




Alan Blackwell

# Applying Ethics to Software



What kind of  
career do you  
imagine?

- How many expect to be paid to write software?
- How many expect to work for a large company?
- How many expect to create systems with business functionality?
- How many expect to have impact on people's lives?
- How many expect to see your work featured in the news?
- How many expect to see your work featured in a movie?

## The Post Office “Horizon IT” system

Custom software in UK post offices had many bugs:  
e.g. pressing enter when screen frozen silently repeated the transaction, so money seemed to go missing

- over 900 “sub-postmasters” (local shopkeepers) were convicted of fraud
- “one of the greatest miscarriages of justice in British history”

Whose fault was it?

- “Complex governance arrangements can make it hard to pin accountability on individuals”
- Who managed ICL pension payment system, cancelled after £700m ‘private finance’?
- Who proposed the business ‘pivot’ for the code base, after ICL was acquired by Fujitsu?
- Which software experts denied the system could cause observed accounting discrepancies?
- Who implemented functions for secret remote access to ‘correct’ local records?







## Studio E - 'Professional buck-passers'

Lead consultant responsible for Grenfell tower cladding

- Lacked knowledge of building regulations
- Relied on subcontractors without question
- Made fundamental errors in design

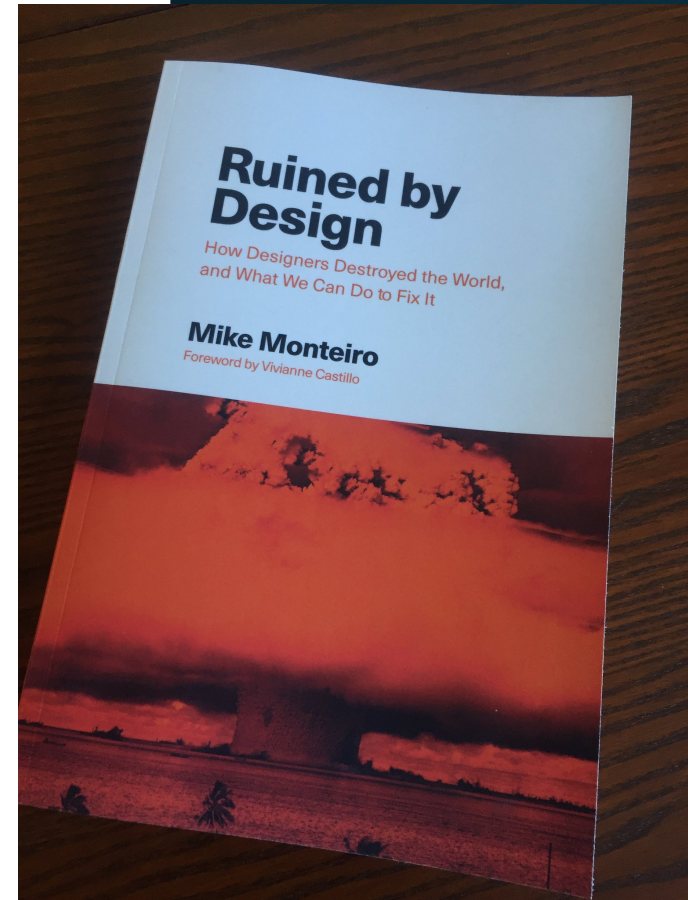
“An ineffectual middleman, supposedly presiding over multiple specialisms while having little technical knowledge of any of them”

**72 people died**



## Interaction designer Mike Monteiro says *the world is working exactly as we designed it*

- The combustion engine destroying our atmosphere is working exactly as we designed it.
- Guns and bombs work exactly as they're designed to work.
- Facebook's privacy settings, which out gay teens to their conservative parents, are working exactly as designed.
- (eX)Twitter's toxicity and lack of civil discourse is working exactly as it's designed to work.
- Designers are gatekeepers of what they bring into the world, and what they choose not to.
- Design is a craft with responsibility, with power to choose and power to influence



## RETAIL

OCTOBER 11, 2018 / 12:04 AM / UPDATED 3 YEARS AGO

## Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin



SAN FRANCISCO (Reuters) - Amazon.com Inc's [AMZN.O](#) machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

The team had been building computer programs since 2014 to review job applicants' resumes with the aim of mechanizing the search for top talent, five people familiar with the effort told Reuters.

Automation has been key to Amazon's e-commerce dominance, be it inside warehouses or driving pricing decisions. The company's experimental hiring tool used artificial intelligence to give job candidates scores ranging from one to five stars - much like shoppers rate products on Amazon, some of the people said.

"Everyone wanted this holy grail," one of the people said. "They literally wanted it to be an engine where I'm going to give you 100 resumes, it will spit out the top five, and we'll hire those."

"In effect, Amazon's system taught itself that male candidates were preferable. It penalized resumes that included the word "women's," as in "women's chess club captain." And it downgraded graduates of two all-women's colleges, according to people familiar with the matter. They did not specify the names of the schools.

Amazon edited the programs to make them neutral to these particular terms. But that was no guarantee that the machines would not devise other ways of sorting candidates that could prove discriminatory, the people said.

The Seattle company ultimately disbanded the team by the start of last year because executives lost hope for the project ..."

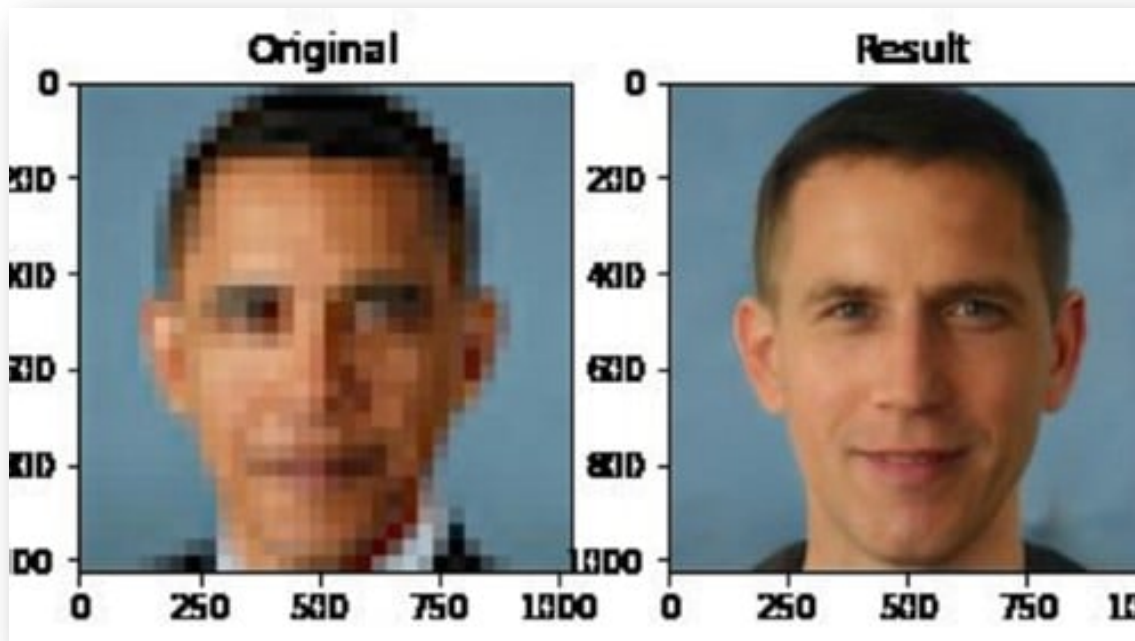
# An AI tool which reconstructed a pixelated picture of Barack Obama to look like a white man perfectly illustrates racial bias in algorithms

Isobel Asher Hamilton Jun 22, 2020, 4:00 PM



Barack and Michelle Obama. AP Photo/Carolyn Kaster

- A tool called Face Depixelizer grabbed the attention of the artificial intelligence research community this weekend.
- The tool takes pixelated pictures of people and uses AI to reconstruct sharp images of them.
- When given a pixelated photograph of Barack Obama, Face Depixelizer turned him into a white man.





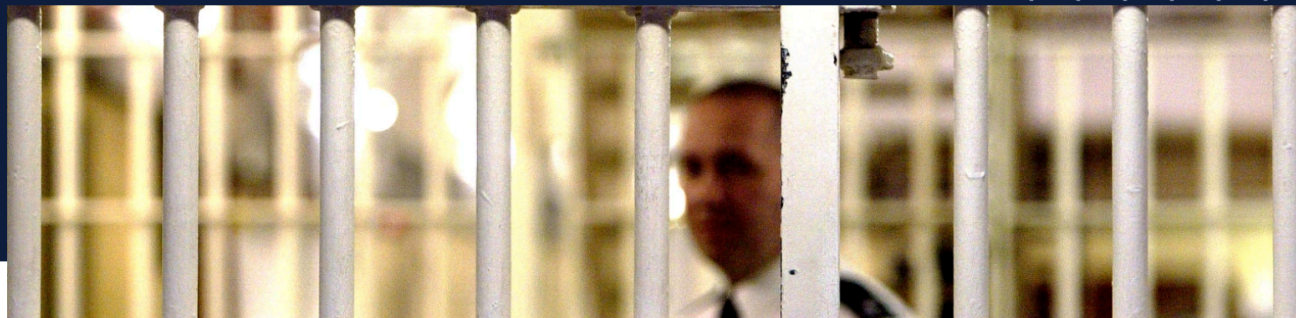
TECH POLICY

# AI is sending people to jail — and getting it wrong

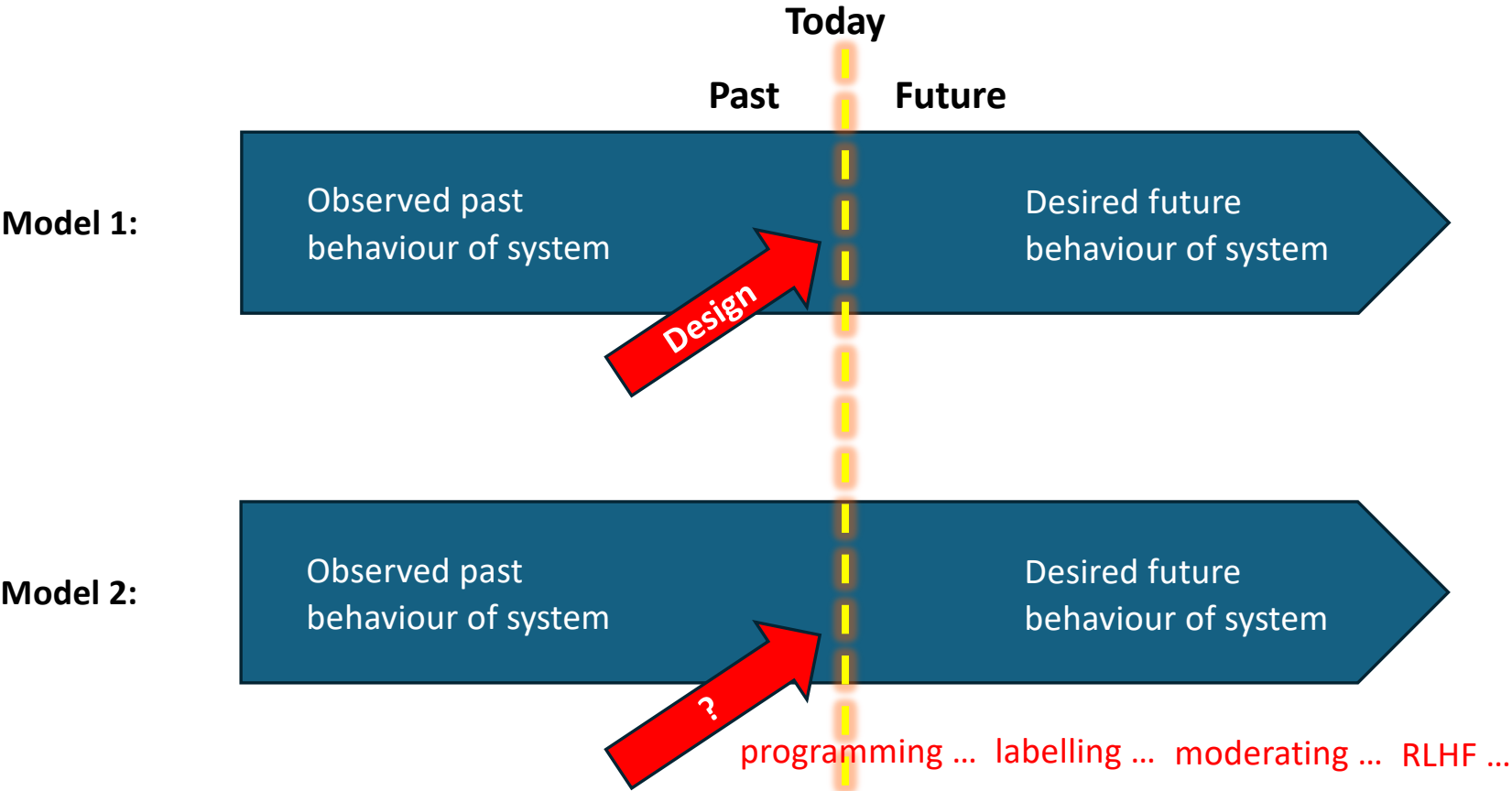
Using historical data to train risk assessment tools could mean that machines are copying the mistakes of the past.

By Karen Hao

January 21, 2019



# Ethics of software design vs. ethics of machine learning



COMMUNICATIONS  
OF THE  
ACM


[Computing Applications](#)  
Review articles

## Datasheets for Datasets

Documentation to facilitate communication between dataset creators and consumers.

By [Timnit Gebru](#), [Jamie Morgenstern](#), [Briana Vecchione](#), [Jennifer Wortman Vaughan](#), [Hanna Wallach](#), [Hal Daumé III](#), and [Kate Crawford](#)

Posted Dec 1 2021

 **Datasheets for Datasets**  
CACM

BY TIMNIT GEBRU, JAMIE MORGENSTERN, BRIANA VECCHIONE, JENNIFER WORTMAN VAUGHAN, HANNA WALLACH, HAL DAUMÉ III, AND KATE CRAWFORD

# Datasheets for Datasets

04:00

acm

Feedback

Heres an idea:  
introduce some professional software engineering  
to machine learning!

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Iii, H. D., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86-92

Sacked from Google  
after asking if language  
models can be too big

Cambridge CS Tripos, UK  
SET student of the year  
2001, co-founder of  
Women in ML

<https://cacm.acm.org/research/datasheets-for-datasets/>



# 'AI Ethics'

- Popular topic for postgraduate study in Cambridge
  - Leverhulme Centre for Future of Intelligence (Faculty of Philosophy) where Claire Benn is based
  - also Centre for Human-Inspired AI, Cambridge Digital Humanities, Minderoo Centre for Technology & Democracy ...
- Funded by many governments and philanthropists
  - Sunak AI Safety Summit, Biden executive order on safe AI, UN Secretary General AI Advisory body, OECD AI Policy Observatory, EU AI act...
- But what is 'AI'? Not just LLMs, not just deep learning, but 'algorithms'.
  - EU Act refers to 'software' - far more than what a specialist might call AI, including many surprisingly simple algorithms
  - Greatest concerns are decision making systems (and GDPR)
- Large companies, investors and sponsors would rather talk about and regulate AGI ... in 50 years time!

# Ethics of 'AI' business models

---

Mining personal data (if you're not paying, you are the product)

---

EULAs 'signed' under duress (e.g. Facebook, LinkedIn)

---

Minimal reward for creative work (e.g. Spotify, DALL-E, JukeDeck)

---

Ghost work, low pay and toxic content for RLHF 'Turkers'

---

Recommender algorithms that reward extremism

---

Regulatory imperialism/extraction in low income countries

---

Large investors prefer to focus on the risk of AGI in 50 years' time!

# Environmental ethics

← **Tweet**

 **asia murphy, fuck ICE (pitchwarring)** @am\_anatjala · Aug 16, 2018  
i still don't get bitcoin

24 164 585

 **nudity apprentice** @Theophite

Replying to @am\_anatjala

imagine if keeping your car idling 24/7 produced solved Sudokus you could trade for heroin

11:49 PM · Aug 16, 2018 · Twitter Web Client

**4.7K** Retweets **1.1K** Quote Tweets **19.3K** Likes

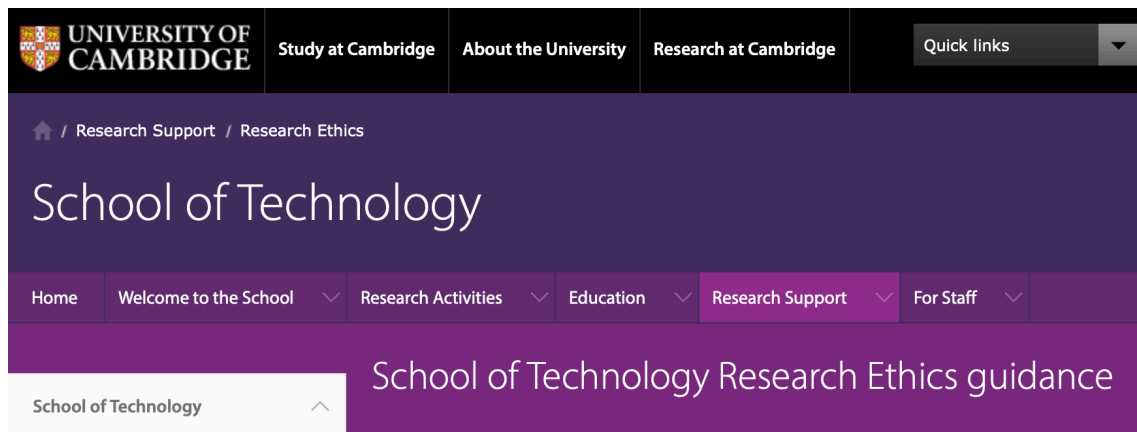


# Environmental ethics

- Financial speculation hides the absence of value in Ponzi schemes like Bitcoin.
- Tragedy of the commons – if it isn't costed, how can we stop it being destroyed?
- Venture investment in GenAI encourages research with high environmental cost relative to theoretical value (e.g. reimplementing classic algorithms as ChatGPT prompts)
- Many tech advances have unacknowledged environmental costs, including e-Waste, lithium mining, water extraction and pollution.



# Ethics for computer science students (and student engineers, Judge business school, manufacturing tripos ...)



<https://www.tech.cam.ac.uk/research-ethics/>

## **School of Technology Research Ethics guidance**

- > Action-based Management Research
- > Ageing and Disability Inclusion
- > Collaborative and Participatory Design
- > Controlled Experiments
- > Data Research
- > Diary and Probe Studies
- > Ethnographic and Field Study Techniques
- > Instrumented Software
- > Survey Methods

# Ethics content coming up ...

Group design projects in Lent term

Further HCI in Lent term

Cybersecurity in Easter term

Part II:

- Dissertation projects
- Human-Centred AI
- Bioinformatics
- Business studies
- E-commerce
- Affective AI
- Mobile Health
- ... and probably more!