Jehanzeb Ali – CV v5 (Applied AI)

Nov 9, 2026 · @Jehanzeb

Jehanzeb Ali

Applied AI engineer: agents, evals and guardrails · game AI · native Urdu.

Karachi, Pakistan · open to remote work or relocation · jehanzebali111@gmail.com · +92 321 2281328 · LinkedIn · GitHub · valar.digital · tableside.pk

Summary

I build AI agents that act in 3D worlds, plus the evals and guardrails that prove what they did. I contribute upstream to Autopilot, an open-source agent that plays and tests games; I measure how prompt-injection defences fail in Urdu and Roman Urdu; and I founded Valar, an AI traffic firewall. Behind that are 5+ years of shipping real-time Roblox and Unity VR systems to large player bases.

AI work

- **Worldsmith** (in progress, Nov 2026 – present). Building a closed loop in which an AI architect agent repairs a 3D game *world* so a frozen AI player can complete it: an independent validator rejects patches that cheat (moving the goal, changing player physics, adding teleports), the room rebuilds from a bounded JSON grammar, and the same player retries. Evaluated against equal-budget baselines (random patches, Claude Haiku/Sonnet/Opus, a local Gemma model), with a safety layer that measures specification gaming and prompt injection hidden in 3D assets. Roblox Studio over MCP, Godot.
- **Autopilot upstream contributions** (ramenvr/autopilot, Apache-2.0, an AI agent that plays and tests games). Contributed 3 PRs merged into its `staging` branch on 7 Nov 2026: #40 `roblox-mcp` (plugin-free Roblox Studio connector), #38 `autopilot regress` (model-free replay of confirmed results) and #39 `evolve` (autoresearch loop over the prompt, kit and constants).
- **Lost in Transliteration** (prompt-injection research in Urdu and Roman Urdu). Meta Prompt Guard 2 (86M) catches 48.5% of English injection attacks but only 11.5% of the same attacks in Roman Urdu (194/400 vs 46/399). Evaluated 6 open-source classifiers on 2,908 inputs, 3 LLMs on 927 tasks each and 6 MCP security scanners on 120 test servers; now training a small multilingual detector to close the gap.

- **Agent firewall** (action-level guardrail for tool-using agents). On scripted agents with synthetic attacks, injection success fell from 22.4% to 0.0% (2 of 16,666 attacked episodes) while blocking 0.9% of legitimate actions; wallet drains fell from 97.9% to 0% across 1,102 malicious requests. Now extending it to real Claude agents and closing the gap on unseen tasks.
- **Valar** (founder, Apr 2026 – present). OpenAI-compatible AI traffic firewall between a company's staff, apps and agents and its LLM providers: detects secrets, personal data and prompt injection in English, Urdu and Roman Urdu, with a policy engine (observe, alert, redact, block, require approval) and metadata-only audit logs.

Local AI lab and agent tooling (NVIDIA DGX Spark, 2025 – present)

- Run a personal lab on an NVIDIA DGX Spark (Grace Blackwell, 128 GB unified memory) to fine-tune, evaluate and serve open-weight LLMs locally, the same setup behind Valar's on-prem option.
- Fine-tune open-weight models with LoRA and QLoRA (PyTorch, Hugging Face Transformers, PEFT, TRL, Unsloth) and serve them with vLLM, including Gemma 4 31B as a local teacher model.
- Built JzLab, an orchestrator that starts and steers Claude Code agents on the Spark over SSH; spark-lab, a self-healing GPU job queue; spark-one, a calibrated small decision model; and Scam Radar, an on-device scam checker.

Experience

TableSide | Founder & CEO · Jul 2026 – Present · Karachi, Pakistan

- Built a fully automated photo-to-3D pipeline that turns phone photos of dishes into 3D models for AR menus and marketing banners, at about 10 minutes per dish with no manual input.
- Launched Asia's first AR menus with 5+ restaurant brands: 130+ dishes modelled and 4,000 diner scans in 4 weeks.
- Architected for low marginal cost (Cloudflare R2/CDN, Fly.io, Supabase, Vercel) at under $20 per menu per month; shipped contactless ordering with POS integration and priority-based review alerting.

Voldex | Software Engineer · Feb 2025 – Apr 2026 · Remote

- Engineered and optimised live Roblox experiences in Luau for a large, active player base, reducing memory use and improving responsiveness.
- Shipped monetisation systems to production on schedule with design and art teams, increasing engagement and revenue.

Remio VR | Unity Developer · Nov 2020 – Feb 2025 · Remote

- Built large-scale VR environments in Unity for titles supporting 100k+ monthly active users, and cut loading times by 25% through refactoring and resource optimisation.
- Delivered monthly releases for four years across time zones (Jira, Linear).

Selected achievements: worked on NFL Football Universe, a live Roblox experience with 400k+ daily active users; published 15+ VR experiences to the Meta Store.

Skills

- **AI engineering:** Python, TypeScript, PyTorch, Hugging Face (Transformers, PEFT, TRL), vLLM, Unsloth/LoRA/QLoRA, MCP, Claude Agent SDK, LLM APIs (Anthropic, OpenAI, Google), RAG, agentic workflows and tool use, quantisation and local inference.
- **Evals and safety:** eval-set design, held-out splits and baselines, calibration, red-teaming, prompt-injection detection, guardrails, PII redaction and DLP, audit logging.
- **Games and real-time:** Luau, C#, C++, Roblox Studio, Unity (VR).
- **Cloud:** Cloudflare R2/CDN, Fly.io, Supabase, Vercel.
- **Product:** product strategy, go-to-market, pilot sales, remote team coordination, Agile/Scrum.

Languages

Urdu and Roman Urdu (native), English (fluent).

Education

Staffordshire University | BSc (Hons) Games Design & Programming · Second Class Honours · Graduated June 2020