IM

CONTACT

- hello@ilyasm.dev
- Veenendaal, Netherlands
- ilyasm.dev
- linkedin.com/in/ilyas-mallah
- github.com/ilyas-mallah

SKILLS

PROGRAMMING

Luau, C++, Rust, CUDA, Python, TypeScript, C#, .NET

AREAS

compilers and JITs, WebAssembly, GPU performance, PostgreSQL and TimescaleDB

TOOLS

Nsight Compute, perf, Git, CMake, Docker

SPOKEN LANGUAGES

Dutch, English

ILYAS MALLAH

SOFTWARE ENGINEER

PROFILE

I am a third-year Computer Science student working on performance and security in compilers, runtimes and AI inference engines. I am open to internships and contract work, part-time during the year or full-time in the summer.

EXPERIENCE

Rolantir, analytics platform for the Roblox games market — Co-founder, June 2026 to present

- I built almost all of the backend (TypeScript, PostgreSQL, TimescaleDB) for a catalog of over 1 million games.
- I moved the live 13.9 GB SQLite database to TimescaleDB with the full history, which cut hourly chart queries from 400 to 900 ms to 3 to 8 ms.

Dig It, Roblox game — Co-founder, October 2024 to March 2025

- I did the programming, design and live operations for a game with 73 million visits and 27,400 players online at its peak.
- I led the negotiation when the game was sold.

OPEN SOURCE

Luau, Roblox's scripting language — 7 PRs merged in 2026

- I gave integer `==` in branches its own IR command in the native code generator (#2571, merged in August 2026). It measured 3.18 ns to 0.63 ns per comparison. Luau has since removed its integer type.
- I fixed an x64 register allocator miscompile when restoring a spilled value (#2987).

Wasmtime, wabt, mruby, wgpu — 12 PRs merged in October 2026

- I fixed Wasmtime's panic on float to 8- and 16-bit integer conversions on x64, an issue open since 2022 (#14778), and I am the credited reporter of CVE-2026-115893.
- I fixed wabt's `wasm2c` on 32-bit hosts (#2874), mruby's `pack` and `unpack` for strings over 2 GiB (#7692) and naga's fuzz build (#10637).

LLVM, libc++ and Clang — 3 merged, eleven PRs in review

- My merged PRs fix the callee pop size of x86 tail calls with aligned stack arguments (#235417). They also make Clang's format attribute warning work without its libc call warning (#232898).
- My PRs in review add hardening assertions to `insert`, `emplace`, `string` and `valarray` (#230490, #234566, #236120).

vLLM — four PRs in review

- I made Hopper pick the block-FP8 kernel by batch size, which gives 17 to 52% more throughput on an H100 NVL (#60464).
- I added a context window for EAGLE-3 drafters, which raises decode at 32k from 16 to 34 tok/s on an A40 (#63245).

llama.cpp — 2 PRs open

- I traced half-speed q4_0 KV cache decode on RTX 50 GPUs to a CUDA 12.8 compiler bug, and my kernel workaround gives up to 2.1x faster decode at 32k context (#30077).

tt-metal — 1 issue with fix

- I found `ttnn.logit` up to 8,767 ULP off near 0.5 with a 54 op simulator sweep, and my fix brings it to 3 ULP (#59930).

PROJECTS

- **Ulana** is an optimizing WebAssembly to Luau compiler of about 47,000 lines of Luau. SQLite, QuickJS, jq and zstd run as generated Luau and match wasmtime.
- **Ganglion** is a Rust and CUDA connectome simulator. It runs a fruit fly nervous system (165,122 neurons) at 0.37 s per simulated second on one RTX 3070 Ti, with the same output as its Python reference.
- **Lute FFI** is a specification and five-milestone implementation of a C foreign function interface for Lute, with full Luau types. It is not yet public.

EDUCATION

Utrecht University, BSc Computer Science — September 2024 to early 2028 (expected)

- My average grade is 8 out of 10, and my minor is Entrepreneurship.