being tracked and then, if they do not opt out, to track their behavior and use that data to market to them with targeted banner ads on other sites as they browse the internet?
- Is it ethical to give someone access to a website only if they consent to being tracked?

Source: 97 Things About Ethics Everyone in Data Science Should Know, Bill Franks. O'Reilly, 2020

Ethical framework for AI

- Non-maleficence. AI should be good, should not cause harm
- Responsibility or accountability. Who should be held responsible when AI causes harm
- Transparency and explainability. Understanding what and why AI does what it does
- Justice and fairness. AI should not discriminate, should be fair
- Respect for various human rights (privacy, security). AI should respect human rights

Conclusion

- Bias exists in AI systems; this can effect people in the real world
- AI use should be fair
- There should be accountability structure in place so people do not lose trust
- Transparency can lead to trust
- AI should be ethical

References

- *Understand, Manage, and Prevent Algorithmic Bias: A Guide for Business Users and Data Scientists*. Tobias Baer. Apress, 2019
- *97 Things About Ethics Everyone in Data Science Should Know,* Bill Franks. O'Reilly, 2020
- https://www.computerweekly.com/feature/Accountability-is-the-key-to-ethical-artificial-intelligence-experts-say
- https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai.html#toc1
- https://hbr.org/2021/08/how-to-build-accountability-into-your-ai
- https://www.computerweekly.com/feature/Accountability-is-the-key-to-ethical-artificial-intelligence-experts-say
- https://www.ibm.com/design/ai/ethics/accountability
- https://www.ibm.com/design/ai/ethics/fairness
- https://homes.cs.washington.edu/~msap/pdfs/sap2019risk.pdf
- https://www.vox.com/recode/2020/2/18/21121286/algorithms-bias-discrimination-facial-recognition-transparency
- https://hbr.org/2019/11/when-algorithms-decide-whose-voice-will-be-heard