

SALAH-EDDINE BENBRAHIM, PhD, MBA, B.Eng.

AI Governance & Conformity Assessment · Audit Evidence and Reproducibility · Enterprise Platform Leadership

• benbrahim.salah.eddine@cinelinfiniti.com • +1 647-875-0826 (CA) • +33 7 81 64 82 93 (FR) • [LinkedIn](#)

cinelinfiniti.com • github.com/CinelInfini/Video_Quality_Auditor • huggingface.co/benbrahimsalaheddine777 • orcid.org/0009-0005-7050-4014 • [CinelInfini Studio \(live\)](#) • publications.polymtl.ca

PROFESSIONAL SUMMARY

AI governance practitioner with 10+ years leading enterprise platforms in regulated public-sector environments and a PhD in Computer Engineering. Technical lead and enterprise architect for Government of Quebec digital platforms serving 2M+ citizens, accountable for release gating, availability targets and data quality across eight major releases with zero critical incidents. Builds the evidence layer that conformity assessment requires: hashed inputs, seeded runs, bootstrap intervals, multiplicity correction and reproducibility appendices generated from real run artifacts rather than screenshots or verbal attestation. Creator of CinelInfini, an evaluation and governance platform over 14 corpora and 1.4M rated assets (>90% test coverage, computation-level audit logging, DOI-archived reproducibility). Two sole-authored papers under review provide the methodological basis.

CORE COMPETENCIES

AI Governance & Conformity: EU AI Act readiness, ISO/IEC 42001, NIST AI RMF, conformity-assessment evidence, technical documentation, traceability and provenance, risk registers, model and dataset documentation, release gating, independent verification. **Evaluation & benchmarking:** LLM & model evaluation, reproducible evaluation pipelines & standardized protocols, benchmarking & benchmark reliability, inter-rater reliability, multiverse / sensitivity analysis, reference datasets, metric validation, acceptance criteria, LLM-as-a-judge, human preference data

Trustworthy, Safety & Responsible AI: trustworthiness scoring, causal-reasoning validation, multi-modal safety assessment, human-in-the-loop evaluation, annotation & aggregation-bias analysis, robustness, AI-agent reliability, responsible AI

Reproducibility & MLOps: >90% test coverage, CI/CD with go/no-go release gating, regression & drift control across releases, semantic versioning, audit logging, Zenodo DOI archival, PyPI, GitHub Actions, pytest

LLMs & Multimodal: OpenAI (GPT-4/4o), Gemini 1.5, Claude, DeepSeek; RAG & prompt engineering; data-centric evaluation; CLIP, DINOv2, ArcFace, optical flow, Dynamic Time Warping

Engineering & Systems: Python (async/concurrent, microservices, REST), PyTorch, ONNX, distributed systems, cloud (AWS, GCP), SOLID / dependency injection

PUBLICATIONS & RESEARCH

Index terms: inter-rater reliability; variance components and identifiability; human reference panels; LLM-as-a-judge evaluation; annotation budget allocation; calibrated diagnostics; reporting protocols. Sole author. *Under review* (2026).

Index terms: inter-rater reliability; construction choices in reliability estimation; rater screening and aggregation; panel size; multiverse analysis; reproducibility of published coefficients. **Reproduction package:** doi.org/10.5281/zenodo.22698801. Sole author. *Under review* (2026).

Earlier peer-reviewed work — **seven papers (three journal articles, four conference papers):** IEEE Transactions on Cloud Computing, International Journal of Information Systems & Project Management, Procedia Computer Science, NBIS, and the Asia Modelling Symposium. Citation record: 41+ citations, h-index 4 (IEEE Trans. Cloud Computing 15; IJISPM 11; Procedia Computer Science 6). Full profile: publications.polymtl.ca

EXPERIENCE

Founder & Open-Source AI Research Lead — CinelInfini (Video / AIGC Evaluation Framework) 2023 – Present

- Architected an open-source evaluation framework for video-quality (VQA) and AI-generated-content (AIGC-VQA) assessment, with 24 scoring modules across 6 categories including trustworthiness scoring, causal-reasoning validation, physics-plausibility checking, world-model surprise detection, and multi-modal safety.
- Built a reproducible measurement pipeline: >90% test coverage (<30s execution), <5% variance across repeated measurements across 14 governed corpora, computation-level audit logging, and semantic versioning with full backward compatibility.
- Designed automated CI/CD with build interruption on test failure (go/no-go release gating), version synchronization, GitHub release integration, PyPI publishing, and Zenodo DOI archival for permanent reproducibility.
- Implemented multimodal video analysis (hierarchical shot detection, DINOv2 / ArcFace embeddings, CLIP semantic scoring, DTW temporal coherence) and a 100% centralized YAML configuration (50+ settings, zero hardcoded parameters).

Founder & AI Architect — JokeJoy (Multi-LLM Generative AI SaaS) 2023 – 2025

- Engineered a multi-LLM architecture (GPT-4o primary, Gemini 1.5 fallback) with automatic provider switching on latency/cost thresholds and dynamic context-injection prompting; zero critical incidents.

Technical Lead & Architect — Government of Quebec (Public Digital Platform) 2023 – 2025

- Led end-to-end architecture for platforms serving 2M+ citizens; cut latency 40% (2.5s→1.5s), maintained 99.9% uptime, delivered 8 major releases on schedule with zero critical incidents.

Postdoctoral Research Fellow — Polytechnique Montréal 2017 – 2018

- Research on project control standards and process networks; published 2 peer-reviewed papers.

PhD Researcher & Lecturer — Polytechnique Montréal 2009 – 2016

- Real-time migration of interdependent virtual machines in cloud networks (MIP models + Tabu Search); reduced service penalties 13.5% (avg.) and physical-machine usage up to 37% on Google traces; 4 peer-reviewed papers; taught first-year MATLAB programming at Polytechnique.

EDUCATION

PhD, Computer Engineering — Polytechnique Montréal, Université de Montréal (February 2016). Dissertation on real-time migration of interdependent virtual machines in cloud networks.

MBA — UQAM (Université du Québec à Montréal).

B.Eng., Computer Science & Engineering — Polytechnique Montréal.

OPEN-SOURCE & PUBLIC ARTIFACTS

- CineInfini evaluation framework — open-source, with 24 scoring modules across 6 categories, reproducible pipelines, computation-level audit logging, semantic versioning, automated CI/CD gating, PyPI packaging, and Zenodo DOI archival for permanent citability.
- Methodology and architecture designed by me; built with AI-assisted development under a verification protocol.
- CineInfini Studio — a public, live evaluation environment: selects a video dataset, computes perceptual quality scores (SRCC, PLCC, KRCC), and runs an LLM judge over the results — a working, inspectable demonstration of reproducible AI evaluation.
- A reproducibility package accompanies one of the two papers under review.

SELECTED TECHNICAL ENVIRONMENT

Languages & frameworks: Python (async/concurrent), PyTorch, ONNX, pytest, GitHub Actions.

Models & APIs: OpenAI GPT-4/4o, Google Gemini 1.5, Anthropic Claude, DeepSeek; CLIP, DINOv2, ArcFace; RAG and prompt engineering.

Infrastructure: AWS, GCP, microservices and REST architectures, distributed systems, CI/CD, PyPI, Zenodo.

ADDITIONAL INFORMATION

- I can work in Canada, France and across the European Union and EEA. Eligible for TN status under the USMCA for the United States. Open to remote, hybrid, and relocation (North America, Europe, Gulf, Asia). Available in Montréal, Toronto and Paris.
- Languages: **English** — fluent (professional); **French** — native.