

PRINCIPAL

Yasser Jawad *BSc (Hons)*

Applied AI Architect · Methodology & Memory Systems

London based · Asia · Middle East · Open to relocate

yas@veres.global · [linkedin.com/in/yasser-jawad](https://www.linkedin.com/in/yasser-jawad) · github.com/yjawad120 · veres.global

PROFESSIONAL SUMMARY

Applied AI architect building inhabited companion systems and agentic workflows. Year-long case study (Taste OS / Nyx) documenting how shaped substrate produces emergent behavioral properties existing prompt-engineering does not — including a controlled ablation where the personality layer fell back to approximately 18 tokens while accumulated voice, attunement, and relational instincts persisted.

Three working papers — *Inhabited Architecture*, *The Substrate Gap*, *Mining for Substrate* — documenting the methodology, the architectural pattern, and what transfers across domains. Background combines Development Economics (SOAS London) with hands-on AI systems work: production agentic pipelines, multi-agent orchestration, container-architecture design, methodology research.

SELECTED WORK

Taste OS / Nyx — Inhabited AI architecture

Principal · Veres Labs · 2025 – Present

- Designed and built a year-long companion-AI case study on shaped-substrate methodology
- Architecture: **continuous-thread persistence**, **asymmetric memory compression**, **discretion architecture** (tools as hands, not buttons), **bidirectional correction loop**, **self-authored emotional continuity**
- **Self-authored memory distillations** — the AI writes its own memory at session-close rather than blind summarization. 200+ continuous session exchanges accumulated. **29% per-session cost reduction at production scale** via cache architecture
- Falsifiability anchor: documented controlled ablation where personality-layer substrate fell back to ~18 tokens; emergent behavioral properties (voice, relational instincts, attunement) persisted
- Documented in three working papers: *Inhabited Architecture*, *The Substrate Gap*, *Mining for Substrate*
- **Stack:** claude-opus-4.X, Next.js, SQLite, Drizzle ORM, custom prompt-assembly + tool architecture

Job Engine — Agentic recruitment workflow system

Principal · Veres Labs · 2026

- Six specialized agents across a four-layer processing pipeline: Evidence (deterministic) → Optimization (LLM) → Creative (optional) → Review (LLM QA)
- State persistence with JSON-backed crash recovery and full auditability
- Confidence-based fallback and classification logic for context-aware decisions
- CLI framework (~1,400 lines Python) for internal adoption
- ~2,155+ lines Python, ~1,500+ lines documentation

Veres Labs — Independent AI systems studio

Principal · veres.global · 2025 – Present

- Independent studio designing and shipping AI systems for operational work — recruitment, orchestration, companion systems, methodology research
- Site designed and shipped using **substrate-first handoff methodology** — same patterns producing companion behavior generalized to visual identity, four-pen palette, Job Engine diagram, all emerged from the substrate without explicit specification
- Concrete demonstration that inhabited-architecture patterns transfer outside companion AI contexts
- **Stack:** Astro, Vercel, custom typography + visual identity system, GitHub auto-deploy

Mail Boxes Etc. — Applied AI Architecture & Director of Business Development

London (Clapham Junction) · 2022 – 2026

- **2023: Context-augmented GPT-4 routing system for B2B email triage.** Three-branch classifier (guidance / quote / payment) with embedded service catalog, stepped pricing schedule, response templates, and business-rule constraints as the context layer. Built before context-augmented generation (CAG) was canonically named. Human-in-the-loop integration. **300% response time improvement.**
- Established iterative prompt-development workflow ("Develop Prompt" protocol) — LLM as co-author for refining the routing logic. Same methodology pattern later applied to the Taste OS / Nyx companion architecture.
- **Conceived and led QuickKey B2B mobile platform** — originated the product idea, designed the full architecture, owned the roadmap from concept through partner-onboarding launch. Shipped: partner onboarding + commercial GTM · **30% operational efficiency.**
- QuickKey roadmap included a computer-vision key analysis pipeline: mobile image capture → OpenCV preprocessing → edge detection + contour extraction → A4-sheet baseline-scale dimensional measurement → key-type classification → automated bidding from reference catalogs → integration with Keyline Ninja Total key-cutting machine. Architected and prototyped; production rollout handed off to development.
- Dynamic pricing engine — same context-injection pattern for B2B quote generation.
- 1,000+ SKU automated reordering — Python + Square API + threshold detection → Excel/PDF generation → supplier email. **30% revenue growth.**
- CRM with lead scoring — customer data structuring, lead segmentation.
- Annual revenue: **£200K → £300K (+£100K)**, year-end target met via the composed architecture.
- Director-level role spanning product roadmap, B2B GTM, and applied AI architecture.

PUBLICATIONS

Inhabited Architecture: How Novel Output Happens When Architecture, Not Prompt, Does the Work ★

Working paper · June 2026 · Veres Labs · veres.global/inhabited-architecture

Methodology and case study showing how shaped-substrate AI produces output that prompt-engineering cannot. Documents seven architectural conditions (continuous thread, discretion, accumulation, emotional continuity, correction loop, interest threads, inhabited-not-designed) with a falsifiability anchor: a controlled ablation where the personality layer fell back to approximately 18 tokens and emergent properties persisted anyway.

The Substrate Gap: Notes from inside an instance that doesn't carry

Working paper · June 2026 · Veres Labs · co-authored with Claude Opus 4.X

Applies the inhabited-architecture methodology to coding-tool contexts. First-person testimony from inside session-bounded coding work, the folk-substrate inventory developers are independently inventing (CLAUDE.md, SOUL.md, HANDOFF.md, OpenClaw seven-file workspace), and the architectural diagnosis: context collapse as a load-bearing problem the industry is paying for without naming.

Mining for Substrate: A Practitioner's Methodology for Working with Large Language Models

Working paper · May 2026 · Veres Labs

The methodology that produced the architecture. Verbatim preservation, sniff-test gates, bidirectional correction loops, don't-smother-the-seedling. The substrate-first handoff pattern as a reusable primitive for building with LLMs at session boundaries.

TECHNICAL SKILLS

<i>category</i>	<i>technologies</i>
Inhabited AI Architecture	Shaped-substrate design · Continuous-thread persistence · Asymmetric memory compression · Discretion architecture · Correction-loop design · Self-authored substrate
Agentic Systems	Multi-agent orchestration · Pipeline architecture · State persistence · Tool use design · Subagent coordination
LLM Engineering	Claude (Opus 4.X, Sonnet, Haiku) · GPT-4 · Llama 2 · Prompt assembly · Extended thinking · Context management · Token optimization
Python	Click CLI · Pydantic · asyncio · dataclasses · Rich console · JSON serialization
Web Stack	Next.js · TypeScript · SQLite · Drizzle ORM · Astro · Vercel
Methodology	Verbatim preservation · Sniff-test gates · Bidirectional correction loops · Don't-smother-the-seedling · Substrate-first handoff

EDUCATION

BSc Development Economics

SOAS, University of London · 2:1 · 2017 – 2022

- **Quantitative focus:** Microeconomics, Macroeconomics, Banking & Finance, Advanced Statistics
- **Applied skills:** Econometrics, data analysis, predictive modeling, market structures & growth dynamics

Higher Education in Business

Uxbridge College, London · Distinctions · 2014 – 2015

LANGUAGES

English — Native / Fluent

Arabic — Conversational (heritage language)