# Anaelia Ovalle

they/them

anaeliaovalle | anaeliaovalle | Website | anaeliaovalle@gmail.com

## RESEARCH INTERESTS

Responsible AI strategy, Inclusive NLP, LLM Adversarial Safety, Low-Resource NLP

## EDUCATION

**University of California, Los Angeles**                           Sept 2019 - Aug 2024
*PhD Candidate in Computer Science, Advised by Professor Kai-Wei Chang*
*M.S. Computer Science*                                            Sept 2021

**University of San Francisco**                                    May 2017
*B.S. Data Science*

## WORK EXPERIENCE

**Global Partnership on AI**

Equality and Diversity in Artificial Intelligence, *Machine Learning Consultant*          Present

– Deliver evidence-based recommendations to regional leadership to incorporate effective diversity approaches throughout the AI lifecycle.
– Partnership with Mila - Quebec Artificial Intelligence Institute & The Organisation for Economic Co-operation and Development (OECD)

**Meta**

Fundamental AI Research (FAIR) , *Research Scientist Intern*          Aug 2023 - Dec 2023

– Conceptualized and experimented with sociotechnical evaluation framework to understand biases in training LLMs from human preferences.

Responsible AI (RAI), *Research Scientist Intern*          June 2021 - Oct 2021

– Created new fairness metric and approach for assessing 1) fair content sourcing from Instagram and 2) unfair content sourcing mitigation which was later productized by Facebook Ads team.

**Amazon**

Trustworthy Alexa, *Applied Scientist Intern*          June 2023-Aug 2023, June 2022 - Sept 2022

– Discovered how LLM misgendering behavior is exacerbated via BPE Tokenization. Accepted to 2024 Neurips Workshop (top 15%).
– Created benchmark dataset TANGO and automatic/human evaluations to evaluate gender inclusivity of generative LLMs. Published in 2023 ACM FAccT Conference.
– Led dialogue across teams outside of Alexa for stakeholder buy-in and interdisciplinary feedback on experiment design.

Prime Video Compliance, *Applied Scientist Intern*          June 2020 - Sept 2020

– Enabled racist content discovery within Amazon Prime Video for improved regional Prime Video compliance and more transparent content flagging for viewer discretion.

**Unity Technologies**

Monetization Team, *Data Scientist*                                      Sept 2017 - Aug 2019

– Implemented reinforcement-learning (RL) based and non-RL-based products to serve mobile game devs. Collaborated with product and engineering teams in fast-paced environment, resulting in >5% improvement on ad spend.

– Led experiments for in-app purchase recommendation and personalized ad placements in a fast-paced and cross-functional team. Designed A/B tests and provided actionable insights to guide product.

Analytics Team, *Data Scientist Intern*                                  May 2017 - Aug 2017

– Built player engagement models with game-play data from the Unity Engine to inform player churn prevention strategy. Used Pyspark and Tensorflow to implement multiple ML algorithms (e.g. logistic regressions, XGBoost, various deep neural network topologies) for deployment within a mobile games.

## SELECTED WORKS

[1]   Anaelia Ovalle, Ninareh Mehrabi, Palash Goyal, et al. "Are you talking to ['xem'] or ['x', 'em']? On Tokenization and Addressing Misgendering in LLMs with Pronoun Tokenization Parity". In: *Neurips 2023 Workshop Queer in AI*. **Top 15% of papers.** 2023.

[2]   Organizers Of Queer in AI, Anaelia Ovalle, Arjun Subramonian, et al. "Queer In AI: A Case Study in Community-Led Participatory AI". In: *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. **Best Paper Award.** 2023, pp. 1882–1895.

[3]   Anaelia Ovalle, Palash Goyal, Jwala Dhamala, et al. ""I'm fully who I am": Towards Centering Transgender and Non-Binary Voices to Measure Biases in Open Language Generation". In: *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 2023, pp. 1246–1266.

[4]   Anaelia Ovalle, Arjun Subramonian, Vagrant Gautam, et al. "Factoring the Matrix of Domination: A Critical Review and Reimagination of Intersectionality in AI Fairness". In: *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. 2023, pp. 496–511.

[5]   Nathan Dennler, Anaelia Ovalle, Ashwin Singh, et al. "Bound by the Bounty: Collaboratively Shaping Evaluation Processes for Queer AI Harms". In: *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. 2023, pp. 375–386.

[6]   Anaelia Ovalle, Davi Liang, and Alicia Boyd. "Should they? Mobile Biometrics and Technopolicy Meet Queer Community Considerations". In: *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. 2023, pp. 1–10.

[7]   Anaelia Ovalle, Sunipa Dev, Jieyu Zhao, et al. "Auditing Algorithmic Fairness in Machine Learning for Health with Severity-Based LOGAN". In: *2023 AAAI Health Intelligence Workshop*. 2023.

[8]   Sunipa Dev, Masoud Monajatipoor, Anaelia Ovalle, et al. "Harms of gender exclusivity and challenges in non-binary representation in language technologies". In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. 2021, pp. 1968–1994.

[9]   Chandra L Ford, Bita Amani, Nina T Harawa, et al. "Adequacy of existing surveillance systems to monitor racism, social stigma and COVID inequities: a detailed assessment and recommendations". In: *International Journal of Environmental Research and Public Health* 18.24 (2021), p. 13099.

# TRAININGS

| | |
|---|---|
| **Intersectionality Training Summer Intensive, Intersectionality Training Institute** | Aug 2022 |
| **Leadership Training, Unity Technologies** | Aug 2019 |

# INVITED TALKS

**Hub de América Latina y el Caribe, Investigación sobre Inteligencia Artificial** — Feb 2024
*Title: Soy Quien Soy: Queer Voices for Inclusive Language Technologies*

**Meta, FAIR Society & AI Team** — March 2023
*Title: Intersectionality as a Means for Interrogating Power Structures in AI*

**UC Berkeley, Data Science for Social Justice Workshop** — July 2023
*Title: Towards Inclusive Harm Evaluation Frameworks: The Case of Gender Minorities*

**University of Michigan, Gender Diversity in Robotics Seminar** — July 2023
*Title: Queer Realities Meet Harm Evaluation Frameworks for Generative AI*

# SERVICE

| | |
|---|---|
| **NAACL Student Workshop Co-Organizer** | 2024 |
| **EACL Reviewer** | 2023 |
| **NeurIPS Ethics Reviewer** | 2023 |
| **KDD EDI Day Co-Organizer** | 2023 |
| **ACL Trustworthy NLP Workshop Co-Organizer and Reviewer** | 2023, 2024 |

# COMMUNITY BUILDING

| | |
|---|---|
| **Queer in AI Core Organizer** | Present |
| **Deep Learning Indaba Mentorship Program Committee** | Present |
| **Deep Learning Indaba Mentor** | Present |
| **ACM FAccT CRAFT Workshop Organizer** | 2022 |

# HONORS AND AWARDS

| | |
|---|---|
| **2023 ACM FAccT Best Paper Award** | 2023 |
| **Ford Foundation Research Grant, Queer in AI, $50k** | 2023 |
| **NSF NRT Program Trainee, $22k** | 2021 |
| **Ford Fellowship Honorable Mention** | 2021 |
| **GEM Full Fellowship, $16k, 1 year** | 2020 |
| **Eugene V. Cota Robles Full Fellowship, $30k, 4 years** | 2019 |
| **The Anita Borg Systers Pass It On Award, $3k** | 2018 |
| **USF Magna Cum Laude** | 2017 |
| **USF Senior Leadership Award** | 2017 |
| **USF Changes the World From Here Academic Research Award** | 2017 |