

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

Optimization & Distributed Systems Researcher · Cloud Resource Management · Operations Research

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

[cineinfinite.com](#) · [orcid.org/0009-0005-7050-4014](#) · [dblp.org/pid/160/4876](#) · [huggingface.co/benbrahimsalaheddine777](#) · [publications.polymtl.ca](#)

PROFILE

Operations-research and distributed-systems researcher with a **PhD in Computer Engineering** from Polytechnique Montréal and more than ten years delivering large-scale production systems. Doctoral and post-doctoral work in **mixed-integer programming, metaheuristics and combinatorial optimization** applied to cloud resource allocation, network design and process analysis, published in *IEEE Transactions on Cloud Computing* and validated at scale (25,000+ VMs). Since 2023 this has extended into applied AI: multi-LLM systems in production and reproducible evaluation methodology. Bilingual: French native, English professional.

AREAS OF EXPERTISE

Optimization & operations research : mixed-integer programming (MIP), linear and non-linear programming, Tabu Search and metaheuristics, combinatorial optimization, mathematical modelling, scheduling and resource allocation under constraints

Cloud & distributed systems : virtual-machine placement and live migration, SLA-aware cost/penalty optimization, capacity and resource planning, large-scale network design, microservices, high availability

Software & data engineering : Python, distributed architectures, REST APIs, CI/CD, automated testing, AWS and GCP, RDBMS and data warehouses, streaming

Applied AI & machine learning : large language models (GPT-4o, Gemini, Claude), multi-LLM orchestration, RAG, PyTorch, multimodal analysis, model evaluation and reproducibility

Analytics & decision support : business intelligence (SAP BI), data governance and quality, simulation and empirical validation, statistical analysis

SELECTED PUBLICATIONS

- [1] Benbrahim, S.-E., et al. (2019). *Live Placement of Interdependent Virtual Machines to Optimize Cloud Service Profits and Penalties on SLAs*. *IEEE Transactions on Cloud Computing*, 7(1), 237–249. — 15 citations
- [2] Perrier, N., Benbrahim, S.-E., et al. (2019). *A comparison of project control standards based on network analysis*. *International Journal of Information Systems and Project Management*, 7(3), 37–62. — 11 citations
- [3] Perrier, N., Benbrahim, S.-E., et al. (2018). *The core processes of project control: A network analysis*. *Procedia Computer Science*, 138, 697–704. — 6 citations
- [4] Benbrahim, S.-E., et al. (2014). *On the Design of Large-Sized Cloud Networks Optimized for Live Migration Using Tabu Search Algorithms*. *8th Asia Modelling Symposium*, Kuala Lumpur.
- [5] Benbrahim, S.-E., et al. (2014). *New Distributed Approach for an Autonomous Dynamic Management of Interdependent Virtual Machines*. *NBiS 2014*, Salerno, 531–534. — 3 citations
- [6] Benbrahim, S.-E., et al. (2014). *Real-Time QoS Issues in Live Migration of Interdependent Virtual Machines*. *8th Asia Modelling Symposium*, Kuala Lumpur.

Under review (sole author) — **reliability and reproducibility of AI evaluation**: Index terms: inter-rater reliability; variance components and identifiability; human reference panels; LLM-as-a-judge evaluation; annotation budget allocation; calibrated diagnostics; reporting protocols. (under review) and Index terms: inter-rater reliability; construction choices in reliability estimation; rater screening and aggregation; panel size; multiverse analysis; reproducibility of published coefficients (under review). Citation record: 41+ citations, h-index 4. Reproduction package: doi.org/10.5281/zenodo.22698801.

RESEARCH & PROFESSIONAL EXPERIENCE

PhD Researcher — Polytechnique Montréal 2009 – 2016

- Doctoral research on the real-time migration of interdependent virtual machines in cloud networks. Formulated mixed-integer programming models for SLA-aware placement and developed Tabu Search metaheuristics to solve them at scale.
- Validated on Google traces: reduced service penalties by 13.5% (avg.) and physical-machine usage by up to 37% against baseline placement strategies. Published in *IEEE Transactions on Cloud Computing*, *NBiS* and the *Asia Modelling Symposium*.
- Taught MATLAB programming to first-year engineering undergraduates at Polytechnique Montréal.

Postdoctoral Research Fellow — Polytechnique Montréal 2017 – 2018

- Network analysis of project-control processes across PMBOK, PRINCE2 and AACE standards, identifying which processes are structurally central. Published in *Procedia Computer Science* and the *International Journal of Information Systems and Project Management*.

Technical Lead & Architect — Government of Quebec 2023 – 2025

- Led end-to-end architecture of public digital platforms serving 2M+ citizens: 40% latency reduction (2.5s → 1.5s), 99.9% uptime, eight major releases delivered on schedule with zero critical incidents. Distributed systems, microservices, cloud infrastructure, CI/CD.

Senior BI Data Analyst — Government of Quebec 2019 – 2020

- Business intelligence and analytics (SAP BI); data governance and quality assurance, reducing data-quality issues by 35%; mentored a team of eight analysts.

Founder & Architect — CineInfini (2023 – Present) and JokeJoy (2023 – 2025)

- CineInfini: open-source evaluation framework built as a production system — >90% test coverage, computation-level audit logging, semantic versioning, automated CI/CD with go/no-go release gating, DOI-archived results, and a public live studio.
- Designed the methodology, architecture and statistical framework myself, and implemented the platform with AI-assisted development under a verification protocol — the same discipline it applies to the systems it evaluates.
- JokeJoy: multi-LLM production system (GPT-4o primary, Gemini 1.5 fallback) with automatic provider switching on latency and cost thresholds; zero critical incidents.

EDUCATION

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

MBA — Université du Québec à Montréal (UQAM). B.Eng., Computer Science & Engineering — Polytechnique Montréal. M.Sc.A. thesis (2011) on defence against traffic-analysis attacks in wireless sensor networks.

SELECTED ACHIEVEMENTS

- Reduced SLA service penalties by 13.5% (avg.) and physical-machine usage by up to 37% in cloud placement, validated on Google traces — published in a Q1 IEEE journal.
- Cut platform latency by 40% and sustained 99.9% uptime for government services used by more than two million citizens.
- Reduced data-quality issues by 35% through governance and QA processes, while mentoring a team of eight analysts.
- Built and released open-source research software with reproducibility guarantees: variance under 5% across repeated measurements across 14 governed corpora.

RESEARCH INTERESTS

Resource allocation and scheduling under uncertainty; energy-aware and cost-aware optimization of cloud and data-centre workloads; SLA-constrained placement and migration; large-scale network design; metaheuristics and decomposition methods for combinatorial problems. A current interest sits at the intersection of my two lines of work: **the computational and energy cost of evaluating AI systems** — how to obtain a reliable measurement of model quality without spending disproportionate compute, and how to schedule evaluation workloads efficiently.

LANGUAGES & ELIGIBILITY

French — native. **English** — professional (fluent). I can work in Canada, France and across the European Union and EEA. Eligible for TN status under the USMCA for the United States. Available in Montréal, Toronto and Paris.