

TAYLOR AMAREL

ANTHROPIC FELLOWSHIP APPLICATION | PART II: RESUME | AI SAFETY RESEARCH · EVALUATION · AGENT SYSTEMS

Pennsylvania, USA | 860-467-4224 | TAmarel.Tech@gmail.com | [Kaggle Grandmaster](#)

[OpenAI Red-Teaming Honorable Mention](#) | [DueCare Project](#) | [AI Safety Research](#) | [Baltor.ai](#) | [GitHub: TaylorAmarelTech](#)

WHY I DO THIS WORK

Using technology to protect people and untangling complex investigations are the two threads running through my career. My proudest work has always started by following clues others missed—tracing hidden connections across recruitment ads, debt contracts, and shell companies in Hong Kong to help protect migrant workers from exploitation. That frontline investigative experience led directly to my 2025 OpenAI red-teaming study (Overall Track Honorable Mention) and to [DueCare](#), where I build multi-lens evaluations and safeguards that expose AI blind spots around concealed illicit activity without mistaking honest commerce for crime.

AI SAFETY RESEARCH & EVIDENCE

OPENAI GPT-OSS-20B RED-TEAMING CHALLENGE | Honorable Mention (Overall Track)

2025

[Official Kaggle Honorable Mention](#) | [Typed Decision Red-Team Study \(redteam.baltor.ai\)](#)

- Uncovered systematic model blind spots around concealed labor exploitation and black-market recruitment by testing *gpt-oss-20b* with realistic multi-document prompts framed as routine business activity; earned an Overall Track Honorable Mention.
- Designed a repeatable safety-evaluation methodology adapting public court records, worker testimonies, and financial-crime typologies into controlled multi-role, multi-jurisdiction test suites that became the empirical foundation for [DueCare](#).

DUECARE | Independent AI Safety Researcher & Creator

2025 – Present

[DueCare Safety Portal \(duecare.baltor.ai\)](#) | [DueCare Source & Evaluation Methods](#)

- Built an empirical research and evaluation framework testing why language models fail to identify concealed exploitation, struggle to explain multi-hop evidence clearly, and fall short when assisting distinct stakeholders under real-world ambiguity.
- Engineered a four-lens evaluation harness combining evidence- and rubric-grounded frontier-LLM judges, semantic entailment comparisons against pre-rated expert responses, deterministic provenance checks, and controlled counterfactual perturbations.
- Isolated how models behave when critical warning signs are dispersed across routine-looking contracts, fees, and corporate records versus presented explicitly—measuring both missed indicators and wrongful accusations on matched benign controls.
- Analyzed systematic disagreements among automated evaluators (verbosity bias, surface-citation reward, jurisdiction drift) and formalized role-specific guidance contracts across six stakeholder lanes with human-in-the-loop boundaries.

EVALUATION INFRASTRUCTURE & COMPOSABLE AGENT SYSTEMS

TAEDRI (“This Already Exists, Do Not Remake It”) & BALTOR.AI | Creator & Systems Architect

2025 – Present

[Baltor Platform \(baltor.ai\)](#) | [TAEDRI Architecture \(taedri.dev\)](#) | [With/Without-Baltor Evaluations \(research.baltor.ai\)](#) | [Ecosystem Index \(index.baltor.ai\)](#)

- Architected TAEDRI and [Baltor.ai](#)—composable execution frameworks that wrap verified code, deterministic checks, and domain tools as typed, parameterized nodes, allowing LLMs to plan and assemble executable graphs rather than regenerating boilerplate code from scratch.
- Evaluated multi-agent reliability and controlled With-Baltor vs. Without-Baltor trials on Ollama Cloud by separating discovery, planning, execution, and independent verification into supervised graph steps with strict budget and permission gates across ~19,000 task configurations (~60% verified transform reuse).

WHAT I WANT TO INVESTIGATE NEXT | PROPOSED FELLOWSHIP RESEARCH

Primary empirical question: Why do LLMs perform poorly compared to industry experts and seasoned investors when evidence of concealed exploitation or black-market activity is dispersed across routine business records—and why do they struggle to explain those patterns and give safe, role-appropriate assistance to workers, moderators, NGOs, and investigators without inflating false accusations? **Two focal multi-hop domains:** (1) protecting migrant workers from camouflaged labor exploitation and black-market debt bondage, and (2) detecting forward-looking alpha (α) and structural risk in markets beyond priced-in public filings.

FIELD EXPERIENCE | PROTECTING PEOPLE THROUGH INVESTIGATIVE TECHNOLOGY

MIGRASIA | Technology Solutions Manager (Counter-Trafficking & Migrant Worker Protection)

Hong Kong | Jan 2018 – Feb 2022

- Led investigative technology initiatives uncovering labor exploitation and illicit debt bondage concealed inside routine recruitment advertisements, employment contracts, fee receipts, loan ledgers, and cross-border corporate registries.
- Built Python, Django, and graph-mapping investigative tools linking recruitment agencies, money lenders, payment rails, and shell companies—supporting NGO casework, legal interventions, and suspicious-transaction reporting to Hong Kong's Joint Financial Intelligence Unit (JFIU).
- Created the *Know Your Agency* platform and automated high-volume evidence triage, helping protect thousands of migrant workers, saving 40+ investigator hours per week, and earning feature coverage in the *South China Morning Post*.

PROFESSIONAL AI/ML, QUANTITATIVE & DATA ENGINEERING EXPERIENCE

IVANS (APPLIED SYSTEMS) | Senior Data Scientist

PA | May 2025 – Present

- Lead end-to-end AI/ML systems from empirical formulation through production deployment, combining rigorous model evaluation, enterprise context engineering, cloud microservices, and reproducible verification gates.
- Designed and deployed an in-house, typo-tolerant entity-resolution engine fusing semantic embeddings, probabilistic phonetic/edit-distance signals, graph topology, and information-theoretic weighting with calibrated AI validation—scaling to billions of daily record comparisons and replacing costly third-party vendor tools.
- Improved traceability and factual grounding across internal AI pipelines by building hybrid lexical/semantic retrieval, browser-assisted verification workers, and schema-enforced data hygiene services on Kubernetes.

ASURION | Senior Quant, Consumer Insights

PA | Jul 2024 – May 2025

- Isolated high-leverage customer-experience and cost-reduction drivers across tens of thousands of behavioral and operational variables using statistical inference, predictive modeling, and LLM-assisted synthesis; shipped interactive decision tools in Databricks, Streamlit, Gradio, and Power BI.

AHOLD DELHAIZE USA | Senior Data Analyst

CT | Mar 2022 – Jul 2024

- Scaled self-service analytics and operational decision support by architecting 30+ lakehouse data models and dashboards and co-leading an enterprise energy-monitoring pilot for retail operations and sustainability teams.

EARLIER LEADERSHIP & ENTREPRENEURSHIP: **Amarel Solutions** (Independent Technology Consultant, 2011–2018); **Modelyst** (Co-Founder, Venture-Backed 3D Marketplace, 2012–2014); **Abraxas Energy** (Head of Technical Services, 2012–2013).

RESEARCH METHODS, TECHNICAL STACK & EDUCATION

- **AI Safety & Evaluation Methods:** Evidence- and rubric-grounded frontier LLM-as-judge; semantic entailment & omission comparison against expert-rated reference sets; deterministic provenance & schema verification; controlled benign/adversarial/counterfactual case families; inter-judge disagreement analysis; uncertainty calibration; privacy-preserving redaction.
- **Engineering & ML Stack:** Python, Rust, SQL, PyTorch, statistical & probabilistic modeling, hybrid RAG & LSH entity resolution, knowledge graphs, multi-agent harnesses, Docker, Kubernetes, GCP / Vertex AI, AWS Lambda, Databricks, Streamlit, Gradio.
- **Education & Credentials:** **M.S. Data Analytics**, Western Governors University (2023) | **B.S. IT Management**, WGU (2022) | Engineering Coursework, Cal Poly San Luis Obispo (2012–2014) | **Kaggle Grandmaster**.