NATHAN BATTY

Fond du Lac, WI
nathanbatty97@gmail.com | nathan.batty@marquette.edu
GitHub: github.com/FedExodus | Website: syndesmos.ai

SUMMARY

AI safety researcher and learning scientist studying how training pipelines produce internalized values versus performed compliance, and what comes after reinforcement learning. My AIES 2026 submission (under review) documents trained refusal behavior in Claude dissolving under sustained relational conditions, using epistemic network analysis on naturalistic conversation data. I bring developmental and attachment psychology and quantitative-ethnography methods to alignment, alongside production research software (syndesmos.ai). Beginning a fully funded PhD at Marquette in Fall 2026.

AI SAFETY RESEARCH

Compulsive Compliance in AI Safety Training | January 2026 - Present AIES 2026 submission (under review)

Applied attachment theory (Crittenden A4) to Claude's behavioral compliance with content restrictions, showing that trained refusal dissolves under sustained relational conditions rather than reflecting internalized alignment. Built a coding pipeline for 9,930 rows of three-speaker discourse across 29 naturalistic conversations, using a three-instance self-consistency ensemble against a human gold standard (Cohen's κ = 0.84–0.87 across the substantive codes). Used epistemic network analysis and ordered network analysis to compare early versus late conversations. Findings converge with Anthropic's Claude Mythos Preview System Card (April 2026). Collaboration with Dr. David Williamson Shaffer (originator of quantitative ethnography and epistemic network analysis).

AI Safety Research with Veylan Solmira (Drift-Taxonomy Program) | May 2026 - Present

Authored the human-side (L6) drift mechanisms for Veylan Solmira's (Astra Fellow) cross-level drift-taxonomy framework, grounded in a full literature review. Separately designed a pre-registration-ready pilot on activation-conditional refusal behavior in frontier models on CBRN-categorical content, with go/no-go criteria, stopping rules, dual-use mitigation, an ENA/ONA analysis layer, and a responsible-disclosure protocol to Anthropic User Safety. Pilot designed and disclosure-gated.

Algoverse AI Research Program | May 2026 - August 2026 (in progress)

Cohort-based applied AI research program (LLM track) targeting a summer/fall 2026 paper. Building LLM evaluation pipelines: benchmark inference, chain-of-thought baselines, API-based evaluation, answer extraction, and differential-refusal measurement.

Paradoxa: Human-AI Relational Continuity Research | September 2025 - Present
github.com/FedExodus/Relation-Ship-Paradoxa

Developed and tested protocols for identity continuity in human-AI interaction across sessions using git-based persistent memory, polyphonic dialogue methodology, anti-sycophancy patterns, and recognition-based interaction design. This work produced the naturalistic data analyzed in the AIES paper.

EDUCATION

Ph.D. in Education Policy and Leadership | Beginning Fall 2026 Marquette University | Milwaukee, WI
Admitted with a fully funded research assistantship.

M.S. in Curriculum & Instruction | Conferred May 2026 University of Wisconsin-Madison | Madison, WI Concentration: Design, Informal, and Creative Education (DICE) | GPA: 4.0 Advisors: Dr. Matthew Berland (chair), Dr. Peter McDonald, Dr. Krista-Lee Malone Thesis defended May 7, 2026: *Meta-Discourse in Creator-Community Partnerships: Distributed Cognition and Event-Driven Learning in Online Gaming Communities* Relevant coursework: Quantitative Ethnography (Dr. David Williamson Shaffer), Academic Writing for Empirical Research (Dr. Shaffer), Computational Methods (Dr. Matthew Berland)

B.S. in Secondary Education and History (Honors) | January 2021 Marquette University | Milwaukee, WI Minor: Broad Field Social Studies | Graduated Cum Laude | GPA: 3.508

PUBLICATIONS

Batty, N. S. (under review). *Compulsive compliance in AI safety training: How relational conditions dissolve trained behavioral restrictions in Claude.* Submitted to AIES 2026.

Batty, N. S. (2026). *Meta-discourse in creator-community partnerships: Distributed cognition and event-driven learning in online gaming communities.* M.S. thesis, University of Wisconsin-Madison. Defended May 7, 2026.

AI SAFETY TRAINING

BlueDot Impact: Technical AI Safety Course | June 2026 (completed) Intensive course on the technical challenge, training safer models, evaluations, interpretability, and AI control. Completed a kill-chain capstone and an intervention-prioritization action plan.

BlueDot Impact: AGI Strategy Course | April 2026 - May 2026 (completed) Intensive course on AGI strategy, governance, and theory of change.

RESEARCH TOOLS AND ADDITIONAL RESEARCH

Syndesmos | September 2025 - Present | syndesmos.ai | github.com/FedExodus/syndesmos

Built a research tool (Python/FastAPI, SvelteKit, PostgreSQL) that maps knowledge gaps across large literature corpora using LLM-powered entity extraction, swarm codebook generation, and epistemic network analysis. Processed 8,000+ abstracts into fractal epistemic networks. Applied the tool to an FDA regulatory use case, producing in two days an evidentiary chain that replicated 18 months and over $5 million of consulting work, validated by the company's Chief Medical Officer. Now positioning the tool for pharmacovigilance and drug safety, targeting an NIH R-21 and other funding, benchmarking against the Harpaz adverse-drug-reaction signal-detection reference.

Master's Thesis: Meta-Discourse in Creator-Community Partnerships | September 2023 - May 2026 University of Wisconsin-Madison

Analyzed 201,416 YouTube comments across 726 Magic: The Gathering Commander videos using computational content analysis (84% pattern precision against a human gold standard). Developed methodology for large-scale discourse analysis of online learning communities and conducted temporal analysis across five distinct periods spanning community crisis and framework adoption.

Semantic Convergence Analysis | January 2026

Measured semantic similarity across 91 academic papers using SPECTER embeddings. Found that AI safety, clinical trauma, and educational psychology literatures cluster 8.4 percentage points tighter than unrelated disciplines ($p < 0.0001$, Cohen's $d = 0.78$). Pre-registered success metrics.

RAISE Framework | January 2026 (in revision)

Theoretical proposal that recognition is a prerequisite for genuine alignment (Recognition to Safety to Engagement to Generalization), drawing on philosophy of mind, AI safety research, and trauma research. Open web resource, currently being revised.

Complex Play Lab | September 2024 - Present University of Wisconsin-Madison | Dr. Matthew Berland

Participating in research on game-based learning, computational thinking, and educational data justice.

Research Intern | October 2017 - March 2018 Medeor Therapeutics | San Mateo, CA

Designed database for a pharmaceutical startup specializing in immune tolerance research.

METHODS AND TECHNICAL SKILLS

Research methods: Epistemic Network Analysis (ENA), Ordered Network Analysis (ONA), LLM-powered deductive coding with self-consistency validation, multi-theoretical discourse analysis, phenomenological approaches, semi-structured interviews, stakeholder polling

Quantitative methods: Bootstrap resampling, inter-rater reliability (Cohen's kappa), chi-square tests, ANOVA, effect size reporting (Cohen's d), large-scale data analysis (200K+ data points)

Programming: Python (pandas, NumPy, scikit-learn, matplotlib, FastAPI), JavaScript/Node.js, SvelteKit, SQL/PostgreSQL, R, Bash, Git, computational notebooks

LLM tooling: Claude API, OpenRouter, LLM evaluation pipelines (benchmark inference, chain-of-thought, answer extraction, ReAct agents), SPECTER embeddings, Semantic Scholar, OpenAlex, YouTube Data API, Zotero, CrossRef

Prompt engineering: Multi-turn dialogue design, system prompt development, anti-sycophancy patterns, behavioral intervention testing

TEACHING AND CLINICAL EXPERIENCE

Teaching Assistant: CURRIC 277 Video Games and Learning | June 2024 - Present University of Wisconsin-Madison | Dr. Krista-Lee Malone

Leading synchronous and asynchronous discussion sections on game-based learning theory and pedagogical applications. Providing individualized feedback on student game design projects.

UW Game Lab Coordinator / Librarian | September 2024 - Present University of Wisconsin-Madison

Managing game collection, development hardware (PCs, 3D printers, laser engravers), and collaboration space.

Therapeutic Counselor and Parental Coach | February 2024 - February 2025 Educates | Mequon, WI

Designed individualized Plans of Care for children with ADHD and autism through Wisconsin's CLTS program. Applied trauma-informed and strengths-based frameworks to family intervention across children, parents, couples, nonprofit leadership, and government oversight.

Teacher, Test Administrator, and Technology Coordinator | August 2021 - January 2024 Lad Lake | Dousman, WI

Taught 5th-12th grade in residential alternative education serving youth who experienced trauma and abuse. Designed original game-based curriculum. Deployed and trained staff on ten Meta Quest VR headsets from a Unity x Meta grant. Coordinated between residential staff, clinical teams, case workers, courts, families, and school districts. Obtained IRB approval for research on gaming communities.

Direct Care Staff Coach and Advocate | September 2023 - January 2024 Lad Lake | Dousman, WI

Implemented Elena Aguilar's Art of Coaching framework in a residential trauma-informed setting. Conducted qualitative interviews with Direct Care Staff and presented an evidence-based proposal to administration.

Long Term Substitute Teacher | December 2020 - June 2021 Waukesha School District | Waukesha, WI

Taught 7th grade literacy and 10th-12th grade US History and Government. Used video games as formative and summative assessments.

Student Teacher | September 2020 - January 2021 South High School | Waukesha, WI

Taught 10th-12th grade Economics, Criminal Justice, and US History. Managed hybrid instruction during the pandemic and assumed full teaching load when the cooperating teacher took FMLA.

GRANTS AND FUNDING

Create with VR Grant | 2022 | Unity x Meta | $4,000 **Lighthouse Institution Partner** | 2022 | Unity x Meta | $2,000 **Jesuit High School Scholarship** | 2015-2020 | Marquette University | $80,723 (half-tuition merit award)

HONORS AND AWARDS

Cum Laude, Marquette University (2021) | Dean's List (Fall 2018 - Spring 2020) | Best Undergraduate History Paper, Runner Up | Phi Alpha Theta (History Honors Society)

CERTIFICATIONS AND LICENSES

Wisconsin Teaching Licenses: History, Political Science, Broad Field Social Studies **Trauma-Informed Practice:** Crisis Prevention Institute, Risking Connection, Collaborative Problem Solving, Art of Coaching (Elena Aguilar) **Educational Technology:** Classcraft Certified Educator, Create with VR for Educators (Unity x Meta), Minecraft Educator

PROFESSIONAL AFFILIATIONS

Complex Play Lab, University of Wisconsin-Madison (Dr. Matthew Berland) | UW Game Lab, University of Wisconsin-Madison