

I build AI that shows its work: Tsetlin Machines whose decisions are logical clauses you can read, guardrails whose safety properties are proved by solvers, and certificates anyone can re-check. Outside the lab I build reinforcement-learning agents, data systems and apps that run in the browser.

EDUCATION	MSc in Artificial Intelligence, University of Agder In progress Research at CAIR, the Centre for Artificial Intelligence Research, under Ole-Christoffer Granmo, who introduced the Tsetlin Machine. Master's research: glassbox-lm, steps toward a language model whose knowledge is a set of readable clauses. BSc in Computer Engineering, cybersecurity and programming, NTNU Gjøvik 2025 Thesis: <i>Automatic Removal of Sensitive Content in Images and Documents for Efficient Archiving</i>, with Rami Amer and Yasin Hessnawi, for Mattilsynet.
AWARD	Eurekaprisen, NTNU 2025 Best bachelor thesis at the Faculty of Information Technology and Electrical Engineering.
EXPERIENCE	Chief Financial Officer, SafeMediaAI Norwegian AI startup behind DeclassifAI.
RESEARCH	CANDOR: interpretability by construction Routes a layer through named concepts and emits a re-checkable certificate on every forward pass. On synthetic ground truth: $\delta \approx 0.002$ with no measurable accuracy cost; backdoor detection at AUC ≈ 0.90. TRACE: a guardrail for AI agents Readable content clauses, bounded-memory temporal monitors and a tamper-evident certificate ledger, on a CPU. F1 0.827 on deepset/prompt-injections at 107 μs per check, in 12.5 KiB. VERITAS: formally verified guardrails Safety properties proved or refuted by SAT and SMT solvers, with bypasses repaired by construction. Retraining stalls at 8 provable bypasses; constructive repair reaches 0, and accuracy goes from 0.69 to 0.71. LUCID and graphtm-cbr A glass-box, safety-shielded agent for ARC-AGI-3 that completes two public levels; and a graph Tsetlin Machine for Ames mutagenicity whose minimal-edit recourse succeeds for 95.5% of positives. NEXUS: reinforcement learning for 2 to 6 players AlphaZero-style Chinese Checkers agent with a 7.5M-parameter network and Gumbel MCTS, plus prototypes for coalition-aware search, decomposed values and a wreath-equivariant network. 324 tests pass on every push.
ENGINEERING	PocketOS and SleepArchitect, both live A desktop environment that runs offline in a phone's browser; and a sleep forecaster built on Borbély's two-process model that reproduces all 11 published findings it was checked against, in 18.7 KB. IDM: an epidemiological data warehouse 473,189 weekly disease reports stored as a PostgreSQL star schema, MongoDB buckets and a Neo4j graph, with Kafka and ksqlDB streaming, FastAPI and React. ThermoGuard, Mneme and vsclaude Indoor heat-risk forecasting with equity-aware triage; textbook-to-Anki on local models; and a desktop IDE that shows what a coding agent is doing.
PAPERS AND REPORTS	1. The Explanation Is the Forward Pass: Interpretability by Construction with Certified Legible Bottlenecks. CANDOR 2. A Tsetlin Runtime Assurance and Certification Engine for Interpretable, Tamper-Evident Guardrails on Agentic AI. TRACE 3. A Temporal Hierarchical Graph Tsetlin Machine with Echo-Trace Automata for Trajectory-Level Verification of Agentic AI. THGTM 4. Formally Verified, Competence-Aware Interpretable Guardrails for Language Models. VERITAS 5. Model-Free, Certifying Least-Cost Repair of LLM-Agent Tool-Call Trajectories. REINS 6. A Glass-Box, Safety-Shielded Agent for Open-World Interactive Reasoning on ARC-AGI-3. LUCID 7. A Discrete-Diffusion Tsetlin Machine for Interpretable Generative Modelling of Tabular Boolean Data. DiTM 8. A Multi-View Hierarchical Graph Tsetlin Machine. Multi-HGTM 9. Technical reports on conformal, federated, robust and counterfactual Tsetlin Machines, LLM agent routing, sleep prediction and indoor heat risk. ConfTM, FedTM, RobTM, CCTM, Clause-Driven-LLM, SleepArchitect, ThermoGuard
TOOLBOX	Python, PyTorch, C++, Rust, TypeScript, React, Next.js, Tauri, Three.js and WebGL, FastAPI, PostgreSQL, MongoDB, Neo4j, Kafka, Docker, Linux, Git and LaTeX.