AI and professional work: The practice of law with automated decision support technologies

Deirdre K. Mulligan
Daniel Kluttz

23rd Annual BCLT/BTLJ Symposium
Governing Machines:
Defining and Enforcing Public Policy Values in AI Systems

April 5, 2019
How was gender used?

Dispute between the parties below

• “criminogenic factor”
  or
• statistical “norming”

_Wisconsin Supreme Court did not resolve_
Solicitor General’s brief Loomis v. Wisconsin

(2011). It is a serious constitutional question, however, the extent to which actuarial assessments considered at sentencing may take account of statistical differences for male and female offenders, such as, for example, in recidivism rates. That question may warrant the Court’s attention in the future in an appropriate case.

This case, however, is not a suitable vehicle. Initially, it is unclear how COMPAS accounts for gender—a fact of relevance to the constitutional analysis.
Inattention to values appears to be the norm

“...governments simply did not have many records concerning the creation and implementation of algorithms, either because those records were never generated or because they were generated by contractors and never provided to the governmental clients. These include records about model design choices, data selection, factor weighting, and validation designs. At an even more basic level, most governments did not have any record of what problems the models were supposed to address, and what the metrics of success were.” – Brauneis and Goodman (2018) at 152
Inattention to values appears to be the norm

“...prediction models are often built and used without key policy decisions ever having been articulated, justified, or recorded.”

(Brauneis and Goodman (2018) at 153)
Professional Judgment and Handoffs
From Expert to Predictive Systems
Somali grocers lose right to use food stamps

Businesses say USDA ban an error based on cultural differences; stores may fail

By CHRIS MCGANN, SEATTLE POST-INTELLIGENCER REPORTER  Published 10:00 pm PDT, Monday, April 8, 2002
Operationalizing the Target Variable
Choice of Model
Thresholds
Training Data

HP Face-Tracking Webcams Don't Recognize Black People

Google Photos Mistakenly Labels Black People ‘Gorillas’
BY CONOR DOUGHERTY  JULY 1, 2015 7:01 PM  41

Amazon Rekognition FALSE MATCHES
28 current members of Congress
WHY?
Failure to learn from the past?

Doug Kerr

https://www.flickr.com/photos/dougtone/4297571254/
https://creativecommons.org/licenses/by-sa/2.0/legalcode
Acquisition?  
Procurement + exceptions to APA

Context of use?  
field; organizations; users
Context of use: Where are the professionals?

• Role of professionals under-examined

• Professionals, organizational environments, and professional ecosystem (professionals, clients, vendors)

• Dearth of empirical data
  • routines, habits, norms, values and ideas and obligations of professional identity effect use (journalism and law) Christin 2017
  • cautious and risk-averse, protective of professional expertise, only handing off routinized work to lit sup Blomberg 1996
Case: AI-based systems for lawyers

- Focus application: predictive coding (aka, technology-assisted review, or TAR)
- Used to classify responsive electronically stored information during discovery phase of litigation
- Most developed and widely used form of machine-learning legal tech
Case: AI-based systems for lawyers

• Predictive coding (EDRM, Duke Law School)

• “An industry-specific term generally used to describe a Technology-Assisted Review process involving the use of a Machine Learning Algorithm to distinguish Relevant from Non-Relevant Documents, based on a Subject Matter Expert’s Coding of a Training Set of Documents.”
TAR 1.0 & 2.0

TAR 1.0
Identify seed documents
• Judgment based (atty ids a set he/she believes relevant)
• Random sampling (then atty ids those relevant and not)
• Combination

Train model
• uses the docs that reflect subject matter expert’s judgments to train a classification/ranking algorithm
• compares its results against the SME-tagged control set to gauge its accuracy

Establish a scoring threshold for production

TAR 2.0
Continuous Active Learning—skip step 1, more like setting up a preference engine, continuous loop of refinement based on feedback
Handing off Judgment

- Learning from *outputs* of lawyer’s judgment, not lawyer’s *reasoning*
- Algorithms are learning to predict *outcome* not *reasoning*
- Developing intuition but not at all clear that it mirrors the logical reasoning of humans—in fact certain it doesn’t because machines and humans “see” differently
Questions of Interest

• What’s driving adoption of TAR?
• How is TAR shaping field?
• How are lawyers evaluating reliability of ML systems to which they are handing off judgments?
• What concerns do legal professionals have about use of ML tools?
Research Design

Semi-structured, in-depth interviews
• 25 respondents to date. ~ 26 hrs of interview data
• Attorneys (12), in-house litigation support technical staff (5), and managers at legal tech companies that provide automated software and consulting services to lawyers (8)

Strategic sample, not representative
• Skews toward large firms (>50 attorneys), corporate defense, significant litigation practices.
• Recruitment email & phone; snowball sampling

Analysis of interview data with MAXQDA software
• Thematic Coding
What’s driving adoption of TAR?
Cost + Data Deluge

• “It is common knowledge now from judges that everyone's using software. No one really expects that I was the one doing it. What's happened is lawyer rates have gone up so high that everyone just assumes that it's being done by somebody else.”

• “It's actually related to spending money. I don't really know if I can put it more gracefully than that, but the reality is if you’re a multi-billion dollar company, do you really want a law firm that charges $1,000 an hour making all of your decisions for you, or do you want to have people internally that they're definitely looking out for your best interests and vetting outsourced vendors accordingly, contracting directly with them, managing that process internally versus going through a law firm?”
How is TAR shaping field?
Restructuring

“We do have a lot of jobs that come to us through law firm relationships, but often times the law firm's just a middle party that was retained by our client, which is a corporation, and we end up working with both the corporations and the law firm, of course, in that scenario. We have tons of projects where the end client is actually the corporation. I mean, maybe half and half, maybe one-third/two-thirds, somewhere in there.”

“I will go and give a presentation to the contract reviewers I will explain to them what the case is about. I will answer questions. I'm going there in person and doing this before the project get started...It does signify a longer-term shift in lawyers and that lawyers need to find other ways to the value-added business partners. I think five years from now, all of this stuff will just get done in the nova. You'll send it over to someone; it'll get done; they'll explain to you how the software works; and the documents will just come back.”
Changing influence of firms over tools and professionals

“The whole basis for needed to be barred in the jurisdiction where you’re practicing law is really just to ensure that you have competence in the laws of the state. That really has no relevance in document review, which is almost exclusively not state-specific. even if it was, everyone would be trained accordingly, how to do the document review, regardless of whether or not they were a barred attorney in that state. I think it’s a particularly flimsy justification.

Again, it’s not really about whether or not you can justify it. It’s about whether or not the outside law firm that is engaging with you and is ultimately responsible for the production, ultimately responsible for the review, for that matter – it’s really whatever they’re comfortable with is what – reality is, it is what it is in that scenario.”
Changing relationships between lawyers and professionals employed by vendors

• “Law firms are always free to tell us that they have other ideas. It’s not like we’re trying to say that there is no one out there who could have a good idea. It’s just who’s going to do the initial set up of how everything’s doing, and who’s going to do the day-to-day work on the ground? The clients hired us because we usually cost less than the law firms do, a lot of that day-to-day work.”

• “It's really a joint effort between them and us to confirm that the population is accurate, doesn't contain frivolous material, is responsive, hasn't been produced before, is not duplicative of anything...It's definitely a joint effort and I think it's incumbent upon both the vendor and the law firm to confirm what they're producing on the other side is what we really want to produce.”
Concerns from the field
Ethical Duty of Technology Competence

“People are wondering in AI, where are the responsibilities if the platform gets screwed up? Or you make mistakes? Or you make a representation that’s belied by the data, that kind of thing.”
Ethical Duty of Good Faith and Disclosure

“It's almost always driven by volume. Attorneys will say here, run this group of search terms, and they want to know how many documents it brings back. Then they say oh, that's too many. We got to change the terms. They've met the terms based on the number of documents that they've returned. Then they get the other side to agree to the search terms we're using and vice-versa. That's usually how that goes It's very funny, and I've never really understood this. How is it that we're determining what to review based on how many documents come back on a given search term set? Either the search terms are perpetually responsive or they're not.”
Loss of Expertise

“...every time we use one of these AI tools, provide feedback and provide inputs to it, we're giving the AI vendor a little bit of our know-how, and in doing that again, and again, and again, and giving away our know-how, we're putting ourselves out of business...the more sophisticated consumers, there's an acknowledgment of that, but it's like privacy with our smartphones. What are you going to do? You've got to choose between giving up all this privacy in order to get the incredible convenience that location tracking and everything else provides. You got to give up a little bit of your know-how in order to be able to do this thing better, faster, cheaper, as the client's expecting to get a little bit of a competitive edge doing that. It's acknowledged, and it's going to be a problem. We're going to be the originators of our own demise by dragging the machines different places.”
Confidentiality and Conflicts

Data

• Privacy
• Security

Models trained on data

Contract attorneys
How are lawyers evaluating reliability of ML systems?
“...the predictive coding stuff is typically done by vendors and staff attorneys who do the review. We don't get that involved with it and continuous active learning, honestly, is something that I've used a lot in the past but here at XXX has not traditionally been used internally, and I'm trying to introduce it.”

“Most of them don’t really even to go as far as to want to talk about the underlying technology, logistic regression. That’s probably a phrase I’ve only said ample times in the last couple of years. People don’t really want to talk about that, and for good reason. It doesn’t really matter. Validation is probably the most important process of all predictive coding projects. At the end of the day, if your project validates the way you expect it to, do you really care how you got there? Do you really care if it was logistic regression, SVM, or whatever it is? That technology decision that the vendor has made that’s providing you with that software – that can definitely have an impact as to how efficient the process went.”
“It's almost always driven by volume. Attorneys will say here, run this group of search terms, and they want to know how many documents it brings back. Then they say oh, that's too many. We got to change the terms. They've met the terms based on the number of documents that they've returned. Then they get the other side to agree to the search terms we're using and vice-versa. That's usually how that goes It's very funny, and I've never really understood this. How is it that we're determining what to review based on how many documents come back on a given search term set? Either the search terms are perpetually responsive or they're not.”
“...the predictive coding stuff is typically done by vendors and staff attorneys who do the review. We don't get that involved with it and continuous active learning, honestly, is something that I've used a lot in the past but here at XXX has not traditionally been used internally, and I'm trying to introduce it.”

“Most of them don’t really even to go as far as to want to talk about the underlying technology, logistic regression. That’s probably a phrase I’ve only said ample times in the last couple of years. People don’t really want to talk about that, and for good reason. It doesn’t really matter. Validation is probably the most important process of all predictive coding projects. At the end of the day, if your project validates the way you expect it to, do you really care how you got there? Do you really care if it was logistic regression, SVM, or whatever it is? That technology decision that the vendor has made that’s providing you with that software – that can definitely have an impact as to how efficient the process went.”
Responsibility in systems

• Algorithmic systems
  “People are wondering in AI, where are the responsibilities if the platform gets screwed up? Or you make mistakes? Or you make a representation that’s belied by the data, that kind of thing.”

• Outsourcing
  “Do they need to be reviewed by attorneys, or can you send them to wherever on the other side of the world to have them reviewed?”
Accountability of Systems

“...when the explainability isn’t there, that’s where we see a red flag...a lot of people walk away thinking it’s a black box scenario...that being said, we still have a pretty robust way to validate whether the tool is doing what it says it’s doing. We can still add a layer of defensibility there...by following industry standard validation processes.”

“measuring recall and precision, and doing that through either creating what they call a control set at the outset and then doing a sample of the documents that are quote left behind, the non-responsive predicted documents, doing a random statistical sample in that set and then measuring recall and precision that way. The other validation step to – it’s often not required on a case, but it’s one that people tend to use in making sure we find key documents. There can be a really highly-targeted search term that you can run and say, are there any important documents that we left behind?... It’s not necessarily a firm policy...”
How to right this?
Professional governance

- Testing
- Validation
- Professional rules
Contestability as Governance Strategy

- **Contestability**: mechanisms for users to understand, construct, shape and challenge model predictions
- Active & dynamic: Help users learn about algorithms as they use them
- Facilitate engagement and reflection between entities in shaping performance
- “Skin in the game” (i.e. accountability/responsibility)
Contestability as Governance Strategy

- Publicizes politics of system
- Avoids delegating some important policy choices even if system is outsourced
- Continuous feedback from domain experts
Handoffs: epistemically responsible knowing
Joint work with:
Daniel Kluttz, Post Doc Algorithmic Opacity and Fairness Working Group, School of Information UCB

Funding
NSF CCESTEM: Emerging Cultures of Data Science Ethics in the Academy and Industry
NSF INSPIRE: Value-Function Handoffs in Human-Machine Compositions SES 1650589
Google through Algorithmic Opacity and Fairness Working Group at UCB (Prof. Jenna Burrell (co-lead))