Gianluca Mazza

Ruoli e responsabilità sono il mio resoconto professionale. Dove esiste un lavoro pubblico, collego l'artefatto per permettere di esaminarlo.
Canonical: https://gianlucamazza.it/it/curriculum
Porto sistemi LLM in produzione: stato, ripristino, eval, costi.
Lavoro su software e infrastrutture dal 2012. Come cofondatore e CTO di Scaling Parrots, ho guidato le scelte tecniche di prodotto e infrastruttura. Oggi aiuto i team tecnici a gestire stato, errori, qualità degli output e costi dei sistemi LLM.

Contatti

- Email: info@gianlucamazza.it
- GitHub: https://github.com/gianlucamazza
- LinkedIn: https://linkedin.com/in/gianlucamazza
- ORCID: https://orcid.org/0009-0005-1462-3019
- Hugging Face: https://huggingface.co/gianlucamazza

AI Lead · progetto riservato

Luglio 2026 – Presente · Italia
Guido lo sviluppo di una piattaforma AI per un progetto riservato, con un piccolo team: architettura, sviluppo assistito da agenti AI, CI/CD e rilasci.

Consulente AI indipendente · gianlucamazza.it

Ottobre 2026 – Presente · Italia
Progetto sistemi LLM per team che devono portarli dal prototipo all'uso quotidiano.
- Progetto sistemi agentici che restano ispezionabili: contratti tipizzati per i tool, stato persistente e percorsi di ripristino che resistono a un riavvio.
- Costruisco eval riproducibili, così l'effetto di un cambio di prompt o di modello è una differenza misurata, non un'impressione.
- Studio la memoria per agenti AI e pubblico benchmark riproducibili nel progetto emotional-memory.

Ricerca e consulenza AI indipendenti — Venere Labs

Settembre 2024 – Settembre 2026 · Italia
Fino a settembre 2026, Venere Labs è stato il nome con cui presentavo il mio lavoro indipendente di ricerca e consulenza AI su sistemi agentici ispezionabili e valutazioni riproducibili.
- Con quel nome ho lavorato anche al progetto emotional-memory.

Lavori pubblicati collegati:
- emotional-memory — DOI 10.5281/zenodo.19972258: https://doi.org/10.5281/zenodo.19972258

Stack: LangChain, LangGraph, MCP, OpenAI/Anthropic

Docente del corso di AI @ ente di formazione per il personale scolastico

Settembre 2026
Ho tenuto questo corso dal 7 al 10 settembre 2026.
- Corso per docenti e personale amministrativo.
- 12 ore / 4 moduli.

Co-fondatore e CTO @ Scaling Parrots

Aprile 2020 – Novembre 2025 · Brescia, Italia
Come cofondatore e CTO, ho diretto le scelte tecniche di prodotto e infrastruttura.
- Ho coordinato sviluppo, prodotto e infrastruttura lungo il ciclo di consegna.
- Ho curato revisione del codice, integrazione continua e preparazione dei rilasci.

Stack: Python, Rust, AWS

Consulente di prodotto e sistemi @ cliente industriale (consulenza esterna)

2025
Come consulente esterno ho progettato l'architettura di un sistema software e ne ho definito il prodotto, dai requisiti alla consegna tecnica.
- Ho fissato i confini del prodotto attorno a quell'architettura.
- Ho tradotto il progetto in un piano tecnico implementabile e ho supportato la consegna insieme al team del cliente.

Stack: Progettazione di prodotto, Architettura di prodotto

Sviluppatore software senior @ Bizonweb

Novembre 2019 – Aprile 2020 · Brescia, Italia
Ho costruito applicazioni web su misura per progetti di clienti, in base ai requisiti di ciascuno.
- Ho scritto in Node.js i servizi backend e le API REST usati da queste applicazioni.

Stack: Node.js, JavaScript, REST APIs

Esperienza precedente, 2017–2019: sviluppo software e infrastruttura, anche per clienti del settore finanziario

Sviluppatore software junior @ MYTI
Gennaio 2015 – Dicembre 2016 · Brescia, Italia
Ho costruito applicazioni web single-page responsive in AngularJS.
▪ Ho implementato componenti di interfaccia responsive con Angular Material.
Stack: AngularJS, Angular Material, REST APIs

Consulente software freelance
Aprile 2013 – Dicembre 2014 · Italia
Ho svolto consulenze software, incarichi brevi, senza nomi di clienti.

DevOps e amministratore di sistema @ Atic
Ottobre 2012 – Marzo 2013 · Brescia, Italia
Ho curato installazione, supporto e sicurezza dei sistemi per stazioni di servizio, incluse procedure di ridondanza, failover e disaster recovery.
Stack: Linux, Ridondanza, Disaster recovery

Sviluppatore Microsoft, tirocinio @ Lutech
Maggio 2012 – Settembre 2012 · Brugherio, Italia
Ho lavorato su tecnologie enterprise e cloud Microsoft durante un tirocinio.
▪ Ho sviluppato applicazioni sullo stack Microsoft .NET.
▪ Ho rilasciato servizi su Azure e integrato flussi aziendali attraverso BizTalk.
Stack: .NET, Microsoft Azure, BizTalk, SQL Server

Progetti pubblici selezionati
Codice, documentazione e risultati di ricerca pubblici.

emotional-memory
Livello di memoria per sistemi LLM con codifica dello stato affettivo, pacchetto PyPI, DOI Zenodo, prove di richiamo pre-registrate e risultati negativi pubblicati su LoCoMo, MADial-Bench ed ES-MemEval.
▪ https://github.com/gianlucamazza/emotional-memory
▪ https://doi.org/10.5281/zenodo.19972258
▪ https://pypi.org/project/emotional-memory/
▪ https://github.com/gianlucamazza/emotional-memory/tree/main/benchmarks
▪ https://github.com/gianlucamazza/emotional-memory#validation--benchmarks
▪ https://github.com/gianlucamazza/emotional-memory/releases

orka
Runtime self-hosted in Rust che riceve lavoro da Telegram, Discord, Slack, WhatsApp e HTTP, assegna priorità e instrada verso workflow LLM vincolati con supporto MCP/A2A. È una versione preliminare, per cui non offro SLA. Un ciclo opzionale che trasforma le esecuzioni passate in regole per i prompt è disattivato di default.
▪ https://github.com/gianlucamazza/orka

mklang
Linguaggio dichiarativo per macchine a stati guidate da LLM. Spec 0.4; l'interprete di riferimento 1.4.0 è stabile, con check, lint, test, doctor e una suite di conformità. L'accuratezza dei gate si misura, non è garantita, e non è garantito che le esecuzioni siano deterministiche. Il judge opzionale, offerto in hosting, è sperimentale ed è l'unica parte indicata come non pronta per la produzione.
▪ https://github.com/gianlucamazza/mklang
▪ https://docs.mklang.dev/
▪ https://huggingface.co/spaces/gianlucamazza/mklang-playground

Competenze tecniche

Sistemi AI
LangChain, LangGraph, MCP (Model Context Protocol), A2A (Agent-to-Agent), OpenAI / Anthropic / Gemini, Sistemi RAG, Database vettoriali, Progettazione dei prompt, Function calling (chiamata di funzioni), CrewAI, Pipeline di agenti, Server MCP, WASM, Inferenza locale (Ollama, llama.cpp), Redis

Dati e ML
PyTorch, Transformers, Hugging Face, Embedding e ricerca semantica, pandas / NumPy, Jupyter, PostgreSQL

Infrastruttura e sicurezza
Cifratura end-to-end (X25519 / Ed25519), Double Ratchet, AWS, GitHub Actions, Docker, Linux / Unix

Linguaggi e framework
Python, TypeScript, Rust, Go, JavaScript, C++, SQL, Bash, React, Node.js, GraphQL, Git / GitHub

Formazione
▪ Informatica, Facoltà di Scienze Matematiche, Fisiche e Naturali — Università degli Studi di Milano-Bicocca, 2009 – 2011, Milano, Italia, Percorso universitario non completato
▪ Liceo scientifico — Liceo Luzzago, 2003 – 2009, Brescia, Italia

Lingue
▪ Italiano — Madrelingua
▪ Inglese — Competenza professionale