

Abandoning Criminal Risk & Recidivism

On Dangerous Goals in ML Scoring-Decision Systems

Benjamin Laufer
NeurIPS 2020 Resistance AI

Formulations of Risk

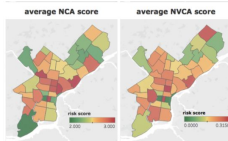
(1) Risk as Prediction

- o Earnest Burgess (1928)
- o Tools explicitly using race to predict recidivism '20s-'70s



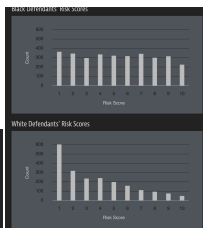
(2) Risk as a Condition to be Treated

- o Barabas et. al., "Interventions over Predictions" (2017)
- o Harcourt, "Against Prediction" (2007)
 - Criminal "Elasticity" to police stops
- o 1980s: "Risk, Needs, Responsivity"



(3) Risk as a Benchmark for Establishing Fairness

- o Northpointe-ProPublica debate
- o Incompatibility studies (Kleinberg, Corbett-Davies)



	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

Critique

- All three formulations (1)(2)(3) adhere to the **risk principle**: that risk should be measured to inform criminal interventions.
 - o Risk is rooted in the **theory of incapacitation**. Greenwood & Abrahamse (1982).
- Faux Exogeneity: Risk is not as scientific as it seems.
 - o Risk and recidivism are assessments made by judges, *related to subsequent decisions made by judges*. Risk is not innate to an individual.
- Perfect prediction (which satisfies fairness defs) can be racist & unjust.
 - o Is it enough that an algorithm is "fair"?



Alternative

Risk is rooted in theories of human delinquency, punitive logic and incapacitation. The goal of reducing **risk** or **recidivism** can justify massive increases in policing or sentences. Instead the goal should be reducing **harm**, which includes violence carried out by the state.

Faulty Goals in Other Domains

- Lending & credit: capital accumulation
- Hiring: capital accumulation; ability
- Education & test scores: ability

Pitfalls of Faulty Goals

- Formalize existing biases
- Legitimize decision-making as data-driven or scientific
- Justify more infrastructure in the name of reducing risk ('war on risk')

Making NeurIPS "Impact" Direct

- It is unclear if CS research helps people.
- Idea: use NeurIPS "Impact" section to center and direct resources to communities that need it, i.e. ongoing direct action and mutual aid efforts.