

OLGA MANAKINA

Bilingual English/French • olga@manakina.com
[LinkedIn](#) • [Google Scholar](#) • [GitHub](#) • [YouTube](#)

Applied AI Researcher | LLM Reliability & Educational Products

LLM Evaluation • Benchmarking • Experimental Design • Educational Product Development

SUMMARY

Applied AI researcher and educational product builder working at the intersection of LLM reliability and education. Research experience includes designing and analyzing large-scale multi-turn evaluations (an ICLR 2026 study of 84,540 trajectories across 8 model families, a NeurIPS 2025 multi-agent reliability benchmark, and adaptive LLM evaluation using multi-armed bandits), with five peer-reviewed publications in 2025–2026. Complementary product experience building digital education at scale: five Udemy courses serving 1,000+ learners and a YouTube channel with 35.2K subscribers. Skilled in turning open-ended questions into controlled experiments, quantitative metrics, and human-validated evaluation pipelines, and in translating complex subject matter into courses, workshops, and learner-facing materials.

PUBLICATIONS

- **ICLR 2026** (LLM Reasoning Workshop). Bogdanov I., **Manakina O.**, Lung CH. "Evaluation of Multi-Turn Consistency in LLM Agents: Survival Analysis and Failure-Rationale Taxonomy"
- **NeurIPS 2025** (MTI-LLM Workshop). **Manakina O.**, Bogdanov I., Lung CH. "Delay-of-Gratification as a Multi-Agent Survival Micro-benchmark for Long-Horizon LLMs"
- **AIED 2025** (main track). **Manakina O.** & Muldner K. "Prompts Eliciting Active and Constructive Engagement Improve Learning Across Traditional and LLM Contexts"
- **IEEE COMPSAC 2025** (QUORS Workshop). **Manakina O.** & Lung CH. "Designing Reusable LLM-Enhanced Assignments: A Quality-Oriented Framework for Software Engineering Education"
- **EDM 2025** (WLIEDM Workshop). **Manakina O.** & Bogdanov I. "Learning to Grade Efficiently: A Bandit-Driven Prompt-Selection Framework for Low-Cost LLM Essay Scoring"

RESEARCH & PROFESSIONAL EXPERIENCE

INDEPENDENT RESEARCHER — LLM EVALUATION & AGENT RELIABILITY — Research collaboration with Systems & Computer Engineering researchers, Carleton University, Ottawa, ON, Canada (2025–Present)

- Authored the failure-rationale taxonomy and led the semantic analysis for a controlled study of judgment consistency in multi-turn LLM agents, categorizing 13,780 model-generated rationales ($\kappa=.83$ against human audit); co-developed the surrounding 84,540-trajectory benchmark across 8 model families using factorial perturbations and survival/hazard modeling. *ICLR 2026 LLM Reasoning Workshop*.
- Co-designed a long-horizon multi-agent benchmark with 19,200 ReAct-agent trajectories across 64 experimental cells, testing social exposure, persona and tool-use effects through survival analysis and targeted ablations. *NeurIPS 2025 MTI-LLM Workshop*.
- Built a multi-armed-bandit controller for adaptive LLM prompt selection in automated scoring, reducing LLM calls by 78.4% and token use by 72.8% versus exhaustive evaluation. *EDM 2025 WLIEDM Workshop*.

GRADUATE AI RESEARCHER, AI FOR LEARNING SYSTEMS — Carleton University, Department of Cognitive Science, Ottawa, ON, Canada (09/2024–Present)

- Designed and evaluated an LLM-based tutoring system across 45 sessions, defining metrics for instruction following, generation fidelity, grading accuracy, and remediation reliability (MA thesis).
- Designed a practical framework to integrate LLMs into Computer Science and Software Engineering curricula (published at IEEE COMPSAC 2025).
- Conducted controlled research on human–LLM interaction ($N=101$ experimental study, published at AIED 2025), combining experimental design, quantitative learning-outcome analysis, and qualitative interaction analysis.

AI & LEARNING CONSULTANT — Self-Employed, Remote (2022–Present)

- Design AI enablement materials — prompt templates, workflow demos, tutorials — for non-technical and semi-technical users; provide private AI-literacy training for language learners and instructors.
- Designed AI-prompting strategies and templates for professional language training; achieved a **90% first-try pass rate** for clients taking French language exams.

TEACHING ASSISTANT — Carleton University, Ottawa, ON, Canada (09/2024–05/2026)

- *Introductory Python (CGSC1005)* & *Introductory Cognitive Science (CGSC1001)*: conducted weekly tutorials for 20+ students; graded and provided feedback for 170+ students.

LANGUAGE TRAINER FOR ADULT LEARNERS — Self-Employed, Remote (2013–2022)

- Designed diagnostic assessments and individualized French training programs for working professionals, including executives in law, technology, and fashion.

UX DESIGNER & CONTENT MANAGER — IBCICO Development (full-cycle software development agency), Moscow, Russia (2014–2017)

- Managed content quality standards, review workflows, and delivery for client software products including financial data systems, e-commerce, and ticket booking platforms.

UNIVERSITY PROFESSOR & PROGRAM LEAD — Moscow State Institute of International Relations, Moscow, Russia (2002–2013)

- Led and mentored a team of 11 language instructors; designed undergraduate and graduate curricula, assessment instruments, and examination standards.
- Co-authored a national prize-winning textbook adopted by 94 universities in Russia.

REVIEWER — ACL CoNLL 2026 (March–April 2026) • NeurIPS 2026 Ethics Reviewer (July 2026) • NeurIPS 2025 Workshop MTI-LLM (September 2025)

DIGITAL EDUCATION PORTFOLIO

- Founded a [YouTube channel](#) on educational technologies; organically built an audience of 35.2K subscribers through 54 instructional videos and audience-informed content refinement (2020–present).
- Authored five French-language online courses for the [Udemy project "Learn French Online"](#), serving 1,000+ students (2018–present).

TECHNICAL & LEARNING DESIGN SKILLS

LLM Evaluation & Agentic AI: Multi-turn LLM evaluation, agent reliability, benchmark design, failure analysis, reasoning-trace analysis, LLM-assisted annotation & human audit.

Experimental & Statistical Methods: Factorial experimental design, ablation studies, Kaplan–Meier survival analysis, discrete-time hazard modeling, logistic regression, ANOVA, inter-rater agreement, qualitative taxonomy development.

Programming & Data: Python, Pandas, NumPy, scikit-learn, PyTorch, multi-armed bandits, Git, API-based LLM experimentation, SPSS, SAS.

Learning Design & Enablement: Adult learning, curriculum and assessment design, virtual facilitation, learning evaluation, instructional content development, learner needs analysis.

Learning Technologies & Content: Udemy, Camtasia, Canva, Figma, Final Cut Pro, LaTeX, Google Slides.

EDUCATION

MA, Cognitive Science & Data Science — Carleton University, Ottawa, ON, Canada (2026)

Thesis on multi-turn LLM evaluation and reliability in an adaptive tutoring environment for novice programmers

BA, Cognitive Science & Computation — Carleton University, Ottawa, ON, Canada (2024)

PhD, Political Science — Moscow State Institute of International Relations, Moscow, Russia (2011)

MA, Philology & Education, French Language — Lomonosov Moscow State University, Moscow, Russia (2002)

ML PORTFOLIO

- **End-to-End ML System for User Behavior Prediction** (2025): scalable machine learning system predicting telecom customer churn with personalized recommendations.
- **Exam Score Predictor & Recommender** (2022): Naïve Bayes classifier predicting language-exam scores with personalized learning recommendations.

AWARDS & SCHOLARSHIPS

Cognitive Science Graduate Scholarship (Carleton University, Ottawa, ON, Canada — 2025)

Dean's Honour List (Carleton University, Ottawa, ON, Canada — 2021, 2022)

Government of the Russian Federation Prize in Education (Moscow, Russia — 2015) — *co-author of the course package "Le Français.ru B1"*

Award "Book of the Year" (Moscow, Russia — 2010) — *co-author of the course package "Le Français.ru", levels A1–B1*