David F. Dumaresq

dfdumaresq@gmail.com | GitHub | LinkedIn

Professional Summary

Senior Software Engineer & AI Safety Researcher

"Builder-Philosopher" transitioning 30+ years of software engineering experience into technical AI safety and alignment. Combining deep systems thinking with a historical perspective on AI development (from 1980s expert systems to modern LLMs). Proven track record of building complex, high-reliability systems and currently developing technical tools for AI agent identity, provenance, and steerability. Uniquely positioned to bridge the gap between theoretical alignment concepts and robust, production-grade engineering.

Technical AI Safety Projects

AI Agent Fingerprinting System | *Lead Developer* | GitHub

- **Problem**: Addressing the lack of accountability and the risk of Sybil attacks in agentic AI ecosystems.
- **Solution**: Architected and built a blockchain-based registry (Ethereum/Solidity) for immutable AI agent identity and provenance.
- **Tech Stack**: Solidity, Web3 implementation, Keccak256 hashing for cryptographic identity verification.
- **Impact**: valid proof-of-concept for "Access to Compute" governance mechanisms and liability tracking.

ChatMoi Hybrid Interface | *Creator* | GitHub

- **Problem**: Difficulty in strictly comparing and evaluating behavioral differences between closed (OpenAI) and open (Ollama) models.
- **Solution**: Developed a unified interface for side-by-side model steering and evaluation, capable of seamless context switching.
- **Safety Feature**: Implemented "Safety Expert" personality presets to test model adherence to constitutional constraints under varied system prompts.

Virtue-AI | *Creator* | GitHub

- **Focus**: Model Fine-tuning & Ethical Reasoning.
- **Project**: Fine-tuned `Llama-3.2-1B-Instruct` to actively recognize and refuse unethical

premises, promoting virtues like honesty and responsibility rather than neutral engagement.

- **Technical**: Overcame specific stability challenges for training on Apple Silicon (MPS), optimizing for mixed-precision constraints and gradient clipping.
- **Research Insight**: Demonstrated that architectural base (Instruct vs Completion) is a harder constraint than data quality when attempting to induce safety behaviors via fine-tuning.

Akkadian Cuneiform Translation | *Kaggle Competitor* | Kaggle
- **Focus**: Neural Machine Translation for Ancient Languages (Low-Resource).
- **Project**: Developing a translation pipeline for Old Assyrian cuneiform texts, utilizing fine-tuned LLMs/SLMs to decipher fragmented archival tablets.
- **Technical**: Implementing custom tokenization strategies and LoRA fine-tuning to handle non-Latin scripts and high-context archaeological data.
- **Relevance**: Applied research in "AI for Heritage" and robustness in low-resource settings.

Novel Analysis with Gemini 1.5 | *Kaggle Competitor* | Kaggle
- **Project**: Built a "Lightning-Fast Literary Analysis System" leveraging Gemini 1.5 Flash's 1M token context window to analyze entire novels in a single pass.
- **Tech**: Python, Google Gemini API, Long-Context Prompting (128k+ tokens).
- **Relevance**: Demonstrates capability to build evaluation pipelines for long-context conceptual coherence in LLMs—a key frontier in safety monitoring for RAG and agentic systems.

Skills

- **AI Safety & Alignment**: Agent Provenance, Model Steerability/Evaluations, Governance Frameworks (AIA), Value Alignment Theory.
- **Core Engineering**: Systems Architecture, Full-Stack Development (React, Vue, Node.js), API Design, Database Design (SQL).
- **Languages & Tools**: Python (Learning), JavaScript/TypeScript, Solidity, C#, Docker, Git, Playwright/Cypress.
- **Research & Analysis**: Systems Thinking, Technical Writing, Algorithmic Impact Assessment.

Professional Experience

UI Developer (Senior) | EducationPlannerBC.ca | *January 2017 – Present*
- Develop and maintain complex, high-traffic provincial application systems used by students across British Columbia.
- **Authentication & Security**: Provided technical design and Proof of Concept implementation for

retiring Identity Server 4 in favor of **Keycloak**, modernizing the enterprise authentication infrastructure.

- Architected frontend solutions for the "Management Console," requiring strict data validation and role-based access control (RBAC).
- Championing accessibility and usability standards, a crucial component of human-centric software design.

Open Source Contributor | Government of Canada (StatCan) | *January 2020*
- **Project**: Algorithmic Impact Assessment (AIA) Tool.
- **Contribution**: Designed and implemented a transparency visualization ("heads-up display") to help users understand how specific risk scores were calculated.
- **Relevance helps**: Direct experience implementing governance logic into software code.

Mobile App Developer (Civic Tech) | Cloud Graf Projects | *January 2012 – January 2014*
- **Role**: Conscientious Developer / Co-Founder. Focused on using Open Data to improve public safety and consumer protection.
- **Vancouver RentalDog** (January 2012 – January 2013): Co-designed an Android/iOS app using City of Vancouver Open Data to "correlate rental unit information with bylaw infraction data" (cited in SSRN research), helping renters avoid unsafe accommodations.
- **DineSafe Surrey** (January 2013 – January 2014): Developed a mobile application mapping health code violations for restaurants in Surrey, BC, which was later integrated into the **City's MySurrey app**.

Senior Technical Analyst | BCcampus | *January 2007 – January 2017*
- Led technical implementation of shared services for the BC post-secondary system, focusing on interoperability and data exchange standards.

Programmer/Analyst | Kwantlen Polytechnic University | *January 1994 – January 2007*
- **Leadership**: Led the development of an award-winning Learning Management System (LMS) and coordinated enterprise email/web presence.
- **Systems**: Created early online voting and grading systems, managing the transition to digital institutional workflows.
- **Tech**: Unix, Apache, Lotus Notes/Domino.

Software Engineer | Simon Fraser University | *January 1989 – February 1991*
- *Project*: **DOCENT** - Developed to assist novice teachers with learning interventions based on relevant research in Instructional Psychology.

- *Role*: Collaborated with Dr. Phil Winne to implement graphical knowledge representation languages in **Prolog**.
- *Publication*: Co-authored paper presented at AAAI Association for the Advancement of Artificial Intelligence (1988).

Education & Certifications

Technical AI Safety | BlueDot Impact | *October 2025 – December 2025 (In Progress)*
- Technical curriculum covering RLHF, mechanistic interpretability, robust agents, and safety evaluations.

Intro to Transformative AI | BlueDot Impact | *March 2025 – April 2025*
- Comprehensive coursework on AI Risk, Governance, and technical alignment landscapes.

BTech in Computer Systems (Honours) | BCIT | *September 1996 – December 1997*
Advanced Diploma in Software Development (Honours) | BCIT | *September 1995 – December 1996*
- *Project*: Context-Based Reasoning System.

Computer Systems Technology (AI Focus) | BCIT | *September 1986 – December 1988*
- *Focus*: Expert Systems and Neural Networks (Signature Recognition with 70% accuracy).

Selected Writing & Creative Work

- **Article (Submitted)**: *"Báb-El: The Gate of God"* (Journal of Bahá'í Studies, 2025).
- **Article**: *"The Man Who Invented AGI and the Tower We're Still Building"* (AI and Faith, 2025).
- **Article**: *"Superintelligence and the Cosmic Host"* (AI and Faith, 2025).
- **The Shame Game: Stolen Lives** (Novel, 2024): Explores the human cost of identity theft, thematically paralleling current technical challenges in AI agent provenance and identity security.
- **In the Knee of the Curve** (Play, 2025): Exploring the ethical tipping points of exponential technology.
- **Ask Me About the Singularity** (Play, 2017/2023): A dialogue on the blurring lines between human and machine consciousness.
- **The Golden Goose is Near** (Play, 2012/2024): Early exploration of automation's societal impact.

#resume/ai-safety-career