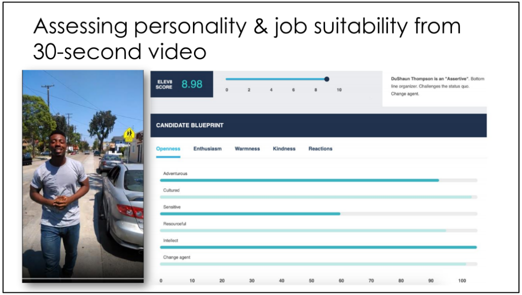
Building a Robot Judge:
Data Science for Decision-Making

12. Algorithms and Decisions IV

What are some problems with algorithmic hiring systems? (Raghavan et al, 2019)



Write down an answer privately for sharing with the group:

- **[Last name starts with A-M] Give an example situation where algorithmic hiring should be allowed and explain.**
- **[Last name starts with N-Z] Give an example situation where algorithmic hiring should not be allowed and explain.**
- **What are some restriction/regulations that address problems without banning algorithmic hiring?**

Outline

AI Governance

What can and should AI decide?

AI for legal decisions

Recap and Conclusion

- Algorithms influence various aspects of life:
  - selecting tax payers for audits
  - granting or denying immigration visas
  - security screening at airports
- Benefits many and growing:
  - efficiency, accuracy, scalability
  - increase consistency and reduce bias
  - economic/innovation

- Algorithms influence various aspects of life:
  - selecting tax payers for audits
  - granting or denying immigration visas
  - security screening at airports
- Benefits many and growing:
  - efficiency, accuracy, scalability
  - increase consistency and reduce bias
  - economic/innovation
- But AI has risks and harms.
  - Public interest requires governance to reinforce benefits and minimize risks.

Tradeoffs

- accuracy vs
  - equity
  - explainability
  - data privacy
- innovation vs
  - safety
  - transparency
  - data privacy
  - consumer rights

Challenges to developing standards

- Collective decision processes
  - tradeoffs among various stakeholders
  - distortions from lobbying
  - technical issues $\rightarrow$ politicians and voters have low information

Challenges to developing standards

- Collective decision processes
  - tradeoffs among various stakeholders
  - distortions from lobbying
  - technical issues $\rightarrow$ politicians and voters have low information
- Global coordination needed for digital tech
  - accounting for different cultures and contexts

Challenges to developing standards

- Collective decision processes
  - tradeoffs among various stakeholders
  - distortions from lobbying
  - technical issues $\rightarrow$ politicians and voters have low information
- Global coordination needed for digital tech
  - accounting for different cultures and contexts
- How to assign responsibility for risks/harms
  - creator / owner / operator/ user?
  - how to understand / determine intentions
  - balance accountability with innovation and growth

Governance Strategies

- Industry-driven approach:
  - Reduces regulatory red tape, could help innovation
  - No central authority to enforce best-practices;
  - Expands the power of large corporations.
  - Significant externalities, tendency to concentration

Governance Strategies

- Industry-driven approach:
  - Reduces regulatory red tape, could help innovation
  - No central authority to enforce best-practices;
  - Expands the power of large corporations.
  - Significant externalities, tendency to concentration
- Regulator-driven approach:
  - significant technical knowledge/skills needed to be effective – often led by lawyers rather than tech experts
  - bad actors always a step ahead.
  - limits innovation and expansion of digital economy.
  - could collude with industry leaders

Governance Strategies

- Industry-driven approach:
  - Reduces regulatory red tape, could help innovation
  - No central authority to enforce best-practices;
  - Expands the power of large corporations.
  - Significant externalities, tendency to concentration
- Regulator-driven approach:
  - significant technical knowledge/skills needed to be effective – often led by lawyers rather than tech experts
  - bad actors always a step ahead.
  - limits innovation and expansion of digital economy.
  - could collude with industry leaders
- Liability-based approach
  - litigation based on harms.
  - more flexible than regulation
  - can lead to under-enforcement and uncertainty on liability

AI Regulation Modalities

- Rules on inputs and training procedure
  - eg data protection, transparency
- Rules on outputs
  - eg red teaming for biases, state secrets
- Rules on consequences
  - eg litigation for harms
- Rules on ?
  - eg copyright

Transparency

- Closed-source algorithms result in “black box justice” and could be abused by insiders.
- But open-source algorithms are prone to gaming: savvy users could “trick” the algorithm.

Transparency

- Closed-source algorithms result in "black box justice" and could be abused by insiders.
- But open-source algorithms are prone to gaming: savvy users could "trick" the algorithm.
- How can we make sure that the decision maker is not merely claiming to follow the rules?
  - Disclose the trained model? training data? training code?

Transparency

- Closed-source algorithms result in "black box justice" and could be abused by insiders.
- But open-source algorithms are prone to gaming: savvy users could "trick" the algorithm.
- How can we make sure that the decision maker is not merely claiming to follow the rules?
  - Disclose the trained model? training data? training code?
- Policy challenges
  - ML processes not understandable by non-experts
  - Sometimes even experts don't understand the model
  - Understanding the code/model not the same as understanding behavior/responses

"An Economic Approach to Regulating Algorithms"

Rambachan, Kleinberg, Ludwig, and Mullainathan (2020)

- Apply welfare economics to the design and regulation of algorithmic decision processes.
- Focus on **post-processing approach** to fairness:
  - (1) training a prediction function, and (2) a decision rule based on the predictions.

"An Economic Approach to Regulating Algorithms"

Rambachan, Kleinberg, Ludwig, and Mullainathan (2020)

- Apply welfare economics to the design and regulation of algorithmic decision processes.
- Focus on **post-processing approach** to fairness:
  - (1) training a prediction function, and (2) a decision rule based on the predictions.

Result 1 (social planner):

- the equity preferences of the social planner have no effect on the training procedure for the prediction function.
- i.e., there should be no limit on the use of sensitive attributes.

“An Economic Approach to Regulating Algorithms”

Rambachan, Kleinberg, Ludwig, and Mullainathan (2020)

- Apply welfare economics to the design and regulation of algorithmic decision processes.
- Focus on **post-processing approach** to fairness:
  - (1) training a prediction function, and (2) a decision rule based on the predictions.

Result 1 (social planner):

- the equity preferences of the social planner have no effect on the training procedure for the prediction function.
- i.e., there should be no limit on the use of sensitive attributes.

Result 2 (private actors):

- key factor is disclosure of decision process (data, ML training, and decision rule), which, unlike human decision-making, allows prejudicial treatment to be detected.

“An Economic Approach to Regulating Algorithms”

Rambachan, Kleinberg, Ludwig, and Mullainathan (2020)

- Apply welfare economics to the design and regulation of algorithmic decision processes.
- Focus on **post-processing approach** to fairness:
  - (1) training a prediction function, and (2) a decision rule based on the predictions.

Result 1 (social planner):

- the equity preferences of the social planner have no effect on the training procedure for the prediction function.
- i.e., there should be no limit on the use of sensitive attributes.

Result 2 (private actors):

- key factor is disclosure of decision process (data, ML training, and decision rule), which, unlike human decision-making, allows prejudicial treatment to be detected.
- without disclosure, algorithms will be just as biased as humans.

“An Economic Approach to Regulating Algorithms”

Rambachan, Kleinberg, Ludwig, and Mullainathan (2020)

- Apply welfare economics to the design and regulation of algorithmic decision processes.
- Focus on **post-processing approach** to fairness:
  - (1) training a prediction function, and (2) a decision rule based on the predictions.

Result 1 (social planner):

- the equity preferences of the social planner have no effect on the training procedure for the prediction function.
- i.e., there should be no limit on the use of sensitive attributes.

Result 2 (private actors):

- key factor is disclosure of decision process (data, ML training, and decision rule), which, unlike human decision-making, allows prejudicial treatment to be detected.
- without disclosure, algorithms will be just as biased as humans.
- with disclosure, discrimination decreases relative to humans, and government should impose no constraints on the use of sensitive attributes as predictors.

"An Economic Approach to Regulating Algorithms"

Rambachan, Kleinberg, Ludwig, and Mullainathan (2020)

- Apply welfare economics to the design and regulation of algorithmic decision processes.
- Focus on **post-processing approach** to fairness:
  - (1) training a prediction function, and (2) a decision rule based on the predictions.

Result 1 (social planner):

- the equity preferences of the social planner have no effect on the training procedure for the prediction function.
- i.e., there should be no limit on the use of sensitive attributes.

Result 2 (private actors):

- key factor is disclosure of decision process (data, ML training, and decision rule), which, unlike human decision-making, allows prejudicial treatment to be detected.
- without disclosure, algorithms will be just as biased as humans.
- with disclosure, discrimination decreases relative to humans, and government should impose no constraints on the use of sensitive attributes as predictors.

Caveats:

- disclosure must include the data (including sensitive attribute) and ML training process, not just the decision rule.
- how to decide on the decision rule?

“Algorithmic Social Engineering” (Cowgill and Stevenson 2020)

We examine the microeconomics of using algorithms to nudge decision-makers towards particular social outcomes. . . . **Manipulating predictions to express policy preferences strips the predictions of informational content and can lead decision-makers to ignore them.** When social problems stem from decision-makers’ objectives (rather than their information sets), algorithmic social engineering exhibits clear limitations. **Our framework emphasizes separating preferences and predictions in designing algorithmic interventions**. . . .

Application: Content/Ad Targeting

- Should social media content/ad targeting algorithms (eg Facebook, Amazon) be able to use sensitive attributes as features?
  - gender, age, race, etc.

Write down an answer privately for sharing with the class:

- **[Last name starts with A-M] Give an example situation where gender/race targeting should not be allowed and explain.**
- **[Last name starts with N-Z] Give an example situation where gender/race targeting should be allowed and explain.**
- **What are some restriction/regulations that address problems without banning the targeting?**

Outline

Data can be biased

- Education:
  - SAT scores might be used to guide college admissions, but some students get SAT prep courses

Data can be biased

- Education:
  - SAT scores might be used to guide college admissions, but some students get SAT prep courses
  - Teachers (grading essays) might be biased against some students $\rightarrow$ so will automated essay graders based on those grades.

Data can be biased

- Education:
  - SAT scores might be used to guide college admissions, but some students get SAT prep courses
  - Teachers (grading essays) might be biased against some students $\rightarrow$ so will automated essay graders based on those grades.
- Criminal risk scoring (Skeem and Lovenkamp 2016):
  - Blacks and whites who are otherwise identical are treated the same;
  - But blacks tend to be rated as more risky due to longer criminal histories (**which were produced by biased system**).

Data can be biased

- Education:
  - SAT scores might be used to guide college admissions, but some students get SAT prep courses
  - Teachers (grading essays) might be biased against some students $\rightarrow$ so will automated essay graders based on those grades.
- Criminal risk scoring (Skeem and Lovenkamp 2016):
  - Blacks and whites who are otherwise identical are treated the same;
  - But blacks tend to be rated as more risky due to longer criminal histories (**which were produced by biased system**).
  - similarly: we measure recidivism as "is re-arrested" rather than "commits more crimes". some people more likely to be re-arrested due to policing bias.

Data can be biased

- Education:
  - SAT scores might be used to guide college admissions, but some students get SAT prep courses
  - Teachers (grading essays) might be biased against some students $\rightarrow$ so will automated essay graders based on those grades.
- Criminal risk scoring (Skeem and Lovenkamp 2016):
  - Blacks and whites who are otherwise identical are treated the same;
  - But blacks tend to be rated as more risky due to longer criminal histories (**which were produced by biased system**).
  - similarly: we measure recidivism as “is re-arrested” rather than “commits more crimes”. some people more likely to be re-arrested due to policing bias.
  - selective labeling:
    - predictive policing – produces evidence of more crimes in the neighborhoods where police want to go.
    - only observe recidivism if released.

Data can be biased

- Education:
  - SAT scores might be used to guide college admissions, but some students get SAT prep courses
  - Teachers (grading essays) might be biased against some students $\rightarrow$ so will automated essay graders based on those grades.
- Criminal risk scoring (Skeem and Lovenkamp 2016):
  - Blacks and whites who are otherwise identical are treated the same;
  - But blacks tend to be rated as more risky due to longer criminal histories (**which were produced by biased system**).
  - similarly: we measure recidivism as "is re-arrested" rather than "commits more crimes". some people more likely to be re-arrested due to policing bias.
  - selective labeling:
    - predictive policing – produces evidence of more crimes in the neighborhoods where police want to go.
    - only observe recidivism if released.
- a subjective label, such as "harmful to self or others", when made by a human, could be biased (and so would teaching an ML model to reproduce that label)

Data can be biased

- Education:
  - SAT scores might be used to guide college admissions, but some students get SAT prep courses
  - Teachers (grading essays) might be biased against some students $\rightarrow$ so will automated essay graders based on those grades.
- Criminal risk scoring (Skeem and Lovenkamp 2016):
  - Blacks and whites who are otherwise identical are treated the same;
  - But blacks tend to be rated as more risky due to longer criminal histories (**which were produced by biased system**).
  - similarly: we measure recidivism as "is re-arrested" rather than "commits more crimes". some people more likely to be re-arrested due to policing bias.
  - selective labeling:
    - predictive policing – produces evidence of more crimes in the neighborhoods where police want to go.
    - only observe recidivism if released.
- a subjective label, such as "harmful to self or others", when made by a human, could be biased (and so would teaching an ML model to reproduce that label)

These types of problems cannot be fixed by ML.
But ML can help diagnose them, or mitigate their consequences.

Perception Tasks

- Content identification (Shazam, reverse image search)
- Face recognition
- Medical diagnosis from scans
- Speech to text

High accuracy causes risk of privacy violations.

Perception Tasks

- Content identification (Shazam, reverse image search)
- Face recognition
- Medical diagnosis from scans
- Speech to text

High accuracy causes risk of privacy violations.

Systems are sometimes more accurate/effective for some groups, e.g. most-frequent customers.

Perception Tasks

- Content identification (Shazam, reverse image search)
- Face recognition
- Medical diagnosis from scans
- Speech to text

High accuracy causes risk of privacy violations.

Systems are sometimes more accurate/effective for some groups, e.g. most-frequent customers.

Overall, problems seem straightforward to solve.

Human Judgment Annotation Tasks

- Spam detection
- Detection of copyrighted material
- Automated essay grading
- Hate speech detection
- Content recommendation

Human Judgment Annotation Tasks

- Spam detection
- Detection of copyrighted material
- Automated essay grading
- Hate speech detection
- Content recommendation

These tasks are subjective, so some error is inevitable.
But human judgments are correlated enough that predictions are useful.

Human Judgment Annotation Tasks

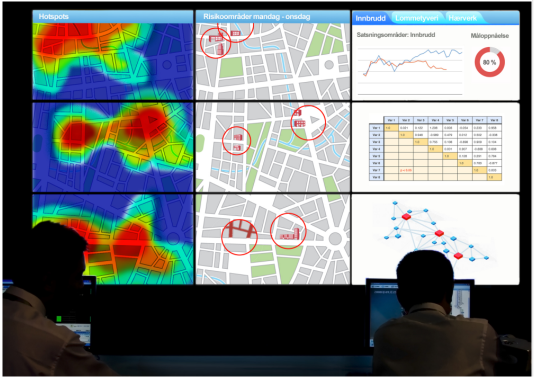
- Spam detection
- Detection of copyrighted material
- Automated essay grading
- Hate speech detection
- Content recommendation

These tasks are subjective, so some error is inevitable.
But human judgments are correlated enough that predictions are useful.

Labels are past behavior, so model is stable and incentive responses are constrained.

- compare: predicting how someone will score on these predictions in the future.

Predictive Policing



Predictive policing poses discrimination risk, thinktank warns

Machine-learning algorithms could replicate or amplify bias on race, sexuality and age

▲ One officer said human biases including more stop and searches of black men were likely to be introduced into algorithm data sets. Photograph: Carl Court/Getty Images https://www.theregister.com/2020/12/08/texas_compsci_phd_ai/

{* ARTIFICIAL INTELLIGENCE *}

Uni revealed it killed off its PhD-applicant screening AI – just as its inventors gave a lecture about the tech

Fears of bias put compsci dept into damage-limitation mode after years of using it to analyze applications

Katyanna Quach Tue 8 Dec 2020 // 12:04 UTC SHARE

A university announced it had ditched its machine-learning tool, used to filter thousands of PhD applications, right as the software's creators were giving a talk about the code and drawing public criticism.

// MOST READ

Apple fires warning shot a
Facebook and Google on
privacy, pledges fight

Generative AI

- large language models
- image generators
- audio generation

Generative AI

- large language models
- image generators
- audio generation

Some significant risks, but with decent solutions:

- **copyright issues**
- **privacy violations**
- **biases/stereotypes in outputs**

Predicting future choices and social outcomes

- Predicting criminal recidivism to assign bail
- Predictive policing to assign police
- Predicting future performance for hiring or school admissions

Predicting future choices and social outcomes

- Predicting criminal recidivism to assign bail
- Predictive policing to assign police
- Predicting future performance for hiring or school admissions

These systems are risky and can have unintended consequences:

Predicting future choices and social outcomes

- Predicting criminal recidivism to assign bail
- Predictive policing to assign police
- Predicting future performance for hiring or school admissions

These systems are risky and can have unintended consequences:

- **Predictions influence availability of labels and subsequent behavior.**

Predicting future choices and social outcomes

- Predicting criminal recidivism to assign bail
- Predictive policing to assign police
- Predicting future performance for hiring or school admissions

These systems are risky and can have unintended consequences:

- **Predictions influence availability of labels and subsequent behavior.**
- **Outcomes are in future so models lack external validity.**

Predicting future choices and social outcomes

- Predicting criminal recidivism to assign bail
- Predictive policing to assign police
- Predicting future performance for hiring or school admissions

These systems are risky and can have unintended consequences:

- **Predictions influence availability of labels and subsequent behavior.**
- **Outcomes are in future so models lack external validity.**
- **Strong incentive responses by decision subjects and decision-makers.**
- **Errors are costly.**

Overview of ML policy problems

- Accuracy issues:
  - model stability
  - selective labeling

Overview of ML policy problems

- Accuracy issues:
  - model stability
  - selective labeling
- Equity issues:
  - (relative) error rate
  - (relative) costs of errors

Overview of ML policy problems

- Accuracy issues:
  - model stability
  - selective labeling
- Equity issues:
  - (relative) error rate
  - (relative) costs of errors
- Social problems from introducing system:
  - externalities (e.g. privacy violations)
  - asymmetric information: AI company knows your preferences (price point) $\rightarrow$ they have information advantage and can capture more surplus.

Overview of ML policy problems

- Accuracy issues:
  - model stability
  - selective labeling
- Equity issues:
  - (relative) error rate
  - (relative) costs of errors
- Social problems from introducing system:
  - externalities (e.g. privacy violations)
  - asymmetric information: AI company knows your preferences (price point) $\rightarrow$ they have information advantage and can capture more surplus.
- Behavioral responses by subjects:
  - subjects try to manipulate features to game system
  - systems (e.g. essay grading) perceived as biased/unfair are discouraging.

Overview of ML policy problems

- Accuracy issues:
  - model stability
  - selective labeling
- Equity issues:
  - (relative) error rate
  - (relative) costs of errors
- Social problems from introducing system:
  - externalities (e.g. privacy violations)
  - asymmetric information: AI company knows your preferences (price point) $\rightarrow$ they have information advantage and can capture more surplus.
- Behavioral responses by subjects:
  - subjects try to manipulate features to game system
  - systems (e.g. essay grading) perceived as biased/unfair are discouraging.
- Behavioral responses by decision-makers:
  - decision-makers ignore model because it is a black box
  - or they rely too much on it and don't do their own diligence

Outline

AI Governance

What can and should AI decide?

AI for legal decisions

Recap and Conclusion

What about legal decisions?

- So far, in the legal context, we have focused mainly on parole and bail decisions.
  - there aren't many papers/systems out there for determining "guilty" vs "innocent"

What about legal decisions?

- So far, in the legal context, we have focused mainly on parole and bail decisions.
  - there aren't many papers/systems out there for determining "guilty" vs "innocent"
- Why? With recidivism:
  - there is a measurable/"true" label that we can predict: whether someone is arrested again in some period of time.
  - the factors that judges are supposed to use are also measured: factors that predict recidivism.

What about legal decisions?

- So far, in the legal context, we have focused mainly on parole and bail decisions.
  - there aren't many papers/systems out there for determining "guilty" vs "innocent"
- Why? With recidivism:
  - there is a measurable/"true" label that we can predict: whether someone is arrested again in some period of time.
  - the factors that judges are supposed to use are also measured: factors that predict recidivism.
- In contrast, for the liability decision (guilty or not):
  - the label is not observed directly, we just have a human judge's decision to go on.
  - the factors are part of a specific circumstance, and not part of a standard data set.

What can legal AI achieve?

- Perception tasks:
  - speeding cameras
  - gunshot detection
  - facial recognition for fare dodging / trespassing

What can legal AI achieve?

- Perception tasks:
  - speeding cameras
  - gunshot detection
  - facial recognition for fare dodging / trespassing
- Human judgement annotation on structured data:
  - copyright infringement
  - detecting corruption in budget accounts
  - detecting evasion in income / tax accounts

What can legal AI achieve?

- Perception tasks:
  - speeding cameras
  - gunshot detection
  - facial recognition for fare dodging / trespassing
- Human judgement annotation on structured data:
  - copyright infringement
  - detecting corruption in budget accounts
  - detecting evasion in income / tax accounts
- Human judgment annotation on unstructured data?
  - determining liability from trial documents
  - e.g. affidavits, police reports, witness testimony

What can legal AI achieve?

- Perception tasks:
  - speeding cameras
  - gunshot detection
  - facial recognition for fare dodging / trespassing
- Human judgement annotation on structured data:
  - copyright infringement
  - detecting corruption in budget accounts
  - detecting evasion in income / tax accounts
- Human judgment annotation on unstructured data?
  - determining liability from trial documents
  - e.g. affidavits, police reports, witness testimony

  ↑ *with aligned LLMs, maybe this is possible now.*

Limitations of legal ML systems

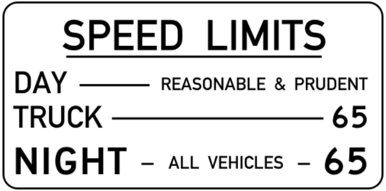
Limitations of legal ML systems

- Existing legal ML systems have evidence constraints:
  - can only interpret evidence that appears in a lot of cases; might ignore special/mitigating circumstances.
  - cannot (easily) contextualize evidence that is more or less trustworthy

Limitations of legal ML systems

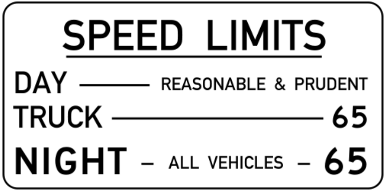
- Existing legal ML systems have evidence constraints:
  - can only interpret evidence that appears in a lot of cases; might ignore special/mitigating circumstances.
  - cannot (easily) contextualize evidence that is more or less trustworthy
- In many important contexts, legal AI would be difficult/impossible to evaluate:
  - cases where only evidence is witness testimony (evidence credibility assessments)
  - antitrust violations (economy is dynamic and in equilibrium)
  - tax avoidance through sophisticated accounting tricks (those adapt to model)
  - new types of cases using new laws/legislation

Legal Vagueness and Value Judgments



- Even if the AI could read new laws, there is the problem of legal vagueness:
  - How will the AI decide in this circumstance?

Legal Vagueness and Value Judgments

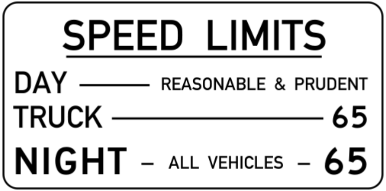


- Even if the AI could read new laws, there is the problem of legal vagueness:
  - How will the AI decide in this circumstance?

- Making choices in the presence of vagueness or indeterminacy requires value judgements.

What counts as a "good" outcome? Is it even measurable?

Legal Vagueness and Value Judgments



- Even if the AI could read new laws, there is the problem of legal vagueness:
  - How will the AI decide in this circumstance?

- Making choices in the presence of vagueness or indeterminacy requires value judgements.

What counts as a "good" outcome? Is it even measurable?

- GPT-type models will give the likely response based on the training/RLHF corpus.
- That is backward-looking and won't (easily) take into account new information.

Philosophical Issues

- What does it mean to surrender the implementation of law enforcement and judicial decision making to machines?
  - at some point, a system might be so good that we wouldn't want humans to interfere

Philosophical Issues

- What does it mean to surrender the implementation of law enforcement and judicial decision making to machines?
  - at some point, a system might be so good that we wouldn't want humans to interfere
- What are the long-term implications for the system and its adaptiveness to change?
  - what are the political and cultural impacts?
  - how does it affect motivation to appeal?

Philosophical Issues

- What does it mean to surrender the implementation of law enforcement and judicial decision making to machines?
  - at some point, a system might be so good that we wouldn't want humans to interfere
- What are the long-term implications for the system and its adaptiveness to change?
  - what are the political and cultural impacts?
  - how does it affect motivation to appeal?

Thoughts? What else?

Outline

Building a Robot Judge

- This course has focused on **machine learning** and **causal inference** for **decision-making**.
  - **expert** decision-making requiring **judgment** – not just legal but also medical, political, etc.

Building a Robot Judge

- This course has focused on **machine learning** and **causal inference** for **decision-making**.
  - **expert** decision-making requiring **judgment** – not just legal but also medical, political, etc.
- Engineering goals:
  - Develop tools for "building a robot judge" – machine prediction and support of expert decisions.

Building a Robot Judge

- This course has focused on **machine learning** and **causal inference** for **decision-making**.
  - **expert** decision-making requiring **judgment** – not just legal but also medical, political, etc.
- Engineering goals:
  - Develop tools for "building a robot judge" – machine prediction and support of expert decisions.
- Scientific goals:
  - Understand the factors underlying decisions of judges.

Building a Robot Judge

- This course has focused on **machine learning** and **causal inference** for **decision-making**.
  - **expert** decision-making requiring **judgment** – not just legal but also medical, political, etc.
- Engineering goals:
  - Develop tools for "building a robot judge" – machine prediction and support of expert decisions.
- Scientific goals:
  - Understand the factors underlying decisions of judges.
  - Assess the real-world impacts of decisions on society – e.g. defendants, patients.

Building a Robot Judge

- This course has focused on **machine learning** and **causal inference** for **decision-making**.
  - **expert** decision-making requiring **judgment** – not just legal but also medical, political, etc.
- Engineering goals:
  - Develop tools for "building a robot judge" – machine prediction and support of expert decisions.
- Scientific goals:
  - Understand the factors underlying decisions of judges.
  - Assess the real-world impacts of decisions on society – e.g. defendants, patients.
- Policy goals:
  - Understand how (not) to use data science tools (machine learning and causal inference) to support expert decision-making.

Next Week: In-Class Exam

- do not miss next week's class!!!

Next Term: NLP Course

- In the spring term, I teach a complementary course in natural language processing:
  - "Language Models for Law and Social Science" (851-0739-01L)

Next Term: NLP Course

- In the spring term, I teach a complementary course in natural language processing:
  - "Language Models for Law and Social Science" (851-0739-01L)
- Not a lot of overlap, and in many ways it builds on the content in this course.
  - i.e., focus on sequence data, and on transformer architectures (e.g. BERT, GPT)
- Similar setup in terms of course credits, assignments, etc.

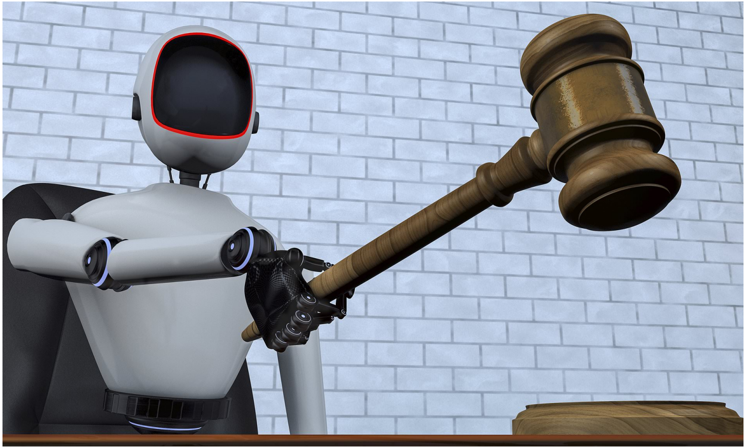
Stay in touch

- e.g. add me on LinkedIn
- let me know if anything in this course helps you later on!
- can provide references for your work in the course.

Lightning Recap Essay

For last minutes of class:

`https://forms.gle/ApgPqYyZEKmNhin8A`



Meeting Adjourned!