DTSC 690: Data Science Capstone: Ethical and Philosophical Issues in Data Science

“The Model Doesn’t Speak My Pain”: Algorithmic Bias in Symptom Interpretation for Marginalized Health Communities

By **Olubunmi Olarinde**

I. Introduction

Slide 1: Title Slide
Hello, my name is **Olubunmi Olarinde**, and today I’m presenting on a critical ethical and philosophical issue in data science titled: **“The Model Doesn’t Speak My Pain”**

This presentation explores *algorithmic bias in symptom interpretation*, particularly how it impacts marginalized health communities, such as patients living with Sickle Cell Disease.

Slide 2: Patient Testimony

Imagine being in excruciating pain, arriving at the hospital, and being told to wait for hours because your pain isn’t believed. That was the experience of Mimi, a Sickle Cell warrior of Arab American background. For her, this wasn’t hypothetical. It was routine.

And sadly, her story reflects what many patients, especially those from marginalized communities, tend to face every day when seeking care.

Slide 3: Problem defined

As data scientists and healthcare technologists, we increasingly rely on natural language processing, or NLP, to understand patient symptoms. Tools like chatbots and digital pain diaries now collect and analyze patient descriptions in real time. But language is not neutral. It is shaped by culture, dialect, and personal experience. And for Sickle Cell Disease patients, whose pain is often invisible and subjective, this can be dangerous. The failure to interpret diverse expressions of pain isn't just a technical flaw, it's a structural injustice.

Slide 4: Thesis Statement

In this presentation, I argue that NLP-based health tools risk reinforcing systemic health disparities if they are not designed to recognize how marginalized communities express their symptoms. If a model cannot understand your pain, it cannot help you. Worse, it may deny you the help you desperately need. As Abeba Birhane (2021) argues, the issue isn't just biased data, it's that these systems emerge from and operate within historical and relational patterns of exclusion.

Slide 5: Supporting Ethos

Scholars and institutions like the West et al., (2019), Abeba Birhane (2021), and journalist Karen Hao (2025) all emphasize that algorithmic bias is not merely a coding problem. The AI Now report warns, “Efforts to mitigate bias must extend beyond technical fixes to include transparency, contextual analysis, and participatory design” (p. 6). Birhane challenges the data science community to shift toward relational ethics, a framework that centers the lived experiences of marginalized communities rather than abstract ideals of fairness. She writes, “The harm, bias, and injustice that emerge from algorithmic systems...disproportionately impact individuals and communities at the margins of society” (Birhane, 2021, p. 5). Finally, Hao (2025) reveals a structural flaw at the core of modern AI: “Even OpenAI does not always know what is in their training sets. The data is just too large to audit manually.” These insights reinforce the urgency of interrogating not only what models do but how they’re built, who they include, and who they forget.

II. Background & Context

Slide 6: What Is NLP in Healthcare?

Natural Language Processing (NLP), is increasingly used to analyze unstructured patient text in healthcare. From symptom checkers to chatbots and patient diaries, these tools promise efficiency and personalization. As part of my capstone project in another course, I’m developing a Smart Sickle Cell Diary App. The goal is to use NLP to detect patterns in how patients describe their pain and emotions over time. But these tools only work well when they understand how patients naturally express themselves. Without context-specific training, NLP models risk overlooking critical cues in language: metaphors, dialects, or culturally shaped expressions. My goal is to move beyond generic language models and build one that actually hears what patients are saying, in the way they say it.

Slide 7: Pain in SCD - Subjective and Dismissed

Pain in Sickle Cell Disease is particularly challenging. It is invisible, episodic, and lacks biomarkers. Patients rely on language to explain what they're going through. Yet studies show that Black patients, especially those with SCD, are routinely dismissed or accused of exaggerating pain (Hoffman et al., 2016). This bias is not just interpersonal, it's institutional. A majority of white medical students surveyed endorsed false beliefs such as "Black people have thicker skin," which translated into lower pain ratings and fewer treatment recommendations for Black patients. These documented disparities in pain assessment highlight why NLP tools must be built with a deep understanding of race, language, and clinical history.

Slide 8: Why It Matters

The stakes are high. Most clinical NLP models are trained on datasets that reflect dominant language patterns, typically Standard American English. Dialects like African American Vernacular English or Nigerian English may not be interpreted accurately, leading to misclassification of symptoms or emotional tone. As Char et al. (2018) point out, algorithms have already shown biased behavior in sentencing and risk scoring. When we apply similar tools to healthcare, especially in underserved populations, the risk of replicating structural discrimination increases. Misinterpretation isn't just a technical error; it's a potential breach of clinical ethics.

III. How Algorithmic Bias Arises

Slide 9A: Technical View - Points of Entry for Bias

Algorithmic bias is not random, it arises at every step of the machine learning pipeline. First, data collection: if your dataset underrepresents SCD patients or fails to include dialectal diversity, the model cannot learn accurate representations. Large models, such as those described in Bender et al. (2021), are often trained on uncurated web text that over-represents hegemonic perspectives and encodes harmful stereotypes. Second, model design: assumptions about what pain "should" sound like often encode dominant norms. As Zhang et al. (2020) show, even embeddings trained on clinical notes reflect racial and gender disparities in healthcare. Third, deployment: many NLP tools are black boxes. There's often no transparency about how symptom-checking models make decisions. And once deployed, these tools may be trusted over patients' own voices, further amplifying harm.

Slide 9B: Thematic View - Representation, Power, Harm

From a thematic lens, algorithmic bias can be understood in terms of representation, power, and harm. First, representation, as Birhane (2021) and Bansal (2022) argue, data science often excludes the voices of those most affected. Dialectal expressions like “my body is on fire” in AAVE or Nigerian English may not map correctly to standardized medical symptoms. Second, power: developers and institutions set the norms, often without input from marginalized users. As Bender et al. warn, even the language of fairness can obscure whose interests are truly served. Third, harm: as NLP systems are integrated into care pathways, they can misclassify emotional distress as aggression or ignore serious pain, with real consequences. The result: patients not only feel unheard, they may also be untreated.

Slide 10: Demo Example - Phrase Misinterpretation

Let’s look at a simple demo. Imagine we input three phrases into an NLP-driven symptom checker. The first is a metaphorical expression common in AAVE: “My body is on fire.” A generic model might flag this as anxiety or dismiss it as vague. The second, in Nigerian Pidgin: “I dey feel like knife dey cut my chest,” may be completely unrecognized. But the third, clinical, textbook phrasing: “Severe chest pain radiating to back” is flagged correctly. All three describe potentially serious SCD symptoms. But only one speaks the model’s language. The others are at risk of being ignored.

IV. Ethical Implications

Slide 11: Epistemic Injustice - Who Gets to Be Believed?

Philosopher Miranda Fricker identifies two forms of epistemic injustice. *Testimonial injustice* occurs when someone is not believed because of who they are. *Hermeneutical injustice* happens when people lack the cultural resources to articulate their experiences. These aren't abstract ideas for SCD patients when their language doesn't fit clinical norms: they're not just misunderstood, they’re discredited. Rae Langton’s (2010) review of Fricker underscores the harm: being disbelieved damages your status as a knower and, by extension, your dignity as a human being.

Slide 12: Structural Harm - More Than Technical Bias

In *Race After Technology*, Ruha Benjamin introduces the concept of the **"New Jim Code"**, a term she uses to describe how emerging technologies can reinforce and legitimize existing racial inequalities under the guise of objectivity. As highlighted in Crutchley's review (2021), Benjamin argues that algorithms do not erase racism; they often encode it, amplifying historical injustice through “neutral” systems. She urges designers to move beyond “empathy” to equity and co-liberation. Racist robots, as she calls them, aren’t outliers, they’re reflections of the biased data and structural power that shape our systems. Healthcare AI must be built not only with better data, but with a deeper understanding of justice.

Slide 13: From Harm to Ethics - What Can We Do?

What do we do with all this? Abeba Birhane reminds us that ethical AI is relational. it must include the communities most affected by its outcomes. The AI Now Institute echoes this: technical fixes aren't enough. We need transparency, interdisciplinary collaboration, and the courage to ask hard questions: Who benefits? Who is harmed? Should this model even exist? Especially in healthcare, where trust and wellbeing are at stake, these are not just academic concerns: they're matters of life and death.

V. Real-World Consequences

Slide 14: When Models Misunderstand Pain

These biases don't just live in code, they shape real lives. Consider a young Black man with Sickle Cell Disease who rates his pain as 9 out of 10 in the ER. The nurse responds: "Really?" , because he isn't grimacing or crying out. Despite the clinical maxim that "pain is whatever the person says it is," patients with SCD are often doubted, delayed, and dismissed.

Now imagine that bias embedded in an algorithm: a model trained on data that doesn't include expressions like "knife dey cut my chest" or dismisses metaphorical language as psychological distress. Misinterpretation here isn't theoretical, it means longer waits, fewer medications, and missed diagnoses. In fact, studies show that 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).

Slide 15: Bias Has a Body Count

This pattern extends beyond individual encounters. A 2019 study by Obermeyer et al. revealed that a commercial algorithm used to manage millions of patients systematically underestimated the needs of Black patients. Why? Because it used future healthcare costs as a proxy for illness. But Black patients often receive less care, not because they're less sick, but because of systemic access barriers. The model, trained on those spending patterns, inferred they were healthier. In reality, they were sicker and less likely to get help.

Worse still, some clinical tools "correct" for race by assigning lower risk scores to Black patients, leading to fewer referrals and fewer resources. Remember, these aren't just numbers, they shape life-and-death decisions.

Slide 16: Structural Stigma in Clinical Practice

Stigma also plays a major role. Over 60% of nurses believe patients with SCD are likely drug-seeking, despite no evidence that they misuse opioids more than others with chronic pain. As a result, patients who request specific medications, because they know what works, are penalized instead of empowered. In a separate survey, 86% of SCD patients reported feeling excluded from decisions about their care, reinforcing their sense of invisibility within the clinical encounter (Jenerette & Brewer, 2011). The pain is real but the system doesn't believe it. And unless our algorithms are built differently, they'll only reinforce that disbelief at scale.

VI. Call to Ethical Design

Slide 17: Principles of Ethical AI Design

So what does ethical design actually look like? First, participatory design, as the AI Now Institute emphasizes, means involving affected communities in every stage of development. Second, relational ethics, as Abeba Birhane urges, asks us to center the lived experiences of people, not just theoretical fairness metrics. And third, transparency in labeling, as highlighted by Obermeyer et al., means we must interrogate what our models are truly predicting. Good intentions are not enough, we must build systems that reflect justice, not just efficiency.

Slide 18: Case Study - When Participatory Design Reduces Harm [Co-developing with the community]

An example of participatory NLP design comes from *TRIM-AI*, a model built to support postpartum triage in Kenya (Penn State News, 2023). This project was co-developed by researchers from Penn State and Jacaranda Health using real SMS messages from new and expecting mothers. The model was tailored to code-mixed English and Swahili, reflecting how women naturally describe symptoms. Because it was built on real-world user input-messages sent through PROMPTS, it avoided many common pitfalls of misinterpretation. The result: a 17% improvement in identifying high-risk messages, 85% of flagged cases led to care, and help-desk workload dropped by 12%. This is ethical design in action.

Slide 19: Counterexample – When Design Leaves Patients Behind

Now let's look at what happens when patients are excluded. In a landmark 2019 study by Obermeyer and colleagues, researchers analyzed a commercial health risk algorithm used to identify high-need patients. The model used healthcare costs as a proxy for health needs. But because Black patients tend to receive less care, even when equally sick, the algorithm underpredicted their needs. At the same risk score, Black patients were substantially sicker than white patients.

This bias wasn't malicious, it was systemic. But it occurred precisely because the design failed to include the lived realities of underserved populations. Participatory design might have caught that flawed assumption before deployment.

Slide 20: What This Means for Us

For those of us training models, writing code, or designing health tech interfaces, this is personal. It's not just about building smarter tools, it's about building *fairer* ones. We must embed ethics into every technical decision. That means asking difficult questions: *Should this model exist? Whose pain does it understand? Whose pain does it ignore?*

Especially in healthcare, where lives are on the line, our models must do more than perform, they must *care*. And this brings us to the final message of this presentation.

VII. Conclusion

Slide 21: Listening, Learning, and Rebuilding

We've seen how language models in healthcare risk reinforcing the very injustices they promise to fix. Not because the algorithms are evil but because they are built within systems that already disbelieve, dismiss, and devalue certain voices.

The model didn't speak Mimi's pain. And it still doesn't for many like her from Arab American women to Black patients and beyond. But it could.

When we include the languages, metaphors, and lived experiences of those historically excluded…when we design with communities instead of around them…when we question not just the code, but the assumptions behind it… perhaps, we begin to move toward justice.

As data scientists, we don't just have the tools, we have the responsibility. To listen better. To model differently. And to ensure that the next generation of NLP systems doesn't silence pain but helps to heal it.

Or in the words of Isaiah 1:17(CSB):
"Learn to do what is good; Pursue justice, Correct the oppressor;..."

May we build tools that do exactly that.

References

1. Bansal, R. (2022). *A survey on bias and fairness in natural language processing*. arXiv. https://arxiv.org/abs/2204.09591
2. 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

3. 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
4. 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
5. 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
6. 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
7. 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
8. 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
9. Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.
10. 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
11. 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
12. 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
13. 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
14. 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
15. 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

16. 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
17. **The Holy Bible, Christian Standard Bible**. *Isaiah 1:17.*
18. 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
19. West, S. M., Whittaker, M., & Crawford, K. (2019). Discriminating Systems: Gender, Race and Power in AI. AI Now Institute
20. 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