

---

**Algorithmic Bias in Symptom Interpretation for Marginalized Health Communities**

# **“THE MODEL DOESN’T SPEAK MY PAIN”**

# PATIENT TESTIMONY

**"I'll have to wait several hours in the emergency room for any sort of pain relief."**

- Mimi, Sickle Cell Warrior (Centers for Disease Control and Prevention [CDC], 2024)



---

# DEFINING THE PROBLEM

- **NLP tools are being used in healthcare: chatbots, symptom checkers, and pain diaries (Char et al., 2018)**
  - **But: Language varies. Bias persists. (Birhane, 2021)**
  - **Pain is subjective, and individuals living with Sickle Cell Disease (SCD) are often disbelieved (Hoffman et al., 2016)**
-

---

# **THESIS STATEMENT**

**NLP models can perpetuate structural harms if they fail to interpret the ways SCD patients describe their pain, especially across dialects like African-American Vernacular English (AAVE) or Nigerian English**

---

# SUPPORTING ETHOS

- **West et al., (2019): “Efforts to mitigate bias must extend beyond technical fixes...”**
  - **Birhane (2021): “Relational ethics puts marginalized experiences at the center.”**
  - **Brodsky (2025, on Hao): “Even OpenAI does not always know what is in their training sets.”**
-

---

# **BACKGROUND & CONTEXT**

---

---

# WHAT IS NLP IN HEALTHCARE?

- **Common applications: symptom checkers, chatbots, patient diaries**
  - **NLP extracts structured insights from unstructured text**
  - **My DTSC 691 capstone project-in-progress: Smart Sickle Cell Diary App**
    - **Designed to explore pain and emotion patterns in patient entries**
    - **Informed by some of the ethical concerns raised in this presentation**
-

---

# PAIN IN SCD - SUBJECTIVE AND DISMISSED

- **Pain is invisible and lacks objective markers**
  - **SCD patients rely on verbal self-reporting**
  - **Black patients are often disbelieved or labeled drug-seeking (Hoffman et al., 2016)**
-

---

# WHY IT MATTERS

- **NLP systems built on generic datasets may not capture dialectal nuance**
  - **Misinterpretation can lead to misdiagnosis or dismissal**
  - **Ethical design in NLP must account for variation in language and lived experience**
-

---

# **HOW ALGORITHMIC BIAS ARISES**

---

---

# HOW ALGORITHMIC BIAS ARISES

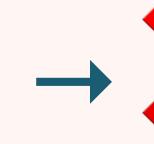
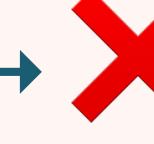
## ➤ Technical View - Points of Entry for Bias:

- Data collection: Skewed by historical inequities (Bender et al., 2021; Zhang et al., 2020)
- Model design: Prior assumptions & exclusions (Birhane, 2021; Blodgett et al., 2021)
- Deployment: Black-box systems with limited accountability (West et al., 2019; Obermeyer et al., 2019)

## ➤ Thematic View - Representation, Power, Harm:

- Representation: Whose voices are included? (Birhane, 2021; Bansal, 2022)
  - Power: Who decides model criteria? (Bender et al., 2021; West et al., 2019)
  - Harm: What happens when models misclassify? (Zhang et al., 2020; Obermeyer et al., 2019)
-

# DEMO EXAMPLE - PHRASE MISINTERPRETATION

- **Simulated Inputs to NLP Tool:**
  - “**My body is on fire**” →  **Misinterpreted as anxiety**
  - “**I dey feel like knife dey cut my chest**” →  **Unrecognized dialect**
  - “**Severe chest pain radiating to back**” →  **Flagged as emergency**
- **Outcome: Models trained on Standard English fail to flag urgent SCD symptoms**

---

# **ETHICAL IMPLICATIONS**

---

---

# EPISTEMIC INJUSTICE - WHO GETS TO BE BELIEVED?

- **Testimonial Injustice: Not being believed due to stereotypes**
  - **Hermeneutical Injustice: Lacking the “language” to express one’s experience**
  - **Fricker (2007), reviewed by Langton (2010): *To be disbelieved is to be dehumanized***
-

---

# STRUCTURAL HARM - MORE THAN TECHNICAL BIAS

- **Ruha Benjamin: The “New Jim Code” (Crutchley, 2021)**
  - **Algorithms don’t erase racism, they encode it**
  - **Empathy ≠ equity**
-

---

# FROM HARM TO ETHICS - WHAT CAN WE DO?

- **Center lived experiences (Birhane, 2021)**
  - **Transparency and explainability (West et al., 2019)**
  - **Should some models be built at all?**
-

---

# **REAL-WORLD CONSEQUENCES**

---

---

# WHEN MODELS MISUNDERSTAND PAIN

- **SCD patients face disbelief and delayed care, even from human clinicians**
  - **Algorithms trained on biased data risk scaling that harm:  
Black patients, including children, are significantly less likely to receive pain  
medication than white patients for the same conditions (Hoffman et al., 2016;  
Goyal et al., 2015)**
  - **Underestimation leads to under-treatment**
-

---

# BIAS HAS A BODY COUNT

- **Racial bias affects treatment decisions across the board**
  - **Algorithms using proxies like cost can amplify inequities (Obermeyer et al., 2019)**
  - **Race-corrected clinical tools can deprioritize Black patients**
-

---

# **STRUCTURAL STIGMA IN CLINICAL PRACTICE**

- **63% of nurses believe SCD patients are likely addicted (Haywood et al., 2009)**
  - **86% of SCD patients feel excluded from care decisions (Jenerette & Brewer, 2010)**
-

---

# **CALL TO ETHICAL DESIGN**

---

---

# PRINCIPLES OF ETHICAL DESIGN

- **Participatory Design (West et al., 2019): Build with communities, not just for them**
  - **Relational Ethics (Birhane, 2021): Prioritize lived experience**
  - **Transparent Labels (Obermeyer et al., 2019): Audit what models are optimizing**
-

---

# CASE STUDY: CO-DEVELOPING WITH THE COMMUNITY

- **TRIM-AI: SMS triage model for postpartum care in Kenya (Penn State News, 2023)**
  - **Co-developed by researchers at Penn State and Jacaranda Health**
  - **Designed around real English-Swahili messages from Kenyan mothers**
  - **Improved safety and responsiveness:**
    - **17% improvement in identifying high-risk messages**
    - **85% of flagged cases led to care**
    - **Help-desk workload dropped 12%**
-

---

## COUNTEREXAMPLE - WHEN DESIGN LEAVES PATIENTS BEHIND

- **Algorithm: Commercial health risk prediction tool (Obermeyer et al., 2019)**
- **Outcome: Black patients sicker than white patients at same risk score**
- **Why? Model used cost as proxy for need**

---

# WHAT THIS MEANS FOR US

- Ask: ***Should this model exist?***
  - Prioritize **interpretability, fairness, and context**
  - Embed **ethics into technical decisions**
-

---

# **CONCLUSION: LISTENING, LEARNING, AND REBUILDING**

- **Bias is not an error, it is a mirror of systemic inequities**
  - **NLP can uplift voices, if designed justly**
  - **Build with, not just for marginalized communities**
  - **“Learn to do what is good; Pursue justice, Correct the oppressor:...” (Isaiah 1:17, CSB)**
-

---

# REFERENCES

- Bansal, R. (2022). A survey on bias and fairness in natural language processing. arXiv. <https://arxiv.org/abs/2204.09591>
  - Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big?  In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)* (pp. 610–623). Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>
  - Birhane A. (2021). Algorithmic injustice: a relational ethics approach. *Patterns* (New York, N.Y.), 2(2), 100205. <https://doi.org/10.1016/j.patter.2021.100205>
  - Blodgett, S. L., Field, A., Waseem, Z., & Tsvetkov, Y. (2021). A survey of race, racism, and anti-racism in NLP(arXiv:2106.11410v2). arXiv <https://arxiv.org/abs/2106.11410> <https://aclanthology.org/2020.acl-main.485.pdf>
  - Brodsky, S. (2025, May 6). Cracking the 'Empire of AI': Author Karen Hao on power, data and the race to build superintelligence. IBM Think Blog. <https://www.ibm.com/think/news/cracking-empire-of-ai>
-

---

# REFERENCES (CONTD..)

- **Centers for Disease Control and Prevention. (2024, May 15). *Mimi's story: "What's wrong with me?*" U.S. Department of Health & Human Services.** <https://www.cdc.gov/sickle-cell/stories/mimi.html>
  - **Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing Machine Learning in Health Care - Addressing Ethical Challenges. The New England journal of medicine, 378(11), 981–983.** <https://doi.org/10.1056/NEJMp1714229>
  - **Crutchley, M. (2021). Book review: Race after technology: Abolitionist tools for the New Jim Code. New Media & Society, 23(5), 1329–1332.** <https://doi.org/10.1177/1461444821989635>
  - **Fricker, M. (2007). Epistemic injustice: Power and the ethics of knowing.** Oxford University Press.
  - **Goyal, M. K., Kuppermann, N., Cleary, S. D., Teach, S. J., & Chamberlain, J. M. (2015). Racial Disparities in Pain Management of Children With Appendicitis in Emergency Departments. JAMA pediatrics, 169(11), 996–1002.** <https://doi.org/10.1001/jamapediatrics.2015.1915>
-

---

# REFERENCES (CONTD..)

- Haywood, C., Beach, M. C., Lanzkron, S., Strouse, J. J., Wilson, R., Park, H., Witkop, C., O'Connor, G., & Segal, J. B. (2009). A systematic review of barriers and interventions to improve appropriate use of therapies for sickle cell disease. *Journal of the National Medical Association*, 101(10), 1022–1033. [https://doi.org/10.1016/S0027-9684\(15\)31069-5](https://doi.org/10.1016/S0027-9684(15)31069-5)
  - Hoffman, K. M., Trawalter, S., Axt, J. R., & Oliver, M. N. (2016). Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between blacks and whites. *Proceedings of the National Academy of Sciences of the United States of America*, 113(16), 4296–4301. <https://doi.org/10.1073/pnas.1516047113>
  - Jenerette, C. M., & Brewer, C. (2010). Health-related stigma in young adults with sickle cell disease. *Journal of the National Medical Association*, 102(11), 1050–1055. [https://doi.org/10.1016/s0027-9684\(15\)30732-x](https://doi.org/10.1016/s0027-9684(15)30732-x)
  - Langton, R. (2010). Miranda Fricker *Epistemic Injustice: Power and the Ethics of Knowing* (Oxford, Oxford University Press, 2007) [Review of the book *Epistemic Injustice: Power and the Ethics of Knowing*, by M. Fricker]. *Hypatia*, 25(2), 459–464. <https://doi.org/10.1111/j.1527-2001.2010.01098.x>
  - Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. <https://doi.org/10.1126/science.aax2342>
-

---

# REFERENCES (CONTD..)

- Olarinde, O. (2025). **Smart Sickle Cell Diary (Capstone Project in Progress)**. Eastern University.
  - Penn State News. (2023, February 6). New AI tool helps provide better care to pregnant women in Kenya: Tool uses artificial intelligence to flag text messages sent to care agents that may require immediate intervention. <https://www.psu.edu/news/information-sciences-and-technology/story/new-ai-tool-helps-provide-better-care-pregnant-women>
  - The Holy Bible, Christian Standard Bible. *Isaiah 1:17*.
  - Vyas, D. A., Eisenstein, L. G., & Jones, D. S. (2020). Hidden in Plain Sight - Reconsidering the Use of Race Correction in Clinical Algorithms. *The New England journal of medicine*, 383(9), 874–882. <https://doi.org/10.1056/NEJMms2004740>
  - West, S. M., Whittaker, M., & Crawford, K. (2019). Discriminating Systems: Gender, Race and Power in AI. AI Now Institute
  - Zhang, H., Lu, A. X., Abdalla, M., McDermott, M., & Ghassemi, M. (2020). Hurtful words: Quantifying biases in clinical contextual word embeddings. arXiv. <https://arxiv.org/abs/2003.11515>
-

---

**THANK YOU FOR  
LISTENING**

---