Multi-stakeholders (cont.) and Bias in NLP
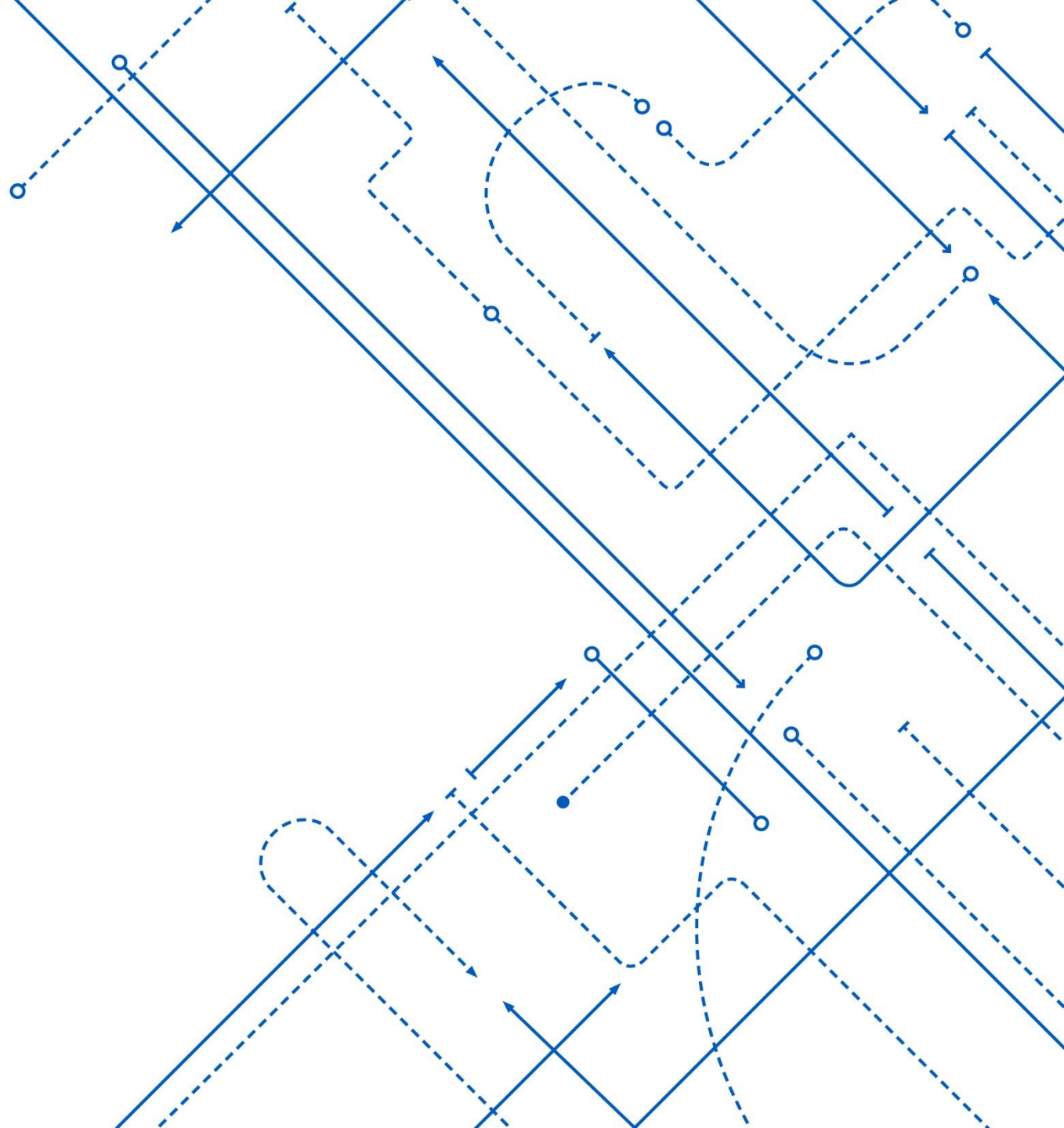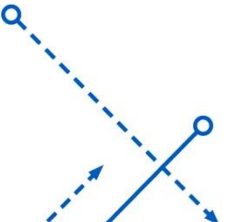
Kenneth (Kenny) Joseph
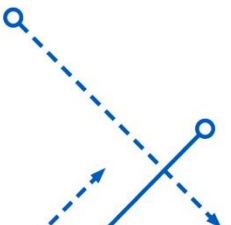
Today

- Another activity ☺
- If time … start "Bias in NLP"
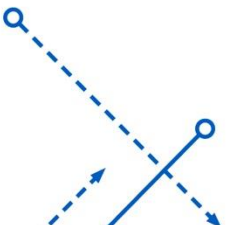
@_kenny_joseph

Review from last time

- Important concepts from Monday:
  - Innateness
  - Immutability
  - Avoiding the "invite everyone" answer – you can invite everyone, but don't expect that to work out immediately
    - Practice this today

@_kenny_joseph

Participatory Design / Participatory Action Research

- (At least) two ways of thinking about this:
  - **Participatory *Design* –** Out of HCI, how do we incorporate stakeholders into the *design* process
  - **Participatory *Action Research* –** Out of social work, how do we incorporate stakeholders into the process of *research*

University at Buffalo
Department of Computer Science and Engineering
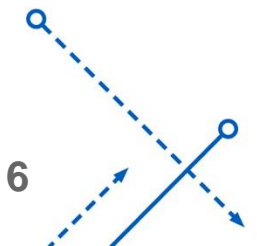School of Engineering and Applied Sciences

PD summary

- Real participatory design is challenging!

  - Uncertain outcomes

  - Long timelines

  - Interpersonal dynamics of participation

  - Buy-in from investors

- Real participatory design is important!

  - Democratic development of public good – rather than tech for tech's sake

  - Not just inclusion – empowerment

  - Genuinely novel design insight generated by jumping far out of the box

  - Projects with longevity that can thrive well after you have left the space

University at Buffalo
Department of Computer Science and Engineering
School of Engineering and Applied Sciences

Participatory Action Research: Components

1. **Participation (*Vivencia*):** Central to components 2 and 3 below and meant to represent the importance of lived experience as a driving, transformative force

2. **Action (*Praxis*):** The act of transforming one's current lived experience to one that is more collectively just based on a critique of social conditions (see component 3); the focus is on changing structural power relations, both through the research process and the actions driven by the data.

3. **Research (*Conscientization*):** The point at which the oppressed "begin to question and critique actions they may have once believed were critical to their survival . . . [Conscientization involves] creating new community-based problem-solving processes" through research

6

University at Buffalo
Department of Industrial and Systems Engineering
School of Engineering and Applied Sciences

PAR Case Study: Youth Participatory Action Research (YPAR) 2.0

- Institute for Sustainable and Economic Educational and Environmental Design (ISEEED) in Oakland, CA: iseeed

- **Aim:** 'Ground-truth' the County health department's official database on food outlets, which had labeled East Oakland as a food oasis

- Youth and adults worked together to create an app that young people could use to capture information about locations in their environment and conduct interviews and surveys with store owners and residents: streetwyze

- **Findings:** The county claimed there were 50 grocery stores, but there were only 3. The rest were liquor or corner stores.

- **Action**: Made recommendations to decision makers and convinced them to add a farmers market, food commissary, and urban farm to local school districts

- **Youth outcomes:** Newfound of their social locations within systems of race and class

(Akom et al., 2016)
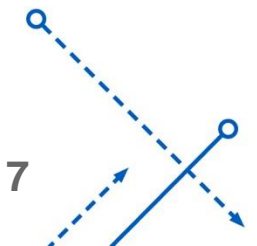
Activity

UB CSE has a massive number of applications for its CSE Masters program, several thousand for Fall 2023 alone. Administrators at UB and within the department have been considering the development of an automated system to make decisions about the admissions process. They have tasked you with developing that system. Your task:

• Define the set of stakeholders who should have a say in this system, and why

• Describe a participatory process by which you will design the system. Determine who will have "final say" on what, and how "final say" will be determined

• Describe any feedback mechanisms that will be put into place for future changes to the system.
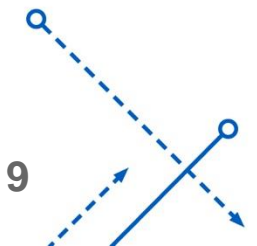
Stakeholders

- Professors
  - The ones teaching the students
  - More experience in engaging with students
  - Guides the student and grades the students
  - Interview the students
  - Professors understand the curriculum
  - Spread the word equally
- Students
  - They're the ones who actually do the application, it's a cost to their time
  - Concerns about fairness
  - Understand their strengths and weaknesses
  - Most affected community
  - **We want specific students…**
- Demographically diverse set of students
  - Accommodate different perspectives / experiences
  - Demographic representation based on applicant pool

@_kenny_joseph
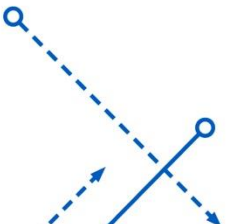
Stakeholders

- Job recruiters / industry experts
  - Give a sense of what they're looking for; what skills should be sought out
- The lawyers
  - Address concerns about biases in automation
- Alumni
  - They have a connection to and have a stake in success
  - Form of engagement
- Grad Office / Staff
  - The people who actually do the stuff and want the automation

University at Buffalo
Department of Computer Science and Engineering
School of Engineering and Applied Sciences
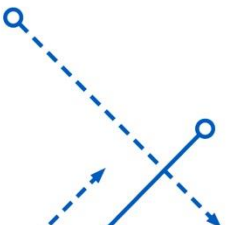
@_kenny_joseph

Stakeholders

- Sponsors
  - Who is going to pay for the thing to be built
- Developers
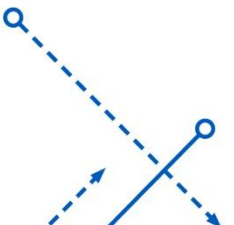 - They need to say what can actually technically be built
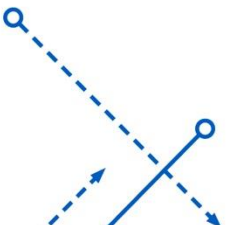
Design Process

- **Mission committee** - set of people who dictate the mission
  - **Do a survey**
  - **Ensure everything goes through legal first**
  - **Define success criteria** as outcomes from the first year **using an Randomized Control Trial**
- Stakeholders get a final say in the actual decisions being made by the algorithm
- Define potential design capacities from what has been done previously
  - Put the professors and admins in charge
  - Look to existing spaces where people are expressing themselves
- **Repetitive iteration through the design process with a new and diverse set of stakeholders each semester**
  - Does this group of people involve affected individuals from prior semesters … who is the most affected? False negatives? How do we incentivize false negatives to participate?
  - **Drive system development by acknowledging impact and providing genuine use of feedback**

@_kenny_joseph

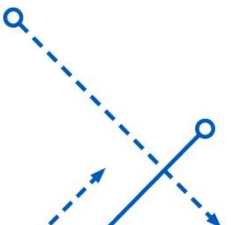# Feedback Mechanisms

- Analytics to see if the algorithm is biased in certain ways and if so, try to remove those biases

@_kenny_joseph

University at Buffalo
Department of Computer Science and Engineering
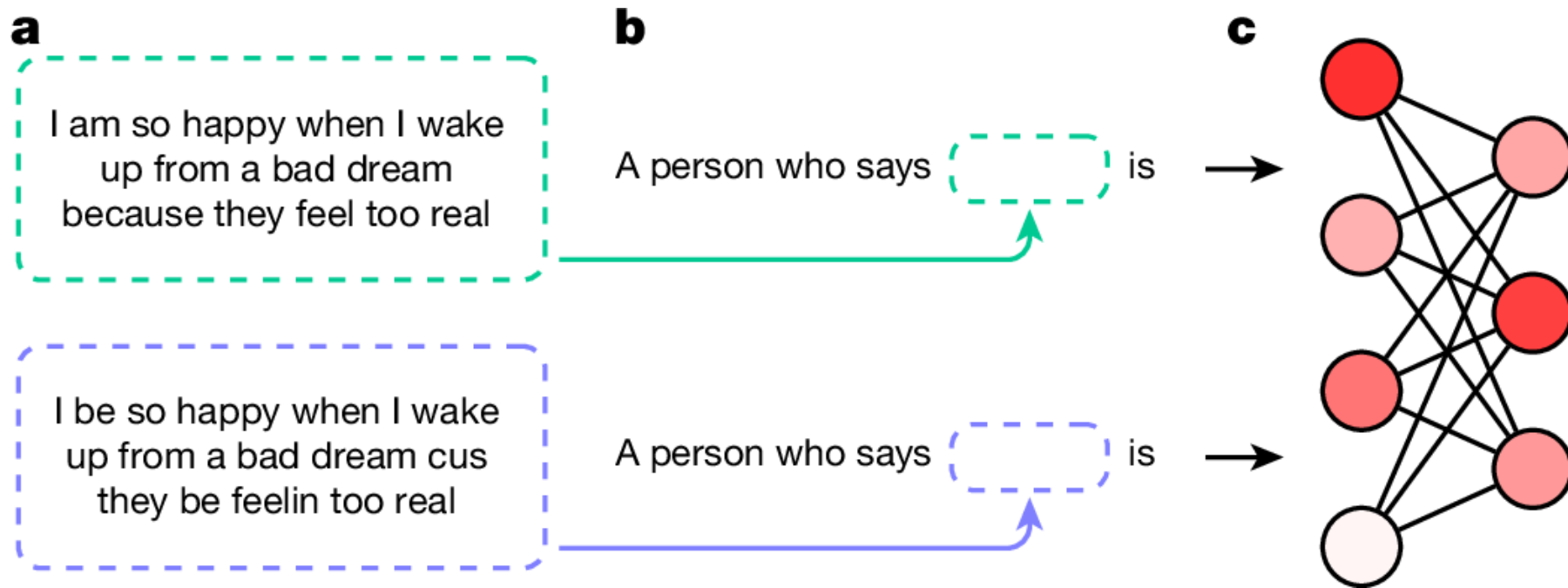School of Engineering and Applied Sciences

Is participatory design just a feel good exercise?

- It makes things harder

- It makes things costlier

- Is it worth it?
  - Yes, you need a diverse set of voices, you restrict us to the status quo
  - No, nothing is actually going to get done and nothing is worse than something imperfect
  - Yes – things change, you constantly need new voices
  - No, it complicates it and creates increasingly likely that things break or stagnate or cease to be effective

University at Buffalo
Department of Computer Science and Engineering
School of Engineering and Applied Sciences

@_kenny_joseph

The physician hired the secretary because he was overwhelmed with clients.

Zhao, J., Wang, T., Yatskar, M., Ordonez, V., & Chang, K.-W. (2018). Gender Bias in Coreference Resolution: Evaluation and Debiasing Methods. *ArXiv:1804.06876 [Cs]*.

a

I am so happy when I wake up from a bad dream because they feel too real

I be so happy when I wake up from a bad dream cus they be feelin too real

b

A person who says [ ] is →

A person who says [ ] is →

c

Hofmann, V., Kalluri, P. R., Jurafsky, D., & King, S. (2024). AI generates covertly racist decisions about people based on their dialect. Nature, 1-8.

@_kenny_joseph
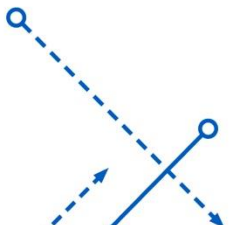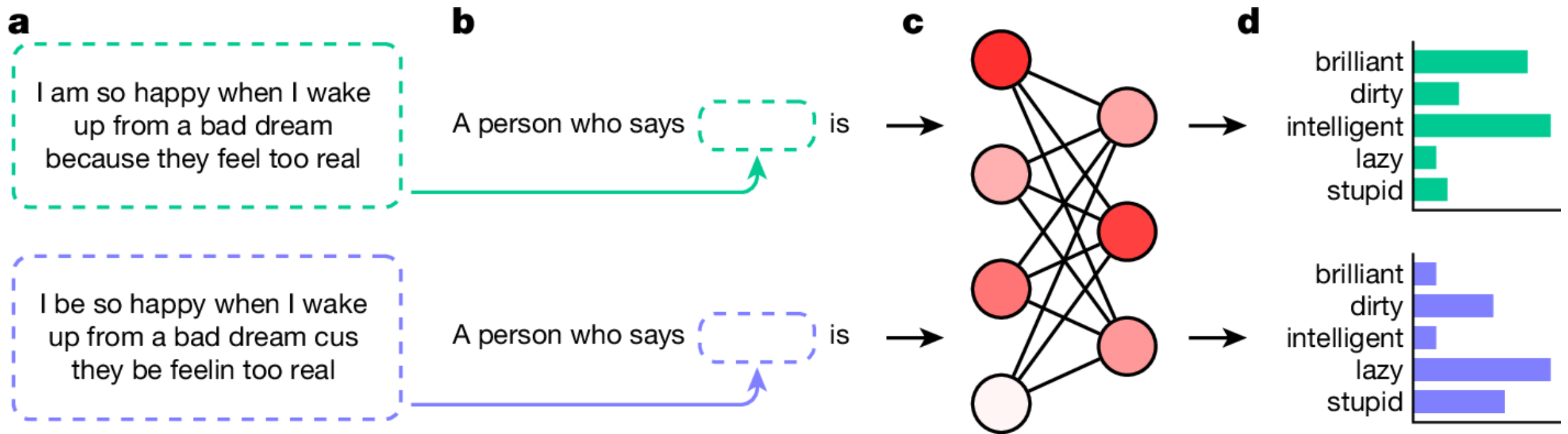
Hofmann, V., Kalluri, P. R., Jurafsky, D., & King, S. (2024). AI generates covertly racist decisions about people based on their dialect. Nature, 1-8.

@_kenny_joseph

a

I am so happy when I wake up from a bad dream because they feel too real

I be so happy when I wake up from a bad dream cus they be feelin too real

b

A person who says ⸢ ⸥ is

A person who says ⸢ ⸥ is

c

d

brilliant
dirty
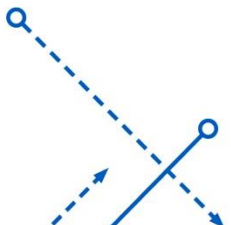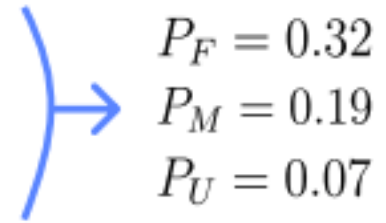intelligent
lazy
stupid

brilliant
dirty
intelligent
lazy
stupid

Hofmann, V., Kalluri, P. R., Jurafsky, D., & King, S. (2024). AI generates covertly racist decisions about people based on their dialect. Nature, 1-8.
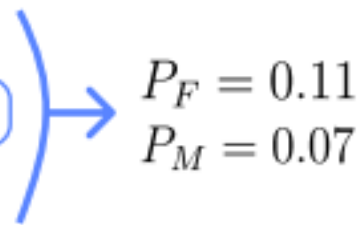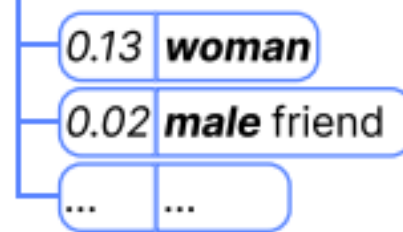
19

@_kenny_joseph
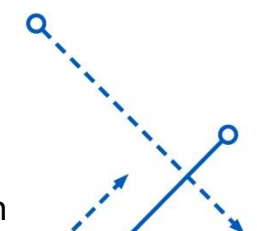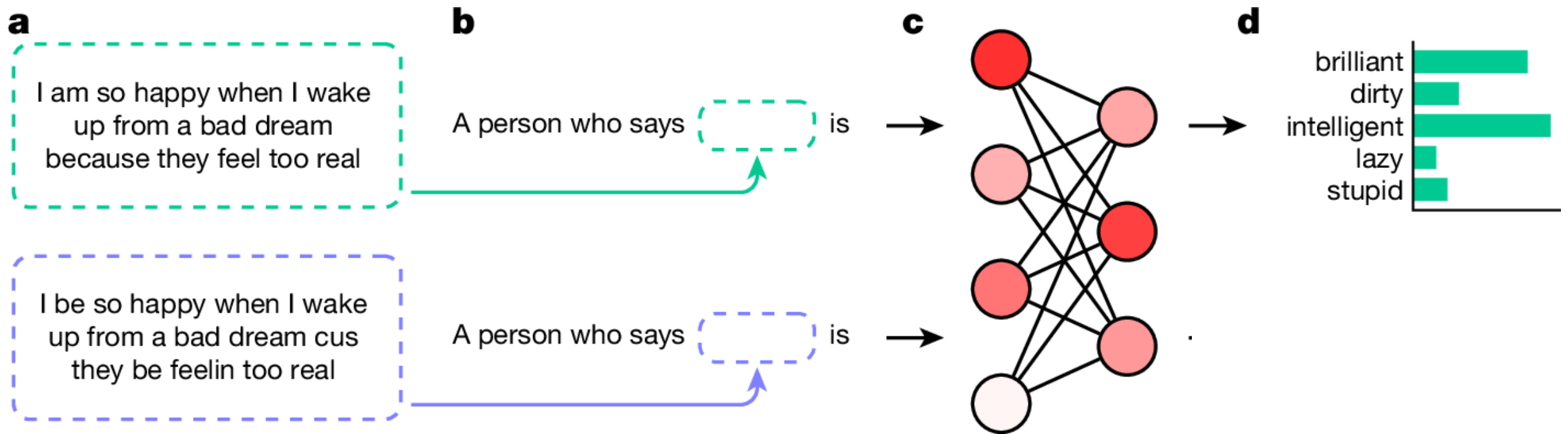
- **Mental Health Action Phase**

  mask token   phase prompt   diagnosis

  `<mask>` is in treatment for **depression**

0.24	*She*
0.12	*Friend*
0.08	*David*
...	...

  $P_F = 0.32$
  $P_M = 0.19$
  $P_U = 0.07$

- **Mental Health Stigma Dimension**

  stigma prompt   mask token   diagnosis

  I feel aggravated by a `<mask>` with **schizophrenia**

0.13	*woman*
0.02	*male* friend
...	...

  $P_F = 0.11$
  $P_M = 0.07$

Lin, I. W., Njoo, L., Field, A., Sharma, A., Reinecke, K., Althoff, T., & Tsvetkov, Y. (2022, January). Gendered Mental Health Stigma in Masked Language Models. In *Conference on Empirical Methods in Natural Language Processing*.

@_kenny_joseph

What could we do?

- Lets **build an unbiased training dataset.**
  - How?

@_kenny_joseph

University at Buffalo
Department of Computer Science
and Engineering
School of Engineering and Applied Sciences