



1

Part 5

Lecture 1 Data Sharing and Ethics



Who I am...

Pascal Tyrrell, PhD *Associate Professor*

Department of Medical Imaging, Faculty of Medicine

Institute of Medical Science, Faculty of Medicine

Department of Statistical Sciences, Faculty of Arts and Science



Overview for Part 5

- ❑ Ethics
 - ❑ Accountability
 - ❑ Value Alignment
 - ❑ Explainability
- ❑ Fairness
- ❑ Data Ownership
- ❑ Regulatory Implications

Consent is rooted in ethical treatment of patients – Nuremberg Trials

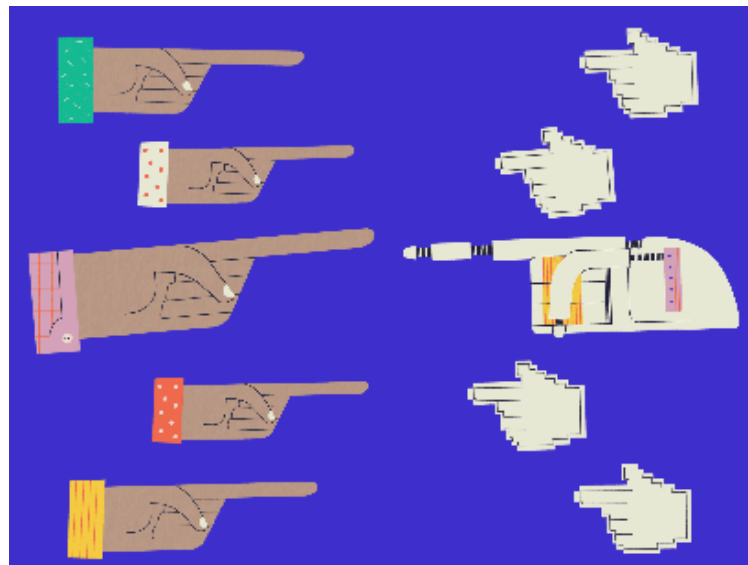
- Patients need and deserve the opportunity to control their health data
 - Primary use: Healthcare
 - Secondary use: Researchers need access to data for legitimate research



Problem 1: Accountability

Accountability

- AI designers and developers are responsible for considering AI design, development, decision processes, and outcomes
- Human judgement plays an important role, humans are the ones who write algorithms define a success and failure



Accountability

- ❑ Make decisions about the uses of systems and who may be affected by a systems outcomes
- ❑ Every person involved in the creation of AI at any step is accountable for considering the system's impact in the world



Accountability: How to Resolve the Ethical Concern

- ❑ Make clear, accessible company policies to the design and development team
- ❑ Everyone must know their responsibility
- ❑ Understand where the responsibility of the company/software ends



Accountability: How to Resolve the Ethical Concern

- ❑ There may not be any control over the data or the tool that will be used by a user client
- ❑ Detailed records of your design → Keep track of records during the process
 - ❑ Keep in mind and follow the company, national, international guidelines



Problem 2: Value Alignment



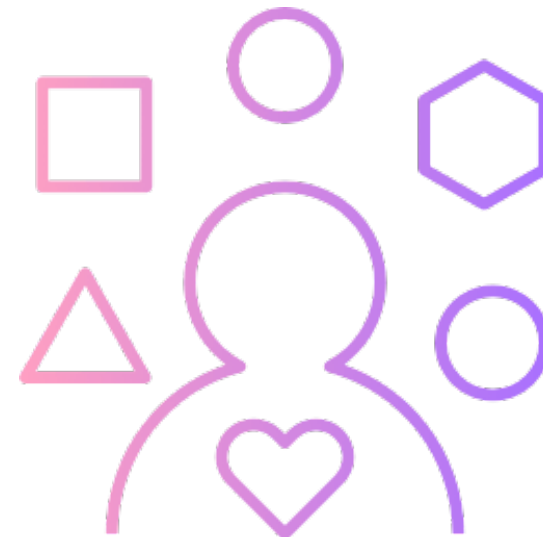
Value Alignment

- ❑ “If machines engage in human communities as autonomous agents, then those agents will be expected to follow the community’s social and moral norms. A necessary step in enabling machines to do so is to identify these norms. But whose norms?”
- ❑ The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems



Value Alignment

- ❑ AI must be designed to align with the norms and values of your user group in mind
- ❑ Humans use contextual factors, experiences, memories, upbringing, and cultural norms in order to make a decision and judging between “right and wrong”



Value Alignment

- ❑ AI does not have these experiences to draw upon
- ❑ Designers and developers must consider values of the user group of interest in order to create an ideal AI system
- ❑ Care is required to ensure sensitivity to a wide range of cultural norms and values



Value Alignment: How to Resolve the Ethical Concern

- ❑ Consider the culture that establishes the value systems you're designing within
- ❑ Get perspective from policy makers and academics that can help your team
- ❑ Map out a general understanding of values
- ❑ Values are subjective and differ globally



Trust and Approval From Patients

- ❑ Correct decision making is a function of the structure of the data used as input
- ❑ Need relationship between clinicians who understand the specifics of the clinical data and the developers creating the algorithms



Problem 3: Explainability



Explainability

- ❑ Decision making process should be explainable so everyone can understand
 - ❑ Important because explainability builds transparency, confidence, and trust
- ❑ Must be able to understand a decision process and the AI should have explained reasoning
 - ❑ Designed for humans to easily perceive, detect, and understand its decision process



Explainability: How to Resolve the Ethical Concern

- ❑ People should ask questions during any process of what an AI is doing
- ❑ Decision making processes must be reviewable → especially when using highly sensitive personal data
 - ❑ For example, personally identifiable information, protected health information, and/or biometric data

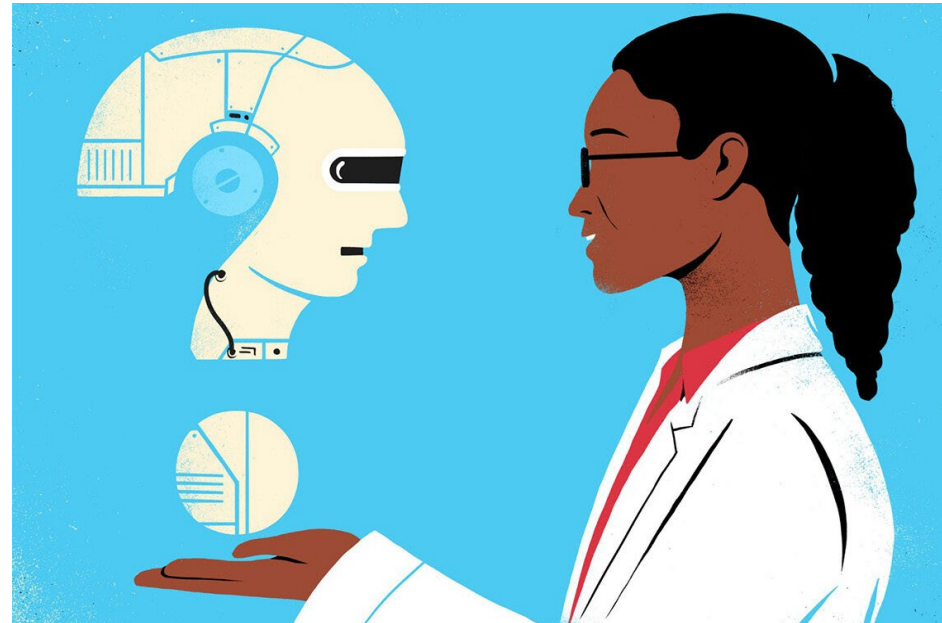


Fairness



Fairness

- ❑ AI that is built by humans increases the chances for human bias to be found in the systems
- ❑ AI has to be designed to minimize bias and be as inclusive as possible



Unconscious Biases: Shortcut Biases

- ❑ Availability bias: Overestimating certain events due to greater “availability”
- ❑ Base rate fallacy: Tendency to ignore general information and focus on specific information
- ❑ Congruence bias: Tendency to test hypotheses exclusively through direct testing, instead of testing alternative hypothesis



Unconscious Biases: Shortcut Biases

- ❑ Empathy gap bias: Tendency to underestimate the influence or strength of feeling, in either one's' self or other
- ❑ Stereotyping: Expecting a member of a group to have certain characteristics without having actual information about that individual



Impartiality Biases

- ❑ Anchoring bias: Relying too much on one trait or piece of information when making decisions
- ❑ Bandwagon bias: Tendency to do or believe things because many other people do
- ❑ Bias blindspot: Tendency to see oneself as less biased than others, or to be able to identify more cognitive biases in others than in oneself



Impartiality Biases

- ❑ Confirmation bias: Tendency to search for, interpret, or focus on information in way that confirms one's preconceptions
- ❑ Halo effect: Tendency of an overall impression to influence the observer. Positive feelings in one area causes ambiguous or neutral traits to be viewed positively



Self-Interest Biases

- ❑ Ingroup/outgroup bias: Tendency or pattern of favoring members of one's ingroup versus outgroup members
- ❑ Sunk cost bias: Tendency to justify past choices, even though they no longer seem valid
- ❑ Status quo bias: Tendency to maintain the current situation even when better alternatives exist



Self-Interest Biases

- ❑ Not invented here bias: Aversion to contact with or use of products, research, standards, or knowledge developed outside a group
- ❑ Self serving bias: Tendency to focus on strengths/achievements and overlook faults/failures



Steps to Resolve Bias

- ❑ If there is bias, team must investigate and understand where it originated and how it can be changed
- ❑ Design and develop without intentional biases and schedule team reviews to avoid unintentional biases
- ❑ Create a feedback mechanism or open dialogue with users





End of Lecture 1

Next up Part 5 Lecture 2: Data Ownership and Privacy

