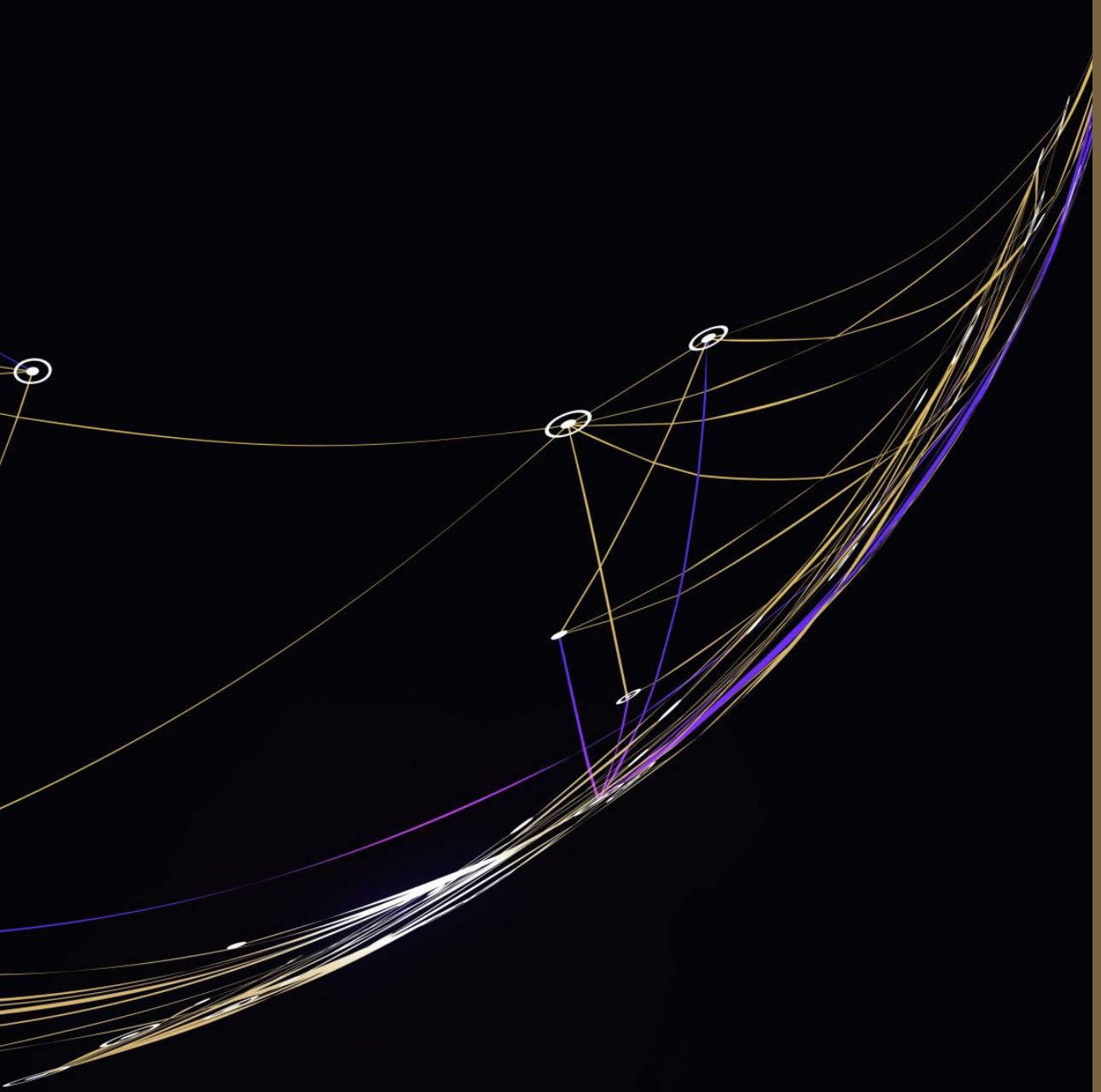


ACCOUNTABLE ALGORITHMS AND CRIMINAL JUSTICE

Rebecca Williams, Oxford

2 ASPECTS OF ACCOUNTABILITY

Procedural

Substantive

Accountability for Procedural issues

Bates v Post Office

- Claimants all former or serving sub-post masters (SPMs) all of whom had contracted with the Post Office to run branch Post Offices across the country.
- Post Office introduced a computer system called Horizon in 2000 across all branches, changed to online version in 2010
- Both systems were or are unreliable, leading to unexplained shortfalls and discrepancies in their bank accounts.
- The Post Office denied this and prosecuted (successfully) several SPMs on the basis that discrepancies or shortfalls were the result of theft. Marriages broke down, the stress is alleged to have led to health conditions, addiction and premature deaths.
- *Bates v Post Office* 2019 EWHC 3408
- In December 2019 the Post Office agreed to settle with 555 claimants, paying £58m in damages, claimants receiving £12m after legal fees.

How could this happen? 2 reasons.
Problem 1 (legal):

Computer Records Evidence

Legal Guidance

Reviewed 30 june 2017

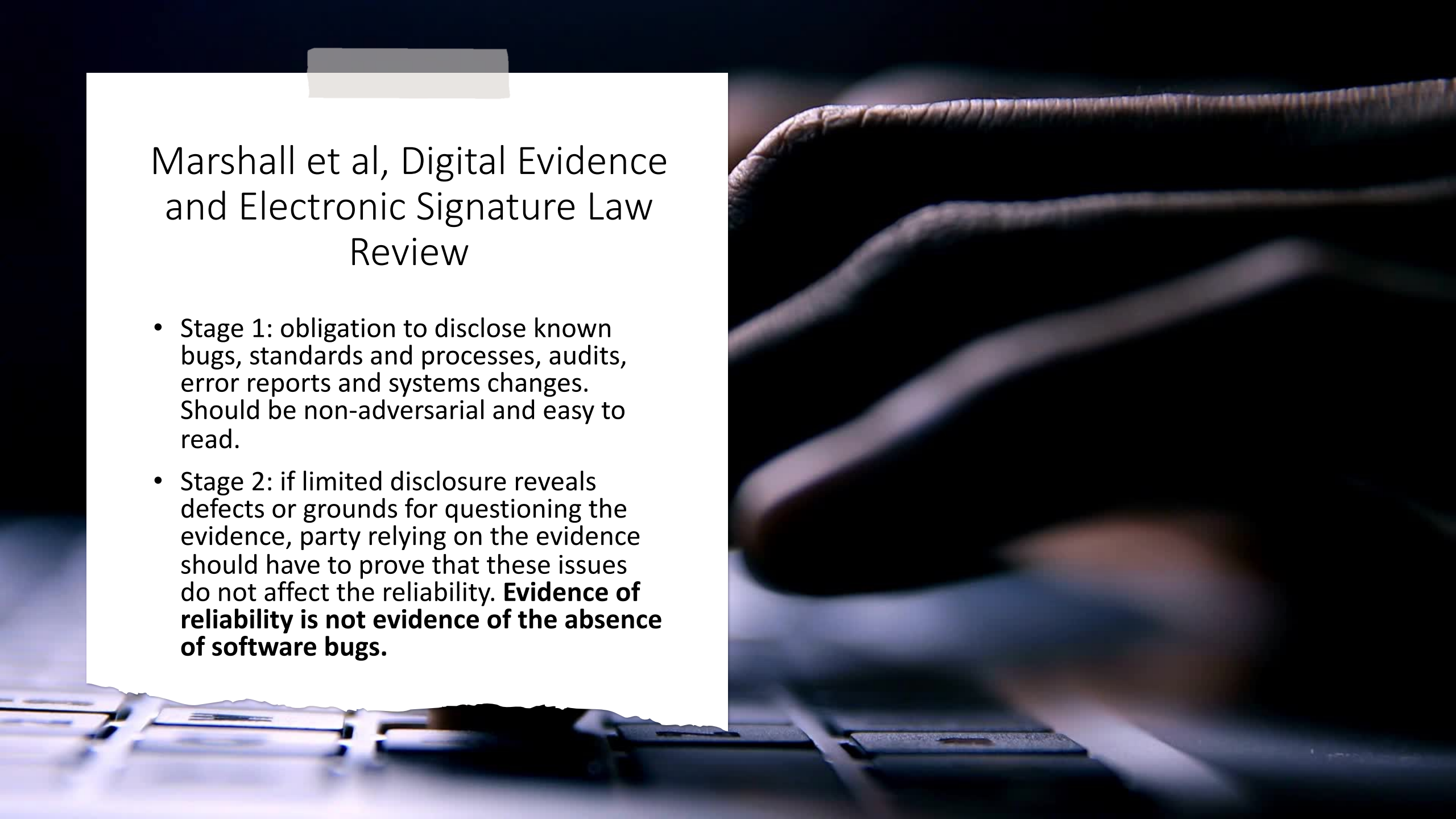
Principle

Rebuttable Presumption of Correct Functioning of a Computer

Prior to the repeal of section 69 of the Police and Criminal Evidence Act 1984 by section 60 of the Youth Justice and Criminal Evidence Act 1999, it was necessary to prove that a computer was operating properly and was not used improperly before any statement in a document produced by the computer could be admitted in evidence.

Computer evidence must now follow the Common Law rule, that a presumption will exist that the computer producing the evidential record was working properly at the material time and that the record is therefore admissible as real evidence.

That presumption can, however, be rebutted if evidence to the contrary is adduced. In that event it will be for the party seeking to produce the computer record in evidence to satisfy the court that the computer was working properly at the material time. For detailed guidance as to the law, see <Archbold 9-11 9-15>.



Marshall et al, Digital Evidence and Electronic Signature Law Review

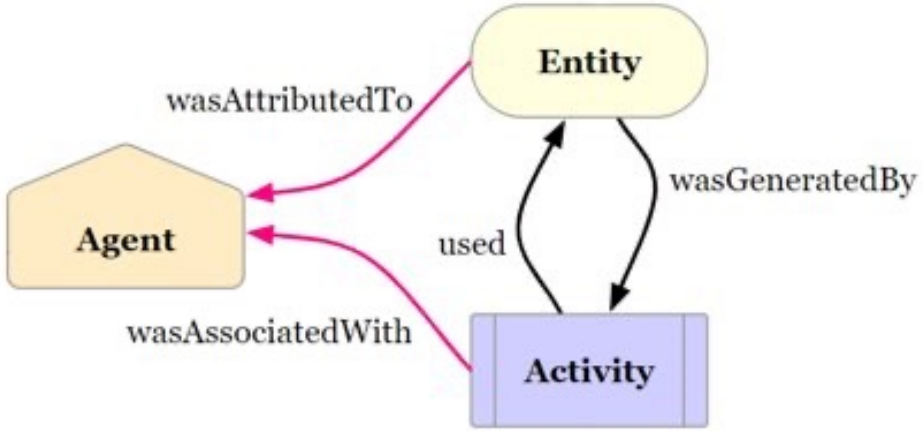
- Stage 1: obligation to disclose known bugs, standards and processes, audits, error reports and systems changes. Should be non-adversarial and easy to read.
- Stage 2: if limited disclosure reveals defects or grounds for questioning the evidence, party relying on the evidence should have to prove that these issues do not affect the reliability. **Evidence of reliability is not evidence of the absence of software bugs.**

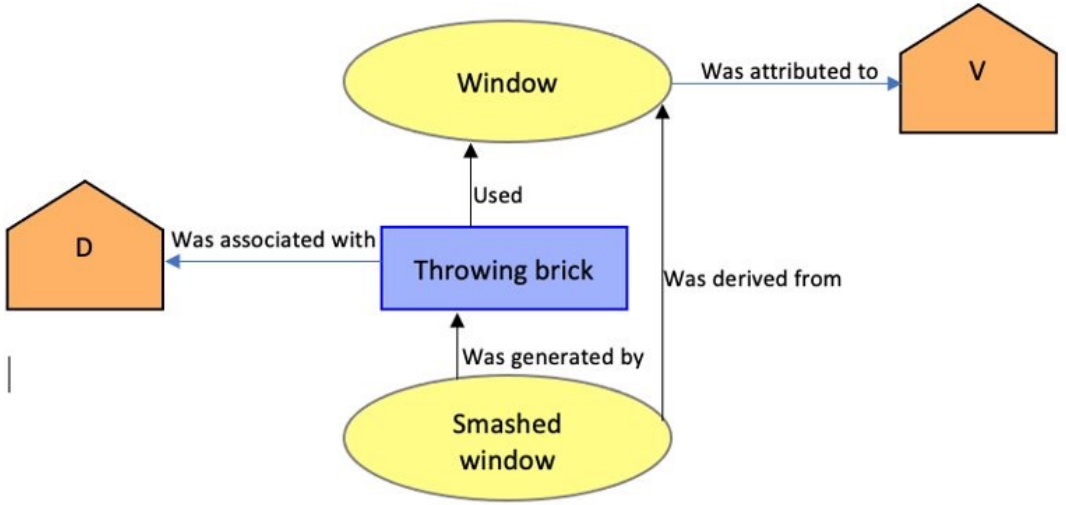
But problem 2 (technical):

- As part of Stage 1, ‘Relevant documents should be routinely kept and easily available’ in a proportionate way.
- Not just in the UK, also, e.g. US Blueprint, AI Act (recital 46 , Art 17(1)(k)).
- This requires a technical solution.

RAInS

- Developed an ‘accountability’ fabric using computational models of provenance for use across the whole lifecycle of an intelligent system
- https://rains-uoa.github.io/ISWC_2021_Demo/

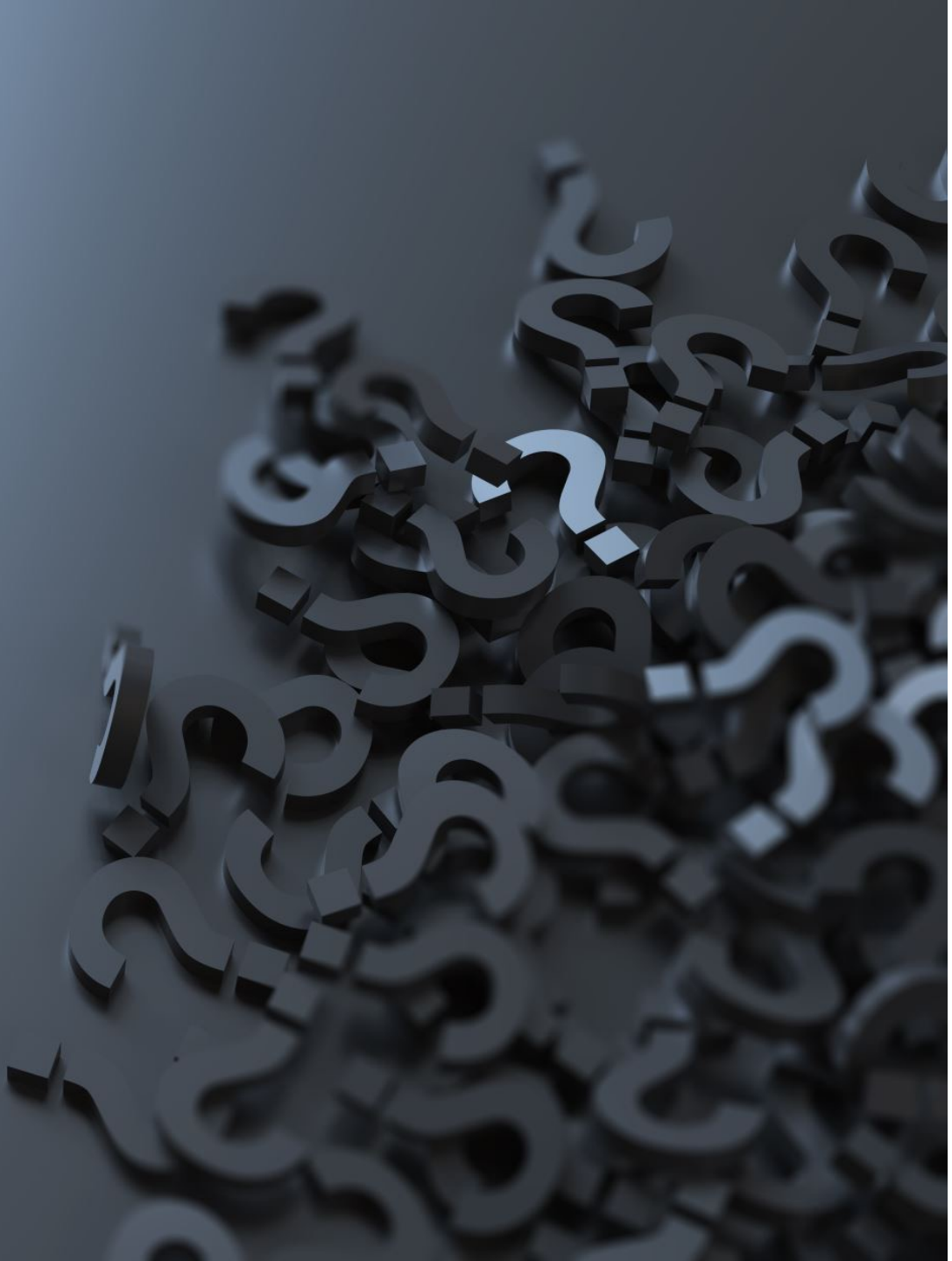




Note the need for interdisciplinarity:

- Fraser J in *Bates v Post Office* [2019] EWHC 3408 (QB) [72] ‘this is not a case that is being tried in a Specialist List, such as the Technology and Construction Court – it is a general Queen’s Bench Division case – but it could readily have been tried in such a list. It contains a great deal of technical subject matter… Such subject matter and such expert evidence is readily suited to analysis by the parties and precision, which is the usual approach of courts generally.’
- Included a ‘Technical Appendix’ with the technical details.
- Marshall et al: Marshall (barrister); Christie (independent testing consultant), Ladkin (Professor of CS,), Littlewood (Emeritus Prof of software engineering), Mason (barrister); Newby (Professor of statistical science); Rogers (legal academic), Thimbleby (Digital Health Fellow); Thomas, Visiting Professor of Software Engineering).
- RAInS as an interdisciplinary process

Substantive Accountability



The problem of metrics

- We know that systems can ‘fail’; they can be insecure, leak data, or contain problems with the software.
- But what the various different metrics show us is that they can also succeed differently.
- Art 15 of the proposed AI Act: ‘an appropriate level of accuracy’ – but which one? Accountability according to what metric?
- This was the focus of the COMPAS Northpointe v ProPublica debate, as well as (to some extent) *Bridges*
- As Berk et al put it,

‘there are complicated tradeoffs between different kinds of fairness and between different kinds of fairness and different kinds of accuracy. You can’t have it all. Computer scientists and statisticians will over time provide far greater clarity about these tradeoffs, but they cannot be (and should not be) asked to actually make those tradeoffs. The tradeoffs must be made by stakeholders through legal and political processes. This will be very challenging’

The law can help

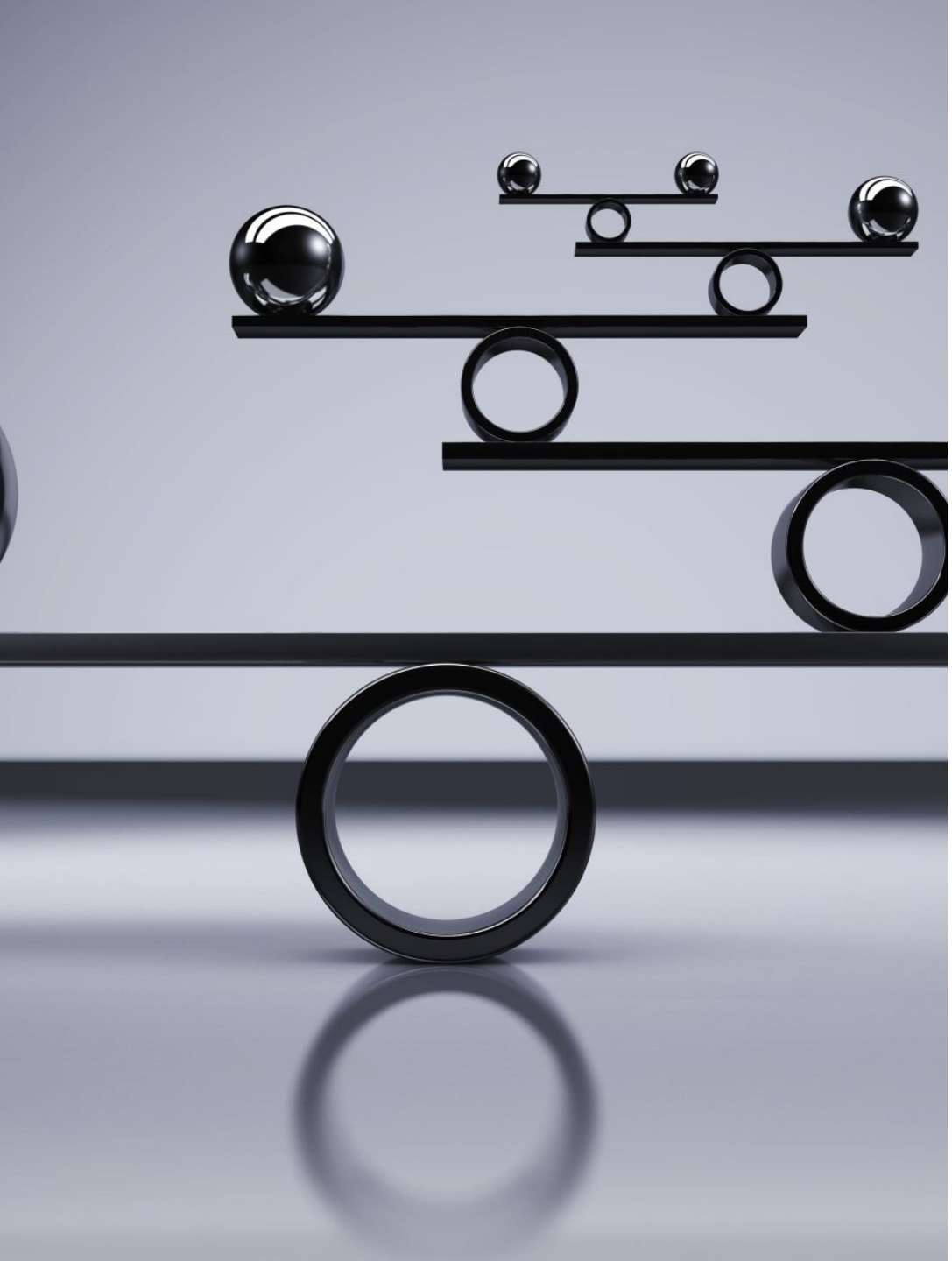
- We do have the tools to ensure that we’re holding systems accountable by the right metric.
- But we need to develop those tools into a full theory of metric choice.

A theory of metric choice

- The beginnings of a theory; contrast Blackstone and Lord Hoffmann in *AF* [74].

‘it is better that 10 guilty persons escape than that one innocent suffer’

“It is sometimes said that it is better for ten guilty men to be acquitted than for one innocent man to be convicted. Sometimes it is a hundred guilty men. The figures matter. A system of justice which allowed a thousand guilty men to go free for fear of convicting one innocent man might not adequately protect the public.”



2 sources of help:

- Both from public law (judicial review), but could be applied outside.

1. Proportionality's 3 questions:
    - Suitability (a connection between means and ends)
    - Necessity (using a sledgehammer to crack a nut?)
    - Fair balance between means and ends.

2. A duty to take certain things into account.

Suitability

- Is the metric chosen to assess the performance of the system suitable to ensure that the decision-making system achieves its overall aim?
- Detecting threats:
  - Sensitivity (how many of the target cases were detected?)
  - Negative Predictive Value (how many negative results were really negative?)
  - False omission/rejection rate (how many false negatives?)
- Pre-screening/triaging
  - Specificity (how many of the rejected samples were correctly rejected?)
  - Negative predictive value
  - False omission/rejection rate.

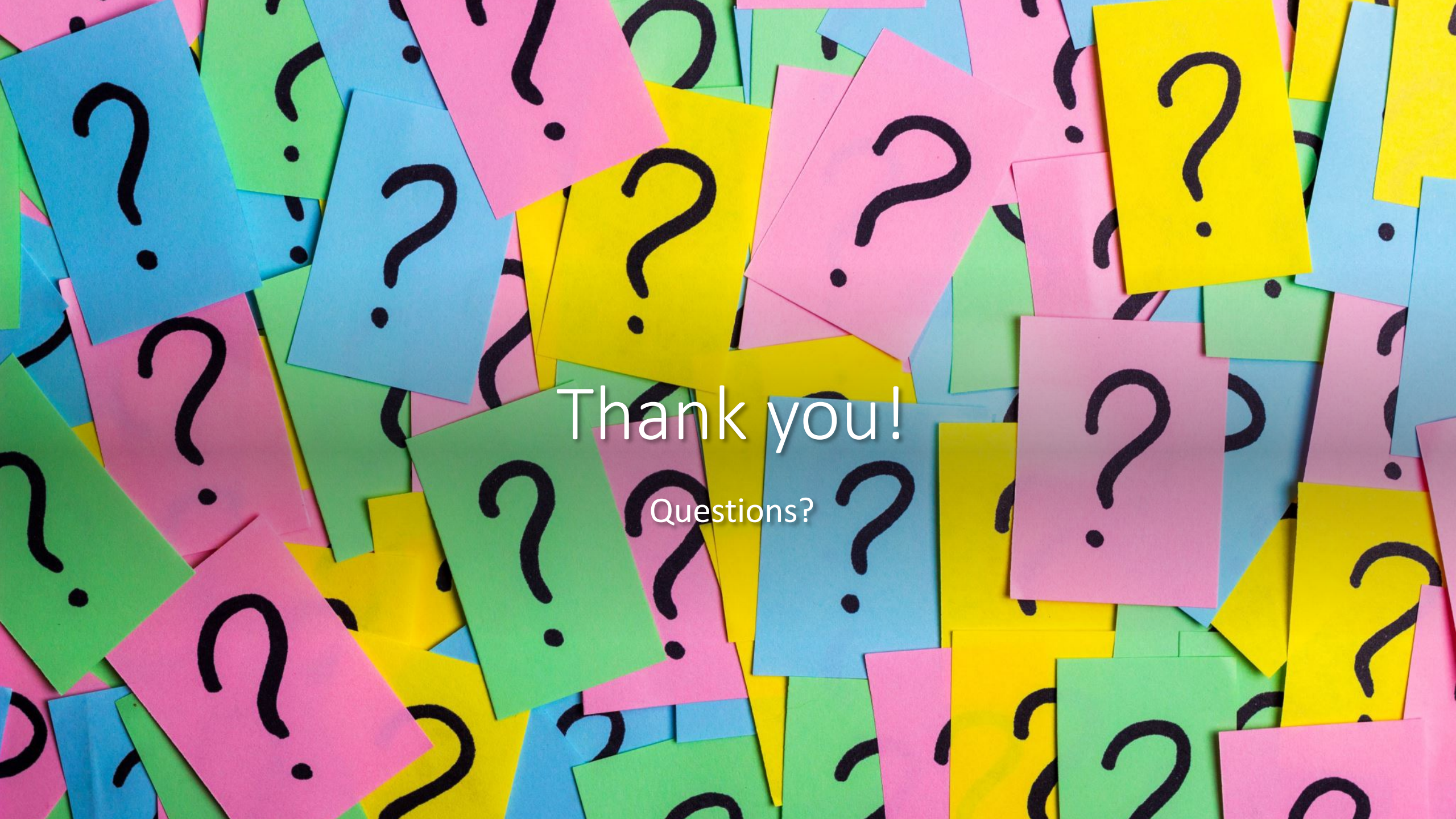
Fair balance

- What are the costs of failure? If a system fails to perform well against one metric, does that disadvantage outweigh the advantages of the metrics against which it performs well?
- Need to weigh the costs of false positives and failure to reject (specificity) against the benefits of picking up cases (sensitivity) and avoiding false negatives. (NB a perfectly sensitive system would just pick up everyone!)

A duty to take into account ground truth and the necessity doctrine

- Where there is a ground truth, even if there is not an inbuilt incentive to check for it, the law should impose one. (For public lawyers this is a more interventionist version of the *Tesco* doctrine).
- The necessity doctrine can then back this up; is it necessary for the decision-maker to act on the prediction before ground truth is established? E.g. (Rodolfa et al) social service interventions flagged but only triggered if those identified do indeed end up back in court. So necessity might require ground truth-matched intervention.

- Sometimes the situation will be more complicated
  - No ground truth
  - Heterogeneous group so that parity is also a concern (COMPAS). (See Wachter et al on metric choice in this context).
- But even here suitability and fair balance can be channels through which we can consider metric choice.
- Specific legislation may help too, but these are common law concepts which the courts can use even in the absence of legislative intervention.
- But again, not possible without complete interdisciplinarity.


Thank you!
Questions?