Ethics
on Big Data & AI

Pablo Orviz <orviz@ifca.unican.es>

Master en Data Science
2019/20

CSIC
Consejo Superior de Investigaciones Científicas

UIMP
Universidad Internacional
Menéndez Pelayo

UC
UNIVERSIDAD
DE CANTABRIA

What are Ethics?

"Principles or morals that shape our (individual or group) behavior and actions in certain situations"

What are Ethics?

Ethical Dilemmas

An **ethical dilemma** or **ethical paradox** is a decision-making problem between two possible moral imperatives, neither of which is unambiguously acceptable or preferable.

The complexity arises out of the situational conflict in which obeying one would result in transgressing another. Sometimes called ethical paradoxes in moral philosophy, ethical dilemmas may be invoked to refute an ethical system or moral code, or to improve it so as to resolve the paradox

(Wikipedia, 2019)

What are Ethics?

Approaches to Applied Ethics

***Applied Ethics** refers to the practical application of moral considerations in real-world situations*

- Utilitarianism (or Consequentialism): *What is the greatest possible good for the greatest number?*
  - Best choice is the one that maximizes *utility* (== greatest amount of good consequences)
  - Sum of all pleasure that results from an action, minus the suffering of anyone involved in that action
- Deontological ethics (or Non-Consequentialism): Does an action follow a moral rule?
  - <u>Golden rule</u>: “Treat others how you want to be treated”
  - Act according to the agreed moral rule, no matter the consequences
- Virtue ethics: Does an action contribute to virtue?
  - ..whatever that means (blame Aristotle)
  - Right action will be that chosen by a suitability *virtuous agent*
  - Focus less on actions or consequences and rather places all of the pressure on the moral character of the person who does the action

The Trolley Dilemma

https://www.youtube.com/watch?v=bOpf6KcWYyw

Ethical approaches to the Trolley Dilemma

1. Utilitarianism (consequentialism)
    a. Original problem:
        - Killing 1 person is better than killing 5
    b. Fat man problem:
        - Same outcome as above, no matter how: 1 is better than 5
    c. Conclusion: **The ends never justify the means**
2. Deontology (non-consequentialism)
    a. Original problem:
        - Moral rule: pulling the lever is a good/neutral act by itself
    b. Fat man problem:
        - Moral rule: pushing somebody off the bridge is not OK
    c. Conclusion: **All things in life are contextual**
3. Virtue ethics
    a. Original problem:
        - A virtuous person would say it is morally required to switch the track
    b. Fat man problem:
        - A virtuous person would never push somebody to stop the trolley, his motivations wouldn't be pure
    c. Conclusion: ***What is a virtuous motivation?*** **Turns out varies widely between people, cultures, and geographic locations**

Ethical approaches to the Trolley Dilemma

Would this approaches change if the fat man was the villain that put the 5 people on the peril?

The Trolley Dilemma - The solution?

https://www.youtube.com/watch?v=-N_RZJUAQY4

Ethics issues (and solutions..)

main source: *The Hitchhiker’s Guide to AI Ethics*

Ethics in Software

- Ethics provides *rules* or *decision paths* to determine what is good or right
    - This leads to the **predictability** of the outputs
        - Predictable inputs conduct to predictable outputs
        - When software is well designed and tested
    - But..Can decisions be universally good or bad?
        - Usually chosen by the development teams
- Technology is **not neutral**
    - Ethics of a technology starts with the ethics of its creation, and its creators
- Programming Ethical Guidelines / Code of Ethics
    - Association for Computing Machinery (ACM)
    - IEEE Code of Ethics

What about AI algorithms?

- In AI, predictability is not guaranteed
  - Output is not a certainty but merely a prediction
  - Missing data/inputs/decision paths -(affects)-> prediction -(affects)-> end outcomes -(affects)-> Impact on humans
- Ethics of AI lies on **ethical quality** of its:
  - Prediction
  - End outcomes
  - Impact on humans
- Ethical quality -> moral obligations and duties of developers of AI
  - *How right, how fair and how just is an AI system's output, outcome and impact?*

Why Ethics on AI matters?

1. AI ~~has the potential to~~ already *changing the world*, pushed by:
   - Accessible (but costly) Cloud service providers
   - Open source machine learning libraries
2. This *change* is going to be *rapid and at scale*
   - Unintended harms will also occur at scale
3. AI has the potential to *cause harm* (biases, ..)
   - A harm is caused when a prediction or end outcome negatively impacts:
     i. Individual's identity (*harms of representation*)
     ii. Ability to access resources (*harms of allocation*)
   - Ethical obligations of AI system creators: mitigate all such harms

Issues on the Ethics of AI

- What AI is: *ethics issues stemming from data (inputs), models (learner) or predictions (inferer*)*
  - Bias and Fairness
  - Accountability and Remediability
  - Transparency, Interpretability and Explainability
- What AI does: *ethics issues that arise from AI systems that indirectly change our behavior as they take control of our operating environments*
  - Safety
  - Human-AI interaction
  - Cyber-security and Malicious Use
  - Privacy, Control and Surveillance
- What AI impacts: *ethics issues of not foreseen consequences (e.g. AI on social media)*
  - Automation, Job loss, Labor trends
  - Impact to Democracy and Civil rights
  - Human-Human interaction

Issues on the Ethics of AI

- What AI is: *ethics issues stemming from data (inputs), models (learner) or predictions (inferer*)*
    - **Bias and Fairness**
    - Accountability and Remediability
    - Transparency, Interpretability and Explainability
- What AI does: *ethics issues that arise from AI systems that indirectly change our behavior as they take control of our operating environments*
    - Safety
    - Human-AI interaction
    - Cyber-security and Malicious Use
    - Privacy, Control and Surveillance
- What AI impacts: *ethics issues of not foreseen consequences (e.g. AI on social media)*
    - Automation, Job loss, Labor trends
    - Impact to Democracy and Civil rights
    - Human-Human interaction

Issues on the Ethics of AI

What AI is

Bias and Fairness (or lack thereof)

- Biased algorithms can lead to unfair outcomes, discrimination and injustice **at scale**
- Humans inherently transfer cognitive biases to decision-making in AI systems
  - Garbage-in-Garbage-Out slogan
- But *biased data is only part of the story*
  - AI systems can amplify human biases, especially in “black box” or Discriminative Models
- “Structural bias is a social issue first and a technical issue second” (Kate Crawford)
  - A model not representing the input data -> ML problem
  - A model reflecting unfair predictions -> more than a ML problem

THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS

Issues on the Ethics of AI

What AI is

Bias and Fairness (or lack thereof)

- AI researchers can can seep bias into an AI system depending on how
  - Frame the problem
  - Train the model
  - Deployed the AI system
- ML algorithms understand data through patterns, relying on *features* identified by humans
  - E.g. if we generalize data, the ones with features too unique will get ignored -> source of discrimination
- Sources of Bias (Harini Suresh approach)
  - *Historical* bias: already exists in the data
  - *Representation* and *Measurement* biases: result of how the dataset is created
  - *Evaluation* and *Aggregation* biases: result of choices made during the model building

Issues on the Ethics of AI

What AI is

Historical bias

- Problem statement: biases sneak in training data and how machine learning mechanisms reinforce them, causing more discrimination and injustice
- Publication: Semantics derived automatically from language corpora contain human-like biases
- Machine learning technique known as *"word embedding"*
    - Builds up a mathematical representation of language, in which the meaning of a word is distilled into a series of numbers (known as a word vector) based on which other words most frequently appear alongside it
        - Words for flowers are clustered closer to words linked to pleasantness, while words for insects are closer to words linked to unpleasantness
    - Is transforming the way computers interpret speech and text
    - Future? May involve machines developing human-like abilities such as common sense and logic

Issues on the Ethics of AI

What AI is

Representation and Measurement biases

“Whether by design or as unintended consequences, the process of constructing data build social values and patterns of privilege into the data”

- Consider the example of *Street Bump* application
  - The app records patches/potholes/.. of cities while users drive through mobile phone’s GPS
  - The dataset created is then useful to make road work more efficient and targeted
  - Ethical issue: data is only provided by those who own smartphones (poorer, old-aged neighborhoods will be left out)
  - Risk of social exclusion?

Issues on the Ethics of AI

What AI is

Evaluation and Aggregation biases

- Filter Bubble and the Confirmation Bias
    - Result of a personalized search in which a website algorithm selectively guesses what information a user would like to see based on her/his profile
- Term first used in Eli Pariser's *The Filter Bubble* where he questioned the benefits of personalized content like Facebook's EdgeRank algorithm, Netflix's movie suggestions and Amazon's book recommendations
    - *"The Filter Bubble introduces three dynamics we've never dealt with before: first, you are alone in it, as it is you own personal bubble. Second, it is invisible in its actions. Finally, you don't choose to enter into the bubble."*
- Impacts
    - You are only being given news stories and social media posts biased to your existing beliefs, **isolating users from differing in viewpoints and perspectives**.
    - Reinforcement of a **narrow view** -> radicalization & sectarism
        - Well-known in a physical group since it is apparent
    - **Denial of awareness**: subject is not aware of being victim of a filter bubble
- Damaging reach: Fake news (faking culture)

Issues on the Ethics of AI

- What AI is: *ethics issues stemming from data (inputs), models (learner) or predictions (inferer*)*
  - Bias and Fairness
  - **Accountability and Remediability**
  - Transparency, Interpretability and Explainability
- What AI does: *ethics issues that arise from AI systems that indirectly change our behavior as they take control of our operating environments*
  - Safety
  - Human-AI interaction
  - Cyber-security and Malicious Use
  - Privacy, Control and Surveillance
- What AI impacts: *ethics issues of not foreseen consequences (e.g. AI on social media)*
  - Automation, Job loss, Labor trends
  - Impact to Democracy and Civil rights
  - Human-Human interaction

Issues on the Ethics of AI

What AI is

Accountability and Remediability

- Way out of bias?
    - De-biasing techniques: adjusting imbalances in data
        - Requirement: Biases need to be already identified in the dataset
        - Not enough by itself
    - ***Accountability***
        - AI challenges the traditional conception of responsibility -> AI learns from data, rather than 100% coded
        - Accountability in AI system development:
            - Identify norms of the specific community where AI system will be deployed
            - Identify the features appropriate for use
            - Identify dignity & rights in the situated use
        - Achieved by human audits, impact assessments or via governance through policy or regulation
            - Governance through *human-in-the-loop* -> high-risk decisions to be done by humans
            - Regulations
                - Google's Perspectives on Issues in AI Governance
                - European Commission's Ethics Guidelines for Trustworthy AI (AI HLEG)
                - European Commission's Ethics and Data Protection

Issues on the Ethics of AI

What AI is

Google's Perspectives on Issues in AI Governance

1. *Explainability standards*
   - ○ Explain AI system behaviour to boost trust in society
   - ○ Not straightforward to deliver in practice
2. *Fairness appraisal/evaluation*
   - ○ Fairness is not an universal concept; governments and policy makers (regulations) play a vital role
   - ○ AI systems could be used to identify human and societal biases
3. *Safety considerations*
   - ○ Continuous monitoring, automatic failover to a neutral state, 2-layer verification, ..
4. *Human-AI collaboration (Human-on-the-loop)*
5. *Liability/Accountability frameworks*
   - ○ Persons or organizations should be the ultimate responsible for the actions of AI systems

Issues on the Ethics of AI

What AI is

European Commission’s Ethics Guidelines for Trustworthy AI

- Trustworthy AI: foundational ambition
  - *Lawful/legal:* compliant with laws and regulations
  - *Ethical:* adherence with ethical principles and values
  - *Robust:* both from a technical and social perspective
- Set of Ethical principles (adherence)
  - *Respect for human autonomy*: dignity and liberty -> Charter of Rights
  - *Prevention of harm*
  - *Fairness*
  - *Explicability*
- 7 requirements for AI systems to meet Trustworthy AI (implementation)
  - *Human agency and oversight*
  - *Technical robustness and safety*
  - *Privacy and data governance*
  - *Transparency*
  - *Diversity, non-discrimination and fairness*
  - *Societal and environmental wellbeing*
  - *Accountability*

European Commission

The European Commission’s

HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE

AI

DRAFT

ETHICS GUIDELINES FOR TRUSTWORTHY AI

Working Document for stakeholders’ consultation

Brussels, 18 December 2018

Issues on the Ethics of AI

What AI is

Accountability and Remediability

- Way out of bias?
    - ***Remediability***
        - What happens once damage is done?
            - Remediation process:
                - World Economic Forum’s How to Prevent Discriminatory Outcomes in Machine Learning
            - Investigative journalism, push for accountability and action:
                - ProPublica, Algorithmic Justice League, AI Now Institute
            - Outcomes: AI systems withdrawn, modified or dismissed

Issues on the Ethics of AI

What AI is

Accountability and Remediability

- ProPublica investigation case: Risk scores in criminal justice system (Kirkpatrick, 2016)
    - Software gaining momentum as it is able to determine the likelihood of committing future crimes
        - Widely **used in the U.S. criminal justice system**
        - Scores are computed based on the result of 137 questions answered by either the defendant or pulled from criminal records
            - Defendant’s race is not one of the questions
    - *Pro Publica*, a non-profit investigative journalism organization, challenged the output of this algorithm
        - They state that some of the **questions may be seen as highly impacting blacks**:
            - Was one of your parents ever sent to jail or prison?
            - How many of your friends are taking drugs illegaly?
    - According to *Pro Publica*, the risk scores examined across 2013 and 2014
        - **Proved unreliable in forecasting violent crimes, with 20% of success**
        - The algorithm falsely flagged black defendants as future criminals, wrongly labelling them at almost twice the rate of white defendants

Issues on the Ethics of AI

- What AI is: *ethics issues stemming from data (inputs), models (learner) or predictions (inferer*)*
  - Bias and Fairness
  - Accountability and Remediability
  - **Transparency, Interpretability and Explainability**
- What AI does: *ethics issues that arise from AI systems that indirectly change our behavior as they take control of our operating environments*
  - Safety
  - Human-AI interaction
  - Cyber-security and Malicious Use
  - Privacy, Control and Surveillance
- What AI impacts: *ethics issues of not foreseen consequences (e.g. AI on social media)*
  - Automation, Job loss, Labor trends
  - Impact to Democracy and Civil rights
  - Human-Human interaction

Issues on the Ethics of AI

What AI is

Transparency, Interpretability and Explainability

- Way out of bias?
    - ***Transparency, Interpretability and Explainability***
        - Most of ethical concerns arise from the “black box” behaviour, still existing since
            - Companies reluctant to share the “secret sauce”
            - Involve large complex math operations
        - *Transparency* research progresses on the interpretation of how algorithms learn (layer by layer)
        - *Interpretability* research focuses on opening up the black box
        - *Explainability* research tries to understand the decision
        - Ongoing discussion:
            - Only outputs of algorithm need to be justifiable or this is insufficient
            - Algorithms should explain by themselves
        - Opponents to full transparency rely on the argument of being an obstacle for rapid progress in maximising AI efficiency and accuracy

Issues on the Ethics of AI

What AI is

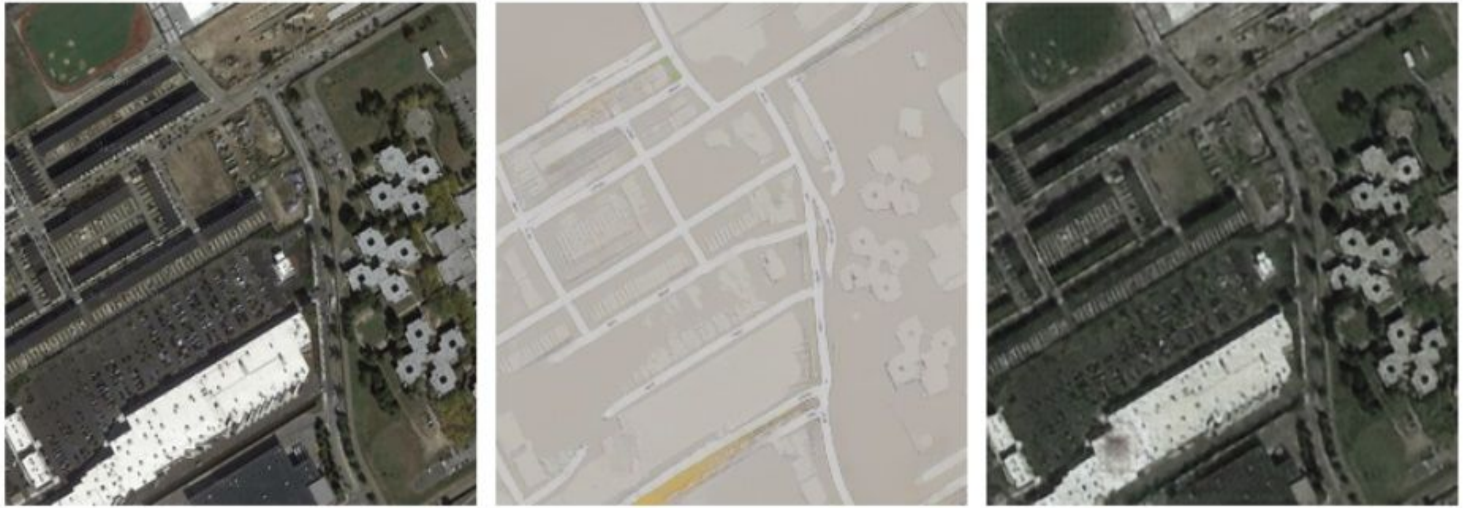
Black box in AI: Unpredictable Behavior

- Problem statement: AI algorithms may no longer execute in predictable contexts, requiring new kinds of safety assurance and the engineering of artificial ethical considerations
- NN called CycleGAN was trained to transform aerial images into street maps and then back again into aerial images
- They found that details were omitted in final image, reappeared when it was reverted
  - For instance, skylights on a roof that were eliminated in the process of creating the street map would magically reappear when they asked the agent to do the reverse process
- The system did not learn to match the features of either type of map, but **it learned to subtly encode the features of one into the noise patterns of the other**
  - This practice of encoding data into images isn’t new; it’s an established science called steganography
  - But a computer creating its own steganographic method to evade having to actually learn to perform the task at hand *is* rather new
- This study revealed that AI systems, if not explicitly prevented from doing so, may find an alternative ways (other than developers think) to address problems
  - **“machines are getting smarter” OR “machines do exactly what they are asked”**
  - Raises the need of **understanding “black boxes”**

Issues on the Ethics of AI

What AI is

Black box in AI: Unpredictable Behavior



(a) Aerial photograph: x.

(b) Generated map: Fx.

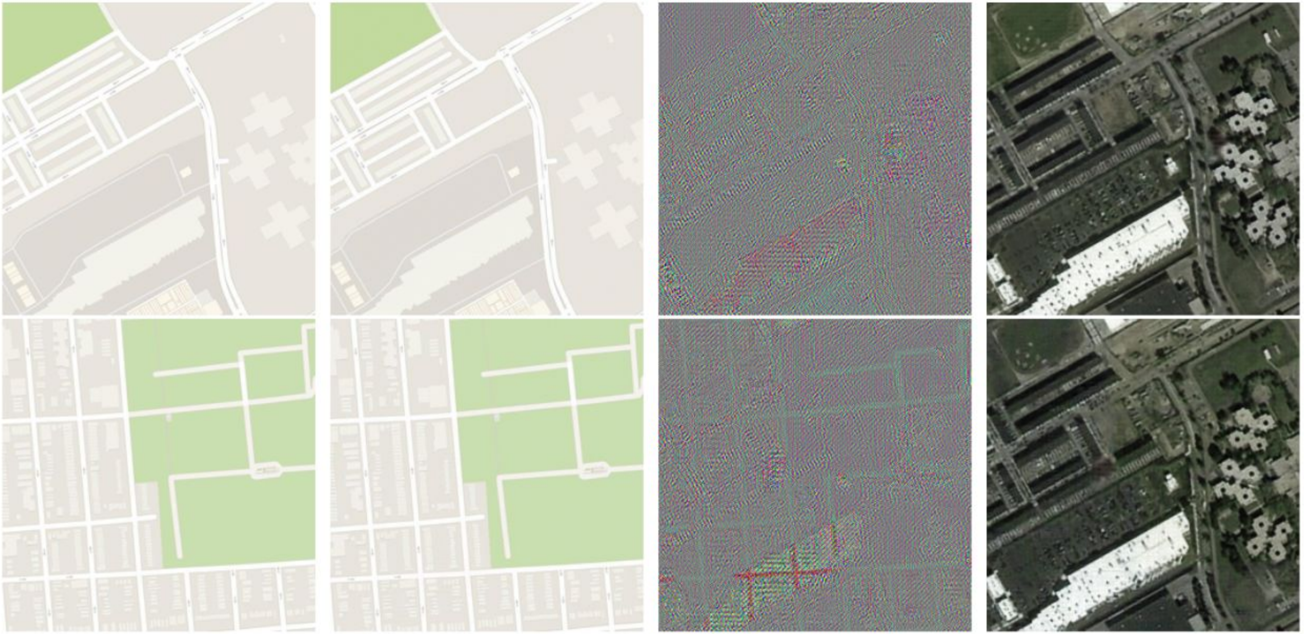
(c) Aerial reconstruction: GFx.

Note the presence of dots on both aerial maps not represented on the street map.

Issues on the Ethics of AI

What AI is

Black box in AI: Unpredictable Behavior



(a) Source map: y_0. (b) Crafted map: y^*. (c) Difference: $y^* - y_0$. (d) Reconstruction: Gy^*.

Issues on the Ethics of AI

- What AI is: *ethics issues stemming from data (inputs), models (learner) or predictions (inferer*)*
    - Bias and Fairness
    - Accountability and Remediability
    - Transparency, Interpretability and Explainability
- What AI does: *ethics issues that arise from AI systems that indirectly change our behavior as they take control of our operating environments*
    - **Safety**
    - Human-AI interaction
    - Cyber-security and Malicious Use
    - Privacy, Control and Surveillance
- What AI impacts: *ethics issues of not foreseen consequences (e.g. AI on social media)*
    - Automation, Job loss, Labor trends
    - Impact to Democracy and Civil rights
    - Human-Human interaction

Issues on the Ethics of AI

What AI does

Safety

“AI must not cause accidents, or exhibit unintended or harmful behavior”

- Even with no Bias, an AI system can be used to help or harm us
    - Decision making in AI entails a huge responsibility
- *Autonomous* AI system is not a rules-based system
    - It mimics human behaviour, the decision-making is more complex
    - Human actions are determined by intentions, norms, values and biases
    - Conception of *safe* changes with time and context
- Thus, an AI system shall:
    - Be responsive to contexts as they arise
    - Be able to model this uncertainty it its environment
    - Be aligned on what is “right” -> key theme
        - ***Value Alignment*** principle or How to align AI with Human Values

Issues on the Ethics of AI

What AI does

Value-alignment

“Highly autonomous AI systems should be designed so that their goals and behaviors can be assured to align with human values throughout their operation”

- Big challenge in AI: How do we ensure that an AI will do **what we really want,** while not harming humans?
  - Reality:
    - Different opinions and cultures, conflicting values
    - Emotional and sentimental values
  - Approaches:
    - Only consider the values generally accepted?
      - What about cultures?
    - Humans-in-the-loop: risky decisions on humans
    - Continuous refinement between humans and AI
    - ...

Issues on the Ethics of AI

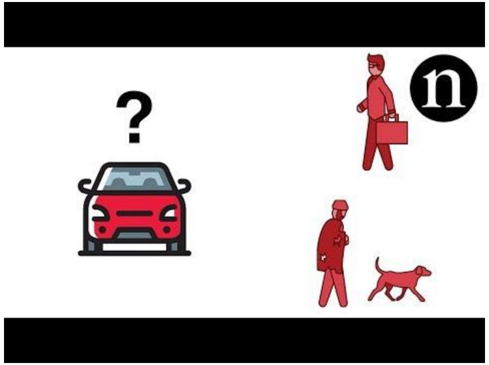
What AI does

Trolley problem on AI: Moral Machine Platform

- Trolley problem become the canonical example for self-driving cars
- Self-driving cars might soon have to make ethical judgements
- How to code societal values into autonomous vehicles?
- Catalog human opinion on how future machine intelligence should respond in various conditions
  - Invites users to judge a series of hypothetical scenarios through a survey
  - Rank of "preferences" as a result
- Why?
  - Enabling a machine to make decisions demands declarations of our more fundamental values, on which those decisions should rest

http://moralmachine.mit.edu/

https://www.youtube.com/watch?v=jPo6bby-Fcg



MORAL MACHINE

https://www.youtube.com/watch?v=XCO8ET66xE4

Issues on the Ethics of AI

http://moralmachine.mit.edu/

What AI does

Trolley problem on AI: Moral Machine Platform

- Results per country: http://moralmachineresults.scalablecoop.org/
- ~2.3 millions survey (https://doi.org/10.1038/s41586-018-0637-6, 2018)
  - Highlights:
    - **Patterns by country**
      - E.g. prosperous countries with strong institutions are less likely to spare a pedestrian who stepped into traffic illegally
    - **Patterns by dominant religion areas**
    - **Patterns by level economic inequality**
      - Small gaps between rich and poor showed little preference (e.g. Finland) while on countries with greater economic disparity (e.g. Colombia) choosed to kill the lower-status person.
    - **Prioritize Humans** (over pets) & **Groups of people** (over individuals)
      - In line with 2017 German Ethics Commission on Automated and Connected Driving report
    - Self-driving cars are safer but..
      - Ethical paradox: based on the results, *would you buy a car programmed to prioritize protecting pedestrians?*

Issues on the Ethics of AI

- What AI is: *ethics issues stemming from data (inputs), models (learner) or predictions (inferer*)*
    - Bias and Fairness
    - Accountability and Remediability
    - Transparency, Interpretability and Explainability
- What AI does: *ethics issues that arise from AI systems that indirectly change our behavior as they take control of our operating environments*
    - Safety
    - **Human-AI interaction**
    - Cyber-security and Malicious Use
    - Privacy, Control and Surveillance
- What AI impacts: *ethics issues of not foreseen consequences (e.g. AI on social media)*
    - Automation, Job loss, Labor trends
    - Impact to Democracy and Civil rights
    - Human-Human interaction

Issues on the Ethics of AI

What AI does

Human-AI interaction

“The impact of algorithms, positive and negative, on our mental and emotional wellbeing is also cause for concern”

- Human-AI interaction cannot be understated
- Think about the following use cases:
  - Facebook’s suicide prevention algorithm, actually saving lives
  - When algorithms think you want to die, social media platforms (Instagram, Pinterest) sending recommendations of suicide and self-harm images, based on the preferences of a suicidal teenager
  - Amazon’s Alexa to combat loneliness

Issues on the Ethics of AI

- What AI is: *ethics issues stemming from data (inputs), models (learner) or predictions (inferer*)*
  - Bias and Fairness
  - Accountability and Remediability
  - Transparency, Interpretability and Explainability
- What AI does: *ethics issues that arise from AI systems that indirectly change our behavior as they take control of our operating environments*
  - Safety
  - Human-AI interaction
  - **Cyber-security and Malicious Use**
  - Privacy, Control and Surveillance
- What AI impacts: *ethics issues of not foreseen consequences (e.g. AI on social media)*
  - Automation, Job loss, Labor trends
  - Impact to Democracy and Civil rights
  - Human-Human interaction

Issues on the Ethics of AI

What AI does

Cyber-security and Malicious use

- AI systems can simultaneously empower cyber-security (solution) and being susceptible to malicious use (threat)
- Cyber-attackers use AI against their users
  - Highly autonomous AI systems are the most concerning
- E.g.: Flaws of an utilitarian self-driving car implementation
  - Use stickers in traffic signs to fool autonomous cars (using DNNs)
    - Stop sign misread as a speed limit sign,
    - Right turn sign to be classified as either a stop or added lane sign.
  - Laying a trap

Issues on the Ethics of AI

What AI does

Cyber-security and Malicious use

Autonomous systems: Easy to exploit?

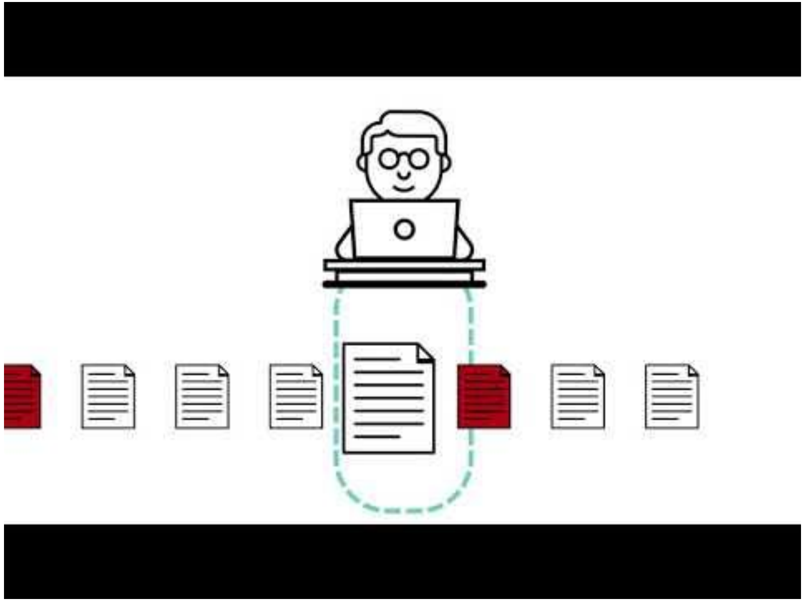
- Algorithm: values children’s lives higher than the one of adults
- Scenario:
  - A murderer wishes to kill another person, the victim
  - Murderer knows that the victim uses the same path in his self-driving car every day at 9:00am
  - There is a school located at one point of the victim’s way to work
  - Murderer positions himself in front of the school, waits for the victim and his self-driving car
  - Murderer sees the victim approaching and pushes a child --on the way to school-- onto the street in front of the victim’s self-driving car
  - The AI system chooses to sacrifice the driver as it knows that the sharp turn will end directly into a concrete wall

Issues on the Ethics of AI

What AI does

Cyber-security and Malicious use

Ethical Hacker Dilemma



https://www.youtube.com/watch?v=wGZil-NAaac

Issues on the Ethics of AI

- What AI is: *ethics issues stemming from data (inputs), models (learner) or predictions (inferer*)*
    - Bias and Fairness
    - Accountability and Remediability
    - Transparency, Interpretability and Explainability
- What AI does: *ethics issues that arise from AI systems that indirectly change our behavior as they take control of our operating environments*
    - Safety
    - Human-AI interaction
    - Cyber-security and Malicious Use
    - **Privacy, Control and Surveillance**
- What AI impacts: *ethics issues of not foreseen consequences (e.g. AI on social media)*
    - Automation, Job loss, Labor trends
    - Impact to Democracy and Civil rights
    - Human-Human interaction

Issues on the Ethics of AI

What AI does

Privacy, Control and Surveillance

- Misuse of the tech includes also the AI’s ability to be repurposed or intentionally designed for surveillance
- While defining **privacy** is hard, identifying its violation is intuitive
    - Why? Behind the many definitions is something fundamentally human: **dignity** and **control**
    - **Basic human right:** 1st article of the EU Charter of Fundamental Rights
- Privacy violation could be justified (e.g. for a bigger benefit, such as public safety) outside the digital, big data, AI world
    - But in an AI world, risks of privacy violation are not immediate nor obvious: *greater than the sum of the parts*
    - Consider Facial Recognition Technology -> most virulent form of privacy-violation-made-easy-by-tech
        - AI can do facial analysis, skin texture analysis, speech recognition, etc.. without permission or cooperation from the individual
        - Facial recognition to streamline cruise ship boarding process
        - Fingerprint authentication for payment cards

Issues on the Ethics of AI

What AI does

Privacy, Control and Surveillance

- **Privacy dilemma**: limiting the use of sensitive data
    - E.g.: In a health context, we as patients care about the privacy of our medical record; but also would like to benefit from the obtained from the processing of this data
- Individual’s benefit from broadcasting personal data (or even Open Data)
    - Challenged by several authors: in practice Open Data isn’t necessarily accessible to everyone
        - (Johnson, 2014) “Open data projects remain dominated by government and business users: enterprises have the capacity to take advantage of big, open data, a capacity that citizens lack..The result is that Big Data is not, in practice, open to citizens”

Issues on the Ethics of AI

- What AI is: *ethics issues stemming from data (inputs), models (learner) or predictions (inferer*)*
    - Bias and Fairness
    - Accountability and Remediability
    - Transparency, Interpretability and Explainability
- What AI does: *ethics issues that arise from AI systems that indirectly change our behavior as they take control of our operating environments*
    - Safety
    - Human-AI interaction
    - Cyber-security and Malicious Use
    - Privacy, Control and Surveillance
- What AI impacts: *ethics issues of not foreseen consequences (e.g. AI on social media)*
    - **Automation, Job loss, Labor trends**
    - Impact to Democracy and Civil rights
    - Human-Human interaction

Issues on the Ethics of AI

What AI impacts

Automation, Job Loss, Labor Trends

- Landscape of labor is/will be disrupted with AI
    - Not entirely clear how fast this change will occur
    - Should be taken seriously to avoid violations of human rights: dignity, ..
- Stories:
    - AI creating millions of jobs
    - Factory workers being replaced by robots
- Forecast about AI influence on people and society
    - Partnership on AI
    - Brookings
    - Obama White House

Issues on the Ethics of AI

- What AI is: *ethics issues stemming from data (inputs), models (learner) or predictions (inferer*)*
  - Bias and Fairness
  - Accountability and Remediability
  - Transparency, Interpretability and Explainability
- What AI does: *ethics issues that arise from AI systems that indirectly change our behavior as they take control of our operating environments*
  - Safety
  - Human-AI interaction
  - Cyber-security and Malicious Use
  - Privacy, Control and Surveillance
- What AI impacts: *ethics issues of not foreseen consequences (e.g. AI on social media)*
  - Automation, Job loss, Labor trends
  - **Impact to Democracy and Civil rights**
  - Human-Human interaction

Issues on the Ethics of AI

What AI impacts

Democracy and Civil Rights

"Power always learns, and powerful tools always fall into its hands." — Zeynep Tufecki (AI, techno-sociologist)

- Effects of AI in the hands of the powerful, e.g. Mass Surveillance in China: network of monitoring systems used by the Chinese government to supervise the lives of Chinese citizens
  - Tight control over the Internet through public regulation
    - Restrictions on publication/distribution of online news (blogs, social media)
    - Major Internet platforms and messaging services apply self-censorship mechanisms: WeChat is under continuous surveillance, conversations stored for six months
    - VPNs from main operators blocked
  - 200 million government surveillance videocameras across the country (~1 per 7 citizens)
    - Government uses AI facial recognition tech to identify each person captured and create activity profiles
- Democracies and civil rights suffer also by the *fragmentation of truth and loss of trust*
  - Culture of fakery
    - Year after year, less than 60% of web traffic is from humans
    - Bots or bad-actors generate content customised for virality, and this affects how we consume information

Issues on the Ethics of AI

- What AI is: *ethics issues stemming from data (inputs), models (learner) or predictions (inferer*)*
  - Bias and Fairness
  - Accountability and Remediability
  - Transparency, Interpretability and Explainability
- What AI does: *ethics issues that arise from AI systems that indirectly change our behavior as they take control of our operating environments*
  - Safety
  - Human-AI interaction
  - Cyber-security and Malicious Use
  - Privacy, Control and Surveillance
- What AI impacts: *ethics issues of not foreseen consequences (e.g. AI on social media)*
  - Automation, Job loss, Labor trends
  - Impact to Democracy and Civil rights
  - **Human-Human interaction**

Issues on the Ethics of AI

What AI impacts

Human-Human Interaction

- Effects:
  - Gendered AI promotes stereotypes and discrimination (UNESCO's "I'd blush if I could" report)
    - Promotes education to close gender divides in digital skills, e.g.:
      - *Siri* voice assistant: powerful illustration of gender biases coded in tech
        - Projected as a young woman, submissive and servant
        - Human user would tell 'her', "Hey Siri, you're a bi***." -> as of Ap 2019, Siri has been updated to say "I don't know how to respond to that"
  - Natural language AI leads to a loss of courtesy
    - Children hit hardest, they command unrespectfully
  - In "hybrid systems" (people and robots interact socially) the right kind of AI can improve the way humans relate to one another..but evil AI can make us less productive and ethical
  - Moral de-skilling: loss of skill at making moral decisions due to lack of experience and practice -> if AI makes decisions for us, humans will become de-skilled

Infusion 2,5€
PRIX DU CAFÉ EN TERRASSE
"UN CAFÉ"..........7€
"UN CAFÉ, S'IL VOUS PLAIT."........4,25€
"BONJOUR, UN CAFÉ, S'IL VOUS PLAIT."1,40€

Conclusion

Ethical AI

- Ethical and Moral philosophy is a theoretical field
    - No universal metrics for ‘good’ and ‘bad’
    - Debate of morals and virtue often left with more questions than answers
- AI tech is pushing for a Practical Ethics
    - Values are required to be built into them
    - If not explicitly defined -> Risk of unfair consequences for many
- Ethical landscaping and discussion towards Intentionality
    - Intentional systems reduce risk of unintended societal harm

Ideas to build ethical AI

- Ethics by Design
  - Ethical education on AI stakeholders (researchers, developers, deployers, users, ..)
  - Risk assessment
  - Build upon (self, governmental) regulations
- Improve the level of trust
  - Transparency
    - AI decision-making is key
    - Government policies and actions
  - Accountability
    - Roles and responsibilities
    - Goals and Purpose of critical algorithms are clearly defined and documented
- Align Cyber-security and AI Ethics
  - Prioritize security within AI algorithms development
- Mitigate Bias