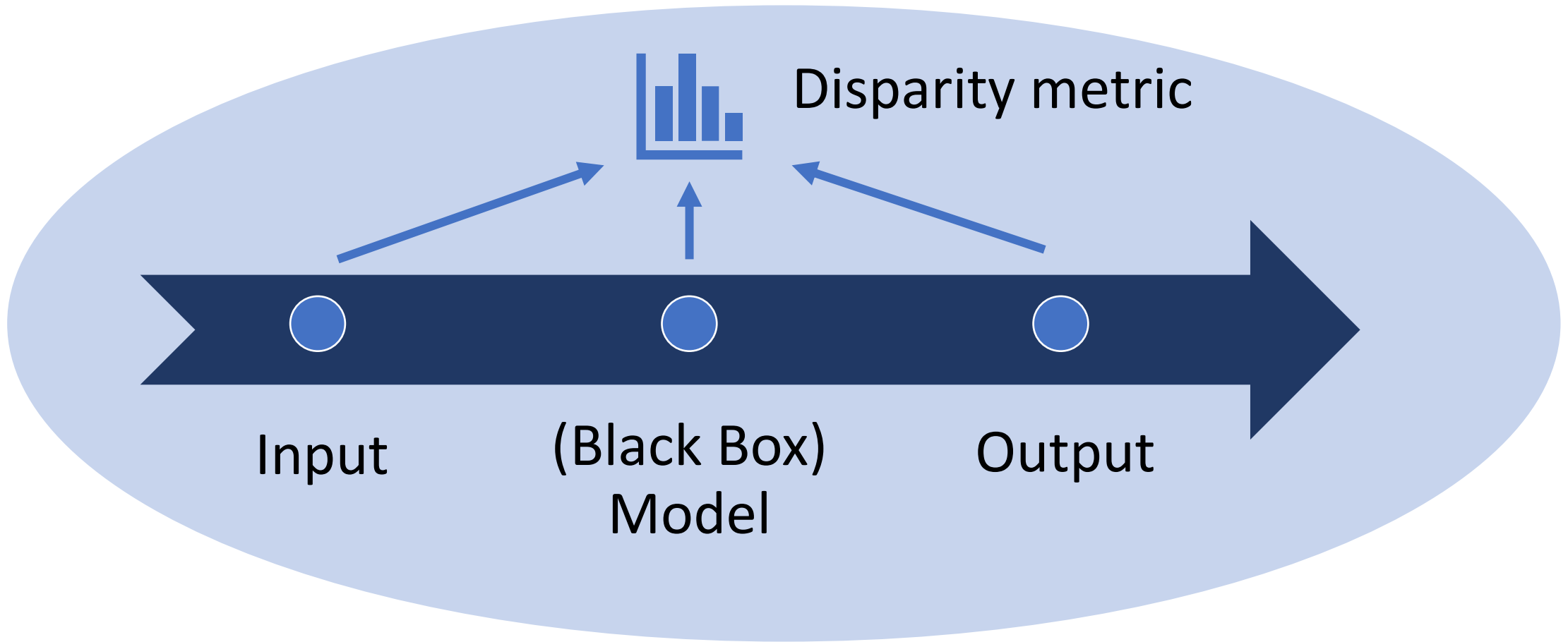


Fairness and Abstraction in Sociotechnical Systems

Andrew D. Selbst, Data & Society Research Institute
danah boyd, Microsoft Research and Data & Society Research Institute
Sorelle A. Friedler, Haverford College
Suresh Venkatasubramanian, University of Utah
Janet Vertesi, Princeton University

ACM Conference on Fairness, Accountability, and Transparency (FAT)*

Abstraction in Fairness-aware Machine Learning



Societal context abstracted away by considering input/model/output

The Abstraction Traps

Framing Trap

Portability Trap

Formalism Trap

Ripple Effect Trap

Solutionism Trap

A sociotechnical perspective

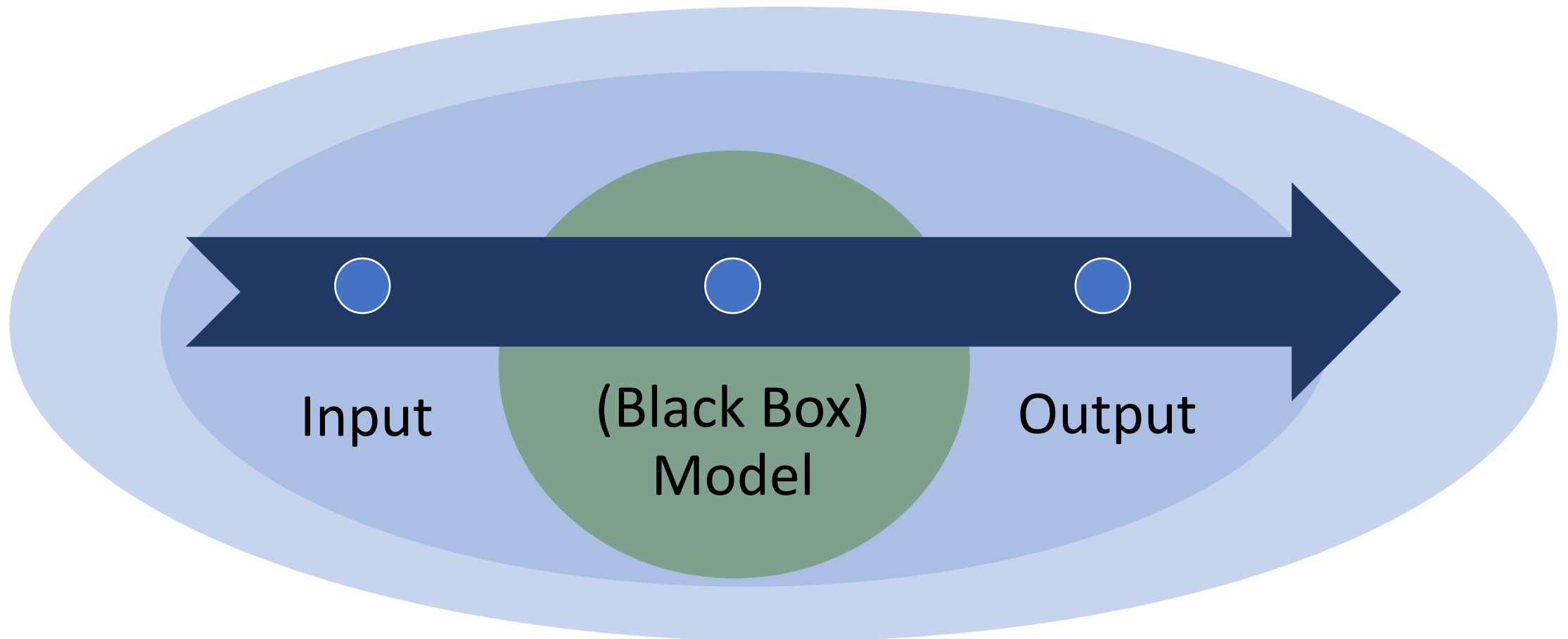
- Consider social & technical
- Sociotechnical systems lens →



- Bruno Latour. 1987. Science in action: How to follow scientists and engineers through society. Harvard University Press.
- Trevor J. Pinch and Wiebe E. Bijker. 1984. The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. Social Studies of Science 14, 3 (1984), 399–441.

Framing Trap

- Failure to model the entire system over which a social criterion, such as fairness, will be enforced



Framing Trap: Pretrial risk assessment



Judge has to decide whether to release (with or without bail) or detain defendant



Automated risk assessment is meant to help judge as a recommendation

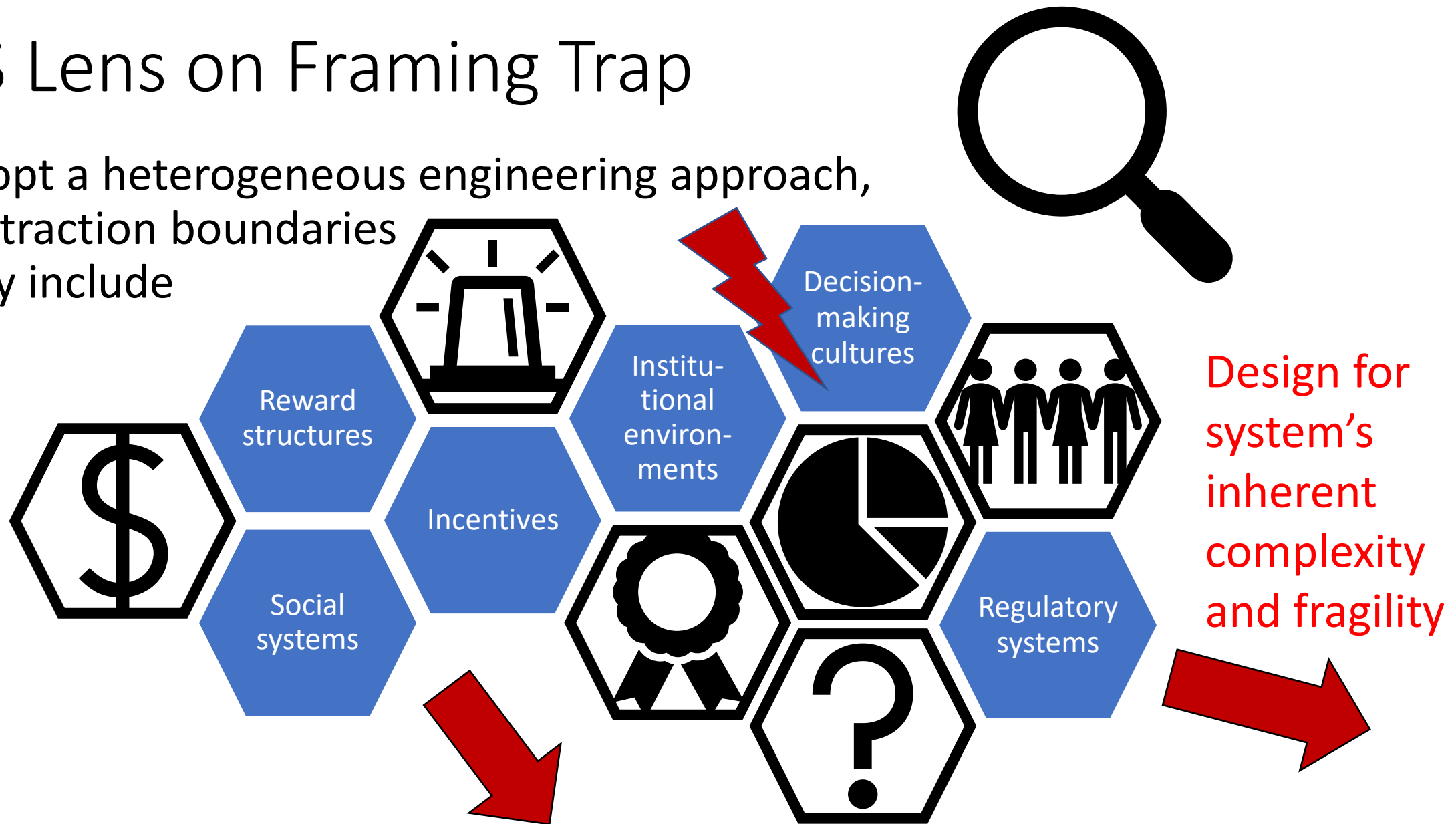


Sociotechnical frame: model judges' automation bias, deviations from risk scores is the only way to get end-to-end guarantees

- Angèle Christin. 2017. Algorithms in practice: Comparing web journalism and criminal justice. *Big Data & Society* 4, 2 (2017).
- Danielle Keats Citron. 2008. Technological due process. *Washington University Law Review* 85 (2008), 1249–1313.
- Linda J. Skitka, Kathleen L. Mosier, Mark Burdick, and Bonnie Rosenblatt. 2000. Automation bias and errors: Are crews better than individuals? *The International Journal of Aviation Psychology* 10, 1 (2000), 85–97.
- Megan T. Stevenson. 2018. Assessing risk assessment in action. *Minnesota Law Review* 103 (2018).

STS Lens on Framing Trap

- Adopt a heterogeneous engineering approach, abstraction boundaries may include



Portability Trap

- Failure to understand how repurposing algorithmic solutions designed for one social context may be misleading, inaccurate, or otherwise do harm when applied to a different context
- Portability is a extremely valuable in software engineering

Portability Trap: Pretrial risk assessment



Assume we avoid the Framing Trap by including all relevant aspects of the context into our model
... then this is now a very specialized solution that is not portable to other domains



BUT TRANSFER LEARNING!

... “not sufficiently expressive to capture the vast changes in social context”



Even within the pretrial risk assessment domain the characteristics of the next courthouse may be very different

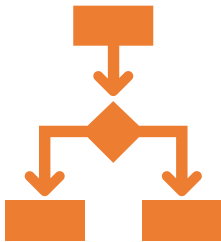
STS Lens on Portability Trap



- Contextualizing user “scripts” that describe how technology is used in a specific context since fairness is tied to that social context e.g. model cards, datasheets, nutrition labels

Formalism Trap

Failure to account for the full meaning of social concepts such as fairness, which can be procedural, contextual, and contestable, and cannot be resolved through mathematical formalisms.



Procedural

e.g. firing based on gender is illegal, but firing is otherwise legal; mathematical definition only captures whether somebody was fired



Contextual

cultural context determines whether discrimination is morally wrong, e.g. a men's clothes manufacturer rejects female model applicant



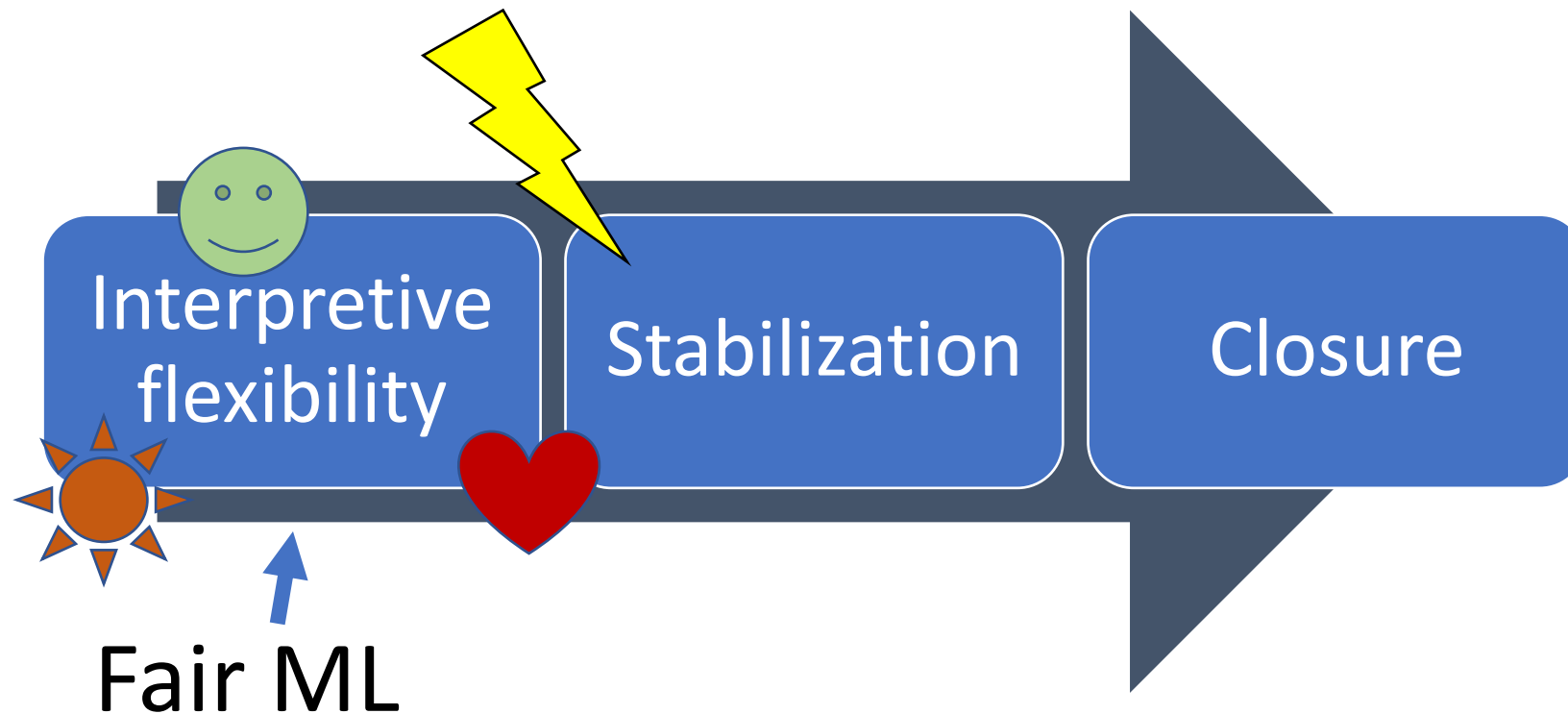
Contestable

legal definitions and societal understanding of fairness & discrimination change over time, and with legislation or court cases

STS Lens on Formalism Trap



Social Construction of Technology (SCOT)
by Pinch, Bijker describes steps towards adoption of technology



Ripple Effect Trap

Failure to understand how the insertion of technology into an existing social system changes the behaviors and embedded values of the pre-existing system



<https://pxhere.com/en/photo/1447321> (CCO)

STS Lens on Ripple Effect Trap



Reinforcement Politics: Introducing a new technology can affect power dynamic between existing groups

Reactivity behaviors can alter the social context that the design planned for and destabilize existing values, incentives and structures

→ Avoid reinforcement politics and reactivity, study extensive literature on ripple effects

- Rob Kling. 1991. Computerization and social transformations. *Science, Technology, & Human Values* 16, 3 (1991), 342–367.
- Wendy Nelson Espeland and Michael Sauder. 2007. Rankings and reactivity: How public measures recreate social worlds. *Amer. J. Sociology* 113, 1 (2007), 1–40.

Solutionism Trap

- Failure to recognize the possibility that the best solution to a problem may not involve technology

TECH

Musk proposes rescuing boys trapped in Thai cave with a 'submarine' made from SpaceX rocket part

PUBLISHED MON, JUL 9 2018•7:01 AM EDT | UPDATED MON, JUL 9 2018•8:39 AM EDT



Ryan Browne
@RYAN_BROWNE_

SHARE



<https://www.cnn.com/2018/07/09/elon-musk-sends-kid-size-submarine-to-help-thai-cave-rescue-mission.html>

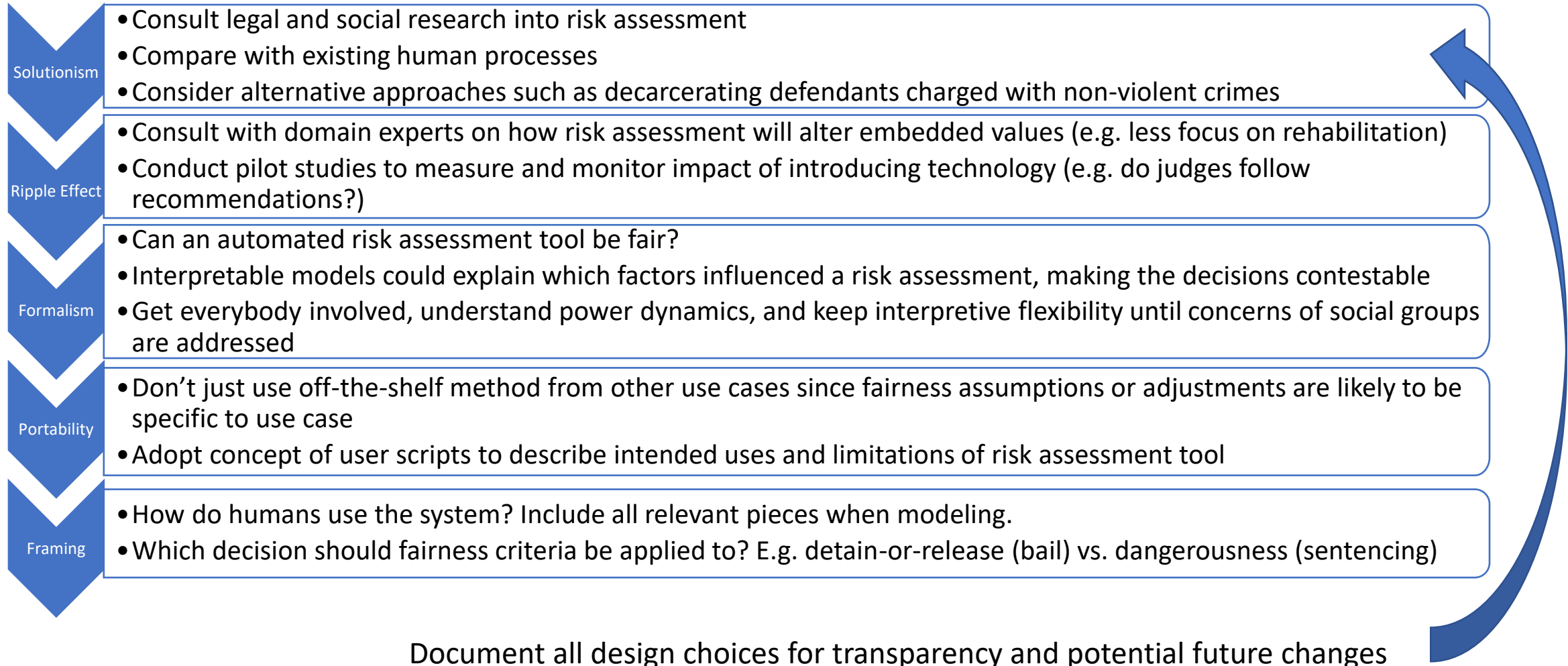
STS Lens on Solutionism Trap



- When not to build solution with technology:
 - Shifting definitions of fairness (see Formalism Trap)
 - Problem is too complex, computationally intractable
- Hippocratic oath: “first, do no harm”
- Takeaway from the field of human-computer interaction: sometimes it is better not to design a solution

An example: Risk Assessment

Goal: reduce pre-trial detention



Sources

- Paper:
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3265913
- Thanks to danah boyd and Suresh Venkatasubramanian for sharing their slides which served as an inspiration for this deck