

Anna-Maria Gueorguieva

agueorg@uw.edu

EDUCATION

UNIVERSITY OF WASHINGTON

Ph.D., Information Science

2024 – Present

UNIVERSITY OF CALIFORNIA, BERKELEY

B.A., Data Science (Honors); B.A., Legal Studies

May 2023

- **GPA:** 3.8 / 4.0
- **Awards:** Dean's List, Benjamin A. Gilman Scholarship, Fulbright English Teaching Award, Cal Alumni Association Leadership Award, Cyril and Virginia Smith Trust Scholarships, Ribbon of Excellence in Undergraduate Research
- **Honors Thesis:** *Artificial Intelligence & Regulatory Actors in the Sociotechnical System of the Department of Veterans Affairs* Advisor: Prof. Deirdre Mulligan, UC Berkeley School of Information

PUBLICATIONS

- **Gueorguieva, A-M.,** Caliskan, A., [*Identifying Features Associated with Bias Against 93 Stigmatized Groups in Language Models and Mitigation with Guardrail Models*](#), Proceedings of the 2026 AAAI Conference on Artificial Intelligence, AI Alignment Track (AAAI AIA)
- Wilson, K., Sim, M., **Gueorguieva, A-M.,** Caliskan, A., [*No Thoughts Just AI: Biased Hiring Recommendations Alter Human Decision Making and Limit Human Autonomy*](#), Proceedings of the 2025 AAAI/ACM Conference on AI, Ethics, and Society (AIES); [*reported on by the Washington Post*](#)
- Sun, Y., **Gueorguieva, A-M.,** Caliskan, A., [*VEAT Quantifies Implicit Associations in Text-to-Video Generator Sora and Reveals Challenges in Bias Mitigation*](#), Proceedings of 2nd Annual Conference International Association for Safe & Ethical AI (IASAI) 2026
- **Gueorguieva, A-M.,** King, J., Panidapu, A., Ho., D, *Privacy Without Remedy: An Assessment of Data Broker Compliance with California Privacy Law*, under submission
- **Gueorguieva, A-M.,** Abdalla, A., Sekhar, C., Zhong, I., Vuong, T., Yang, I., Saxon, M., Caliskan, A., *Location Persona-Assignment in Large Language Models Increases Implicit Bias and Reflects Geographic Political Biases*, under submission
- **Gueorguieva, A-M.,** Sun, Y., Tran, S., and Caliskan, A., *Text-to-Video Generator Sora Amplifies Gender and Racial Stereotypes and Perpetuates Sexualization in Short-Form Videos*, under submission

- Guldemann, P., Spiridonov, A., Staab, R., Jovanović, N., Vechev, V., **Gueorguieva, A-M.**, Balunović, M., Konstantinov, N., Bielik, P., Tsankov, P., Vechev, M. [COMPL-AI Framework: A Technical Interpretation and LLM Benchmarking Suite for the EU Artificial Intelligence Act](#) October 2024; [website for open access](#); reported on by [Reuters News](#) and [TechCrunch](#)
- **Gueorguieva, A-M.**, Rakhmetulla, G., Sabbir Arif A., [Enabling Input on Tiny/Headless Systems Using Morse Code](#), ArXives arXiv:2012.06708, December 2020

POLICY and PUBLIC COMMENT CONTRIBUTIONS

- Kallina, E., Safir A.H., Soulimani, A., Makina, A., **Gueorguieva A-M.**, Neumann, A., Borda, A., Bar, A., Konda, C., Vergara, C., Lucchini, F., Moses, M.D., Çağlar, O., and Bowies, W., [AI & EQUALITY Input Comment OHCHR's Call for input for EMRTD study "Artificial Intelligence, Cultural Rights, and the Right to Development"](#) <AI & Equality>, submitted to Office of the High Commissioner for Human Rights (OHCHR)'s on their thematic study on "Artificial Intelligence, Cultural Rights, and the Right to Development", January, 2026.
- **Gueorguieva, A-M.**, [The Effectiveness and Future of United States Executive Orders to Promote Human-Rights Based Artificial Intelligence](#), Re-visions of Now & Future: Commentary & Research on Contemporary AI and a Way Forward, Vol. 2, Chapter 2, <AI&Equality>, October 2024
- **Gueorguieva, A-M.**, Cruz, S., Kallina, E., Kraft-Buchman C., , [Human Rights, Transparency, and Stakeholder Involvement in AI Accountability: Public Comment Submitted to the United States National Telecommunications and Information Administration](#), <AI & Equality>, June 2023. (Cited by NTIA's AI Accountability Report)

RESEARCH EXPERIENCE

GRADUATE STUDENT RESEARCHER

Project: Evaluating Usage of AI in Election Campaigns and Policy Implications

Advisor: Prof. Nicolas Weber, University of Washington October 2025 – Present

- Building large-scale and open-source dataset through identification and scraping pipelines with computational methods and MCP servers
- Investigation of AI usage on text, image, and video in the context of political campaigns using open-source detection methods
- Empirically testing the impact of state AI disclosure and prohibition laws on the usage of AI in political campaigns

Project: Evaluating Data Broker Compliance to California Privacy Laws on Transparency and Consumer Privacy Rights

June 2025 – Present

Advisor: Prof. Dan Ho, Stanford Law School; Dr. Jennifer King, Stanford Institute for Human-Centered AI

- Manual review of privacy policies of all registered data brokers in California to determine compliance to California privacy laws on transparency
- Conducting qualitative coding on privacy rights request process to identify non-compliance and features increasing friction in the exercising of privacy rights
- Applying data analysis to advocate for user interfaces and tools that increase consumer agency and transparency

Project: Human-AI Bias Interaction in Decision Making in Multimodal LLMs

Advisor: Prof. Aylin Caliskan, University of Washington Sept 2024 – June 2025

- Applied taxonomy from psychology literature to ground investigation into model attitudes towards stigmatized groups
- Evaluated text and video generation models using embeddings, statistical analysis, language generation, and LLM personalization
- Conduct research on the performance of moderation and “guardian” open source models in identifying risk against stigmatized groups in user interactions
- Used high-performance cloud computing clusters to run scripts to call models, clean outputs, and organize data

UNDERGRADUATE RESEARCH ASSISTANT

Project: COMPL-AI Framework: A Technical Interpretation and LLM Benchmarking Suite for the EU Artificial Intelligence Act

Advisor: Prof. Nikola Konstantinov, Institute for Computer Science, Artificial Intelligence, and Technology (INSAIT) , Sofia University Aug 2023 – Oct 2024

- Legal project lead to LLM benchmarking project, which interprets the EU AI Act into necessary benchmarks to require AI deployers and developers to rigorously measure the impact of their model
- Bridged the gap between abstract legal requirements and technical compliance by organization of 144 pages of EU AI Act into 18 compliance categories upon which to use technical measurements

Project: Quantifying LLM Bias via Visibility in Ranking-Based Outputs

Advisor: Prof. Nikola Konstantinov, Institute for Computer Science, Artificial Intelligence, and Technology (INSAIT) , Sofia University Aug 2023 – Present

- Conducted experiment to identify how prompt engineering to request ranked outputs “jailbreaks” language model’s anti-bias measures
- Used Python to create scripts to prompt language models to rank individuals, collected data in CSVs and applied statistical measures of “exposure in rankings” to quantify bias in outputs
- Applied paired sample t-Test to verify that models had outputted biased information in a group fairness scenario

Project: Tools for Integrating Social and Ethical Understanding within Data Science Research

Advisor: Dr. Ari Edmundson, UC Berkeley Data Science May 2020 – Dec 2022

- Contributed to developing and improving the Data Ethos Lifecycle as a tool to promote ethical and social consciousness into data science research
- Conducted IRB certified ethnographic research to create a case study of how to integrate the Data Ethos Lifecycle into an actual data science research project
- Presented findings at Research Symposium for Center for Science, Technology, Medicine, and Society at UC Berkeley

Project: Operationalizing Responsible AI: Workshop Development;

Advisor: Prof. Umang Bhatt, University of Cambridge, Leverhulme Centre for the Future of Intelligence May – Oct 2022

- Worked in partnership with the Responsible AI Institute to define their Pillars of Responsible AI in an accessible way for technical professionals
- Developed materials for workshop aimed at AI researchers and practitioners to bridge gap between abstract AI ethical guidelines and actual implementation

Project: Evaluating Bias in Google Search Results

August 2021 – February 2022

Advisors: Prof. Deirdre Mulligan and Emma Lurie, UC Berkeley School of Information

- Created Python script for applying ranking bias measure to discern the source of bias in Google Search results
- Identified significant bias in search results presenting information on Proposition 22 in the 2020 California Ballot Propositions
- Presented findings at UC Berkeley Data Science Undergraduate Research Symposium

Project: Student Success Through Predictive Analytics

June 2020 – August 2021

Advisor: Jacqueline Change, UC Berkeley College of Letters & Sciences

- Applied feature selection and feature engineering to large datasets of UC Berkeley undergraduates to identify features that best predict which students are in need of Advisor assistance
- Used Python and Jupyter Notebooks to perform logistic regression, KNN, Naive Bayes, Decision Trees, and SVM to provide predictive analytics for students in need of assistance
- Performed interviews to translate stakeholder needs into actionable technological solutions and presented findings to stakeholders through data visualizations

PRESENTATIONS and TALKS

- *Identifying Features Associated with Bias Against 93 Stigmatized Groups in Language Models and Mitigation with Guardrail Models*, Poster at the 2026 AAAI Conference on Artificial Intelligence

- *Data Privacy Laws and Compliance*, Guest Lecture for University of Washington's INFO 351 "Information, Ethics, and Policy" Course, November 2025
- *Application of Psychology Taxonomy Reveals Differences in Language Models' Biases Towards 93 Stigmatized Groups*, Poster at Responsibility in AI Systems and Experiences, University of Washington, March 2025
<https://ischool.uw.edu/news/2025/03/raise-expo-highlights-research-responsible-ai>
- *Artificial Intelligence & Human Rights: Understanding Measures of Fairness*, Presentation at AfricAI Conference in Kigali, Rwanda, June 2023
- *Sociotechnical Systems of Artificial Intelligence and Regulations in Federal Government*, Poster and Oral Presentation at the Women in Data Science Conference, UC Berkeley, March 2023
- *Symposium on Surveillance and Education*, Panelist, Algorithmic Fairness and Opacity Group, UC Berkeley, April 2022
<https://afog.berkeley.edu/programs/symposium-on-surveillance-and-education>
- *Ethics Working Group: Putting Ethics for Data Science Into Practice*, Moderator for Academic Data Science Alliance, October 2020
- *Evaluating Bias in Google Search Results*, Poster and Oral Presentation at the Data Science Undergraduate Studies Discovery Research Program Showcase, UC Berkeley, December 2021
- *Anthropology and Algorithms in Administrative Systems*, Presentation at the Center for Science, Technology, Medicine, and Society Research Symposium, UC Berkeley, November 2021
- *Performing Participant-Observer Data Science Research*, Oral Presentation at the Data Science Undergraduate Studies Showcase, UC Berkeley, May 2021

PROFESSIONAL EXPERIENCE

<AI & Equality> Community Engagement Lead, November 2022 – January 2024
Women at the Table, Switzerland

- Established and grew community for the AI Equality & Human Rights Initiative at Women at the Table (NGO) through online and offline modes (view more [here](#))
- Creator and editor for annual AI & Human Rights publication and monthly newsletter
- Organized conference "Artificial Intelligence Frameworks & Implementations" in Geneva, Switzerland

- Team Lead, Human Context and Ethics of Data, UC Berkeley May 2021 – May 2023
- Managed a team of undergraduate and graduate students to collaborate with faculty and administration to promote ethics and justice within data science at UC Berkeley through human contextualized technical curriculum
 - Applied technical and social science skills to create curriculum about sentiment analysis, classification tasks, and data collection / gathering for UC Berkeley upper division courses
- Data Science Fellow, United States Department of Defense May – August 2021
- Worked at the United States Special Operations Command to improve civic tech infrastructure through creation of efficient database and website development
 - Developed database tools and web interface for the government to address safe and fair national security
- Curriculum Development Intern, UC Berkeley January – June 2020
- Developed Data Science curriculum with the Human Contexts and Ethics of Data team under the UC Berkeley Division for Computing, Data Science and Society
 - Researched and implemented concrete concepts of social implications into technical course curriculum at UC Berkeley

PROFESSIONAL SERVICE

- Reviewer, *Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence (AAAI)* 2025 – Present
- Reviewer, *International Association for Safe & Ethical Artificial Intelligence (IASAI)* 2025 – Present
- Reviewer, *ACM Conference on Artificial Intelligence, Ethics, and Society (AIES)* 2024 – Present
- Reviewer, *ACL Conference on Empirical Methods in Natural Language Processing (EMNLP)* 2025 – Present
- Editor, *Re-visions of Now & Future: Commentary & Research on Contemporary AI and a Way Forward*, Volume I, <AI & Equality>, <https://aiequalitytoolbox.com/wp-content/uploads/2023/12/Full-Publication.pdf> December 2023

TEACHING EXPERIENCE

Teaching Assistant, “Core Methods in Data Science”, September – December 2025
University of Washington

- Leading discussion sections on statistical methods in data science and relevant topics on data and social impact

Fulbright English Teacher, Kyustendil, Bulgaria September 2023 – July 2024

- Grant funded by competitive U.S. State Department Fulbright English Teaching Award
- Taught grades 8th - 12th at Dr. Petar Beron Foreign Language High School

Code Coach, The Coder School, Berkeley January 2022 – January 2023

- Taught children ages 6 to 15 the foundations of computing through Scratch, Python, and Java programming

Course Assistant, “Data, Prediction, and Law”, UC Berkeley January 2021 – May 2022

- Provided technical and curriculum support, in-class help, and hosted office hours

Writing Tutor, Student Learning Center, UC Berkeley January – June 2022

- Assisted students in developing their critical reading and writing skills through peer tutoring

Teacher, Opportunity Through Data Program, UC Berkeley August 2020 – May 2021

- Developed course curriculum and taught introductory Data Science course to high school students from underrepresented communities

VOLUNTEER and COMMUNITY INVOLVEMENT

Disabled Students’ Program Notetaker, UC Berkeley May 2020 – January 2023

- Take in-class notes to share with peers that require alternative learning

Information Intern, Impact Justice Organization, Washington D.C. January – May 2021

- Performed data collection and analysis on the state of restorative justice New Orleans Parish to contribute to reforms in criminal justice system

Renaming UC Berkeley Buildings Project, YWCA January – May 2020

- Coordinated data gathering efforts and analysis through archival research to provide evidence for the unnamings of buildings on campus that hold harmful legacies

Philanthropy Committee, Alpha Delta Pi Sorority, UC Berkeley January – May 2020

- Co-led and organized fundraiser events for the Ronald McDonald House Charities and facilitated communications with local businesses and donors

American Association of University Women February 2013 – Present

Mother/Daughter Science Camp, AAUW Merced Chapter

- Performed chemistry demos and assisted with hands-on experiments for 4th and 5th grade girls to inspire their interest in STEM subjects

SKILLS

- **Technical:** Python, Java, Javascript, SQL, HTML, CSS, NLP via NLTK, transformers, geo-spatial analysis, scikit-learn, NumPy, Pandas, Matplotlib, frequentist and bayesian statistical modeling, Django, machine learning models, feature engineering & extraction, high-performance cloud computing
- **Language:** Native in English and Bulgarian