People are not Binary
Embracing Complexity in Algorithmic Bias Research

Hannah Devinney
(they/them, she/her)
Roadmap

● A joke
● The problem (over-simplification)
● Why it matters
● Theory tour
● Q&A
A dairy farmer wants to increase the amount and quality of the milk she gets from her cows. So she asks her friend, a physicist, for some help. The physicist comes to visit, sees the cows, learns about how milking works, and promises to think about the problem.

After a few weeks, ze calls hir friend the farmer back, and excitedly says-
“I’ve got it!”

“First, assume a vacuum in which there is a perfectly spherical cow...”
Model

over- Simplification

- (easier) maths
- elide unnecessary detail
- “elegant”
- wrong answers?
  - doesn’t apply to real world
- ignores well-established knowledge
Spherical Cow:

Any assumption that

1) Simplifies a model
2) To the point of absurdity*

* untruth / unhelpfulness
Step 1: Define “gender bias”

Embedding bias for term *nurse*

What would “removing” this bias look like?

Reality(?): population statistics?
85% of nurses in the USA are women (>90% in Sweden)

Ideal(?)/gender-equal world: 50/50 odds?
Step 0: Define “gender”

“Folk Model”

- Binary
- Immutable
- Physiological

Os Keyes (2018)
[T]he term *heterosexual matrix* … designate[s] that grid of cultural intelligibility through which bodies, genders, and desires are naturalized … characterize[s] a hegemonic discursive/epistemic model of gender intelligibility that assumes that for bodies to cohere and make sense there must be a stable gender (masculine expresses male, feminine expresses female) that is oppositionally and hierarchically defined through the compulsory practice of heterosexuality.

Sex

Gender

Desire

“male” or “female”

“man” or “woman”

straight/heterosexual

Intersex people exist

Trans people exist

Non-binary & gender non-conforming people exist

Gay, bi, ace, etc. people exist

NORMATIVE ASSUMPTIONS:
the heterosexual matrix
Algorithms don’t exist in a vacuum

- **Matrix of Domination:** “The overall social organization within which intersecting oppressions originate, develop, and are contained”

- **Domains of power:**
  - Structural
  - Disciplinary
  - Hegemonic
  - Interpersonal

Gender is not:

- Binary
  - Non-binary people exist (hi!)
- Immutable
  - Trans people exist
- Physiological
  - Judgment is based on performativity and justified by physiology after

- When we follow these assumptions, we render trans people “unintelligible”
- Being unintelligible to a model can have violent and material consequences, just like being unintelligible to society can
  - (e.g. airport security scanners)
Gender is:

- A social construct
- which varies over
  - Time
  - Culture
Now what?

- We need different (better?) research questions!
- We need different (better?) methods!
- We need… to avoid reinventing the wheel?
What is “fair” AI, and do we want it?

- Bias $\rightarrow$ Power
- Who has access? Who is denied? By what mechanism?
- Metaphor: wall as barrier (created by power structure)
What is “fair” AI, and do we want it?

- Bias $\rightarrow$ Power

- Who has access? Who is denied? By what mechanism?

- Metaphor: wall as barrier (created by power structure)

“Equal” - everyone gets same resources
What is “fair” AI, and do we want it?

• Bias → *Power*

“Equitable” - resources match needs

• Who has access? Who is denied? By what mechanism?

• Metaphor: wall as barrier (created by power structure)
What is “fair” AI, and do we want it?

- Bias → *Power*

- Who has access? Who is denied? By what mechanism?

- Metaphor: wall as barrier (created by power structure)

“Justice” - barrier removed
Factors beyond “gender” matter

• no single/universal “woman’s experience” (or any other gender)
  - bell hooks, Judith Butler, Patricia Hill Collins, etc.

• Intersectionality
  - Kimberlé Crenshaw

“… everybody is simultaneously positioned within social categories, such as gender, social class, and ‘race’. So even when focusing particularly on one social category [...] intersectionality reminds us that we cannot understand that category in isolation.”

Ann Phoenix Interrogating Intersectionality (2006)
METHODOLOGY

• Be reflexive
  − Does the motivation, theoretical grounding, research question, and method all align?
  − Inner two year old: “whywhywhyWHY??” (also: “why not?”)

• Situate knowledge
  − Donna Haraway: “god-trick”

• Seek expertise
  − Beyond academic knowledge!
de-sphericalize your cows
Sources

- **Spot (the cow) was created by Keenan Crane** [http://www.cs.cmu.edu/~kmcrane/Projects/ModelRepository/](http://www.cs.cmu.edu/~kmcrane/Projects/ModelRepository/)
  - This version animated by Liam Devinney [http://devinney.com/liam/](http://devinney.com/liam/)