Machine Learning in the Judicial Process

Yutong Dai and Yunjie Su

Introduction

- Machine learning algorithms are rising in popularity across the judicial process:
  - gathering legal evidence (Relativity)
  - pretrial and bail (PSA)
  - trial and final sentencing (COMPAS)
- However, the algorithms currently employed often perpetuate historical bias.
- Our project analyzes features of major algorithms currently in use to propose the most ethical places of implementation and regulation.
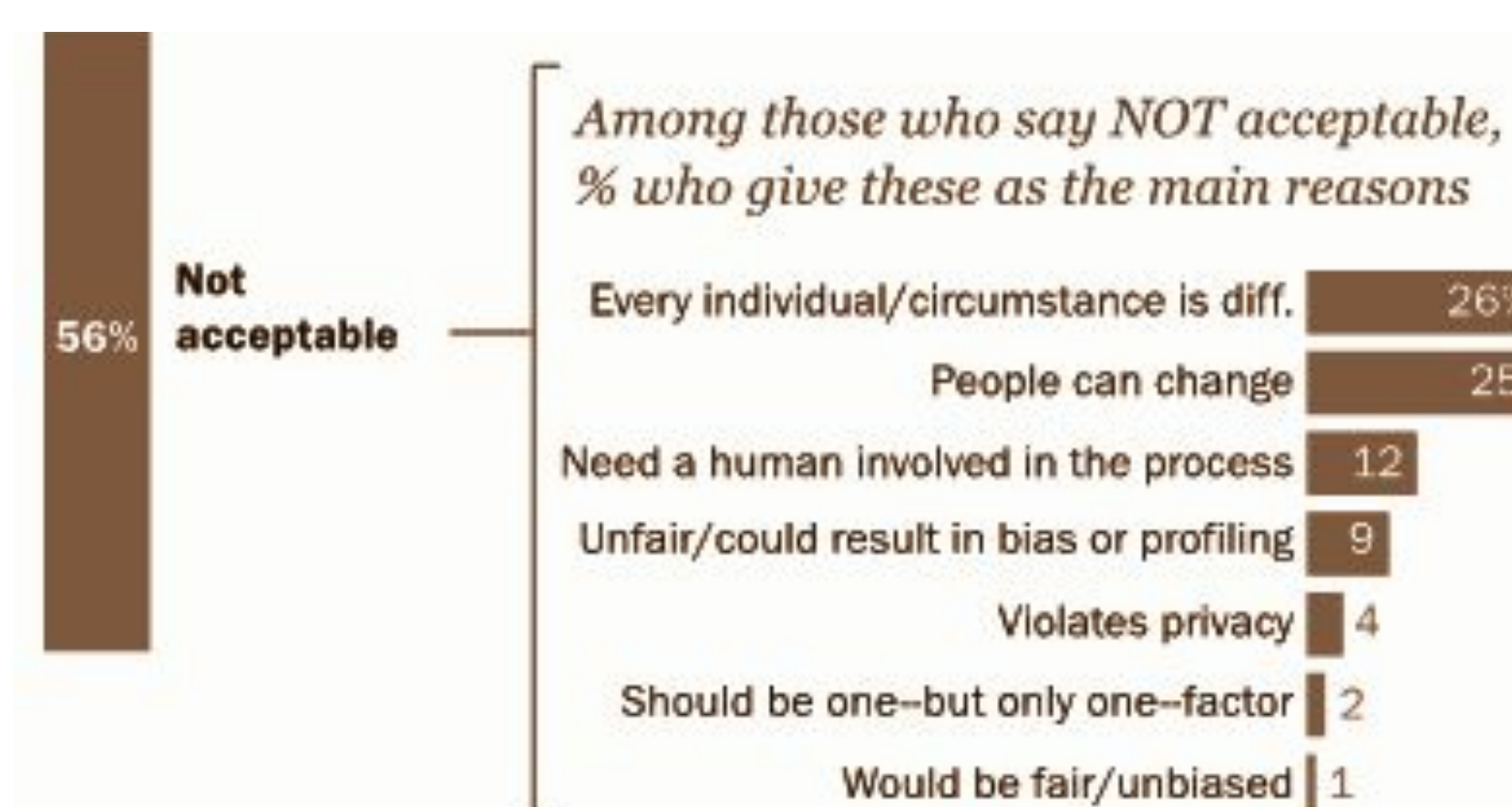
Major Positions

Supporting Use: Efficient, Logical, Corrects Human Bias

- From 2013-2017, NYC courts consistently took an average of 1-2 years to reach a verdict for criminal court cases. [3]

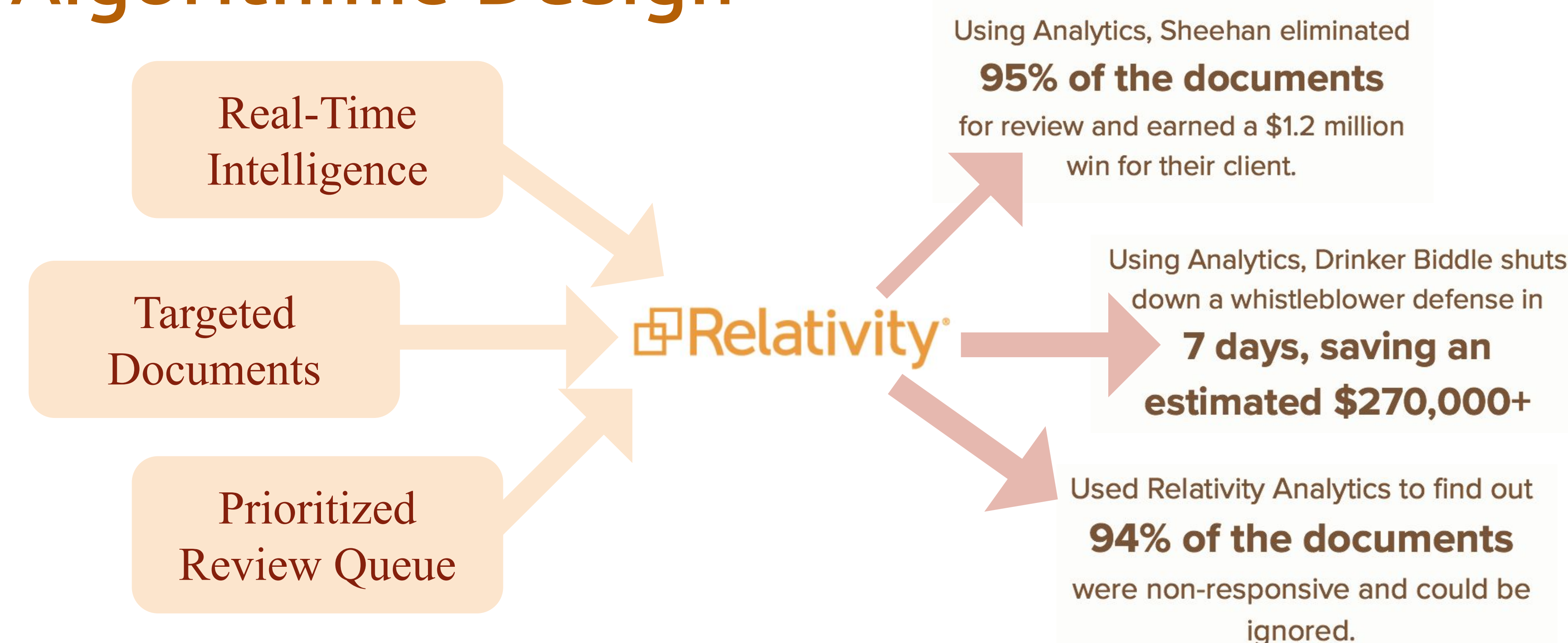### Against Use: Lack of Transparency & Accountability, Unfair

- 2018 PEW Research Survey found that 56% of Americans thought automated criminal risk scores not acceptable

"..it seems like you're determining a person's future based on another person's choices." [4]

Findings

Algorithmic Design

Real-Time Intelligence → Relativity

Targeted Documents → Relativity

Prioritized Review Queue → Relativity

Using Analytics, Sheehan eliminated **95% of the documents** for review and earned a $1.2 million win for their client.

Using Analytics, Drinker Biddle shuts down a whistleblower defense in **7 days, saving an estimated $270,000+**

Used Relativity Analytics to find out **94% of the documents** were non-responsive and could be ignored.
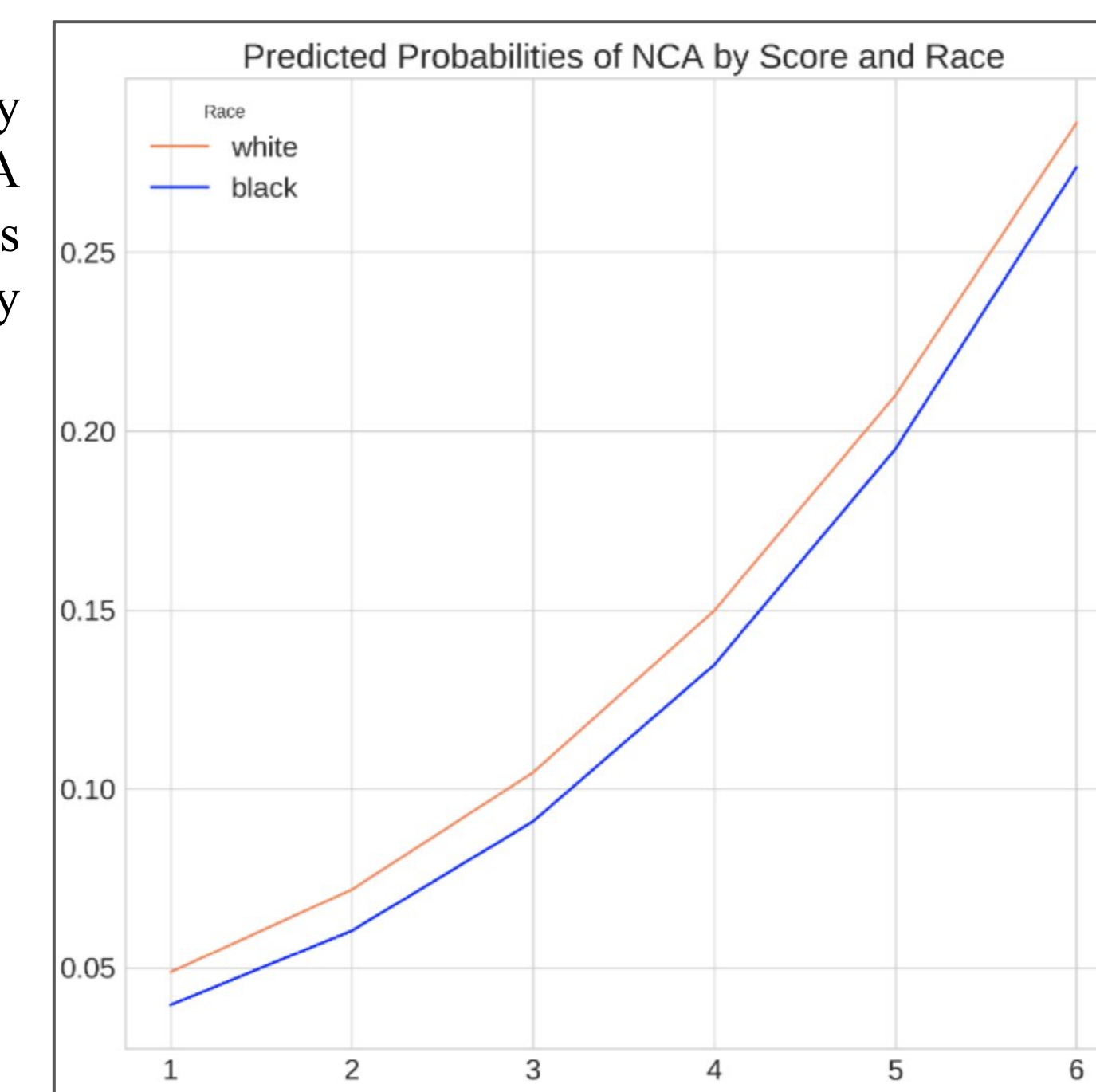
While the Public Safety Assessment clearly states its factors and relative weights, there remains a measurably significant difference in failure rate based on race.

NCA chart from the PSA, which also includes FTA and NVCA charts

Accuracy of the PSA's NSA calculations as separated racially

Although COMPAS considers factors that exhibit significant correlation to further arrest within two years, its predictions of recidivism in trials were found to be racially biased.

COMPAS Scale	Point-Biserial Correlation	AUC	Odds Ratio
General Recidivism Risk	0.34	0.70	3.31
Criminal Involvement	0.20	0.61	1.10
Noncompliance History	0.16	0.61	1.11
Violence History	0.11	0.58	1.06
Current Violence	−0.05	0.52	0.92
Criminal Associates	0.14	0.60	1.09
Substance Abuse	0.02	0.51	1.02
Financial Problems	0.09	0.55	1.08
Voced Problems	0.17	0.61	1.11
Family Crime	0.07	0.54	1.11
Social Environment	0.10	0.56	1.12
Leisure	0.11	0.57	1.07
Residential Instability	0.08	0.55	1.04
Social Adjustment	0.15	0.60	1.10
Socialization Failure	0.18	0.62	1.13
Criminal Opportunity	0.19	0.62	1.10
Social Isolation	0.04	0.52	1.02
Criminal Thinking	0.11	0.57	1.04
Criminal Personality	0.13	0.58	1.05
Cognitive Behavioral	0.23	0.63	1.05

With n=23,635, a correlation of .013 is significant at $p < .05$ (2-tailed).

Table depicts Association Between COMPAS Core Scales and Any Arrest Within Two Years [1]
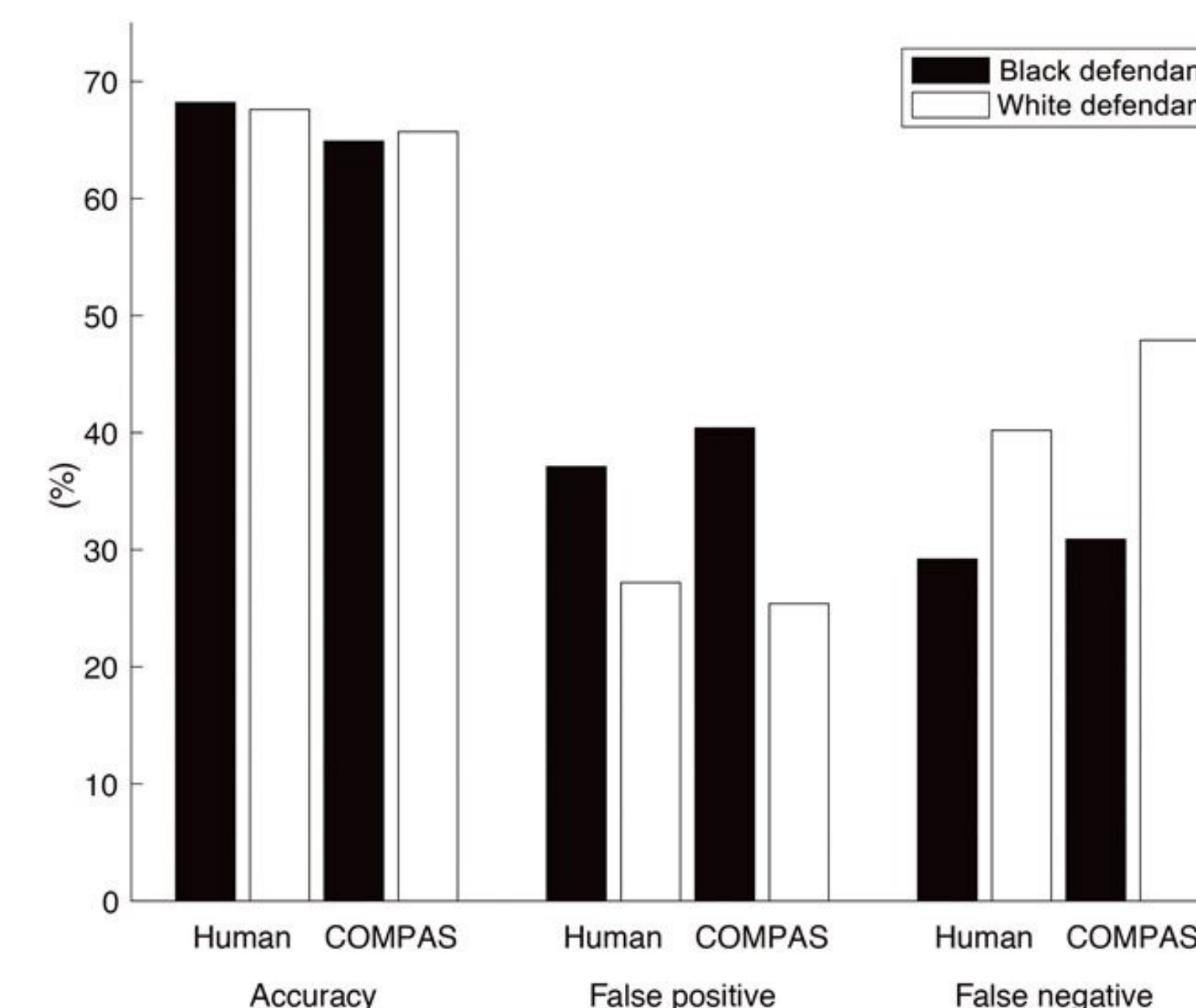
Regulation

- *Gardner v. Florida* established precedent against using confidential passages against a defendant in capital cases.
- **S.3649 - First Step Act**, approved late 2018, gives the Attorney General seven months to develop a risk/needs assessment system for inmates that could potentially shorten their prison sentences

Bill Votes

House
Senate
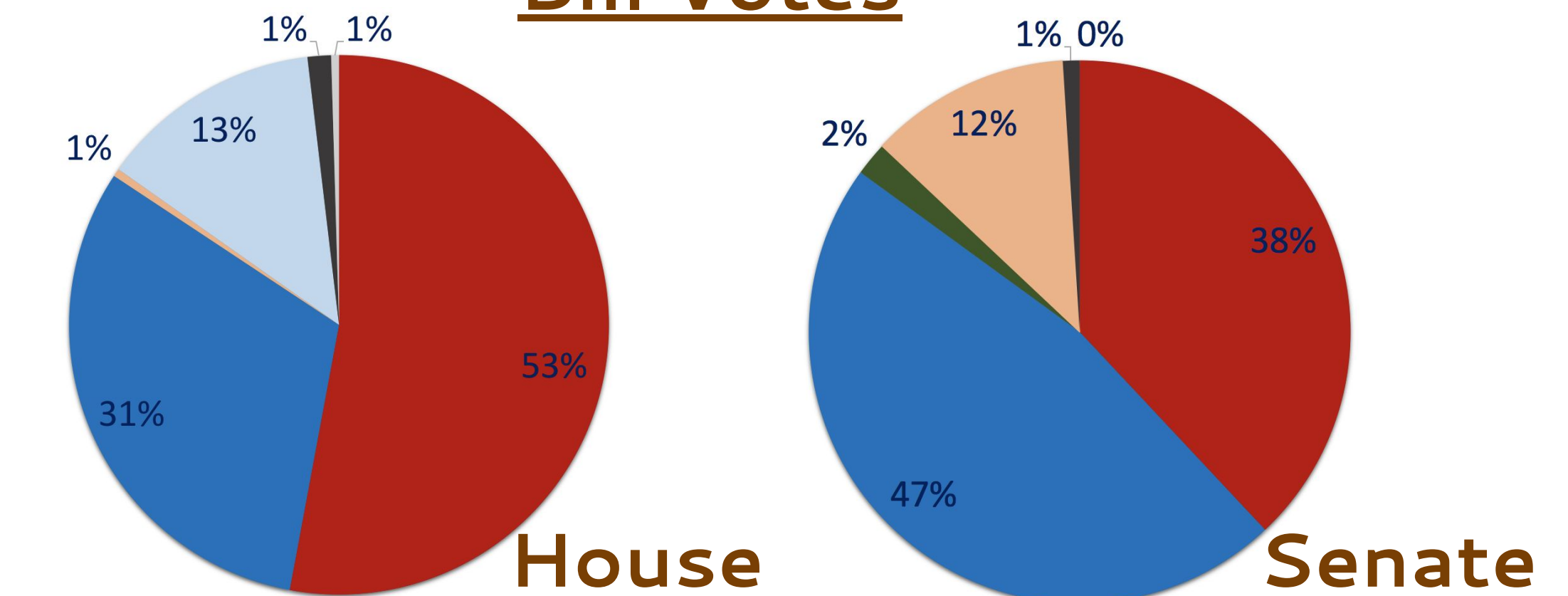
Conclusion

- At present, it is most ethically sound to restrict the implementation of machine learning algorithms to the initial legal evidence gathering phase, which involves more direct interaction with the lawyer, increasing efficiency while minimizing unfairness.

- Current algorithms are not sophisticated enough to accurately perform their intentions of reducing bias in judges' final verdicts; hence, their influence of final verdicts should be highly monitored.

References

[1] Northpointe, Inc. (2015). Practitioner's Guide to COMPAS Core. Retrieved April 25, 2019.

[2] Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, 4(1). Retrieved April 25, 2019.

[3] Amaker, Tamiko (2018, October). Criminal Court of the City of New York: Annual Report 2017. Retrieved April 25, 2019.

[4] Smith, Aaron (2018, November 16). Public Attitudes Towards Computer Algorithms. Retrieved April 25, 2019.