



ALAN BLANCHET



Grenoble, France

contact@alan-blanchet.fr

+33 7 81 86 62 87

alan-blanchet.fr

github.com/AlanBlanchet

linkedin.com/in/alanblanchet

AI Engineer — Agents, LLM Systems & ML

Keen to join Mistral — to help strengthen its sovereign-AI offering.

French — native English — bilingual (3 yrs UK)

5+ years in AI / ML

~€900k+ across delivered projects

Builds & runs the team's AI-agent stack

PROFILE

Largely self-taught AI/ML engineer, 5+ years — production computer-vision, speech and LLM systems, now centred on AI agents and LLM apps end-to-end, with a habit of digging into how models actually work. I hold myself to clean, heavily-generalized, strongly-typed code — built with agents, but not vibe-coded. I care about AI being trustworthy and well-governed, and about Europe not being squeezed between the US and China — hence a strong drive to help build a serious sovereign offering. Pragmatic about it, though: validate with the best-performing models first to remove risk, then move toward sovereign ones where they hold up.

CORE STACK

- Agents & LLMOps**
 - LiveKit Agents
 - LangGraph
 - OpenAI Agents SDK
 - MCP (builds servers)
 - tool-calling
 - LLM-as-judge
 - Langfuse
 - LiteLLM
 - Ollama
 - LM Studio
 - AGENTS.md / CLAUDE.md
- LLMs**
 - Qwen
 - Llama
 - DeepSeek
 - Mistral
 - GPT-OSS
 - Claude / GPT / Gemini (API)
 - RAG
 - fine-tuning / LoRA
 - MoE
 - KV-cache
 - quantization
 - vLLM
- Speech — STT**
 - Whisper
 - faster-whisper
 - NVIDIA Parakeet
 - Kyutai
 - Silero
 - AssemblyAI
 - Deepgram
- Speech — TTS**
 - ElevenLabs
 - Cartesia
 - Deepgram Aura
 - Kokoro
 - Piper
 - Voxtral
 - Kyutai / Moshi
- Audio & VAD**
 - Silero VAD
 - EBEN
 - HiFi-GAN / ++
 - MP-SENet
 - xLSTM-SENet
 - VoiceFilter
 - DTLN
 - RNNNoise
 - SpeechBrain
 - PESQ / STOI / SI-SDR
- Vision**
 - RT-DETR
 - D-FINE
 - YOLO (v8/NAS)
 - DETR
 - SAM
 - SegFormer
 - U-Net
 - DINOv2/v3
 - iBOT
 - SwAV
 - BYOL
 - EquiMod
 - ViT / Swin
 - CLIP / SigLIP
 - Mask R-CNN
- Image matching**
 - GIM
 - RoMa
 - DKM
 - LoFTR
 - LightGlue
 - SuperPoint
 - OmniGlue
 - RANSAC / MAGSAC
 - homography / TPS
- Inference · RL**
 - TensorRT
 - ONNX / graph surgeon
 - INT8 / FP8 quant.
 - CUDA / Triton
 - batching
 - einops
 - REINFORCE / VPG
 - DQN family
 - R2D2
 - PPO
 - RND
 - NGU
 - PER
- Systems & cloud**
 - Python
 - Rust
 - Scaleway (GPU)
 - AWS
 - Linux / systemd
 - Docker / K8s
 - SLURM (Jean Zay)
 - Nginx / Caddy / Traefik
 - MLflow / W&B

AI meter — how each item below was built: ■ ■ ■ ■ hand-written ■ ■ ■ ■ built with my orchestrated AI agents

Both modes in play — and the agents themselves are built in-house.

EXPERIENCE — R&D ENGINEER · NEOVISION, GRENOBLE · 2022–PRESENT (WORK-STUDY 2022–24, THEN PERMANENT)

AI / ML at the core, with regular cross-functional support — networking & infra, full-stack, account administration (admin).

Conversational voice-AI agent — real-time telephony **LEAD TECH** €300k+ project 2025–present AI ■ ■ ■ ■ ■

Real-time phone agent, medical appointment scheduling · pipeline built from scratch → migrated to LiveKit (with the team) · per-call LLM/STT/TTS with fallbacks · tool-calling · LLM safety check · full observability · in production · in-house demo of the full range, incl. live voice cloning on a call.

- LiveKit
- LangGraph
- multi-provider STT/TTS
- SIP telephony
- Langfuse

Internal AI enablement & agent tooling

2024–present AI

Help teammates test or experiment with an AI via API · in-house tooling — MCP servers · multi-provider routing · self-hosted **LM Studio** server (2× RTX 3090, on-prem/confidential) · a prompt/standards system that governs how agents write code · partially automated the company's **R&D tax-credit (CIR)** report generation (2025, reused 2026).

MCP servers local LLM server (on-prem) Ollama / LM Studio AI enablement coding standards as prompts

Multi-head image-classification system **LEAD TECH** €65k project

2026 AI

Multi-head image classification (DINOv2 + VLM), delivered for a client · what mattered most: the **home-built agentic system** used to build it.

DINOv2 multi-head classification agent-built FastAPI / Gradio

Speech enhancement & reconstruction framework **LEAD TECH** €85k project

2025 AI

Rebuilds clean speech from in-ear microphones · several enhancement models (EBEN · HiFi-GAN/++ · MP-SENet · xLSTM-SENet · VoiceFilter) behind one pipeline · mouth-to-ear acoustic-transfer data simulator · PESQ/STOI/SI-SDR · shipped (FastAPI · Gradio · Docker).

PyTorch Lightning torchaudio GANs audio denoising FastAPI / Gradio

Technical mentoring & teaching — internship supervision & student coaching

2025

Supervised a Master's research internship on **LLM-based agents** (defended, Grenoble INP/UGA) · B2B company-data extraction & validation agent · benchmarked LLMs × MCP search tools · coached **5 engineering-student teams at ESISAR** (Grenoble INP) on industrial AI projects — incl. a real-time speech-transcription tool needing domain-vocabulary performance.

internship supervision student coaching LLM agents MCP search ASR / domain vocabulary

Image matching & registration — visual / template matching **LEAD TECH** ~€80k project

2025 AI

Image registration / template matching for product authentication · keypoint matching (GIM-DKM · RoMa · LoFTR · LightGlue · SuperPoint · OmniGlue) → RANSAC/MAGSAC → homography/TPS warping · robust to smartphone capture (lighting · blur · perspective) · custom CVAT keypoint-matching annotation plugin · GTE/CTE accuracy metrics.

keypoint matching RANSAC / homography CVAT plugin GTE / CTE PyTorch

Internal ML-demonstrator platform — hosts demos & internal solutions **LEAD TECH**

2024–26 AI

Hosts the company's client demonstrators + internal solutions · 25+ apps (classification · detection · forecasting · recommendation · photogrammetry) · each deployed/managed from Docker images, shareable links · per-demo orchestration · reverse proxy · role-based access.

Next.js PostgreSQL / Prisma Docker orchestration Scaleway

Real-time object detection — industrial inspection

2024 AI

In-depth study + ablation of SOTA detectors (RT-DETR · YOLO-NAS · YOLOv8) → selected **RT-DETR** (then SOTA) · edge deployment (ONNX · TensorRT · INT8/FP8 quant.).

RT-DETR TensorRT / ONNX quantization MLflow / DVC

DATAWISE — self-supervised-learning benchmark — R&D, on a national supercomputer

2024–25 AI

LEAD TECH ~€400k programme

Benchmark of modern self-supervised vision models (DINOv2 · DINO · iBOT · SwAV · Barlow Twins) across ImageNet-1K/22K · CIFAR · Food-101 · SUN397 · COCO · clean modular backbone/neck/head · trained on the **Jean Zay** national supercomputer (SLURM · Hydra-Submitit sweeps · Lightning DDP) · also reproduced EquiMod (BYOL + LARS). A **~€400k** programme funded by the **Région Auvergne-Rhône-Alpes** — currently unfinished and on hold.

SSL PyTorch Lightning (DDP) SLURM · Jean Zay

Image + text search platform — pattern catalog **€500k project**

2022–24 AI

Lead front-end, €500k production platform · visual pattern extraction (**Mask R-CNN**) · image-similarity + free-text search (CLIP/OpenCLIP) · also built parts of the embedding-search backend.

React / TypeScript Mask R-CNN CLIP / OpenCLIP AWS semantic search

Clinical bacteria-classification R&D — co-author on the resulting paper

2023–24 AI

Co-lead ML · segmentation + classification of bacterial colonies vs classical baselines (SVM · K-fold) · scientific report · co-author on the peer-reviewed SPIE paper (below).

segmentation SVM baselines cross-validation scientific writing

~200

"R&D News" — model watch, briefed to the whole team

A short briefing every 2 weeks on the notable new models, architectures & techniques. Close follow of the SOTA — LMArena · ARC-AGI · SWE-bench...

SELECTED PROJECTS — mostly personal & open-source; more on GitHub

Multi-agent prompt & standards system



The project I'm most attached to: it compiles my engineering standards (generic, flexible, tidy code) into reusable skills, sub-agents and prompts across Claude Code / Codex / Cursor.
personal · the core of how I build

interact



An MCP server that lets AI agents see and act on a screen — browser and desktop — by what's visible, not brittle selectors. Built to make my own agents better.
github.com/AlanBlanchet/interact

From-scratch deep-learning framework



ResNet / ViT / DETR (with Hungarian matching) and the DQN / PPO family, reimplemented by hand with a hand-written autograd engine — and several of them re-built in Rust on LibTorch.
github.com/AlanBlanchet/AI-4-Alan

Cross-platform learning app



Flutter + Rust app with an LLM content pipeline — local-embedding dedup, batched and cached inference.
personal · full-stack

Rust compute & visualization engine



A framework-agnostic engine with GPU/SIMD kernels and Python / JS / WASM bindings — built almost entirely with my AI agents.
github.com/AlanBlanchet/any-compute

Recruitment-triage agent



Parses CVs & cover letters, classifies candidates with an LLM, files them and notifies the team — the kind of agent that may well be reading this CV.
internal · scheduled agent

Agent-built bilingual CMS site



A bilingual content site a non-technical team edits themselves — scaffolded and filled in about a day, almost entirely with Claude agents. A small marker of how much agents speed me up.
client · agent-built in ~a day

linux-commands



My published Linux toolkit — self-documenting shortcuts, incl. an interactive multi-branch rebase.
github.com/AlanBlanchet/linux-commands

HOW I LEARN & PUBLICATIONS

HOW I LEARN

Largely self-taught — papers, online courses, strong technical communities, coding certificates. Hands-on: CUDA versions · ONNX & graph nodes · quantization · inference optimization · thousands of hours on Linux, servers & deployment. A compulsive tool-builder who likes helping the team.

PUBLICATIONS

Co-author (2nd of 6) on a peer-reviewed SPIE paper — CNN / SVM classification on biomedical multispectral imaging (2025). DOI: [10.1117/12.3097794](https://doi.org/10.1117/12.3097794).

EDUCATION

- **ML Engineer** (work-study) — OpenClassrooms, 2022–24
- **AI-oriented robotics** (work-study) — IMERIR, 2021–22
- **DUT Computer Science** — IUT Montpellier-Sète, 2019–21
- **Baccalauréat S, mention Bien** — 2019 · Cambridge B2

Online ML degree chosen deliberately — local engineering schools weren't AI-focused enough; the online route freed the time to experiment widely and level up far faster.

Open to Paris — relocation or weekly commute.