

A Validity Perspective on Evaluating the Justified Use of Data-driven Decision-making Algorithms

Amanda Coston, Anna Kawakami, Haiyi Zhu, Ken Holstein, & Hoda Heidari

Problem: Decision-making algorithms often prove unsuitable in real-world.



Child welfare



Healthcare



Welfare fraud



Criminal justice

➤ Often due to problem formulation & latent construct operationalization

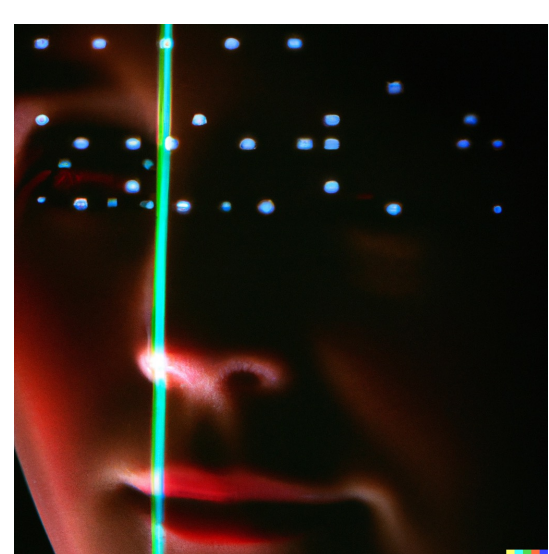
Our solution: Center *validity* in deliberation & evaluation of decision-making algorithms

A measure is **valid** if it accurately reflects the concept we intend to measure [1].

An algorithm is **valid** if it predicts the quantity that we think it does.

What threatens the validity of algorithms?

1. Attribute misalignment



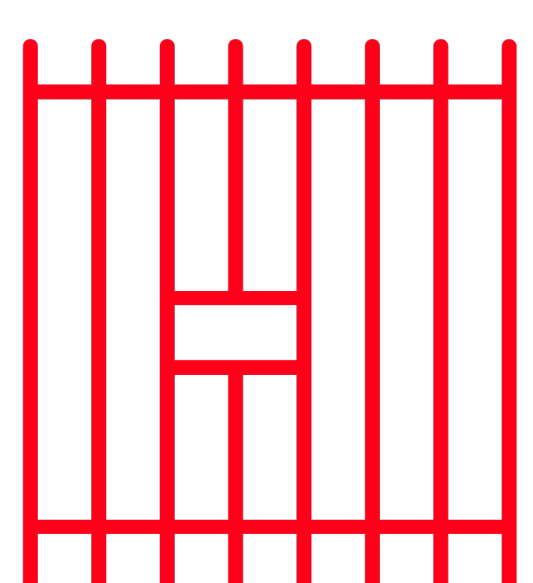
"Predicting" criminality from faces **fails** content & face validity.

2. Target misalignment



Predicting arrests as a proxy for crime can **fail** construct validity.

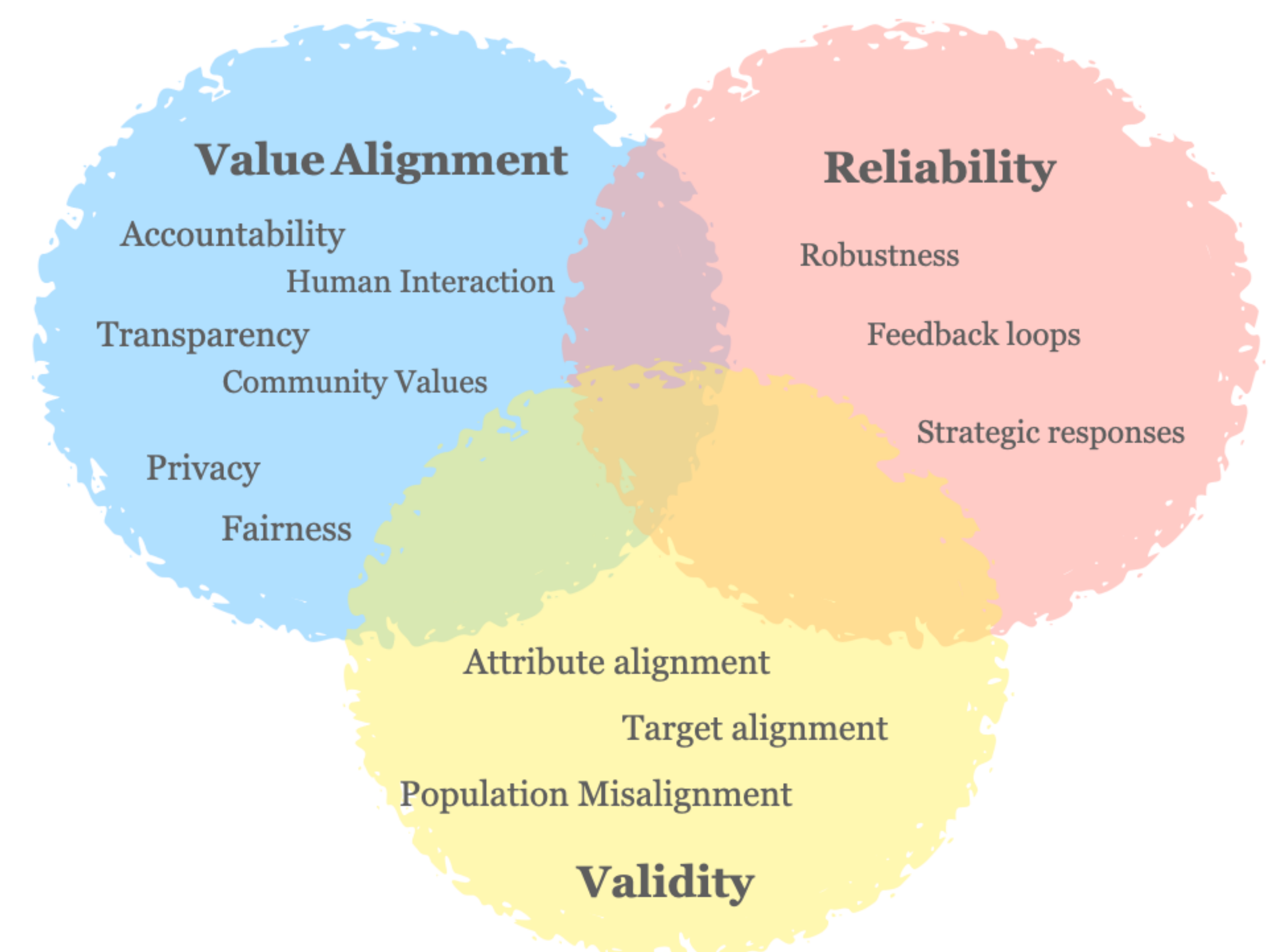
3. Population misalignment



Missing arrest outcomes for defendants who were not released can **fail** external validity

Our contributions:

1. Connect validity theory to problem formulation & data issues that jeopardize validity of algorithms
2. Propose a working taxonomy of criteria for the justified use of algorithms



3. Structure a protocol to promote deliberation on validity

Reliability is the extent to which the output of a measurement is repeatable, consistent, and stable [1]

Value-alignment requires that the goals & behavior of the system comply with values of stakeholders & communities [2]

See paper for our lit review on validity, reliability, and value-alignment.

References

[1] Ellen A Drost. Validity and reliability in social science research. Education Research and perspectives, 38(1):105–123, 2011

[2] Carles Sierra, Nardine Osman, Pablo Noriega, Jordi Sabater-Mir, and Antoni Perelló. Value alignment: a formal approach. arXiv preprint arXiv:2110.09240, 2021

