AI Agent

简介

张益新 男\1991

E-**mail:** zyxcambridge@gmail.com **| TEL：**18810901685 **|** Location：Shanghai

2010.09-2014.06 **|** 北华航天工业学院 **|** 全日制本科

AI Agent项目

- 百川智能特别奖-美国AI中医获客Agent
    - 私人诊所的AI经营助手Agent
        - 1、负责接听电话，负责回复微信
        - 2、回答患者问题，与患者交流，完成约诊
        - 3、端到端和传统混合架构，多Agent
- 找工作copilot-世界人工智能大会Hackathon 二等奖
    - 求职者Agent 和招聘方agent 自动匹配，完成初筛
    - 提高面试次数，每天完成百万次初筛面试
    - Web Agents ,自动投递简历
- 自动驾驶项目代码补全助手
    - 基于star code 微调自动驾驶项目，进行代码补全
    - 公司内网部署自己微调的模型，多部门使用
- **Blender 3d 建模Agent**
    - **开发3d建模 AI agent，对话方式完成建模调整，精细化程度超过diffusion**
    - https://zyxcambridge.github.io/blender_llm/

工作经历

安波福中央电气（上海）有限公司 2024.8 - 2025.2

Machine Learning Engineer

- 参与 L2++ 无图 端到端网络设计方案

- 负责芯片选型，三种芯片的 benchmark，设计端到端网络多任务调度框架

- 拆分模型部署工作形成标准和 work flow,独立完成 OD 和 OC 分支在 2 种芯片 AI 模型部署

图达通智能科技(上海)有限公司 2022.6-2024.04

Algorithm Engineering Leader /Senior Software Architecture

- 招聘组建高效的深度学习部署团队，合作完善算法工程落地流程

- 基于 Jetson xavier，orin 实现 AI LiDAR 模型实时推理，集成,时间延迟降低 4 倍

- 大模型推理部署，AI Agent 在公司内部落地探索

  - 成就：开发专用版本的代码补全 Copilot，荣获公司创新之星奖

  - 技术：成功部署 Ollama、llama.cpp、vllm、TensorRT-LLM、mlc-llm 等模型

- 基于 NVIDIA Drive Smi 完善 MLOps 流程，产生 10W 帧仿真数据集

  - 使用 RoadRunner，SimReady、 Replicator 建立自动化的模型训练、评估和部署流程

  - 实现模型仿真数据和真实数据的联合训练,精度提高 10 个点

Shanghai Xuehu Information Technology Co., Ltd 2019.05-2022.05

Senior Algorithm Engineer

- 从 0 到 1 搭建算法团队，构建 V2X 感知算法工程化落地流程

  - 挑战：激光雷达点云数据量巨大，FPGA 算力低，只有整数类型算力；落地环节太多

  - 行动：设计 FPGA 混合量化的方案，拆分网络结构每一层，精细到乘法和加法运算级别

  - 行动：搭建一整套的算法落地流程，出差实际项目现场驻场开发；

  - 结果：MEC 设备量产 100 台以上，软硬一体变现 200 多万+300W，客户近 10 家。

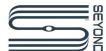**DeepBlue Technology (Shanghai) Co., Ltd   2018.05-2019.2**

Algorithm Engineer

- 新零售业务上密集场景商品关键点检测算法，优化推理过程

  - 挑战: 2 张 P100 要能支持客户不断增加的模型种类和数量

  - 行动：设计多 GPU 的预加载方案，结合线程池

  - **结果** :推理时延控制在 500ms 以内，单卡推理 28 个模型,服务 1000 台设备。

作品集-在多个场合进行20多场公开演讲，影响超过1W人

Date	Lecture Topic	Audience	Description and Links
May 18, 2024	Build With AI 2024 Shanghai Phase II Successfully Closed	100 Trained	https://mp.weixin.qq.com/s/bXNZlkXZgGJOzg9UIM6l2Q
Apr 25, 2024	Build your own private big model based on Gemma in three steps	100 Attendees	https://ctalks.gdgcn.net/blog/develop-1/2024nian-4yue-25ri-qing-liang-aimo-xing-xin-dong-xiang-gemmaji-qi-qian-yan-ying-yong-chang-jing-42
Dec 23, 2023	LLM from 0 to 1	200 Attendees	https://mp.weixin.qq.com/s/46NTXfpV0t9J_ST_oVtdxA
Oct 21, 2023	what is the pattern in AI era ?	196 Attendees	https://1131329524641.huodongxing.com/event/2724257198300
May 27,2023	Google Bard use skills	300 Attendees	https://mp.weixin.qq.com/s/42ZsF9-Ag8OX9Lv8_Zp64Q
Oct 25, 2022	Google Startup Training Camp: Regenerative Doctor X Mentor Zhang Yixin	6 Trained	
Sep 24, 2022	As a GDE Interviewed by Zhejiang TV	Many	http://tv.cztv.com/vplay/1273834.html
Sep 21, 2022	Data - driven algorithm project landing	370 Attendees	https://mp.weixin.qq.com/s/IKB_CImTe7qmxCrV0zlq8w
Mar 30, 2022	V2X-Data-driven algorithm engineering landing	1.1k Attendees	https://www.huodongxing.com/event/7642881313923
Dec 25, 2021	AlphaFold2 TensorFlow	719 Attendees	https://mp.weixin.qq.com/s/gjXz9JKrOihKNZmIuEzEZA
Sep 19, 2021	Data-driven algorithm engineering landing	1.4k Attendees	https://live.bilibili.com/21482458?share_source=wechat
Aug 1, 2021	2021 Google Google Entrepreneurship Bootcamp		Served as a technical instructor in the google entrepreneurship training camp and provided technical guidance for 2 months
Apr 29, 2021	TensorFlow create an Op	900 Attendees	https://live.bilibili.com/22357137?share_source=wechat
Sep 26, 2020	3D LiDAR Deep Learning End-to-Side Solution	5k Attendees	https://mp.weixin.qq.com/s/P8TiYof0q4vYp-BLGmMvZg
Mar 6, 2021	Machine learning Study Jam Season 3 Office Hours	662 Attendees	https://live.vhall.com/v3/lives/watch/483151815?invite=
Jan 8, 2021	TensorFlow Model Optimization: Model Quantization	1.9k Viewers	https://mp.weixin.qq.com/s/9QeVVESP3_rBZ6n_D96lwg
Feb 22, 2020	TensorFlow implements FRCN	500 Attendees	https://edu.csdn.net/learn/27847/378997
Sep 8, 2018	Application of human key point detection in object detection	50 Attendees	https://gdg.community.dev/events/details/google-google-developer-group-gd

【Selected Achievements and VP Recommendation Letter】

SEYOND

Seyond
160 San Gabriel
Drive Sunnyvale
CA 94086
USA
(650) 963-9573

Re: Yixin Zhang
Date: Mar 15 2024

To Whom It May Concern,

I am writing to wholeheartedly recommend Mr. Yixin Zhang for a position with your organization. As Mr. Zhang's supervisor in the software development team at Seyond China for the past two years, I have been consistently impressed by his proactive approach, comprehensive technical expertise, and leadership abilities.

Mr. Zhang has played a pivotal role in leading his agile deployment group, addressing and resolving complex user requirements with efficiency and skill. His comprehensive experience in AI deployment is particularly noteworthy. He was instrumental in integrating our company's first Autonomous Drive Perception System into our test vehicles—a challenging project that he managed with remarkable competence. Furthermore, Mr. Zhang has effectively trained members of the AI deployment group, significantly enhancing our team's capabilities and performance.

His dedication and meticulous attention to detail have been invaluable to our team's success. Mr. Zhang's ability to handle demanding tasks and his unwavering commitment to achieving excellence are traits that will undoubtedly make him a great asset to any team.

I recommend Mr. Zhang without any reservations. Should you require further information, please do not hesitate to contact me via email at

张益新

机器学习工程师，GOOGLE DEVELOPER EXPERT

他/他

g.dev/updatezhang

Copy link

🌐 Public

城市/城镇

📍 Shanghai, China

经验等级

中等（6 - 10 年）

统计信息

⊛ **20**·赢得的徽章

📄 **4**·保存的页面

已收藏的徽章 ⓘ

2024 年 Google 开发者专家	GCP Google 开发者专家	2025 年 Google 开发者专家	2022 年 Google 开发者专家	使用 Gemini API 构建面向未来的产品
2025年4月17日	2025年4月17日	2025年4月17日	2025年4月17日	2024年5月16日

徽章

排序方式 ↑↓

2023 年 Google 开发者专家	学习	Cloud Shell in Cloud 文档	Gemini Code Assist 代理（非公开预览版）	Google Cloud Innovators	Google 开发者计划高级方案
2025年4月17日	2024年11月13日	2025年4月6日	2025年4月17日	2021年10月15日	2023年8月21日

2022 年 Google 开发者专家	Google 开发者专家成员	Firebase Studio 开发者社区	2021 年 I/O 大会的参加者	完成了 1 项以上的 Codelab	2025 年 Google 开发者专家
2025年4月17日	2023年6月22日	2025年4月17日	2021年4月9日	2021年6月25日	2025年4月17日