Building a Robot Judge: Data Science for Decision-Making

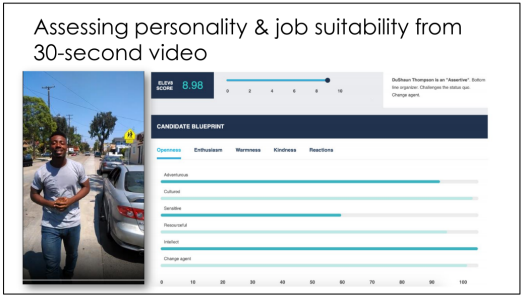
13. Algorithms and Decisions IV

Q&A Page

https://padlet.com/eash44/dn2c5exyl1lx0wnd

https://padlet.com/eash44/of2j80mt3hwnikh4

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



Write down an answer privately for sharing in the zoom chat: Identify a problem with algorithmic hiring, explain why, and what would have to change to fix that problem.

Outline

Perspective | Published: 13 May 2019

Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead

Cynthia Rudin

Nature Machine Intelligence **1**, 206–215 (2019) | Cite this article

https://www.nature.com/articles/s42256-019-0048-x

Simple Rules for Complex Decisions (Jung et al 2017)

Table 1: A defendant's flight risk is obtained by adding the scores for age and prior failure to appear (FTA).

Feature	Score	Feature	Score
$18 \leq$ age < 21	8	no prior FTAs	0
$21 \leq$ age < 26	6	1 prior FTA	6
$26 \leq$ age < 31	4	2 prior FTAs	8
$31 \leq$ age < 51	2	3 prior FTAs	9
$51 \leq$ age	0	4 or more prior FTAs	10

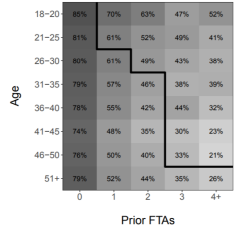


Figure 1: Graphical representation of a simple rule for release decisions, based on setting a release threshold of 10.5 on the risk scores described in Table 1. Groups to the left of the black line are those that would be released under the rule; for comparison, the shading and numbers show the proportion of defendants that are currently RoR'd in each group.

Simple Rules for Complex Decisions (Jung et al 2017)

(1) **Select.** From the full set of features, select k features via forward stepwise regression. For fixed k, we note that standard selection metrics (*e.g.*, AIC or BIC) are theoretically guaranteed to yield the same set of features.
(2) **Regress.** Using only these k selected features, train an L^1-regularized (lasso) logistic regression model to the data, which yields (real-valued) fitted coefficients $\beta_1, \ldots, \beta_k$.
(3) **Round.** Rescale the coefficients to be in the range $[-M, M]$, and then round the rescaled coefficients to the nearest integer. Specifically, set

$$w_j = \text{Round}\left(\frac{M\beta_j}{\max_i |\beta_i|}\right).$$

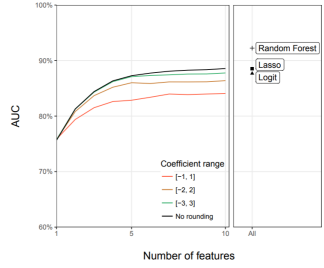


Figure 5: Mean test AUC of decision rules over 22 datasets.

Kleinberg and Mullainathan, "Simplicity Creates Inequity" (2019)

- Individuals have characteristics X and group membership A.
- Algorithm approximates score $f(X, A)$,
  - decide outcome (e.g. admit to college) based on threshold on $f(\cdot)$

Kleinberg and Mullainathan, "Simplicity Creates Inequity" (2019)

- Individuals have characteristics X and group membership A.
- Algorithm approximates score $f(X, A)$,
  - decide outcome (e.g. admit to college) based on threshold on $f(\cdot)$
  - A is correlated with X (a "disadvantaged" group has lower-scored attributes, e.g. less extracurricular activities) but A doesn't independently affect suitability.

Kleinberg and Mullainathan, "Simplicity Creates Inequity" (2019)

- Individuals have characteristics X and group membership A.
- Algorithm approximates score $f(X, A)$,
  - decide outcome (e.g. admit to college) based on threshold on $f(\cdot)$
  - A is correlated with X (a "disadvantaged" group has lower-scored attributes, e.g. less extracurricular activities) but A doesn't independently affect suitability.
- A "simple model" or "approximator" partitions X into cells, and scores each cell.
  - e.g. decision tree.

Kleinberg and Mullainathan, "Simplicity Creates Inequity" (2019)

- Individuals have characteristics X and group membership A.
- Algorithm approximates score $f(X, A)$,
  - decide outcome (e.g. admit to college) based on threshold on $f(\cdot)$
  - A is correlated with X (a "disadvantaged" group has lower-scored attributes, e.g. less extracurricular activities) but A doesn't independently affect suitability.
- A "simple model" or "approximator" partitions X into cells, and scores each cell.
  - e.g. decision tree.

Result 1:

- assume a non-trivial (e.g. real-world) dataset (see paper)

Kleinberg and Mullainathan, "Simplicity Creates Inequity" (2019)

- Individuals have characteristics X and group membership A.
- Algorithm approximates score $f(X, A)$,
  - decide outcome (e.g. admit to college) based on threshold on $f(\cdot)$
  - A is correlated with X (a "disadvantaged" group has lower-scored attributes, e.g. less extracurricular activities) but A doesn't independently affect suitability.
- A "simple model" or "approximator" partitions X into cells, and scores each cell.
  - e.g. decision tree.

Result 1:

- assume a non-trivial (e.g. real-world) dataset (see paper)
- starting from a simple model, there exists a more complex model (smaller cells) that improves both efficiency *and* equity.
  - Efficiency = average $f(\cdot)$ of admitted candidates.
  - Equity = relative share admitted for the disadvantaged group.

Kleinberg and Mullainathan, "Simplicity Creates Inequity" (2019)

- Individuals have characteristics X and group membership A.
- Algorithm approximates score $f(X, A)$,
  - decide outcome (e.g. admit to college) based on threshold on $f(\cdot)$
  - A is correlated with X (a "disadvantaged" group has lower-scored attributes, e.g. less extracurricular activities) but A doesn't independently affect suitability.
- A "simple model" or "approximator" partitions X into cells, and scores each cell.
  - e.g. decision tree.

Result 1:

- assume a non-trivial (e.g. real-world) dataset (see paper)
- starting from a simple model, there exists a more complex model (smaller cells) that improves both efficiency *and* equity.
  - Efficiency = average $f(\cdot)$ of admitted candidates.
  - Equity = relative share admitted for the disadvantaged group.

Result 2:

- with a simple model (relative to a complex model), info on group membership is more likely to help the decision-maker select candidates with higher $f(\cdot)$.

"Predictive power at what cost? Economic and racial justice of data-driven algorithms" (Jabri 2019)

This paper studies how algorithms use variables to maximize predictive power at the cost of group equity. Group inequity arises if variables enlarge disparities in risk scores across groups. I develop a framework to examine a recidivism risk assessment tool **using risk score and novel pretrial defendant case data from 2013-2016 in Broward County, Florida**. **I find that defendants' neighborhood data only negligibly improve predictive power, but substantially widen disparities in defendant risk scores and false positive rates across race and economic status**.
. . .

(good paper for RE03 or RE04)

Outline

- 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
  - bad actors always a step ahead.
  - limits innovation and expansion of digital economy.
  - could collude with industry leaders

Transparency

- Closed-source algorithms result in "black box justice" and could be abused by insiders.
- But open-source algorithms are prone to gaming: savvy attorneys 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 attorneys 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 attorneys 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.
- Algorithmic decision-making has two components:
  - (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.
- Algorithmic decision-making has two components:
  - (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.
- Algorithmic decision-making has two components:
  - (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.
- Algorithmic decision-making has two components:
  - (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.
- Algorithmic decision-making has two components:
  - (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.
  - caveat: disclosure must include the data and ML training process, not just 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**. . . .

(another good paper for RE03 or RE04)

Application: Content/Ad Targeting

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

Write down an answer privately for sharing in the zoom chat:

1. **Give at least one reason that gender/race targeting should be allowed.**
2. **Give at least one reason it should not be allowed.**
3. **Propose a restriction/regulation that addreses your problem without banning the targeting.**

Outline

Perception Tasks

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

Perception Tasks

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

High accuracy causes risk of privacy violations.

Perception Tasks

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

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
- Deepfakes

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

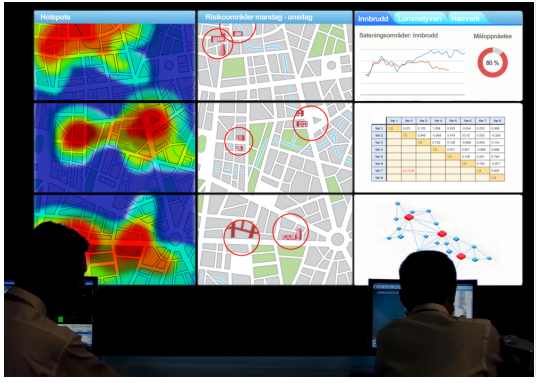
- Spam detection
- Detection of copyrighted material
- Automated essay grading
- Hate speech detection
- Content recommendation

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.

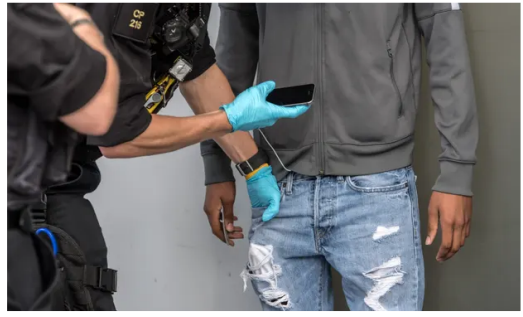
- compare: predicting how someone will score on these predictions in the future.

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

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

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
  - ↑ *would require a lot of (sophisticated) NLP tools*

What can legal AI not achieve

What can legal AI not achieve

- ML system severe evidence constraints:
  - can only use evidence that appears in a lot of cases; it ignores special/mitigating circumstances.
  - cannot (easily) contextualize evidence that is more or less trustworthy

What can legal AI not achieve

- ML system severe evidence constraints:
  - can only use evidence that appears in a lot of cases; it ignores special/mitigating circumstances.
  - cannot (easily) contextualize evidence that is more or less trustworthy
- Would not work in many important types of cases:
  - eg cases where only evidence is witness testimony (evidence credibility assessments)
  - antitrust violations (economy is dynamic)
  - tax avoidance through sophisticated accounting tricks (those adapt to model)

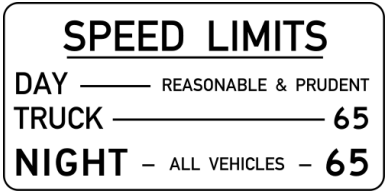
What can legal AI not achieve

- ML system severe evidence constraints:
  - can only use evidence that appears in a lot of cases; it ignores special/mitigating circumstances.
  - cannot (easily) contextualize evidence that is more or less trustworthy
- Would not work in many important types of cases:
  - eg cases where only evidence is witness testimony (evidence credibility assessments)
  - antitrust violations (economy is dynamic)
  - tax avoidance through sophisticated accounting tricks (those adapt to model)
- Would not work on new types of cases.
  - In particular, would not account for new laws/legislation.

What can legal AI not achieve

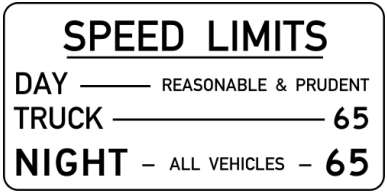
- ML system severe evidence constraints:
  - can only use evidence that appears in a lot of cases; it ignores special/mitigating circumstances.
  - cannot (easily) contextualize evidence that is more or less trustworthy
- Would not work in many important types of cases:
  - eg cases where only evidence is witness testimony (evidence credibility assessments)
  - antitrust violations (economy is dynamic)
  - tax avoidance through sophisticated accounting tricks (those adapt to model)
- Would not work on new types of cases.
  - In particular, would not account for new laws/legislation.
- Teaching the algorithm to understand rare evidence, discount suspicious evidence, and to understand new laws, would require something much closer to **legal artificial intelligence**.

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?

- at a minimum, would require an unrealistic number of empirical policy analysis studies.

Philosophical Issues

- What does it mean to surrender the implementation of law enforcement and judicial decision making to machines?
- 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?
- 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.

Final Assignment

`https://bit.ly/BRJ-exam-guide`

- If you haven't done so yet, sign up for a final assignment cohort by 5pm, otherwise you are in Cohort 2.

Next Term: NLP Course

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

Next Term: NLP Course

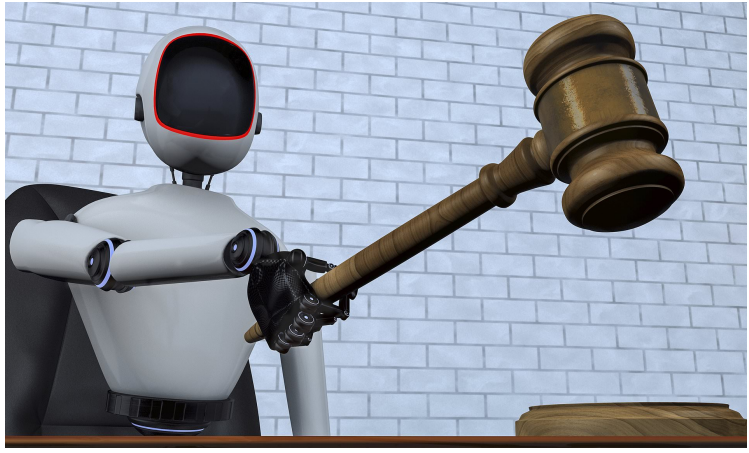
- In the spring term, I teach a complementary course in natural language processing:
  - "Natural Language Processing for Law and Social Science" (851-0739-01L)
- Not a lot of overlap, and in many ways it builds on the content in this course.
  - i.e., focus on sequence data, and on transformer architectures (e.g. BERT, GPT-3)
- Similar setup in terms of course credits:
  - 3 credits for the lectures/assignments, 2 additional credits for a project.

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.

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.



Meeting Adjourned!