PC: https://xkcd.com/1838

Debugging Data & Models

Eric Wong

8/30/2022

Course website

https://www.cis.upenn.edu/~exwong/debugml/

Location

Tuesday - Rittenhouse Laboratory 4C6

Thursday - Chemistry Library 514

1:45 – 3:15 PM

What is in this course?

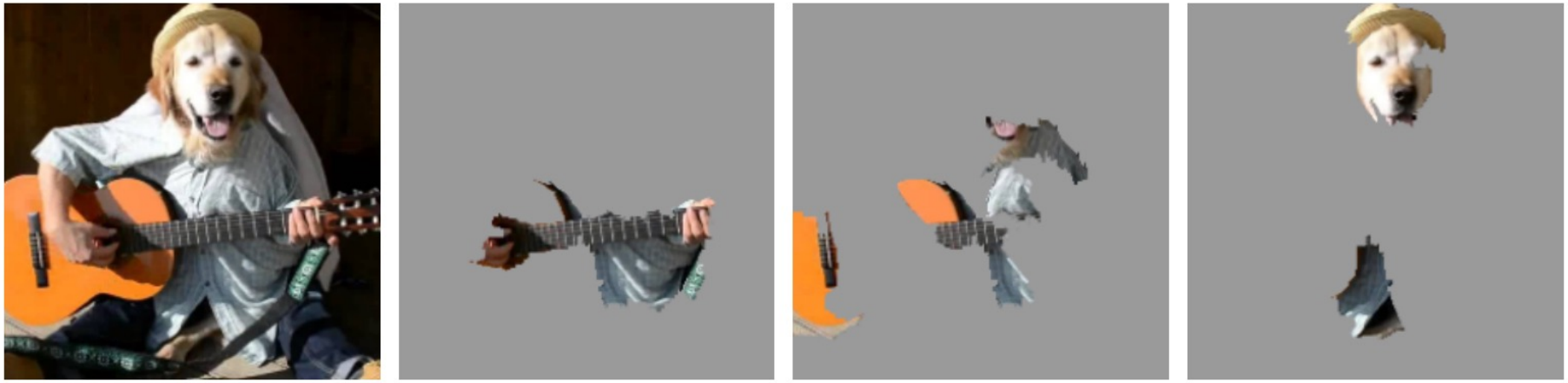
Research topics in debugging ML

- Failure modes
- Debugging tools
- ML repair

Failure modes

Biases, distribution shifts, adversarial changes

Debugging tools

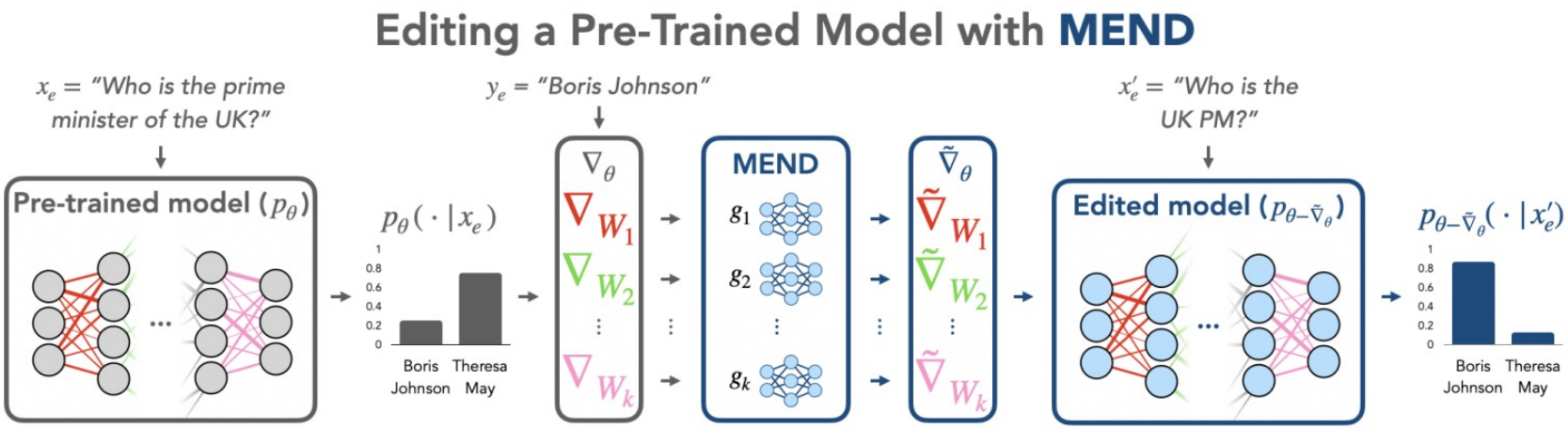


(a) Original Image (b) Explaining *Electric guitar* (c) Explaining *Acoustic guitar* (d) Explaining *Labrador*

Explainability, verification, scientific discovery

Ribeiro et al 2016. "Why should i trust you?" Explaining the predictions of any classifier.

ML repair



Robust training, data interventions, model adjustments

PC: Mitchell et al 2022. Fast Model Editing at Scale.

We will cover:

- Known problems with data/models
- Assumptions and limitations of solutions
- Mathematics, statistics, optimization
- Primarily vision & language

Prerequisites

- ML intro course (i.e. CIS 519/520)
- Linear algebra (i.e. Gilbert Strang's 18.06)
- Statistics (i.e. STAT 512)
- Bonus: optimization/deep learning (?)

We **will not** cover:

- How to fix your model
- Practical, general purpose debugging tools

But: could be a course project!

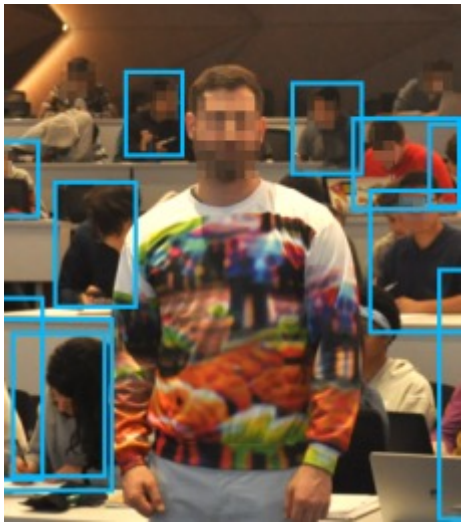
Goals (PhD centric)

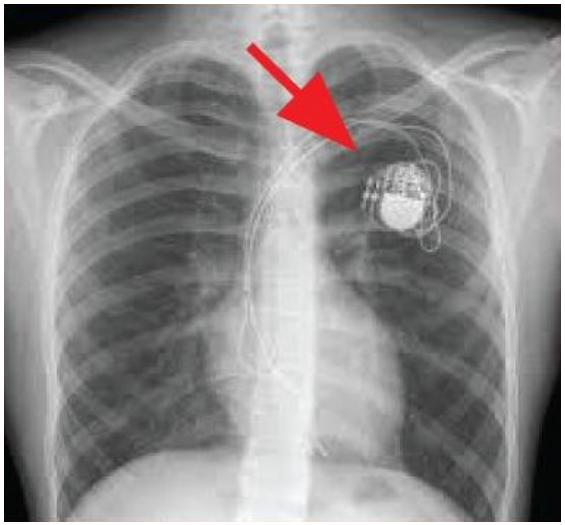
1. Primer on research topics in debugging ML

2. Research experience

3. Communication

1. Debugging ML





Overview of open problems & critical discussion of papers

Lectures

- Not taking attendance
- No recorded lectures
- Masks required (may revisit later)
- Laptops in the back

2. Research experience

Gain experience by doing

Project

- Tackle a research problem in trustworthy ML

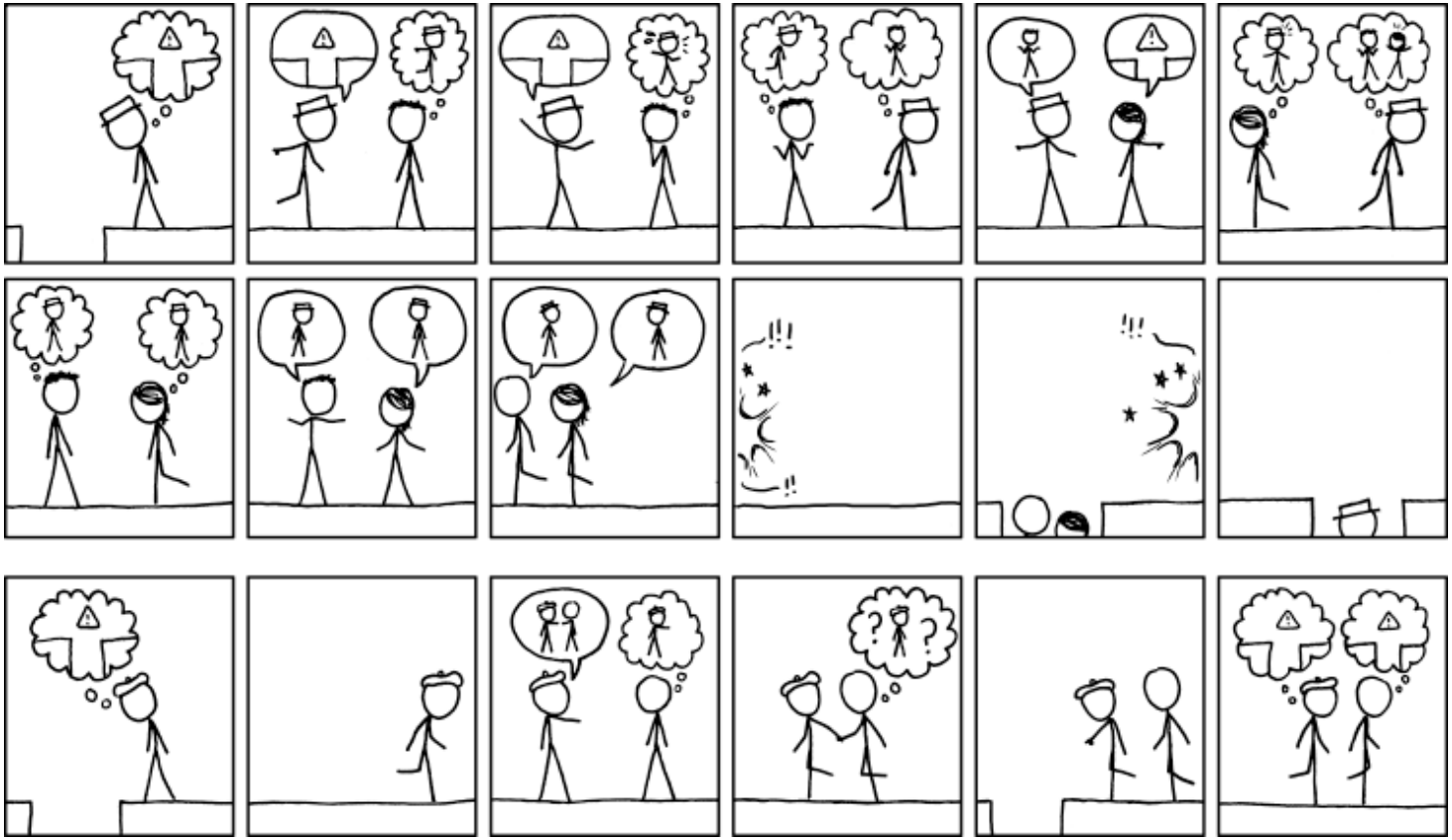
Examples:

- Debug a problem in an ML application (medical, genetic, etc.)
- Analyze or create debugging tools
- Audit a ML pipeline

Project checkpoints

- Proposal (Oct 4th)
- Mid-report (Nov 1st)
- Final report (Dec 15th)
- Groups of 1-3

3. Communication



Practice explaining and conveying ideas

Presentations

- 2 paper talks
- Project checkpoint talk
- Final project talk
- Peer evaluation

Readings

- Weekly readings (you can suggest)
- Post a thought, question, or observation in Ed Discussion before class (skim)
- Sign up to present two papers over semester (in-depth)

Project presentations

- Checkpoint presentation (late October-early November)
- Final presentation (Dec 6/8)
- Sign up for a time slot

Class structure

Approximately:

- 30-60m of “lecture”
- 30-60m of reading discussion
- 15m of project updates

Mask policy

To ensure that health reasons do not deter anyone from coming to class, **masks are required**.

Masks are not provided by the university on a weekly basis. Come by my office if this is an issue.

If you are sick, consider staying home.

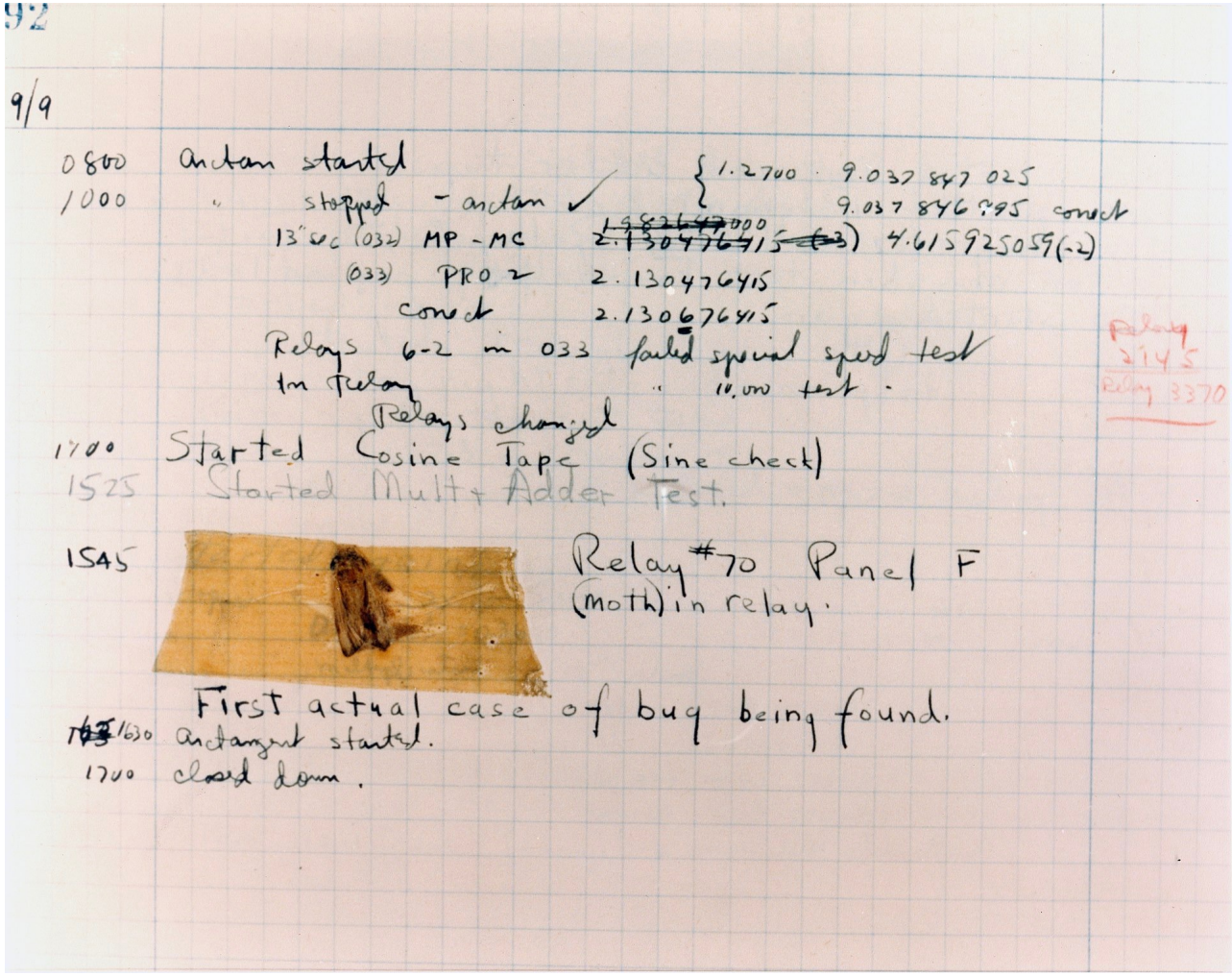
No recordings, but...

- Lecture notes
- Slides
- Jupyter notebooks
- Office hours

Disclaimer

- New course
- Feedback form on website
- Adjustments can be made

What is debugging?



9/9

0800 Arctan started { 1.2700 9.037 847 025

1000 " stopped - arctan ✓ 9.037 846 995 correct

13" sec (032) MP - MC ~~1.9826447000~~ ~~2.130476415 (-3)~~ 4.615925059(-2)

(033) PRO 2 2.130476415

correct 2.130676415

Relays 6-2 in 033 failed special speed test

in relay " 10.00 test.

Relays changed

Relay 2145

Relay 3370

1700 Started Cosine Tape (Sine check)

1525 Started Mult + Adder Test.

1545 Relay #70 Panel F (moth) in relay.

First actual case of bug being found.

~~1615~~ 1630 Arctangent started.

1700 closed down.