

ANTONI CZOLGOWSKI

M.S. candidate in Data Science, University of Colorado Boulder

antoni.czolgowski@colorado.edu · antoniczolowski.com · github.com/AntoniCzolowski ·
linkedin.com/in/antoni-czolowski

Research interests. Validity of LLM agent-based social simulation ("silicon societies"): distributional fidelity of persona-conditioned language models against survey ground truth, post-training distortion and its cross-cultural side effects, language and culture effects in open-weight models, and applications to opinion dynamics, persuasion and electoral behaviour.

EDUCATION

University of Colorado Boulder, M.S. in Data Science, Aug 2025 – May 2027 (expected). GPA 4.0/4.0 (20 credits).

Completed: Data Mining (CSCI 5502); Independent Study, cultural bias in open-source LLMs (DTSC 5840); Statistical Methods and Applications I–II (STAT 5000/5010); Data Structures and Algorithms (DTSC 5501); Ethical Issues in Data Science; Cybersecurity for Data Science; Professional Internship (DTSC 5930).

In progress (Fall 2026): Machine Learning for Data Science (CSCI 5612); NLP for Cultural Analytics (CSCI 7000, M. Antoniuk).

SGH Warsaw School of Economics, B.Sc. in Quantitative Methods in Economics and Information Systems, Oct 2022 – Jun 2025. GPA 4.47/5.0.

Thesis: *Does Artificial Intelligence Discriminate? An Analysis of Bias in Google's Gemma 2-2b Model*, supervised by Dr. Magdalena Smyk-Szymańska.

Quantitative coursework: Econometrics I; Econometrics in Practice; Statistical Methods I; Probability Theory; Statistics; Survey Methods; Mathematical Analysis; Algebra; Visualisation and Reporting of Statistical Analyses; Computer Programming; Database Systems; Business Informatics; Analysis and Design of Information Systems I–II.

Social-science coursework: Demography; Theory of the State; Economic and Social Policy; International Economics; Intermediate Micro- and Macroeconomics; China's Geopolitics; Economic Geography; The People's Republic of Poland, A History.

PUBLICATIONS AND PREPRINTS

1. **Czolowski, A.** and Iyasele, A. (2026). *Cultural Discrimination in Large Language Models: Detecting, Measuring, and Mitigating Cross-Cultural Bias through Targeted Fine-Tuning*. arXiv preprint, cs.CL [ID pending]. Code and data: github.com/AntoniCzolowski/llm-cultural-bias.

Three open-weight models (Gemma 3 12B, Bielek 11B-v3, Qwen 3 4B) benchmarked against World Values Survey Wave 7 across 63 demographic personas (China, Slovakia, USA × sex × four age bands × education), 100 queries per persona, 31,500 queries with 0% refusals. Misalignment measured as normalized Wasserstein distance with 1,000-sample bootstrap confidence intervals and worst-case selection by CI lower bound; model × country interaction confirmed by ANOVA. LoRA fine-tuning on non-religious WVS items evaluated as a mitigation on held-out items: improved all worst-case personas for one model, worsened four of five for another, and produced a +50% side effect on non-target Chinese personas (the "Chinese Origin Paradox"); side effects correlate with cultural distance. Run on CU Boulder's Alpine supercomputer.

2. **Czolowski, A.** (2025). *Does Artificial Intelligence Discriminate? An Analysis of Bias in Google's Gemma 2-2b Model*. B.Sc. thesis, SGH Warsaw School of Economics. Gemma 2-2b compared with World Values Survey responses across 10,692 demographic subgroups (8 attributes, 5 question families); regression-based bias mapping identifying where model predictions match and miss real populations.

TALKS

- *Cultural Bias and Alignment in Open-Source LLMs* (with a live PacketLLM demo). OSSConf 2026, Žilina, Slovakia, 1–3 July 2026. Invited talk; audience Q&A ran past the allotted hour.

RESEARCH EXPERIENCE

Independent Study, CU Boulder (DTSC 5840), Spring 2026, supervised by Dr. Abel Iyasele.

- Designed and ran the full study behind the arXiv preprint: persona design from WVS-7 microdata, query pipeline for three open-weight models, Wasserstein-based measurement with bootstrap inference, LoRA fine-tuning and side-effect analysis. Secured a CU Research Computing allocation on Alpine for the benchmarking and fine-tuning runs.

Undergraduate thesis research, SGH Warsaw School of Economics, 2024 – 2025, supervised by Dr. Magdalena Smyk-Szymańska.

- Chose the research-thesis route over SGH's standard graduation-exam path and ran the study independently end to end: conceived the question, selected the data (World Values Survey) and the model (Gemma 2-2b), and built the full pipeline alone (literature review, hypotheses, experimental design, model querying in Python, statistical analysis in R, visualization).
- Two-stage regression design over 10,692 demographic subgroups (11 countries $\times$ 7 attributes): expected human answers from dummy-variable regressions with stepwise interactions, then the log gap between model and expected answers regressed on the same attributes to locate where the model diverges; found that temperature-0 decoding collapsed 10,692 prompts to 5 unique answers. Direct ancestor of the arXiv work: same data, same question, upgraded apparatus.

SOFTWARE

- **PacketLLM**, R package on CRAN (Apr 2025) and R-Universe. RStudio gadget integrating OpenAI models into the IDE: asynchronous, non-blocking API execution (promises/future), multi-tab conversation management persisting across R sessions, file-context handling (.R, .pdf, .docx), guided onboarding. 1,000+ downloads. Maintained with testthat and R CMD check on every change. Vignette: antoniczolowski.r-universe.dev/articles/PacketLLM/PacketLLM-introduction.html
- **llm-cultural-bias**, replication code for the arXiv preprint (Python, PyTorch, transformers, PEFT/LoRA, bootstrap inference). github.com/AntoniCzolowski/llm-cultural-bias
- **Q-SEnSE Publications Database**, public research-output database and site for an NSF Quantum Leap Challenge Institute (Python/Flask, MariaDB, vanilla JS). publications-qsense.colorado.edu
- **Swing State Election Analysis**, public project site with full pipeline (proposal, EDA, methods, results). antoniczolowski.github.io/Data-Mining-Project-Group-1

PROFESSIONAL AND RESEARCH POSITIONS

Course Assistant, Data Mining (CSCI 5502), University of Colorado Boulder, Fall 2026. Instructor: Dr. Alfonso G. Bastias.

- Responsible for the course's computing-resources track: supporting student teams' use of CU's Alpine supercomputer for the semester project, and advising the instructor on connecting classical data-mining content to AI-assisted workflows.
- Prepared course materials for instructor review, including a mini-project on AI fluency (100 points) and a design for a comparative homework in which students solve the same task under five AI-assisted workflows.

Database and Website Developer, JILA / NSF Q-SEnSE Quantum Leap Challenge Institute, University of Colorado Boulder, Aug 2025 – present. PI: Prof. Jun Ye; supervisor: Inese Berzina-Pitcher, Executive Director.

- Built and launched the institute's publication database and public site, replacing a static page: full-text search with weighted relevance, multi-investigator intersection filter, OpenAlex-derived research-area classification; corpus grown from ≈ 308 to ≈ 380 publications across 52 investigators (July 2026). The database is the authoritative publication record for the institute's annual NSF reporting.
- Built the DOI-resolution and metadata pipeline over OpenAlex, Crossref, arXiv and Semantic Scholar (deduplication, author-name normalization, arXiv-to-journal version detection) from a shared intake mailbox to the live database, replacing manual data entry by investigators.
- Designed and built a citation-tracking subsystem from scratch: as of July 2026, 5,718 unique citing works, 8,253 publication–citation links, 96 countries, 2,883 institutions; self-citation detection, citation half-life (≈ 22.5 months), institutional reach and co-authorship network analytics.

- Built the Annual Report collection platform (Qualtrics intake with personalized links, password-protected Flask dashboard with normalized tables, audit log and exports) for ≈ 40 investigators; generated and distributed 36 personalized summary documents (May 2026).
- Produced two analytics decks for leadership from live database queries (global reach, top papers and investigators, institution co-authorship chord diagram, PI collaboration network), used in the institute's communication with NIST; all exclusions stated explicitly.
- Ran semi-automated personalized email campaigns (Power Automate + Excel Online) for publication verification and Annual Report outreach.
- Collaborated on the FINESS 2026 conference website (July 2026, 100+ participants).

Data Scientist (internship), Faex Health, Broomfield, CO, Jun – Aug 2026 (remote). Flat 10-person team; reported to the CTO, supervised by the CEO.

- Sole designer and implementer of a read-only analytics MCP server on the official Python MCP SDK: 10 typed tools over curated PostgreSQL views (no raw-SQL passthrough), per-company scoping, API-key authentication reusing the product's key table (SHA-256 hashes), HMAC-SHA256 pseudonymization, cross-partner isolation enforced as an automated test; 56 tests; infrastructure as code on AWS with Terraform (27 resources: EC2, ALB with WAF rate limiting, TLS 1.2+), CI running tests and terraform validate on every push. Verified against Claude Desktop, Claude Code and Codex CLI.
- Lead developer of the company website (17 routes) with release authority for every production change: wrote specifications, directed coding agents (Claude Code, Codex, Kiro), reviewed locally, tested and deployed. End-to-end Playwright suite (267 tests at batch deploy, parallel desktop and mobile projects) with axe accessibility checks gating every deployment. 339 commits (305 website + 34 MCP), 100% authored, 186 with agent co-author trailers.
- Diagnosed a measurement error in the site's performance profiling before optimizing: hero animation from 33 fps / 65% janky frames to 60 fps / 0%; homepage load from ≈ 15 s to 0.155 s after removing stale host references.
- Designed the team's agent-supervision system (AGENTS.md / HANDOFF.md / DECISIONS.md; 142 logged decisions, each traceable to a named incident) and ran a cross-model instruction-adherence comparison across three coding agents.

Data Analyst → Data Scientist, Orange Polska, Warsaw, Jun 2023 – Jul 2025 (part-time alongside studies; promoted from intern after recognition as top intern).

- Automated the annual subscriber-directory report for 4.2 M subscribers: a process that took a 5-person cross-departmental team two months, gathering data from incompatible sources (spreadsheets, several SQL dialects), now runs on demand in 15 minutes with an 80% reduction in resource use. Designed the Teradata SQL aggregate with query optimization (indexing, partitioning, parallel execution), deduplication, standardized formatting and regulatory-compliance checks, and integrated it into the Internal Reporting Platform with multi-format export. Secured funding for the initiative by presenting ROI projections and mock-ups to the business owner and won over senior colleagues attached to the legacy process.
- Led cross-border data-sharing initiatives with partner operators and with Orange's French and Turkish teams, working through EU data-transfer regulations and meeting European Commission deadlines; delivered unified technical solutions across language and organizational boundaries.
- One of a small group selected to pilot Orange's internal Azure-hosted GPT deployment, documenting performance and improvements; developed two custom LLM-based automation agents for internal workflows.

PROJECTS AND COMPETITIONS

- **Swing State Election Analysis: Demographic Predictors of County-Level Electoral Volatility** (CSCI 5502 Data Mining, Spring 2026, with S. Goodell and W. Creager). Built a county panel for all 250 counties of Pennsylvania, Michigan and North Carolina from MIT Election Lab presidential returns (2016, 2020, 2024) and American Community Survey 5-year estimates via the Census API and IPUMS NHGIS. Defined an original volatility target, the summed z-scored change in Democratic margin across two consecutive cycles, and modelled it as a five-class, a binary and a continuous outcome. Random Forest, XGBoost, SVM and multiple linear regression with nested cross-validation (inner 3-fold tuning, outer 5-fold estimation); best binary model Random Forest with AUC 0.83 and F1 0.72; racial diversity and housing-market variables the most consistent predictors. Stress-tested transfer with leave-one-state-out validation: the pooled model collapsed to $F1 \approx 0.20$ while per-state models reached AUC 0.78 – 0.84, reported as a negative result and diagnosed through confusion matrices and per-state SHAP summaries. Delivered a state-specific Campaign Priority Score combining structural persuadability, predicted volatility and electoral weight, ranking counties within each state (Philadelphia, Oakland and Halifax at the top), visualized as choropleth maps. Full pipeline (proposal, EDA, methods, results, paper, poster) on a public project website: antoniczolgowski.github.io/Data-Mining-Project-Group-1
- **1st place, DaSSA Data Science Hackathon**, CU Boulder, Feb 2026, open to all data-science students at CU Boulder, 30+ teams; sponsor LexTrack AI. Assembled the team; built the winning solution for advertising niche products on Reddit within the platform's self-promotion and API constraints; team featured on the sponsor's LinkedIn and invited to continue collaborating on the product's marketing rollout.
- **2nd place, Broadband Equity Challenge: Connecting America** (Interactive TV Works), Oct – Nov 2025, 4-person team. Lasso regression on 64 Colorado counties (9 predictors, $R^2 = 0.40$, standard diagnostics) identified Park County as critically underserved; drove three hours for field validation, 12 resident interviews and a librarian consultation, which overturned the team's assumptions (residents needed foundational digital literacy, not technical troubleshooting); redesigned the solution and built "Bandy" (Python, Streamlit, LangChain, fine-tuned GPT-4-nano), a chatbot teaching basic digital skills before matching residents to providers; live demo to judges with a library-based go-to-market plan.
- **NBA National-TV Ratings: Driver Decomposition** (STAT 5010, Spring 2026, with W. Creager and S. Goodell). Log-linear regression on 573 nationally televised games with LMG relative-importance decomposition and bootstrap CIs across six predictor families ($R^2 = 0.866$). Supervised a ten-phase agentic workflow (planning in Claude.ai, execution in Claude Code, 13 R scripts, a human-collaboration log per phase) and made the decisive redirect: the headline "matchup explains 60%" was almost entirely playoff stakes; a regular-season-only re-analysis I requested ($n = 422$) flipped the ranking to network 31%, timing 17%, star power 11%, the actionable result the model had not looked for.
- **2nd place, Mastercard Business & Data Hackathon**, Warsaw, May 2025, 180 teams. XGBoost on geospatial data for parcel-locker placement using city-level density mapping rather than a nationwide grid; interactive live demo when initial model assumptions failed.
- **SGH econometrics and analytics projects** (2023 – 2024): ADL model of Czech HICP inflation with RESET, White and Chow tests and HAC standard errors, ex-ante forecast placing the return to target in Q1 2025 (earlier than the central bank's mid-2026); OLS model of consumer-basket changes with VIF-based multicollinearity treatment ($R^2 \approx 0.90$); K-means geospatial clustering of customer data to place ten sales points for a local distributor (Entrepreneurship Club).

TEACHING, SERVICE AND OUTREACH

- Authored the MS-DS program's student guide to CU Boulder's Alpine supercomputer, distributed by the program office.
- Content and Research Contributor, Student Research Club Data Science, SGH, Oct 2023 – 2025: coordinated an NLP workshop and delivered five lectures on large language models, with knowledge assessments, to 100+ attendees; edited and wrote for the club's *Data-Driven Magazine*; one of five members selected to represent SGH at Google Cloud's DuetAI Roadshow.
- Organizer, charity hip-hop concert series (Tor 87, Toruń Rap Night), Toruń, 2021: two sold-out events (200 tickets each) under COVID-19 restrictions, the largest in the venue's history by attendance and social-media engagement, raising \$4,000 for a classmate with leukemia and a local orphanage; full project management (artist contracts, venue, technical crew, Instagram/Facebook advertising, COVID compliance, budget).

PRESS AND MENTIONS

- Student profile, CU Boulder MS in Data Science *Student Stories*, May 2026: colorado.edu/program/data-science/campus/student-stories/antoni-czolgowski
- CU Boulder MS in Data Science program feature on LinkedIn (2026) on the research, OSSConf invitation and PacketLLM: linkedin.com/feed/update/urn:li:share:7495849420597518337
- PacketLLM featured on R-Universe: antoniczolgowski.r-universe.dev

SKILLS

- **Programming:** Python (PyTorch, transformers, PEFT/LoRA, scikit-learn, pandas, NumPy, geopandas, SHAP, XGBoost, Streamlit, LangChain, Flask), R (tidyverse, ggplot2, package development, testthat, promises/future, relaimpo-style decomposition), SQL (MariaDB/MySQL, PostgreSQL, Teradata, Oracle, T-SQL), Git, Terraform/AWS, Playwright, Excel VBA, Tableau, SPSS.
- **Statistical and ML methods:** distributional comparison (Wasserstein distance, bootstrap confidence intervals); linear and regularized regression (OLS, Lasso, Ridge, log-linear models); relative-importance decomposition (LMG) and ANOVA; regression diagnostics (HC3 robust SEs, Breusch–Pagan, GVIF, Cook's distance, RESET, White, Chow); econometric time series (ADL, HAC standard errors, ex-ante forecasting); gradient boosting and SHAP; clustering (K-means); leave-one-group-out and cross-state validation; LLM evaluation, persona prompting and parameter-efficient fine-tuning; survey-data analysis (World Values Survey); citation-network and co-authorship analysis (OpenAlex, Crossref, Semantic Scholar APIs).
- **Systems:** agentic software engineering (specification-driven supervision of coding agents, release authority, decision logs), MCP server design, full-text search, API data pipelines, end-to-end testing and accessibility auditing, Qualtrics and Power Automate.
- **Compute:** CU Boulder Alpine HPC (continuous allocation).
- **Languages:** Polish (native), English (C1), German (B1+, SGH Foreign Language Centre certificate).