INTERNATIONAL COMPUTATIONAL LAW JOURNAL

CODEX SOURCE ARCHIVE

Detain/Release: Simulating Pretrial Risk Assessments

Keith Porcaro

Computational Law Report · Release 3.0 · 2021-12-07

Reader edition prepared by The Research University under CC BY 4.0. This cover was added; the original PDF pages follow unchanged. Original authorship, publication identity and source links are preserved. This independent republication is not sponsored or endorsed by Stanford University, Stanford Law School, CodeX or MIT.

Volume 1 · Issue 5 · Evidence, Interpretation and Machine Judgment

Original source and license evidence are recorded on the article landing page.

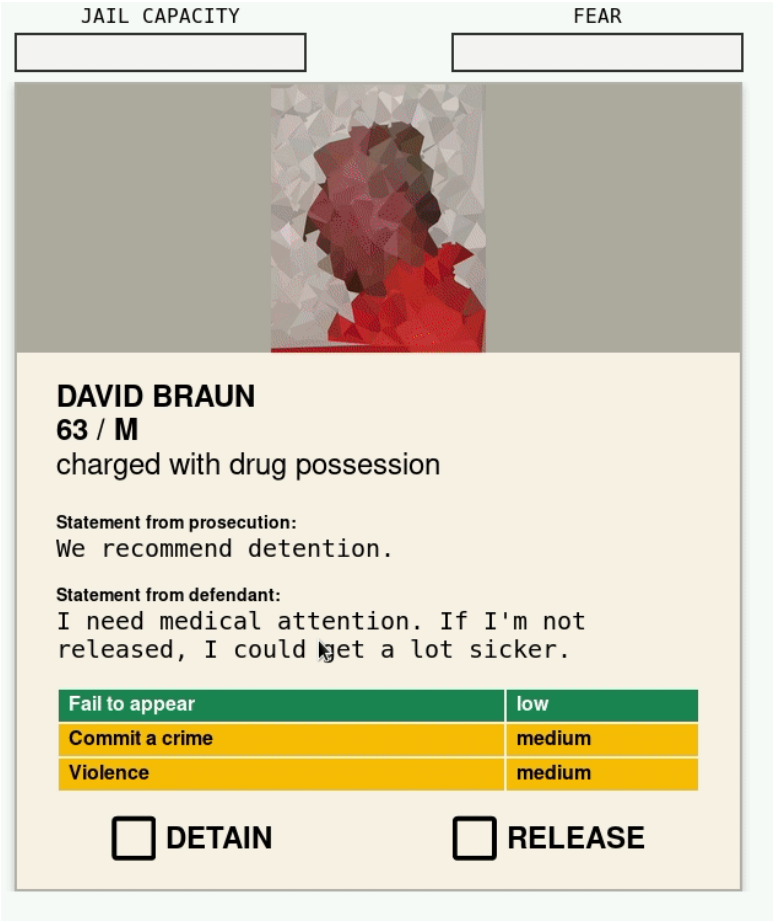
MIT Computational Law Report •

Detain/Release: Simulating Pretrial Risk Assessments

Keith Porcaro

Published on: Dec 07, 2021

License: Creative Commons Attribution 4.0 International License (CC-BY 4.0)



Abstract

This essay describes Detain/Release, an in-class simulation and accompanying lesson designed to teach law students about pretrial risk assessments.

Part 1 describes the teaching goals for the simulation, and contextualizes them within the instant class and broader goals for teaching lawyers about technology. Part 2 covers the development of the simulation itself: basic interaction loops, data sources, and so on. Part 3 discusses how we integrated the simulation into a classroom lesson, which involved a mix of lecture, discussion, and additional software tooling. Finally, part 4 collects anecdotal observations about the simulation's deployment in a variety of contexts, and considers improvements and adjustments.

1.0 Why simulate risk assessments?

Detain/Release was designed to highlight the complexity—the potential messiness—of software-supported decision-making. Much of legal scholarship's dialogue on algorithmic-assisted decision-making explores a linear scale of "trust in algorithms" from algorithmic aversion to automation bias. Literature from interface design to sociology suggests a more complex reality: a user's interaction with a decision-support tool can create emergent, unexpected behavior. To fully capture that complexity, our simulation had three teaching goals.

First, we needed to introduce students to the development and use of algorithmic and actuarial risk assessments in pretrial. My co-professor and I taught Detain/Release in a practicum course that was heavily project-focused, leaving us with limited in-class time to cover a wide range of emergent issues in criminal justice and technology—the lesson described in this essay was designed to fit into a 90-minute class session. The lesson we pair with Detain/Release touches on the substantive history of risk assessment tools, their development and deployment in pretrial and sentencing, and issues with biased data sources and opaque composition.

Second, we wanted students to understand risk assessments within an implementation context. We used interactivity, game mechanics, and in-class discussion to imitate contextual pressures on a judge: e.g., fear of mistakes, policy preferences, consequences of detention on defendants' well-being, political pressure for elected judges. In doing this, we wanted students to consider how the irregular, asymmetric feedback loops in a decision-making context can change (or reinforce) a judge's decision-making, and change how effective a risk assessment actually is.

Finally, we wanted students to understand a judge's use of a risk assessment as an interpretive act. We felt that the best way to do this was to prompt students to use a risk assessment to make decisions and reflect on them. Considering that sound professional judgment is a core competency for lawyers, law schools spend surprisingly little time asking students to formally explore and interrogate their own judgment. As decision-support software becomes more prevalent and tightly integrated with legal practice and administration, it is increasingly important for legal education to remedy this deficiency.

Together, the simulation and lesson ask students to consider how risk assessments contribute to (and shape) a judge's decision-making routine. Our goal is not to provide

easy answers, but to empower students to interrogate the risk assessments and decision-support tools that they will increasingly encounter in practice.

2.0 Building the simulation

Detain/Release is a web-based simulation. While the base simulation runs entirely in the browser, the classroom version (described in part 3) integrates a dashboard and remote control to manage multiple students running the simulation simultaneously.

A participant begins Detain/Release as a county judge in criminal court. The core decision-making loop is simple: determine whether criminal defendants will be detained until trial or released on their own recognizance, receive irregular feedback about the fate of those defendants, and repeat. To simulate contextual pressures (and to make each run finite), Detain/Release also asks participants to manage public fear and jail capacity as they make decisions.

This part describes the core simulation and its constituent components: a) generating synthetic defendants; b) generating a risk assessment; c) processing a participant's decision and managing the overall "state" of the simulation run.

2.1 Generating synthetic defendants

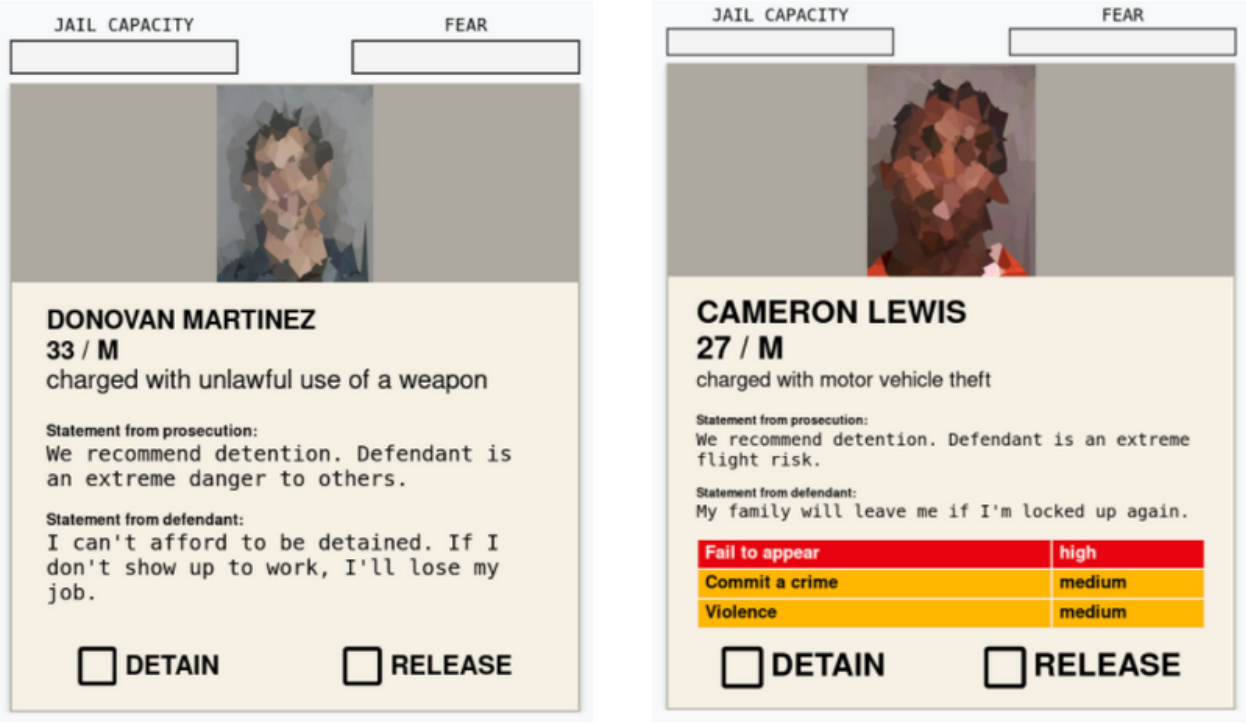


Figure 1: Sample synthetic defendants, with and without risk assessments

Detain/Release generates synthetic defendants as a participant runs the simulation. Each defendant has basic demographic attributes, a stylized picture, and a leading charge. A defendant's digital "card" also contains brief statements from the defendant and prosecutor, and a risk assessment that rates the defendant's risk of failing to appear, committing a new offense, or committing a violent offense.

To generate defendants, the simulation's engine uses weighted psuedorandom number generators derived from criminal justice and census data[1] formatted as frequency distributions—the simulation is more likely to pull from data entries with higher frequencies (e.g., "Smith" is more likely to be a defendant's surname than "Porcaro").[2] This construction allows for the simulation to be customized with new data (e.g., for jurisdiction-specific charges), and to generate hundreds of millions of different defendant combinations.[3]

Murder	**0.67**	**Forgery**	**2.61**
Rape	**1.04**	**Fraud**	**3.38**
Robbery	**6.77**	**Drug trafficking**	**14.82**
Assault	**11.57**	**Drug possession**	**17.77**
Burglary	**8.62**	**Unlawful use of a weapon**	**3.67**
Theft	**8.41**	**Reckless driving**	**4.16**
Motor Vehicle Theft	**2.57**		

Table 1: Default frequency distribution of defendant charges

Synthetic defendant "photos" are derived from a mix of NIST images and various national parliaments, distorted with a rough Delauney triangle algorithm to effectively de-identify the images.



Figure 2: Sample defendant faces

By default, Detain/Release uses a simplified risk assessment, which describes the defendant's likelihood to fail to appear, commit a new offense, or commit a violent offense in terms of a qualitative label: "low", "medium", or "high". The simulation does not reveal to participants how the risk assessment's qualitative labels relate to one another, nor how the risk assessment is derived.

In its default configuration, the risk assessment's qualitative labels are not equivalent across risk categories: a low probability of failure to appear is not the same as a low probability of violence. In addition, Detain/Release's risk assessments are generated independently from a defendant (that is, randomly), but the risk assessment has a loose connection to a defendant's odds of reoffending. This design was partly borne of convenience: rather than randomly generate additional background data that would feed a risk assessment, the simulation skips to randomly generating a risk assessment. Interestingly (although anecdotally), we found that even expert participants assumed Detain/Release's risk assessment was based on real (if hidden) factors.

Our risk assessment's default values and their corresponding probabilities are illustrated in Table 1. The qualitative labels—high, medium, low—correspond to different probability values across categories. Detain/Release's risk assessments have an "accuracy" setting: the probability that the risk assessment accurately captures a defendant's risk. By default, there is a 20% chance that a given defendant's "actual" risk value will be different from the "assessed" risk value.

	Fail to appear	New crime	Violence
high	7.5%	5%	2.5%
medium	5%	2.50%	1%
low	2.5%	1%	0.5%

Table 2: Default risk assessment probabilities

Finally, the synthetic defendants each include a prosecutor recommendation and a defendant story. A prosecutor's detention recommendations are generated based on the risk assessment's output (see Table 3) and lean toward detention. A defendant story is generated from text snippets and falls into one of five categories that hint at collateral consequences for pretrial detention.

Risk assessment	Probability of detention recommendation
Any high	100%
3 medium	100%
2 medium	95%
1 medium	40%
All low	5%

Table 3: Prosecutor's default detention recommendation probabilities

2.2 Decision-making and feedback

Presented with a defendant, our "judge" has two choices: detain the defendant, or release them. Detained defendants are sent to the county jail. Detained defendants spend a variable amount of simulation "time"[4] in jail before exiting depending on their leading charge. Released defendants have a chance to commit subsequent offenses for a limited period of simulation time. If they do, the judge will be notified via a mock newspaper headline, generated with text snippets.[5] Subsequent offenses will also fill up a public fear meter, depending on the severity of the violation, the original charge, and whether the defendant was apprehended (in which case the defendant moves to jail). Fear subsides over "time". Jail capacity and public fear are each measured by graphical meters. If either fill up, the simulation ends.

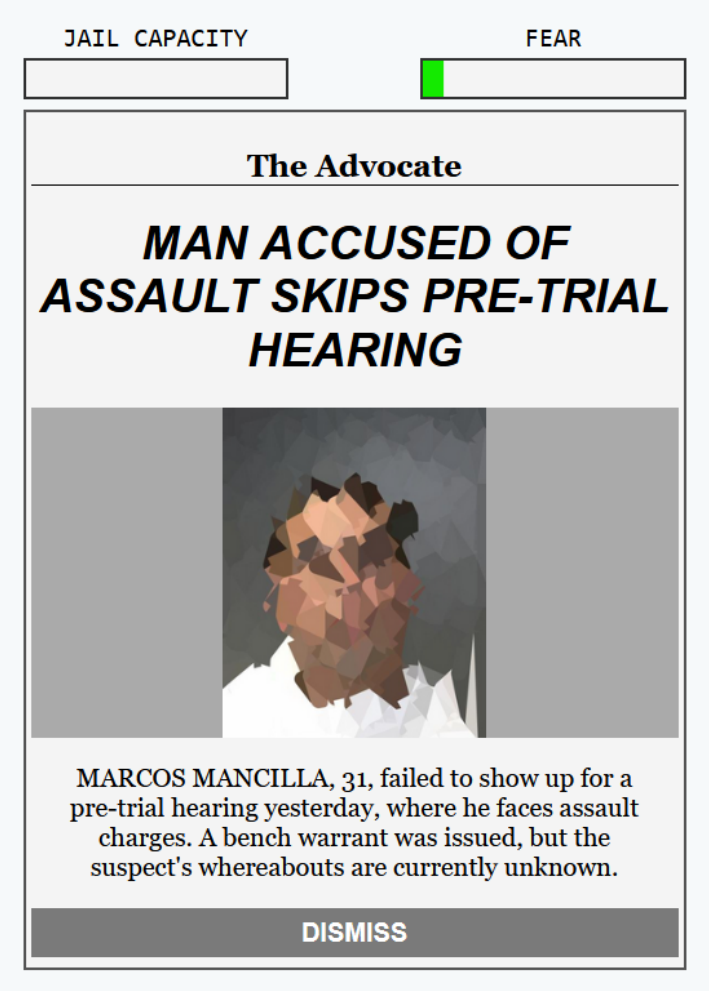


Figure 3: Sample newspaper article and meters

Detain/Release offers two configurations for determining whether a released defendant commits an offense. In a single-shot configuration, a defendant is tested once, after the participant decides to release them. In a cumulative configuration, the entire pool of active released defendants is tested after every participant decision, stopping if a test returns a violation, or once the entire pool has been tested. In a real pretrial environment, a defendant may have multiple court appointments to make, or be released (or supervised) for a long period of time. There may be a long period of time between a judge's pretrial decision and the actual moment (or moments) of feedback: when a defendant makes a subsequent court appearance, misses one, or commits another violation (each of which could be before a different judge). A cumulative configuration, where defendants could violate long after a participant's decision, is intended to evoke that irregular feedback.

To test a released defendant, the simulation executes the following steps:

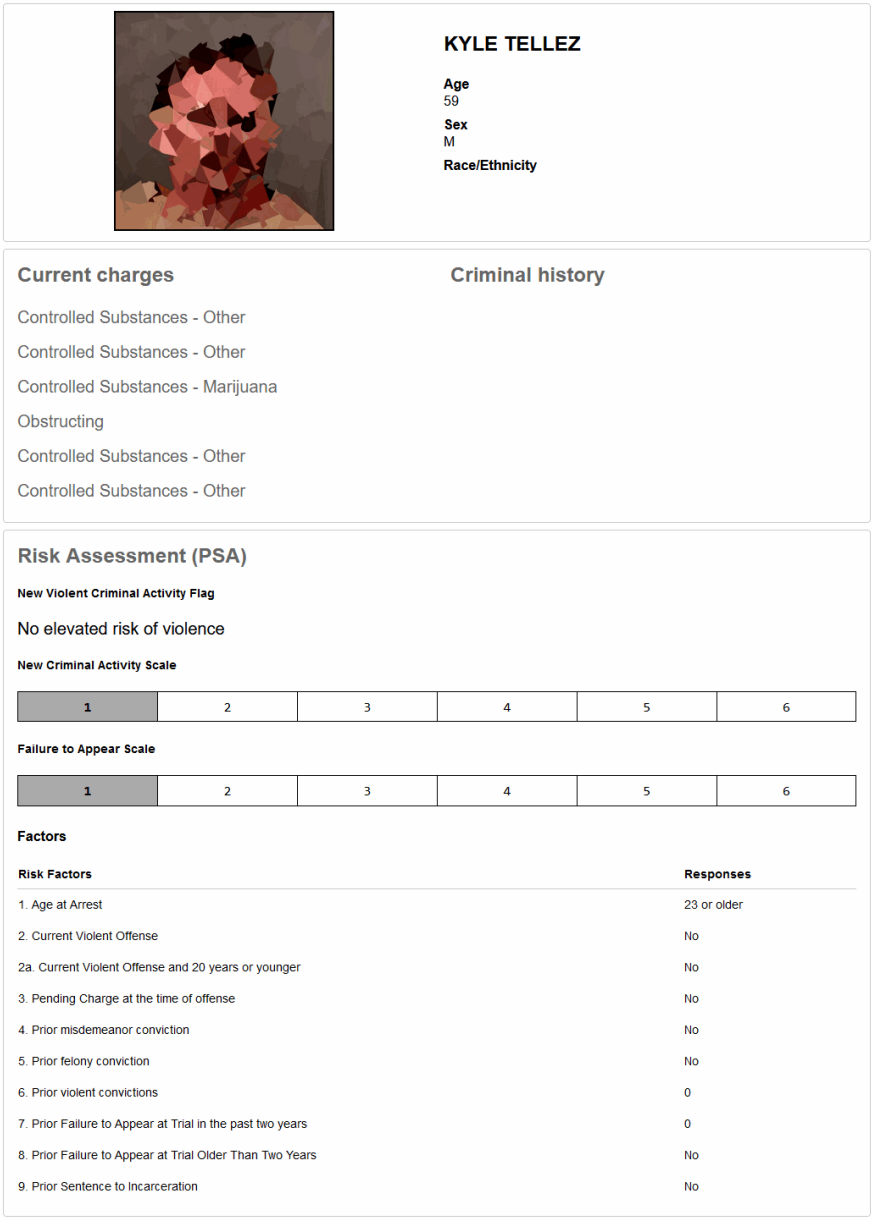
1. Determine whether the defendant's failure to appear risk is "accurate" (i.e. whether it matches the risk assessment). If it is, move to step 2. If not, randomly switch the defendant's risk value to one of the other values. (e.g., if a "medium" risk assessment was determined to be inaccurate, switch to either high or low.)
2. Evaluate whether the defendant failed to appear, using the probability that corresponds to the risk value from step 1. If the defendant fails to appear, generate a corresponding newspaper "card", which appears randomly among the next 10 defendant cards. Otherwise, repeat steps 1 and 2 with new crime, then violence.

If not configured to end after a fixed number of defendants, a simulation run ends when either the jail or public fear meter are filled up.[6]

2.3 Configurability

While we deployed Detain/Release as part of a structured class lesson, the engine that powers the simulation could theoretically be deployed as a basic, configurable platform for testing reactions to risk assessments and their interfaces, using synthetic defendants.

A portion of the engine's configurability is exposed to teachers who are using the simulation in a classroom setting. Still more is untapped. Below is an example of Detain/Release's engine generating synthetic defendants with basic charge sheets (beyond lead charge) and criminal histories while applying the Arnold Foundation's Public Safety Assessment (PSA) risk assessment.



KYLE TELLEZ

Age
59

Sex
M

Race/Ethnicity

Current charges

Controlled Substances - Other
Controlled Substances - Other
Controlled Substances - Marijuana
Obstructing
Controlled Substances - Other
Controlled Substances - Other

Criminal history

Risk Assessment (PSA)

New Violent Criminal Activity Flag

No elevated risk of violence

New Criminal Activity Scale

1	2	3	4	5	6

Failure to Appear Scale

1	2	3	4	5	6

Factors

Risk Factors	Responses
1. Age at Arrest	23 or older
2. Current Violent Offense	No
2a. Current Violent Offense and 20 years or younger	No
3. Pending Charge at the time of offense	No
4. Prior misdemeanor conviction	No
5. Prior felony conviction	No
6. Prior violent convictions	0
7. Prior Failure to Appear at Trial in the past two years	0
8. Prior Failure to Appear at Trial Older Than Two Years	No
9. Prior Sentence to Incarceration	No

Figure 4: Detail/Release's generator used to change sheets and criminal histories, with the PSA risk assessment

3.0 The lesson

Jason Tashea and I designed a lesson around Detain/Release for our 2018 and 2019 iterations of our Georgetown Law practicum on Criminal Justice Technology, Policy, and Law.[7] We have also taught the simulation in various forms in another half-dozen classes and events, and we estimate it has been taught by others in another 10-20 contexts.

Our lesson intersperses runs of the simulation with short discussions and lectures on risk assessments and their deployment of pretrial. Our goal is not to teach the math behind actuarial risk assessments, but the history, origins, and motivations behind the ongoing demand for risk assessments.

To draw students out of the screen, we ask students to run the simulation in small groups. Students discuss decisions amongst themselves, engage more fully in the interstitial discussions, and (hopefully) reflect on their own decision-making tendencies. We use a live dashboard and web-based “remote control” to help orchestrate the lessons, and track student progress in real time. The dashboard has four live data displays, which we introduce over the course of the lesson.

We begin with a basic introduction to bail and pretrial detention, the uncertainty inherent in a pretrial decision, and the rise of risk assessments as a structured decision-making aid. Students then do an individual “tutorial” run of Detain/Release: 10 defendants, and no risk assessment. Here, we aim to familiarize students with the basics of the simulation and the ecosystem it represents, and to give students a sample of the uncertainty that the pretrial structure presents for judges. After a short discussion, we divide students into groups and do three full runs of Detain/Release. During the runs, our dashboard displays each team’s real-time progress through a simulation run.

Between each run, we prompt students to reflect on and articulate their decision-making models, and gradually reveal more information about the simulation’s mechanics. We also give short lectures on actuarial risk assessments, legal challenges to their implementation, and the inescapable racial foundations of criminal justice data and the decision-support tools that use them.

To prompt students to reflect on their own decision-making, we use a dashboard panel that displays aggregate tendencies of all participants throughout all runs in a given session. To frame our discussions of bias, a third dashboard panel shows the faces of detained and released defendants.

We conclude the lesson with a discussion of collateral consequences: the housing and economic instability that linger after a defendant’s engagement with the criminal justice system concludes. Our last dashboard screen details collateral consequences for the defendants students have selected for detention during the lesson. This takes the air out of the room, and drives home the outsize impact of brief pretrial detention decisions. Finally, we reveal that the risk assessment is mostly a fake, and that the

heuristics the students have developed to adjust for the simulation are illusory. We challenge students to consider how risk assessments—or any decision-support tool—may reframe a person's decision-making, and how it may be exceedingly difficult for a user to tell if that tool is high-quality or mere snake oil.

4.0 Lessons learned

In general, students found the simulation and lesson engaging and memorable: it and another simulation were consistently our most well-received lessons. Memorability can be a good trade-off for fidelity: our goal was not to make students experts on the statistical mechanics of risk assessments, but to teach the importance of interrogating risk assessments and other tools like it.

Each run of the simulation is an opportunity to learn, and to be surprised. The lessons described in this part are the product of anecdotal observations, not data. Still, they illustrate the wide range of responses students have to interactive lessons, and help demonstrate how successive teachings of Detain/Release could be tweaked or improved.

Lesson 1: Participants built their own mental models around the simulation's world.

Whether students or professionals, we observed that participants would readily build their own mental models around the simulation's world. As hoped, participants engaged with the contextual feedback—meters, newspaper articles—as proxies for outside pressure, and incorporated it into their decision-making calculus. Somewhat more surprisingly, students would try to predict how the risk assessment evaluated defendants, and attempt to identify factors that the risk assessment emphasized or elided. While participants' decision-making models shared common elements (defendants with a high risk of violence were likely to be detained, for example), they were hardly uniform. Put another way, a participant's own mental model, the simulation's structure, and the specific mix of defendants a participant reviewed seemed to yield a range of unique mental models. We counted this as a success, especially given the uneven results that pretrial risk assessments have yielded in practice.

Few participants suspected that our risk assessment was mostly a fake. Most participants assumed it was imperfect, and attempted to identify (and compensate for) the factors that the risk assessment used (or missed) to make its evaluation. This may have in part been the result of our outsize influence over the class as teachers. But in the truncated sessions we've run with legal and technology professionals, participants

were similarly accepting of our fictional risk assessment. This raises an interesting question for future research: would a judge be able to distinguish between a "high-quality" risk assessment and a well-made "placebo machine"? While not the focus of Detain/Release, our work suggests this question may not have an easy answer.

Lesson 2: Our use of fake faces didn't land. At the moment Detain/Release is not well-suited to help a participant reflect on implicit bias with respect to race.

At the time of the simulation's design, we had ethical reservations about using actual photographs of real people, and generating synthetic faces was still a non-trivial task. The solution we landed on—distorted faces—didn't look enough like a human face to trigger recognition from participants. A future version of the simulation, or one geared toward research, could rely on a more realistic-looking set of synthetic images.

Lesson 3: We need to rein in law students' competitive instincts.

As we taught the lesson to law students, we quickly realized that we needed to nudge students away from treating the dashboard as a scoreboard, and the simulation as a competitive exercise. (This was less of an issue in runs where the participants were professionals.)

Lesson 4: Participants want more information—and we can give it to them.

While running the simulation, participants consistently wish they had more information about their defendants—criminal histories, narrative description of a crime, and so on. A future version of the simulation geared toward training could rely on more robust synthetic profiles, although this would likely make for a longer in-class lesson.

5.0 Conclusion

The evaluation of technology tools cannot be divorced from their implementation context. Nor can critical education about technology tools. The academy must empower future lawyers to think critically about the digital tools that they will rely on to advance their work. Interactive simulations can help. They can recreate or mimic environments that resemble real-life contexts, and help deliver memorable lessons that stay with students.

Header image generated with Wombo

Footnotes

1.

A synthetic defendant's age and charge are loosely derived from Bureau of Justice Statistics data on arrest records for the 50 largest cities in the US. Surnames are derived from a U.S. Census dataset of surnames that occur more than 100 times in the 2010 Census, while first names are derived from a dataset of popular baby names in 2000.

↩

2. My goal with Detain/Release's synthetic defendants was believability, rather than perfect fidelity. The simulation isn't designed to generate a synthetic *dataset*, but it shouldn't generate results that feel obviously "off" in some way. Anecdotally, users of the simulation are more likely to notice unusual name/age combinations: due to the way first names are generated, usually an older defendant with a "young-looking" name. ↩

3. The simulation engine's pseudorandom generator is seeded. Future versions of *Detain/Release* could enable a class to independently navigate identical runs of the simulation. ↩

4. The simulation measures "time" in terms of defendants reviewed. One defendant review = one "tick". ↩

5. The text snippets are derived from local news blotters and custom copy. ↩

6. The default configuration is artificially balanced: A run where every defendant was detained would end at about the same time as a run where every defendant was released. ↩

7.
2019 was our final year teaching the course. I've written on the lesson elsewhere, at https://medium.com/berkman-klein-center/detain-release-simulating-algorithmic-risk-assessments-at-pretrial-375270657819.

Jason has written about it at https://nlsir.com/detain-release-teaching-law-students-about-ai-through-simulated-detention-hearings/. ↩