Daubert meets Dependability

Roy Maxion

Dependable Systems Laboratory
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213
Email: maxion@cs.cmu.edu

22 January 2022

IFIP WG 10.4
Virtual Workshop

Today's goals

- A different way of thinking
- Introduce the Daubert Supreme Court case (1993)
  - A ruling on standards of evidence
- Show its relevance to dependable systems
- Provide some examples
- Not to persuade you to become competent in the law; rather, to become competent in designing systems that will satisfy the law
  - Particularly with regard to methodologically correct and admissible evidence
- Note: this talk isn't meant to be comprehensive

Daubert's origins

- **1993 US Supreme Court case**:
  - Daubert v. Merrell-Dow Pharmaceuticals, Inc., 509 U.S. 579, 584-587 (1993)
- **The case**:
  - The parents of two children with serious birth defects sued Merrell Dow Pharmaceuticals, alleging that the birth defects developed because of an anti-nausea medication marketed by the company.
  - Expert witnesses testified on behalf of both plaintiff and defendant.
- **The issue**:
  - What is the standard for admitting expert scientific testimony in federal court?
- **The outcome**:
  - Overturned the 1923 *Frye* test; set new guidelines for what kind of scientific evidence would be admissible in court.
- **Why it mattered**:
  - The ruling established the "*Daubert* standard" for admissible scientific evidence.
  - This standard governs how judges evaluate scientific testimony given by expert witnesses, and requires scientifically valid reasoning that applies to the facts of the case at hand. Judges must weigh the scientific validity of evidence in order to decide whether to allow juries to hear it.
  - Also affirmed: Rule 702 of the Federal Rules of Evidence prevails for expert-witness testimony; and, provided non-exclusive guidelines to help gatekeeper judges assess the reliability of expert testimony.

 4

Federal rule of evidence # 702

- A witness [who is] qualified as an expert by knowledge, skill, experience, training, or education may testify … in the form of an opinion or otherwise if:

  (1) The testimony is based on sufficient facts or data;

  (2) The testimony is the product of reliable principles and methods; and

  (3) The expert has reliably applied the principles and methods to the facts of the case.

Title 28, Appendix – Rules of Evidence, Rule 702, pp. 393

 5

Daubert factors

For trial courts to use in assessing the reliability of scientific expert testimony

(1) whether the expert's technique or theory can be or has been tested;

(2) whether the technique or theory has been subject to peer review and publication;

(3) the known or potential rate of error of the technique or theory when applied;

(4) the existence and maintenance of standards and controls;

(5) whether the technique or theory has been generally accepted in the scientific community.

Title 28, Appendix – Rules of Evidence, Rule 702, pp. 393

 6

Legal issues accrue when …

- … the output of an ML system is used as substantive evidence in criminal prosecution.
- … an ML decision exposes a user to liability.
- … an ML system acts as an expert whose conclusions serve as substantive evidence in court.
- … there is liability for the system having decided incorrectly.

Nutter: Machine Learning Evidence: Admissibility and Weight
Journal of Constitutional Law, Vol 21, No 3, 2019

 7

Who is the expert witness?

There are two experts …

(1) The decision-making system, presumed to render expert judgment.

(2) The programmer or other expert, presumed to be able to explain the decision-making process of (1).

Exmpls: AI/ML, mission-critical systems

- Autonomous vehicles
- Bank loans
- Authentication / identification
- Behavioral biometrics
- Forensics
- Questioned documents / handwriting
- Face recognition
- Bail eligibility
- Image/scene classification
- Recidivism risk
- Camera ID
- Pipelines
- Chemical and power plants
- Medical
- Etc.

Aspects of ML-based decision makers

- Training data
- Testing data
- Data quality (ill attended to)
- Number of samples
- Classification algorithm
- Distance/similarity metric

- Decisions on all of these (as well as other) factors must be methodologically correct.

Unbiased data sets - issues

- General factors
  - quality of the frame, accuracy, cost, bias,
- Sample frame
  - Population from which the sample is drawn
  - Examples: Survey outside the big-and-tall shop; 1936 presidential election
- Representativeness
- Batch effects (e.g., morning vs evening)
- Instrumentation
  - Adequate for the undertaking?
- Data labeling/cleaning
- Sampling technique
  - Probability sampling
    - Simple random, systematic, stratified, cluster, multi-stage, etc.
  - Nonprobability sampling
    - Convenience, snowball, quota, purposive, etc.
- Number of samples
  - Power analysis

- These are just some of the factors important for data sets
  - All of them can/should invite scrutiny

Disaster: 1936 US presidential election

- Literary Digest predicted the presidential election results (Landon vs Roosevelt), published winner in morning New York Times.
  - But the prediction was wrong.
- Sample frame
  - Phone books, magazine subscribers, etc.
  - Mailed out 10 million mock ballots
  - Received back 2.4 million
- Problems:
  - Selection bias – wrong sample frame
    - Skewed toward middle/upper-class voters
  - Voluntary / nonresponse bias
    - Non-responders are systematically different from responders
      - E.g., people who were out of work

The price of failure

- Huge embarrassment to polling organization
- Went out of business
- (Flummoxed the bookies)

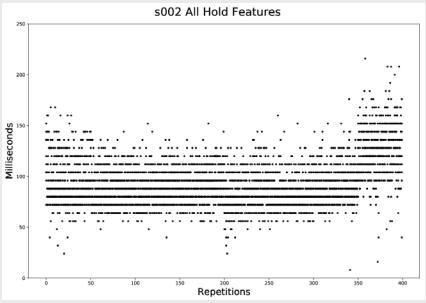
What is a Daubert challenge?

- A Daubert challenge occurs when opposing counsel disputes the validity of an expert's testimony/decision (human or machine) based on … the methodology used to arrive at the expert's opinion/decision.
- Price of failure?
  - Not positive; not a good thing
  - Wrong decisions, perhaps undetected

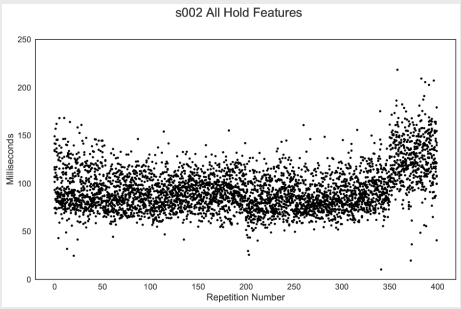
 14

Data-collection bias/artifacts

Corrupt



Correct



1% of the data being "corrupt" changed a classifier's decision by ~20 points, wrongly reversing a distinction between fraudulence and legitimacy.

 15

Daubert's relevance to dependable systems

- Systems should be built using vetted and dependable, not ad hoc, methodologies.

  Hint: Don't make stuff up.
- That way it's unlikely that you'll be subject to Daubert challenges, but if you are, you'll be well prepared.
- Helpful for acceptance and quality testing.
- Aids in reproducibility.

 16

Lesson

- Use the spectre of a Daubert challenge to get it right the first time.
- Build the system so that its methodology is correct, so that it (and you) can withstand a Daubert challenge.
- To do this, one needs to know what the rigorous methodologies are, and to be able to distinguish them from methods that will die on the stand (or peer review).
- Side note: many *published* papers use subpar methodologies; why?

 17

Conclusion

- Ignorance is a voluntary misfortune.
- You may forget a few details, but they won't forget you.