BOARD # 28: Work in progress: Preparing Biomedical Engineers to Tackle Biases in Machine Learning

Dr. Xianglong Wang, University of California, Davis

Dr. Xianglong Wang is an Assistant Professor of Teaching in Biomedical Engineering (BME) at the University of California, Davis, and the program coordinator of the BME Quarter at Aggie Square clinical immersion program. Dr. Wang leads the cube3 lab, an engineering educational lab focused on community building and pedagogical innovations in BME. As a steering committee member, he helps shape the educational programs offered by the Center of Neuroengineering and Medicine at UC Davis. Before joining UC Davis, he was a career-track Assistant Professor at Washington State University (WSU). Dr. Wang is the recipient of the 2024 ASEE-PSW Section Outstanding Early Career Teaching Award, 2023 UC Davis Biomedical Engineering Excellence in Teaching Award, and 2022 WSU Reid Miller Teaching Excellence Award.

Tiffany Marie Chan, University of California, Davis

Tiffany Chan is a 4th-year undergraduate student in biomedical engineering at UC Davis and the recipient of the 2024 ASEE-PSW Section Undergraduate Student Award. She actively contributes to the cube3 Lab, where her interests lie in community building and inclusive practices. Tiffany is involved in various DEI (Diversity, Equity, and Inclusion) research initiatives within the lab, including organizing student-faculty lunches and participating in the gender equity first-year seminar program. Additionally, she serves as the chair of the undergraduate subcommittee for the department's Health, Equity, and Wellness committee and holds the position of president in the BMES student chapter at UC Davis.

Angelika Aldea Tamura, University of California, Davis

Angelika Tamura is a third-year undergraduate student pursuing a B.S. in Biomedical Engineering at the University of California, Davis. She is a research assistant for the Cubeˆ3 lab, which primarily does research in engineering education. She is also deeply engaged in the Biomedical Engineering Society chapter at UCD, where she serves as the graphics designer and actively contributes to the Outreach and Fundraising committees. Alongside her involvement in BMES, Angelika is an enthusiastic member of B-Hours, a student-run organization dedicated to projects benefiting clinics in Sacramento. Focusing her course studies in cell and tissue engineering, Angelika is currently seeking research opportunities to further explore her passion in bioprinting and regenerative medicine.

Work in progress: Preparing Biomedical Engineers to Tackle Biases in Machine Learning

Introduction

From just 21 FDA-authorized (including approvals and 510(k) clearances) artificial intelligence (AI) and machine learning (ML)-enabled medical devices by January 2014 to over 1,000 by September 2024 [1], machine learning has seen explosive growth in biomedical engineering (BME). Besides AI/ML-enabled medical devices which focus on biomedical signals and imaging, ML is actively influencing BME research in areas such as drug design [2], tissue engineering [3], biomaterials [4], and medical diagnostics [5]. AI/ML-based products, especially in large language model (LLM)-based chatbots, are quickly integrated into the current educational environment [6]. Although initial investigations of using these chatbots such as ChatGPT or Perplexity AI in an academic context seemed underwhelming [7], rapid improvement in the performance of these products, as reflected by one faculty's experience of Perplexity AI scoring 80% on their multiple choice-based engineering quiz, accentuated the need for BME educators and students to improve AI literacy and cultivate responsible use of AI.

ML algorithms are computer programs that improve their performance with more experience (data) [8]. Therefore, problems in the data used to train ML algorithms, such as demographic biases, can be reflected in the performance of ML algorithms. In a BME context, GPT-4, which powers ChatGPT and Perplexity AI, showed strong ethnic biases when assigning medical conditions such as HIV/AIDS [9], while GPT-4 and Gemini (also powers AI-enabled notebook, NotebookLM) showed negative perception towards disability in general public and patient populations [10]. Development of fair AI/ML-enabled medical devices and performing bias-free research of ML is significantly challenging the applicability of AI/ML in BME. [11] The U.S. Food and Drug Administration (FDA) recognized the necessity of addressing bias in clinical machine learning systems, first in the proposed regulatory framework published in April 2019 [12] and later as a guiding principle in October 2021 [13].

However, ML courses in BME programs around the U.S. are still rare, and teaching of bias in ML systems remains largely scattered in computer science and ethics departments, which often focus on privacy [14]. At the BME department of UC Davis, we recognize the importance of arming our students with technical proficiency in ML, especially with a rise in AI/ML-focused senior design projects over the past few years. During our pilot offering of the BME ML course in Spring 2024, we designed and taught a one-week module about diversity, equity, and inclusion (DEI) problems in ML with the following learning objectives (LOs):

1. Assessing whether a question solved by ML exacerbates existing society bias.
2. Understanding "distribution shift" from training data to real-life usage, which is the source of bias in ML systems.
3. Designing equitable ML systems with considerations of the "right" problem and dataset.

Methods

Our one-week module consists of one lecture (80 minutes) and one hands-on lab (approximately 90 minutes). The decision to integrate a dedicated hands-on lab is to improve the outcome of LO3 (designing equitable ML systems), which can potentially be better achieved through hands-

on learning [15]. We decided to implement this module immediately after students learned their first classification algorithm, logistic regression, to raise awareness in the challenges before the students become too experienced in designing their own ML algorithms without the awareness of DEI. Students were properly warned about the content in both the lecture and the lab.

The lecture started with a short introduction (25 minutes) on understanding theories such as "garbage-in, garbage-out", "bad" questions to ask ML, different classes of "bad" data, distribution shift, and class imbalance in machine learning. We followed the theoretical portion of the lecture with a series of case studies (55 minutes). These case studies began with discussions on stereotype-reinforcing ML systems including ML-based gender-identification systems for recommenders of online ads [16], reflection of society bias in ML-based word embeddings trained on news articles for LLMs [17], and a startup claiming to identify a person's personality (including labeling a person as a terrorist) using facial recognition [18]. We then discussed more case studies on distribution shift, which include gender/racial performance gaps in facial recognition systems [19], geographical/cultural performance gaps in object recognition [20, 21], and racial biases in ML algorithms for allocating healthcare resources [22].

In the lab, students were tasked with training a machine learning algorithm on an unknown dataset, only to realize that the resulting algorithm associated male names with career words (salary, office), and female names with family words (marriage, relatives). Unlike other labs in this course, which students were graded on the performance of their algorithms, the students were graded on participation only for this lab. The lab concluded with reinforcing the necessity of asking "good" questions with ML, diversifying points of view in ML engineering teams, and avoiding propagation of bias from the training data into real-life performance of ML systems.

To assess the learning outcomes of the module, we designed a 7-item survey, including five (5) 6-point Likert-scale questions on various confidence measures of DEI in ML, and two free-response questions asking about the clearest and muddiest points from the module. The survey was distributed on paper before the lab, and the students completed the survey after the lab. Students were instructed to not include their names in the survey and informed that completion of the survey would not hurt or help their grade in the course. To ensure anonymity, the collected surveys were transcribed into digital format by the primary author's research assistants, who are not affiliated with the course. These research assistants analyzed the quantitative data in Excel, identified central themes by hand, and then reported the aggregated results to the primary author. The study was designated as Non-Human Subject Research by our IRB office.

Results

Nineteen (19) of 22 enrolled students attended the module, and all attendees responded to the survey. The quantitative results of the survey are shown in Table 1. Overall, the students agreed that the learning module increased their sensitivity to bias and DEI problems in ML, with the highest agreement being in providing equal opportunities for ML (5.28/6) and the lowest being in taking actions to reduce bias in ML (4.74/6).

The qualitative results echoed what we observed in the quantitative results. Fourteen (14) students responded to the question about the clearest point. Eleven (11) of the 14 students wrote about ML algorithms needing to be trained on unbiased data, and one student wrote about asking

the right questions with ML algorithms. Six (6) students responded to the question about the muddiest points. Five (5) of the six students expressed concern about what they, as future ML engineers, could do to reduce bias in the ML algorithms, if the bias is rooted in the overall society. One student was curious about potential measures to quantify bias in ML systems. One student foresees difficulty in convincing people to train bias-free ML models especially if the people in question have not experienced such bias.

Table 1: Reported agreement levels on 6-point Likert-scale questions (1=strongly disagree, 6=strongly agree). The average (avg) and standard deviation (std) of the responses are reported below. N=19 (except for statements 1 and 3, where N=18).

From this week's module, I became more confident in…

Statement	Avg ± Std
1. Providing equal opportunities of ML-based medical devices to all groups of people.	5.28 ± 0.67
2. Taking action to prevent reproduction/maintenance of inequalities in machine learning.	4.74 ± 0.93
3. Designing, implementing, and assessing ML plans with a DEI perspective.	5.11 ± 0.90
4. Conveying values in DEI issues in ML.	5.16 ± 0.69
5. Educating ML engineers on DEI issues.	5.05 ± 0.85

Discussion

We successfully taught the first iteration of the bias in the ML module in our BME ML course. Overall, the module successfully planted the seed to become aware of bias in ML among our first cohort of students. The quantitative and qualitative evaluations revealed that students who took this module achieved better outcomes in LO1 and LO2 (understanding biases in ML) than LO3 (taking actions to prevent/reduce biases in ML).

To address this limitation, we reflected on the lecture portion of the module. Our current case studies strongly focused on conveying real-life impact of bias in ML but are relatively lacking in BME-specific case studies and guidelines. In our next iteration, we will integrate more BME-specific discussions with a stronger focus on the FDA's regulatory framework for AI/ML-enabled devices [12] and the associated guidelines [13]. These documents contain actionable prompts for creating equitable ML systems (such as matching the population within the dataset with the intended use group) and evaluating equitable ML systems. These content modifications would adequately address the confusion of how the students should act as a ML engineer and better serve as their technical preparation for their AI/ML-related senior design projects. We also intend to search BME-specific open-source datasets that can be trained to demonstrate disparity in clinical outcomes among certain demographics but are simple enough to train with an elementary level of ML knowledge in the allocated lab time.

Our assessment of the outcomes from the study was largely based on students' perceptions and reflections, which were indirect evidence of the outcomes. For our future iterations, we will work on developing direct assessments for the module, perhaps in the format of an in-class quiz or homework assignments, to assess the understanding of bias in ML more accurately. Our future iterations of this study would also incorporate pre-/post-surveys to assess students' perceptions and confidence instead of asking for increased confidence levels at the end of the module. This form of assessment will help us gain a clearer picture of the efficacy of our module.

Disclaimer

Any product that may be named or evaluated in this article, or claims that may be made by its manufacturer, is not guaranteed or endorsed by the author(s).

Acknowledgement and Material Availability

We would like to thank Julia Chamberlain, Kathleen Cruz, Sara Dye, Rob Furrow, Irene Joe, Bwalya Lungu, Hannah Minter Anderson, Ali Moghimi, and Patricia Turner from the PCI Diversity, Equity, and Inclusion Faculty Learning Community at UC Davis. We will make our current learning module available at https://cube3.engineering.ucdavis.edu.

References

[1] U.S. Food and Drug Administration. Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices [Online] Available: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices

[2] V. Binson, S. Thomas, M. Subramoniam, J. Arun, S. Naveen, and S. Madhu, "A Review of Machine Learning Algorithms for Biomedical Applications," *Annals of Biomedical Engineering,* vol. 52, no. 5, pp. 1159-1183, 2024.

[3] C. Wu, B. Wan, A. Entezari, J. Fang, Y. Xu, and Q. Li, "Machine learning-based design for additive manufacturing in biomedical engineering," *International Journal of Mechanical Sciences,* vol. 266, p. 108828, 2024.

[4] H. Chen, Y. Liu, S. Balabani, R. Hirayama, and J. Huang, "Machine learning in predicting printable biomaterial formulations for direct ink writing," *Research,* vol. 6, p. 0197, 2023.

[5] P. G. Jacobs *et al.*, "Artificial intelligence and machine learning for improving glycemic control in diabetes: best practices, pitfalls, and opportunities," *IEEE reviews in biomedical engineering,* vol. 17, pp. 19-41, 2023.

[6] M. Farrokhnia, S. K. Banihashem, O. Noroozi, and A. Wals, "A SWOT analysis of ChatGPT: Implications for educational practice and research," *Innovations in Education and Teaching International,* vol. 61, no. 3, pp. 460-474, 2023, doi: 10.1080/14703297.2023.2195846.

[7] M. Deike, "Evaluating the performance of ChatGPT and Perplexity AI in Business Reference," *Journal of Business & Finance Librarianship,* vol. 29, no. 2, pp. 125-154, 2024.

[8] T. M. Mitchell and T. M. Mitchell, *Machine learning* (no. 9). McGraw-hill New York, 1997.

[9] T. Zack *et al.*, "Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: a model evaluation study," *The Lancet Digital Health,* vol. 6, no. 1, pp. e12-e22, 2024.

[10] J. T. Urbina, P. D. Vu, and M. V. Nguyen, "Disability Ethics and Education in the Age of Artificial Intelligence: Identifying Ability Bias in ChatGPT and Gemini," *Archives of Physical Medicine and Rehabilitation,* vol. 106, no. 1, pp. 14-19, 2025, doi: 10.1016/j.apmr.2024.08.014.

[11] R. J. Chen *et al.*, "Algorithmic fairness in artificial intelligence for medicine and healthcare," *Nature biomedical engineering,* vol. 7, no. 6, pp. 719-742, 2023.
[12] U.S. Food and Drug Administration, "Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) - Discussion Paper and Request for Feedback," 122235, 2019. [Online]. Available: https://www.fda.gov/media/122535/download?attachment
[13] U.S. Food and Drug Administration, Medicines and Healthcare products Regulatory Agency, and Health Canada, "Good Machine Learning Practice for Medical Device Development: Guiding Principles," 2021. [Online]. Available: https://www.fda.gov/media/153486/download
[14] J. Saltz *et al.*, "Integrating ethics within machine learning courses," *ACM Transactions on Computing Education (TOCE),* vol. 19, no. 4, pp. 1-26, 2019.
[15] M. Schwichow, C. Zimmerman, S. Croker, and H. Härtig, "What students learn from hands-on activities," *Journal of research in science teaching,* vol. 53, no. 7, pp. 980-1002, 2016.
[16] T. Gebru, "Race and gender," *The Oxford handbook of ethics of AI,* pp. 251-269, 2020.
[17] T. Bolukbasi, K.-W. Chang, J. Y. Zou, V. Saligrama, and A. T. Kalai, "Man is to computer programmer as woman is to homemaker? debiasing word embeddings," *Advances in neural information processing systems,* vol. 29, 2016.
[18] F. Pasquale, "When machine learning is facially invalid," *Communications of the ACM,* vol. 61, no. 9, pp. 25-27, 2018.
[19] J. Buolamwini and T. Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," in *Conference on fairness, accountability and transparency*, 2018: PMLR, pp. 77-91.
[20] T. De Vries, I. Misra, C. Wang, and L. Van der Maaten, "Does object recognition work for everyone?," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2019, pp. 52-59.
[21] S. Shankar, Y. Halpern, E. Breck, J. Atwood, J. Wilson, and D. Sculley, "No classification without representation: Assessing geodiversity issues in open data sets for the developing world," *arXiv preprint arXiv:1711.08536,* 2017.
[22] Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, "Dissecting racial bias in an algorithm used to manage the health of populations," *Science,* vol. 366, no. 6464, pp. 447-453, 2019, doi: 10.1126/science.aax2342.