Negative: COMPAS / Black Box Judiciary

By “Coach Vance” Trefethen, Benjamin Lu, Oliver Lippard

***Resolved: The United States federal government should substantially reform the use of Artificial Intelligence technology***

Case Summary: The AFF plan bans the use of “Black Box” (non-transparent, mysterious algorithms) AI systems used in the judicial system to recommend bail, parole and sentencing. The most notable one (and the only one specifically mentioned in AFF evidence) is known by its acronym COMPAS (Correctional Offender Management Profiling for Alternative Sanctions).  
  
NOTE: Despite AFF saying so, their evidence will never say (because it is not true) that COMPAS is used to determine guilt or innocence. AI is never used for that. AI/COMPAS is used to estimate the likelihood that a person will skip out on bail (to help make the decision whether to grant bail and if so, how high it should be), or whether a convict will likely recidivate (to help the judge decide how long a sentence he should get, or to guide the parole board on whether/when to grant parole). On some occasions, COMPAS results might be used by a witness to present character testimony at trial, but COMPAS does not take the place of the jury. It also does not “decide” the sentence, it merely gives input to the judge and the judge makes the decision.  
  
“Loomis” was a criminal in Wisconsin who appealed his 6-year state prison sentence because he couldn’t determine how COMPAS came up with its recommendation to the sentencing judge (the algorithm isn’t fully disclosed). AFF will cite him as a living example of someone who was harmed by COMPAS. He’s actually a good source for Negative material if you dig into his situation. He pled guilty to 2 felonies related to a drive-by shooting. Judge gave him the maximum, a 6-year sentence. But if you look at his awful criminal record, you’ll come away wishing he could have gotten more. And courts have ruled that there’s no constitutional violation when a judge gives a criminal the maximum sentence. Nobody has a “right” to a lighter sentence and judges have a constitutionally protected role to use whatever input they want in sentencing. Nobody’s rights were violated and Loomis wasn’t “harmed.”

Negative: COMPAS / Black Box Judiciary 3

NEG PHILOSOPHY 3

1) Compare to humans, not perfection 3

2) Limit AFF’s claims 3

INHERENCY 3

1. Problem already solved: No Black Box in use 3

COMPAS algorithms are transparent and were published in a peer-reviewed article 3

COMPAS isn’t a “Black Box”- It’s fully transparent to the agencies using it 3

Other agencies can check the quality of the data whenever they want 4

HARMS / SIGNIFICANCE 4

1. Rudin Study flawed 4

AFF study by Rudin, Wang & Coker about harms of COMPAS has misrepresentations, sensationalism and factual errors 4

Context of Jackson & Mendoza evidence: It refers to Rudin, Wang & Coker 2020 4

2. A/T “COMPAS makes lots of errors” 5

AFF study by Rudin, Wang & Coker used wrong data to find “errors” in COMPAS criminal risk results 5

3. A/T “AI no better than human volunteers at predicting recidivism (Dartmouth Study)” 5

California study finds COMPAS is much better than humans, using more realistic data than Dartmouth 5

California Study: When given extra information, algorithms did better than humans (refuting Dartmouth study) 5

Dartmouth Study flawed: More realistic study finds AI is better than humans 6

What Dartmouth got wrong: They immediately told participants the results after each evaluation, which doesn’t happen in real life. Under real life conditions, AI does much better 6

90% accuracy for AI versus 60% for humans when real-world tests are used 6

4. No sentencing harm 7

Judges have discretion to use any information they want, up to the statutory maximum. There’s no constitutional violation 7

5. A/T “Loomis case in Wisconsin” – AFF’s example of someone “harmed” by COMPAS. Loomis wasn’t harmed. 7

First, no trial. Loomis wasn’t convicted by COMPAS. He pled guilty to 2 charges related to a drive-by shooting 7

Second, his complaint. He didn’t like being given the maximum sentence for his crimes 7

Third, No harm. There’s no 6th Amendment violation if someone gets a longer sentence than they hoped for, as long as it’s within the statutory maximum 8

Fourth, what would you have done differently from COMPAS? Loomis violated parole more than 5 times, and was arrested 12 times before. 8

6. A/T “No way to correct mistakes in COMPAS” 8

Loomis refutes. Wisconsin Supreme Court found ways Loomis could challenge COMPAS mistakes 8

7. COMPAS is input, not decision-making 9

COMPAS is input to the sentencing process, not the decision of what sentencing is 9

Judges and other officials make the final decision, not AI 9

8. A/T “PATTERN algorithm is inaccurate” 9

One of the highest levels of predictive performance 9

SOLVENCY 10

1. No good replacement for Black Box 10

AFF experts Rudin & Radin 2019 say we should wait until an interpretable model exists to replace Black Box, and they admit we don’t have one yet 10

AFF claims transparent models exist, but never names one 10

No COMPAS replacement algorithm could ever prove itself to be “fair” 10

2. Replacing one “Black Box” with another: the Judge 11

AFF evidence says removing COMPAS without alternatives gives us another black box 11

AFF plan doesn’t provide any alternative 11

DISADVANTAGES 11

1. Worse decision-making 11

If you think AI is bad, the only alternative, human judgement, is worse. Example: Thousands falsely convicted 11

Impact: Harsher sentences (or “more harm” if you believe AFF’s view that harsh sentences are bad). Human assessments of recidivism are harsher than AI 11

2. Court clog 12

Link: COMPAS is used to ease the administrative burdens of a busy system 12

Brink: Criminal justice system is operating at administrative capacity – that’s why algorithms are used 12

Impact: Justice Denied. Overloaded judges pressed for time make fast decisions inconsistent with justice 12

Works Cited 13

Negative: COMPAS / Black Box Judiciary

NEG PHILOSOPHY

1) Compare to humans, not perfection

The alternative to AI recommendations is human recommendations. The question isn’t whether AI is perfect, the question is whether it makes fewer mistakes than humans, because that’s the alternative if you throw away AI. The Affirmative’s claim that since AI makes mistakes, therefore we shouldn’t use it, would also eliminate all humans from the judicial process, since humans make a lot of mistakes too.

2) Limit AFF’s claims

Don’t let the Affirmative get away with claiming that COMPAS or any other AI system is deciding guilt or innocence or replacing trials in the criminal justice system. COMPAS is used as input for setting bail, length of sentences, and parole decisions. Guilt or innocence is not at stake in this debate round.

INHERENCY

1. Problem already solved: No Black Box in use

COMPAS algorithms are transparent and were published in a peer-reviewed article

Eugenie Jackson and Christina Mendoza 2020. (Jackson and Mendoza are both Research Scientists at the  Northpointe Inc. Research Department) “Setting the Record Straight: What the COMPAS Core Risk and Need Assessment Is and Is Not” 31 Mar 2020 <https://hdsr.mitpress.mit.edu/pub/hzwo7ax4/release/4> (accessed 20 Jan 2022) (brackets added)

We come next to the highly significant assumptions that the GRRS *[General Recidivism Risk Scale]* and VRRS *[Violent Recidivism Risk Scale]* must be complex, difficult to interpret, and entirely opaque. None of these assumptions is warranted. As we have stressed earlier, the GRRS and VRRS are fully interpretable and transparent to the agencies that use them. In Brennan et al. (2009), a publicly available peer-reviewed article in a relevant criminal justice journal, the risk variables, statistical methods (logistic regression and survival analyses) used for the Core RNA, and validation results are explicitly described. Similar descriptions are provided in training sessions to all of Northpointe’s users.

COMPAS isn’t a “Black Box”- It’s fully transparent to the agencies using it

Eugenie Jackson and Christina Mendoza 2020. (Jackson and Mendoza are both Research Scientists at the  Northpointe Inc. Research Department) “Setting the Record Straight: What the COMPAS Core Risk and Need Assessment Is and Is Not” 31 Mar 2020 <https://hdsr.mitpress.mit.edu/pub/hzwo7ax4/release/4> (accessed 20 Jan 2022) (brackets added)

The authors *[Cynthia Rudin, Caroline Wang, and Beau Coker]*  have assumed that the Core RNA *[COMPAS Core Risk and Needs Assessment]* uses complex methods and/or models, is hard to interpret, and is opaque to users. None of this is correct. The Core RNA was developed using widely established methods yielding simple models that are directly interpretable and fully transparent to the agencies that use it.

Other agencies can check the quality of the data whenever they want

Eugenie Jackson and Christina Mendoza 2020. (Jackson and Mendoza are both Research Scientists at the  Northpointe Inc. Research Department) “Setting the Record Straight: What the COMPAS Core Risk and Need Assessment Is and Is Not” 31 Mar 2020 <https://hdsr.mitpress.mit.edu/pub/hzwo7ax4/release/4> (accessed 20 Jan 2022)

All risk and need data belong to the agencies who collect it. Many agencies have their own research and IT departments, and much of their time is spent monitoring data quality and ensuring confidentiality. Unlike credit agencies who may own clients’ data, Northpointe owns no data and has no “control over criminal risk scores” (Rudin et al., 2020). This is a fiction the authors propagate. Northpointe complies with all agency requirements to ensure the security and privacy of the data it works with. Background checks and data-sharing agreements are usually required before work can begin on any project.

HARMS / SIGNIFICANCE

1. Rudin Study flawed

AFF study by Rudin, Wang & Coker about harms of COMPAS has misrepresentations, sensationalism and factual errors

Eugenie Jackson and Christina Mendoza 2020. (both are Research Scientists at equivant/Northpointe Research Department) “Setting the Record Straight: What the COMPAS Core Risk and Need Assessment Is and Is Not” 31 Mar 2020 <https://hdsr.mitpress.mit.edu/pub/hzwo7ax4/release/4> (accessed 20 Jan 2022) (brackets added)

To begin with, we expect the authors to be above using sensationalist tactics like suggesting that the Core RNA has anything to do with crimes like the murder of the man in Westervelt (2017).undefined A similar tactic is employed when the authors write, “In the past, there have been documented cases where individuals have received incorrect COMPAS scores based on incorrect criminal history data (Wexler, 2017a, b) and have had no mechanism to correct it after a decision was made based on that incorrect score.” The use of the plural “cases” and two citations belie the fact that there is a single individual being described. Moreover, the item discussed in those stories is not part of either risk assessment, but part of a need scale. Another sensationalist tactic is the authors’ repeated claim that the General and Violent risk scales have 137 items when their own table shows that together they only have 40.

Context of Jackson & Mendoza evidence: It refers to Rudin, Wang & Coker 2020

Eugenie Jackson and Christina Mendoza 2020. (both are Research Scientists at equivant/Northpointe Research Department) “Setting the Record Straight: What the COMPAS Core Risk and Need Assessment Is and Is Not” 31 Mar 2020 <https://hdsr.mitpress.mit.edu/pub/hzwo7ax4/release/4> (accessed 20 Jan 2022)

“The Age of Secrecy and Unfairness in Recidivism Prediction” (Rudin, Wang, & Coker, 2020) treats a wide variety of topics. It includes a clever approach to determining the age dependency of a COMPAS Core Risk and Needs Assessment (hereafter Core RNA). Regrettably, we find that several mistakes mar this effort.

2. A/T “COMPAS makes lots of errors”

AFF study by Rudin, Wang & Coker used wrong data to find “errors” in COMPAS criminal risk results

Eugenie Jackson and Christina Mendoza 2020. (both are Research Scientists at equivant/Northpointe Research Department) “Setting the Record Straight: What the COMPAS Core Risk and Need Assessment Is and Is Not” 31 Mar 2020 <https://hdsr.mitpress.mit.edu/pub/hzwo7ax4/release/4> (accessed 20 Jan 2022)

We note next a pair of erroneous assumptions the authors appear to make. It appears the authors conflate number of arrests with number of charges. These are two very different kinds of things. A single arrest can be associated with many charges. Additionally, the authors appear to conflate long criminal history with extensive criminal involvement. Again, these are two different kinds of things, as can be seen by considering a case of high age with early first arrest yet low criminal involvement. In Section 3, the authors present low-risk scores that are associated with high-charge tallies (or “counts”) in the context of discussing putative misattribution of high criminal involvement as low risk. Nowhere in this discussion is there any direct evidence of high criminal involvement presented. Apparently, counts are being mistaken as evidence of high criminal involvement, which in turn is mislabeled as a long criminal history.

3. A/T “AI no better than human volunteers at predicting recidivism (Dartmouth Study)”

California study finds COMPAS is much better than humans, using more realistic data than Dartmouth

Edward Lempinen 2020. (writer and media relations specialist with the UC Berkeley Office of Communications and Public Affairs. He has extensive experience in science and science policy communication.) “Algorithms are better than people in predicting recidivism, study says” 14 Feb 2020 <https://news.berkeley.edu/2020/02/14/algorithms-are-better-than-people-in-predicting-recidivism-study-says/>  (accessed 19 Jan 2022)

In addition to the five risk factors used in the Dartmouth study, they added 10 more, including employment status, substance use and mental health. They also expanded the methodology: Unlike the Dartmouth study, in some cases the volunteers would not be told after each evaluation whether their predictions were accurate. Such feedback is not available to judges and others in the court system. The outcome: Humans performed “consistently worse” than the risk assessment tool on complex cases when they didn’t have immediate feedback to guide future decisions. For example, the COMPAS correctly predicted recidivism 89% of the time, compared to 60% for humans who were not provided case-by-case feedback on their decisions. When multiple risk factors were provided and predictive, another risk assessment tool accurately predicted recidivism over 80% of the time, compared to less than 60% for humans. The findings appear to support continued use and future improvement of risk assessment algorithms.

California Study: When given extra information, algorithms did better than humans (refuting Dartmouth study)

Zhiyuan Lin, Jongbin Jung, Sharad Goel and Jennifer Skeem 2020. (Lin is a PhD candidate in computer science at Stanford University, Jung is a data scientist with a PhD from Stanford University, Goel is an assistant professor at Stanford University and executive director of the Stanford Computational Policy Lab, and Skeem is a professor of social welfare and public policy at University of California at Berkeley and directs Berkeley’s Risk-Resilience Lab.) “In the U.S. criminal justice system, algorithms help officials make better decisions, our research finds” 2 Mar 2020 <https://www.washingtonpost.com/politics/2020/03/02/us-criminal-justice-system-algorithms-do-help-officials-make-better-decisions-our-research-finds/> (accessed 19 Jan 2022)

Humans were also worse than algorithms at exploiting additional information — something that criminal justice officials have in abundance. In yet another version of our experiment, we gave humans and algorithms detailed vignettes that included more than the five pieces of information provided about a defendant in the original Dartmouth study. The algorithms that had this additional information performed better than those that did not, but human performance did not improve.

Dartmouth Study flawed: More realistic study finds AI is better than humans

Zhiyuan Lin, Jongbin Jung, Sharad Goel and Jennifer Skeem 2020. (Lin is a PhD candidate in computer science at Stanford University, Jung is a data scientist with a PhD from Stanford University, Goel is an assistant professor at Stanford University and executive director of the Stanford Computational Policy Lab, and Skeem is a professor of social welfare and public policy at University of California at Berkeley and directs Berkeley’s Risk-Resilience Lab.) “In the U.S. criminal justice system, algorithms help officials make better decisions, our research finds” 2 Mar 2020 <https://www.washingtonpost.com/politics/2020/03/02/us-criminal-justice-system-algorithms-do-help-officials-make-better-decisions-our-research-finds/> (accessed 19 Jan 2022)

The Dartmouth findings do not hold in settings that are closer to real criminal justice situations  
The problem isn’t that the Dartmouth study’s specific results are wrong. We got very similar results when we reran the study by asking our own participants to read and rate the same defendant descriptions that their researchers used. It’s that their results are limited to a narrow context. We repeated the experiment by asking our participants to read descriptions of several new sets of defendants and found that algorithms outperformed people in every case.

What Dartmouth got wrong: They immediately told participants the results after each evaluation, which doesn’t happen in real life. Under real life conditions, AI does much better

Zhiyuan Lin, Jongbin Jung, Sharad Goel and Jennifer Skeem 2020. (Lin is a PhD candidate in computer science at Stanford University, Jung is a data scientist with a PhD from Stanford University, Goel is an assistant professor at Stanford University and executive director of the Stanford Computational Policy Lab, and Skeem is a professor of social welfare and public policy at University of California at Berkeley and directs Berkeley’s Risk-Resilience Lab.) “In the U.S. criminal justice system, algorithms help officials make better decisions, our research finds” 2 Mar 2020 <https://www.washingtonpost.com/politics/2020/03/02/us-criminal-justice-system-algorithms-do-help-officials-make-better-decisions-our-research-finds/> (accessed 19 Jan 2022)

After each question, the Dartmouth researchers told participants whether their prediction was correct — so we did that, too, in our initial experiments. As a result, those participants were able to immediately learn from their mistakes. But in real life, it can take months or years before criminal justice professionals discover which people have reoffended. So we redid our experiment several more times without this feedback. We found that the gap in accuracy between humans and algorithms doubled, from 12 to 24 percentage points. In other words, the gap increased when the experiment was more like what happens in the real world. In fact, in this case, where immediate feedback was no longer provided, our participants correctly rated only 47 percent of the vignettes they read — worse than simply flipping a coin.

90% accuracy for AI versus 60% for humans when real-world tests are used

*Edward Lempinen 2020. (Edward Lempinen is a writer and media relations specialist with the UC Berkeley Office of Communications and Public Affairs. He has extensive experience in science and science policy communication, including support for global high-level science diplomacy initiatives.) “Algorithms are better than people in predicting recidivism, study says” 14 Feb 2020* [*https://news.berkeley.edu/2020/02/14/algorithms-are-better-than-people-in-predicting-recidivism-study-says/*](https://news.berkeley.edu/2020/02/14/algorithms-are-better-than-people-in-predicting-recidivism-study-says/) *(accessed 19 Jan 2022)*

When assessing just a handful of variables in a controlled environment, even untrained humans can match the predictive skill of sophisticated risk-assessment instruments, says the new study by scholars at Stanford University and the University of California, Berkeley. But real-world criminal justice settings are often far more complex, and when a larger number of factors are useful for predicting recidivism, the algorithm-based tools performed far better than people. In some tests the tools approached 90% accuracy in predicting which defendants might be arrested again, compared to about 60% for human prediction.

4. No sentencing harm

Judges have discretion to use any information they want, up to the statutory maximum. There’s no constitutional violation

Gregory Garre, Grace Becker, Jessica Silver and Angela Miller 2008. (Garre - Acting Solicitor General. Becker - Acting Assistant Attorney General. Silver and Miller - Attorneys with US Dept of Justice) August 2008 brief for the U.S. to the Supreme Court in the case of Marlowe v. U.S. <https://www.justice.gov/sites/default/files/osg/briefs/2008/01/01/2007-1390.resp.pdf> (accessed 18 Dec 2021)(ellipses and brackets in original)

Because the Guidelines were advisory, the district court had the discretion to sentence petitioner to any term of imprisonment up to the statutory maximum of life authorized by the jury’s finding that petitioner’s violation of Kuntz’s civil rights resulted in his death. See 18 U.S.C. 242. Under those circumstances, the factual findings made by the district court in exercising its sentencing discretion raise no Sixth Amendment concerns. Booker and cases elaborating on that decision make clear that, under an advisory Guidelines regime, judicial fact-finding that supports a sentence within the statutory maximum set forth in the United States Code does not violate the Sixth Amendment. As the Court explained in Booker: We have never doubted the authority of a judge to exercise broad discretion in imposing a sentence within a statutory range.

5. A/T “Loomis case in Wisconsin” – AFF’s example of someone “harmed” by COMPAS. Loomis wasn’t harmed.

First, no trial. Loomis wasn’t convicted by COMPAS. He pled guilty to 2 charges related to a drive-by shooting

Atlas Lab copyright 2020 (an educational resource for lawyers to learn about automated decision making technologies) “Loomis v. Wisconsin - *COMPAS risk scores on trial “* <https://www.atlaslab.org/post/loomis-v-wisconsin-compas-risk-scores-on-trial#:~:text=This%20case%20concerned%20Eric%20Loomis,prepared%20a%20Presentencing%20Investigation%20Report>. (accessed 23 Jan 2022)

In the US, risk assessment tools have been adopted across the criminal justice system. Initially they were primarily used to inform decisions on supervision and treatment of offenders. However, these assessment tools started to expand into decisions on sentencing. This case concerned Eric Loomis, who was charged in 2013 with five offences related to a drive-by shooting. He entered a guilty plea for some of these charges, and others were dismissed.

Second, his complaint. He didn’t like being given the maximum sentence for his crimes

Dr. Anne L. Washington 2018 (PhD; Assistant Professor of Data Policy, Dept of Applied Statistics, Social Science, and Humanities. Steinhardt School, New York Univ.) “How to argue with an algorithm” COLORADO TECHNICAL LAW JOURNAL <http://ctlj.colorado.edu/wp-content/uploads/2021/02/17.1_4-Washington_3.18.19.pdf> (accessed 23 Jan 2022)

The sentencing of Mr. Eric L. Loomis of Wisconsin was based in part on a predictive algorithm that classified him as a high-risk defendant. The algorithm at issue, the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), is a proprietary algorithm sold by equivant, a private company that was doing business as Northpointe before 2017. The Wisconsin circuit court sentenced Mr. Loomis to the maximum penalty on two counts after reviewing the predictions derived from the COMPAS risk-assessment algorithm, despite the defendant’s claims that using a proprietary predictive risk assessment in sentencing violated his due process rights.

Third, No harm. There’s no 6th Amendment violation if someone gets a longer sentence than they hoped for, as long as it’s within the statutory maximum

Judge Thomas B. Griffith 2014. (US Court of Appeals for the District of Columbia Circuit) 14 Mar 2014 opinion of the Court in the case of “U.S. versus Joseph Jones, also known as Jo Jo” <https://www.cadc.uscourts.gov/internet/opinions.nsf/EF97673D47DDE5AE85257C9B004E5DA0/%24file/08-3033-1483944.pdf> (accessed 18 Dec 2021)(brackets and ellipses in original)

No Supreme Court majority has ever recognized the validity of such challenges, and among the courts of appeals the consensus is clearer still: every circuit to have considered such challenges has rejected them as inconsistent, in principle, with the post-Booker rule that “[f]or Sixth Amendment purposes, the relevant upper sentencing limit established by the jury’s finding of guilt is . . . the statutory maximum, not the advisory Guidelines maximum corresponding to the base offense level.” Settles, 530 F.3d at 923; see United States v. Norman, 465 F. App’x 110, 120-21 (3d Cir. 2012) (collecting cases). And though our circuit has not specifically considered such challenges, our precedent is equally categorical: judicial fact-finding does “not implicate the Sixth Amendment even if it yield[s] a sentence above that based on a plea or verdict alone.”

Fourth, what would you have done differently from COMPAS? Loomis violated parole more than 5 times, and was arrested 12 times before.

Wisconsin State Supreme Court Justice Ann Walsh Bradley 2016. Decision of the court in State v. Loomis 13 July 2016 <https://www.leagle.com/decision/inwico20160713i48> (accessed 23 Jan 2022) (brackets added)

The COMPAS report attached to Loomis's PSI [Pre-Sentencing Investigation] contains a list of 21 questions and answers regarding these static factors such as:  
• How many times has this person been returned to custody while on parole? 5+  
• How many times has this person had a new charge/arrest while on probation? 4  
• How many times has this person been arrested before as an adult or juvenile (criminal arrest only)? 12

6. A/T “No way to correct mistakes in COMPAS”

Loomis refutes. Wisconsin Supreme Court found ways Loomis could challenge COMPAS mistakes

Wisconsin State Supreme Court Justice Ann Walsh Bradley 2016. Decision of the court in State v. Loomis 13 July 2016 <https://www.leagle.com/decision/inwico20160713i48> (accessed 23 Jan 2022) (brackets added)

Thus, to the extent that Loomis's risk assessment is based upon his answers to questions and publicly available data about his criminal history, Loomis had the opportunity to verify that the questions and answers listed on the COMPAS report were accurate. Additionally, this is not a situation in which portions of a PSI [Pre-Sentence Investigation] are considered by the circuit court, but not released to the defendant. The circuit court and Loomis had access to the same copy of the risk assessment. Loomis had an opportunity to challenge his risk scores by arguing that other factors or information demonstrate their inaccuracy.

7. COMPAS is input, not decision-making

COMPAS is input to the sentencing process, not the decision of what sentencing is

Eugenie Jackson and Christina Mendoza 2020. (both are Research Scientists at equivant/Northpointe Research Department) “Setting the Record Straight: What the COMPAS Core Risk and Need Assessment Is and Is Not” 31 Mar 2020 <https://hdsr.mitpress.mit.edu/pub/hzwo7ax4/release/4> (accessed 20 Jan 2022)

A widely accepted position is that courtroom decisions should not be based solely on a recidivism risk score. The guidelines issued by the National Center for State Courts (NCSC) regarding how risk assessments are to be used in evidence-based sentencing are clear on this (Casey, Warren, & Elek, 2011). NCSC’s first guideline states: “Risk and need assessment information should be used in the sentencing decision to inform public safety considerations related to offender risk reduction and management. It should not be used as an aggravating or mitigating factor in determining the severity of an offender’s sanction” (Casey et al., 2011, p. 11). That is, a person’s risk scores should not have any effect on the sentence imposed. The second guideline states: “Risk and needs assessment information is one factor to consider in determining whether an offender can be supervised safely and effectively in the community” (Casey et al., 2011, p. 14). Other specific factors must be considered when determining “whether an offender is a good candidate for community supervision” (Casey et al., 2011, p. 14).

Judges and other officials make the final decision, not AI

Edward Lempinen 2020. (writer and media relations specialist with the UC Berkeley Office of Communications and Public Affairs) “Algorithms are better than people in predicting recidivism, study says” 14 Feb 2020 (brackets added) <https://news.berkeley.edu/2020/02/14/algorithms-are-better-than-people-in-predicting-recidivism-study-says/>  (accessed 19 Jan 2022)

The findings appear to support continued use and future improvement of risk assessment algorithms. But, as [Jennifer] Skeem [psychologist who specializes in criminal justice at UC Berkeley] noted, these tools typically have a support role. Ultimate authority rests with judges, probation officers, clinicians, parole commissioners and others who shape decisions in the criminal justice system.

8. A/T “PATTERN algorithm is inaccurate”

One of the highest levels of predictive performance

Crime and Justice Research Alliance 2021. (A centralized resource of authoritative experts and relevant research on crime and criminal justice issues.) “Tool to predict recidivism in federal inmates could make more prisoners eligible for early release” 30 Apr 2021 <https://phys.org/news/2021-04-tool-recidivism-federal-inmates-prisoners.html> (accessed 21 Jan 2022) (brackets added)

The study concluded that PATTERN *[Prisoner Assessment Tool Targeting Estimated Risk and Needs]*  demonstrates one of the highest levels of predictive performance, outpacing that of all contemporary assessments researchers reviewed. In addition, PATTERN was validated on the same population on which it was developed and tested on a large development sample, yielding more stable and reliable estimates. Finally, the tool was tailored to specific outcomes (general and violent recidivism) and groups (males and females), which improves its predictive performance. Tests for bias revealed that PATTERN further reduced race/ethnicity disproportionality.

SOLVENCY

1. No good replacement for Black Box

AFF experts Rudin & Radin 2019 say we should wait until an interpretable model exists to replace Black Box, and they admit we don’t have one yet

Cynthia Rudin and Joanna Radin 2019. (Experts cited by AFF for their plan: Rudin is Professor of Computer Science at Duke Univ. Radin is an Associate Professor in Yale's Program in History of Science and Medicine.) “Why Are We Using Black Box Models in AI When We Don’t Need To? A Lesson From An Explainable AI Competition” 22 Nov 2019 <https://hdsr.mitpress.mit.edu/pub/f9kuryi8/release/6> (accessed 20 Jan 2022) (brackets added)

Let us insist that we do not use black box machine learning models for high-stakes decisions unless no interpretable model can be constructed that achieves the same level of accuracy. It is possible that an interpretable model can always be constructed—we just have not been trying.

AFF claims transparent models exist, but never names one

If it’s so easy to replace COMPAS and other Black Box AI with better models, AFF should be able to name one that can replace it. But they never do. That’s because even their own experts admit it doesn’t exist yet.

No COMPAS replacement algorithm could ever prove itself to be “fair”

Dr. Anne L. Washington 2018 (PhD; Assistant Professor of Data Policy, Dept of Applied Statistics, Social Science, and Humanities. Steinhardt School, New York Univ.) “How to argue with an algorithm” COLORADO TECHNICAL LAW JOURNAL <http://ctlj.colorado.edu/wp-content/uploads/2021/02/17.1_4-Washington_3.18.19.pdf> (accessed 23 Jan 2022)

The ProPublica-COMPAS debate questioned what fairness means and how each definition could be mathematically specified. Fairness could be defined as treating everyone the same or it could be defined as giving everyone similar outcomes. Similar outcomes may require that statistical treatments vary. Variation by race or gender may improve statistical outcomes but raises equal protection concerns. **[END QUOTE**] The research underscores the multiple ways that it is possible to describe treatment of people within sets and call it fair. The research led by Jon Kleinberg presents three conditions that could denote fairness: (1) calibration; (2) balancing negative impact; and (3) balancing positive impact. Kleinberg includes mathematical proofs that show that it is not possible to simultaneously have all three conditions at once. The research led by Sam Corbett-Davis considers what fairness means by running tests that avoid race-specific characteristics or including them. They also discuss the problem of giving special treatment in the database to protected classes. Alexandra Chouldechova gives a well-argued comprehensive view of ways to define fairness mathematically, providing more alternatives than Kleinberg et al. does. Chouldechova provides the proofs along with citations to a wide range of literature that discusses each idea further. The team lead by Sarah Tan developed techniques to detect bias by evaluating statistical differences in outcome variables. [**SHE CONCLUDES LATER IN THE SAME CONTEXT QUOTE:**] There is no single mathematical definition of fairness. The people developing a “fair” algorithm must decide on the uniformity or variation that is necessary for a functioning system. Data science experts conclude that the people who control the algorithms define fairness.

2. Replacing one “Black Box” with another: the Judge

AFF evidence says removing COMPAS without alternatives gives us another black box

Cynthia Rudin, Caroline Wang, and Beau Coker 2020. (AFF evidence source: Rudin is Professor of Computer Science at Duke University, Coker has a Masters Degree in Statistical Science at Duke University) “The Age of Secrecy and Unfairness in Recidivism Prediction” 31 Mar 2020 <https://hdsr.mitpress.mit.edu/pub/7z10o269/release/4>  (accessed 20 Jan 2022)

Interestingly, a system that relies only on judges—and does not use machine learning at all—has similar disadvantages to COMPAS; the thought processes of judges is (like COMPAS) a black box that provides inconsistent error-prone decisions. Removing COMPAS from the criminal justice system, without a transparent alternative, would still leave us with a black box.

AFF plan doesn’t provide any alternative

Nothing in AFF’s mandates requires use of any AI alternative at all – no surprise since there aren’t any better than COMPAS. That leaves everything in the hands of the Judge, who, unlike our judge in this debate round, is an imperfect Black Box himself.

DISADVANTAGES

1. Worse decision-making

If you think AI is bad, the only alternative, human judgement, is worse. Example: Thousands falsely convicted

Wayne Housley 2017. (Criminal defense attorney; over 20 years of experience) “HOW MANY INNOCENT AMERICANS ARE JAILED EACH YEAR?” 9 Feb 2017 <https://www.housleylaw.com/blawg/many-innocent-americans-jailed-year/> (accessed 19 Jan 2022)

Obtaining an exact number of wrongly convicted Americans is a difficult task that involves sifting through federal, state, and county court and prison systems. Even with all of the data that is retrievable, a number of conviction cases never go to trial leaving virtually no traceable records behind. It's estimated that approximately 95% of convictions in the U.S. are by plea bargain with no trial. The only quantifiable data that can be gathered on wrongful convictions comes from exonerations and death penalty records. There have been a total of 1,962 nationwide exonerations since 1989. Additionally, it's believed that roughly 4.1% of all defendants sentenced to death are later proven to be innocent.

Impact: Harsher sentences (or “more harm” if you believe AFF’s view that harsh sentences are bad). Human assessments of recidivism are harsher than AI

Zhiyuan Lin, Jongbin Jung, Sharad Goel and Jennifer Skeem 2020. (Lin is a PhD candidate in computer science at Stanford University, Jung is a data scientist with a PhD from Stanford University, Goel is an assistant professor at Stanford University and executive director of the Stanford Computational Policy Lab, and Skeem is a professor of social welfare and public policy at University of California at Berkeley and directs Berkeley’s Risk-Resilience Lab.) “In the U.S. criminal justice system, algorithms help officials make better decisions, our research finds” 2 Mar 2020 <https://www.washingtonpost.com/politics/2020/03/02/us-criminal-justice-system-algorithms-do-help-officials-make-better-decisions-our-research-finds/> (accessed 19 Jan 2022)

Why was human performance so poor? Our participants significantly overestimated risk, believing that people would reoffend much more often than they actually did. In one iteration of our experiment, we explicitly and repeatedly told participants that only 29 percent of the people they were assessing ultimately reoffended, but our recruits still predicted that 48 percent would do so. In a courtroom, these “judges” might have incorrectly flagged many people as high risk who statistically posed little danger to public safety.

2. Court clog

Link: COMPAS is used to ease the administrative burdens of a busy system

Dr. Anne L. Washington 2018 (PhD; Assistant Professor of Data Policy, Dept of Applied Statistics, Social Science, and Humanities. Steinhardt School, New York Univ.) “How to argue with an algorithm” COLORADO TECHNICAL LAW JOURNAL <http://ctlj.colorado.edu/wp-content/uploads/2021/02/17.1_4-Washington_3.18.19.pdf> (accessed 23 Jan 2022)

The incarcerated population in 1983 was 438,830, while in 2014 it was over 1.5 million. Algorithms and data-driven technology help to ease the administrative burdens of these growing systems. In some cases, courts pay external vendors to produce information the court needs, such as classifying people with predictive assessments of recidivism. COMPAS is a brand of risk-need assessment tools designed to provide decisional support through classification.

Brink: Criminal justice system is operating at administrative capacity – that’s why algorithms are used

Dr. Anne L. Washington 2018 (PhD; Assistant Professor of Data Policy, Dept of Applied Statistics, Social Science, and Humanities. Steinhardt School, New York Univ.) “How to argue with an algorithm” COLORADO TECHNICAL LAW JOURNAL <http://ctlj.colorado.edu/wp-content/uploads/2021/02/17.1_4-Washington_3.18.19.pdf> (accessed 23 Jan 2022)

The controversy over risk assessment algorithms hints at whether procedural due process is the cost of automating a criminal justice system that is operating at administrative capacity.

Impact: Justice Denied. Overloaded judges pressed for time make fast decisions inconsistent with justice

Dr. Cristoph Engel and Prof. Keren Weinshall 2020. (Engel - Chair of the Advisory Board, Amsterdam Center for Law and Economics; Director of Max Planck Institute for Research on Collective Goods.. Weinshall is Edward S. Silver Chair in Civil Procedure, Hebrew University, Jerusalem) 24 Nov 2020 “Manna from Heaven for Judges: Judges’ Reaction to a Quasi-Random Reduction in Caseload” JOURNAL OF EMPIRICAL LEGAL STUDIES <https://onlinelibrary.wiley.com/doi/full/10.1111/jels.12265> (“heuristics” – in this context, it refers to deciding things based on past personal experience rather than looking independently at the facts of the present case)

Judges are not only rational actors striving to optimize their use of time. Although most of the aforementioned studies focus on judges’ strategic choices of the less time-consuming legal outcome, their decisions may also be affected by the physical and emotional fatigue, decline in cognitive performance, and elevated stress levels associated with high workloads.  Research has shown that under time pressure, judges are more vulnerable to heuristics and biases. For example, rulings were found to be more inconsistent when judges face a high caseload (Norris [2018](https://onlinelibrary.wiley.com/doi/full/10.1111/jels.12265#jels12265-bib-0055)) and busy judges were found to expend less effort by according higher weight to non-legal cues, such as litigants’ race or gender, to determine case outcomes (Guthrie et al. [2000](https://onlinelibrary.wiley.com/doi/full/10.1111/jels.12265#jels12265-bib-0030), [2007](https://onlinelibrary.wiley.com/doi/full/10.1111/jels.12265#jels12265-bib-0031); Rachlinski et al. [2008](https://onlinelibrary.wiley.com/doi/full/10.1111/jels.12265#jels12265-bib-0061)).

Works Cited

1. ABC News 2018. (writer is Julia Macfarlane. She is a London-based reporter covering global affairs for ABC News. She began her career as a freelance photojournalist working in Indonesia, where she was born. She then moved to Beirut, Lebanon and worked for the BBC News bureau as a producer and TV journalist. A graduate of University of St. Andrews in Scotland, Macfarlane received her MA in English Literature.) 2 February 2018 “Why is the US still in Afghanistan?” https://abcnews.go.com/International/us-afghanistan/story?id=52763044