



我们在招聘

一起做 端侧大模型推理芯片

公司成立于2026年7月，已完成数千万元融资。NPU/AI 芯片团队在杭州招聘正式岗位与实习岗位，覆盖架构、RTL、验证、编译、Runtime、SoC 和模型部署，欢迎愿意把软硬件细节真正做通的人。

2026.07

杭州主体成立

数千万元

融资已完成

Qwen-8B

完成 FPGA 端到端验证

杭州

正式岗位 + 实习岗位

招聘方向

工作地点杭州 · 正式岗位 + 实习岗位

架构 / 算法

NPU 架构与微架构、AI/HW-SW Co-design、性能建模与架构仿真。

RTL / SoC / Memory

RTL/NPU 设计、NoC/DMA、AXI/CHI、SRAM/LPDDR/HBM 与系统集成。

验证 / 性能

UVM 验证、覆盖率与断言、cycle-level simulator、工作负载与瓶颈分析。

编译 / Runtime / Driver

Graph/IR、算子融合、Tiling、CodeGen、Runtime、Driver 与 Firmware。

后端 / DFT / 原型

综合、STA、P&R、DFT、FPGA 原型、Emulation 与系统级验证。

AI Runtime / HR / 实习

模型部署、算子库与 Profiling；HR，以及全部研发方向实习岗位。

我们在做什么

FusionCore

面向大模型推理的多精度矩阵、向量与片上存储计算 IP。

ScaleNoC

连接计算核心、存储与接口，按数据流组织片上通信。

TileFlow

完成分块、布局、调度、代码生成和运行时适配。

VLX

面向已有工控机、算力盒与电子设备的推理加速产品。

我们期待你

- 基础扎实，能阅读代码并独立完成小模块开发。
- 愿意主动定位问题、调试并推进工程闭环。
- 对 NPU、AI 芯片、体系结构、编译器或系统软件有真实兴趣。
- 课程、科研、开源、竞赛、FPGA、RTL 或 Compiler 项目均可作为证明。

简历投递

careers@veloxis-tech.com

邮件标题建议：姓名 + 应聘方向 + 正式/实习。欢迎芯片、编译器、AI Infra、系统软件和硬件工程背景候选人交流。