

简历

个人信息

- 姓名: 潘海涛 (Pan Haitao)
- 位置: 中国上海
- 当前状态: 离职
- 电话: 19286470192
- 邮箱: manbuzhe2009@qq.com
- **LinkedIn**: www.linkedin.com/in/haitaopan
- 语言: 中文 (母语), 英语 (会话)
- 个人主页: <https://www.svc.plus> (已上线 **ASKAI 助手** Demo)

教育背景

时间范围	最高学历	学校	专业
2006.9 ~ 2010.6	本科	长春工程学院	电气工程及其自动化

期望职位

- AI 应用工程师 / LLM Engineer / AI Agent Developer
- AIOps / MLOps 工程师 (偏应用侧)
- 平台工程师 (LLM 应用与 DevOps 一体化)

技能概述

- 14 年基础设施与平台工程经验 (制造业/金融/运营商/互联网), 近年聚焦 **AI + 运维/业务** 场景落地。
- 具备 **端到端 AI 应用工程** 能力: 数据采集 → 检索索引 → RAG/Agent 设计 → API/服务化 → 观测与灰度 → GitOps/CI/CD 上线。
- 精通 **Kubernetes、GitHub Actions、GitOps、Terraform/Pulumi**, 能把 **LLM 应用** 与 **云原生平台工程** 无缝打通 (开发 → 测试 → 部署 → 监控)。
- 在 **可观测性与 AIOps** 场景有实践: eBPF / OpenTelemetry / 向量检索 (pgvector) + LLM 做 **根因分析、Runbook 问答、告警去重与联动**。

核心技能

- **LLM & Agent**
 - 应用框架: LangChain / LlamaIndex / OpenAI Assistants API / function calling
 - Agent 设计: 工具编排、检索链路、长短期记忆、任务分解、Plan & Execute
 - 模型与服务: OpenAI / Claude / 本地 Qwen/Llama / Ollama / vLLM / Text & Embeddings API
 - **RAG 与检索**
 - 嵌入与重排: bge-m3 / e5 / 常用 rerankers
 - 向量库: PostgreSQL + pgvector
 - 文档工程: 切分策略、元数据索引、段落级召回、热/冷数据分层
 - **MLOps / AIOps (偏应用)**
 - 数据/特征: ETL、日志/指标/链路到知识库的近线导入
 - 评测与对齐: 检索命中率、答案正确率、延迟/成本监控、提示词回归测试
 - 线上治理: Prompt 版本化、Safety Guardrails、灰度与回滚
 - **平台与自动化**
 - 云原生: Kubernetes / Helm / ArgoCD / FluxCD
 - CI/CD: GitHub Actions / GitLab CI / Jenkins (AI 应用构建、评测、发版流水线)
 - IaC: Terraform / Pulumi (AWS/GCP 多云)
 - **编程与后端**
 - Python、Shell (熟悉 Go、Rust、JavaScript)
 - FastAPI / Flask / gRPC / WebSocket、任务队列与缓存 (CRedis)
 - 监控与日志: Prometheus / Grafana / OpenTelemetry / DeepFlow
-

代表性 AI 项目与成果 (精选)

1) AIOps Runbook 智能问答 (RAG + Agent)

- **场景:** 把告警、Runbook、变更记录、仪表板说明等知识整合为检索库, 支持 SRE 查询与自动操作建议。
- **技术:** LangChain、pgvector、bge-m3、BM25 混合检索、ArgoCD 只读工具、Grafana API。
- **流水线:** GitHub Actions 每日增量索引; PR 合并触发评测 (检索/答案正确率回归)。
- **效果:** 常见故障定位时间下降 30%+; 值班交接问答平均响应 < 1.5s (缓存命中场景)。

2) Infra 文档—代码—配置联动的 DevRel Copilot

- **场景:** 查询 Terraform 模块、K8s Helm Values、平台 FAQ, 生成变更 MR 草稿。
- **技术:** LlamaIndex + pgvector、代码/配置分块策略、Schema 约束、GitHub Actions PR Bot。
- **效果:** 规范性变更的 MR 准备时间从小时级降到分钟级。

3) 观测数据问答 (日志/链路/指标统一侧写)

- **场景**: 对接 OpenTelemetry/DeepFlow 索引, 做「自然语言 → 查询面板/Trace 链路」的导航。
 - **技术**: Router + Toolformer 思路, 工具函数: `query_metrics()` / `search_traces()` / `get_logs()`。
 - **效果**: 一线同学可自助拉取上下文, 减少 SRE 手工查询。
-

个人作品与代码片段 (可展示)

- **ASKAI (在线 Demo)**: <https://www.svc.plus> 一对话式 AI 助手演示, 支持知识问答与工具编排。
 - **XScopeHub (AIOps 套件)**: Vector/OpenTelemetry/PG 集成的近线 ETL 与检索服务 (展示 RAG 数据入口与评测脚本)。
 - **XCloudFlow / XConfig**: CI/CD + IaC 模块化, 适配 LLM 应用的灰度/回滚与资源治理 (展示 GitHub Actions 工作流)。
 - **Navi (桌面 AI 助手)**: 任务引导、知识问答、工具编排 (展示 Agent 工具函数与对话记忆管理)。
-

工作经历

北京云杉世纪网络科技有限公司 高级售后服务工程师 2024.11-2025.09

- **技术栈**: Linux、Kubernetes、DeepFlow、NPM & APM、eBPF、混合云
- **职责**:
 - 负责客户现场网络可观测性系统的部署、调优与运维
 - 跟进采集器、DF Server 与存储查询集群的优化与迁移
 - 协助客户完成新技术 PoC (AI Agent、混合云环境观测)
- **重点项目成就**:
 - 甜橙金融项目: 优化采集器性能, 解决高并发丢包, 支撑南京区域 DF Server 迁移与扩容, 信创改造, 完成硬件兼容性调优与标准化维保手册输出
 - 中远海运 AI Agent POC: 完成模型接入、容器化部署与 QA 测试方案, 交付复用文档
 - 其他客户项目, 上海电力, 浙江移动, 江苏电信, 麦糖 (互联网客户)

特斯拉 (上海) 有限公司 站点可靠性工程师 2024.01-2024.05

- **技术栈**: Linux, Kubernetes, GitHub Actions, GitOps (ArgoCD), Ansible, Helm, Jenkins
- **职责**:
 1. 通过监控维护产线工业控制主机、虚拟机和 Kubernetes 集群, 确保核心生产线的稳定性和高可用性。
 2. 快速响应并解决生产软件和基础设施中的故障, 最大限度地减少生产停机时间。

3. 实施监控和警报机制，快速检测 and 解决问题。
 4. 与开发团队合作，使用 GitHub Actions 和 ArgoCD 优化部署流程。
 5. 利用 Ansible、Helm、Jenkins 和 GitHub Actions 进行软件和系统的常规更新，提升安全性和性能。
- **具体事项:**
 1. **开发 Jenkins 流水线，提升 Ignition 软件和边缘系统的部署与升级效率:** 负责开发 Jenkins 自动化流水线，专注于边缘软件及工业控制系统（如 Ignition）的部署和升级。该流水线覆盖了从编译、打包到部署和回滚的全过程，自动化了边缘设备的更新操作。此改进显著提高了系统升级的速度与可靠性，减少了人为操作错误，并确保工业生产线的持续稳定运行。
 2. **搭建 ArgoCD Server 和开发流水线模板:** 独立搭建 ArgoCD Server，并设计和开发 GitHub Actions 与 ArgoCD 的自动化流水线，用于 ITSM（IT 服务管理）系统的 DMP（数据管理平台）项目。该流水线模板涵盖从代码提交到生产环境的全流程交付，显著减少了部署时间，并提高了部署过程的一致性。
 3. **编写 Ansible Playbook，将 Grafana 配置代码化:** 编写面向 Grafana 的 Ansible Playbook，自动化配置警报仪表盘和监控系统。将 Grafana 的告警规则、仪表盘配置等以代码方式管理，实现了监控配置的自动化和可重复部署，减少了人工操作带来的配置差异和错误。
 4. **通过 GitHub Actions 和 ArgoCD 重新设计发布流程，优化 ITSM 系统 CI/CD:** 通过整合 GitHub Actions 和 ArgoCD，优化 ITSM 系统的 CI/CD 流水线，使自动化部署过程更加顺畅和安全，减少了人工干预和潜在的发布故障。重新设计的发布流程覆盖了从代码开发到生产环境部署的各个环节，提升了发布效率和可靠性。

华讯网络系统有限公司 高级云解决方案架构师 2022.03-2023.10

- **技术栈:** AWS/Alicloud, IAC(terraform), Gitlab CI / GitOPS(FluxCD)
- **职责:**
 1. 参与云迁移项目、自动化基础设施部署及混合云解决方案的设计与实施。
 2. 带领团队实施 DevOps 和 GitOps 实践，确保持续交付和稳定性。
 3. 开发和优化 Terraform 和 Pulumi IAC 模板，管理云基础设施。
 4. 指导团队进行监控和故障排除，使用 Prometheus 和 Grafana 进行实时监控。
- **成就:**
 1. 成功交付多个云迁移项目，帮助客户在 AWS 和 AliCloud 上建立混合云环境。
 2. 实施 GitOps (FluxCD)，显著提高了部署效率，并减少了人为干预的错误。
- **客户项目:**
 1. **Roche MSP 维护项目:**
 - 使用 GitOps 工具 FluxCD 管理 EKS Add-on、Datadog Agent 等组件，确保集群的高效运行。
 - 结合 RunDesk 自动化工具和 GitLab CI 流水线，实现 AliCloud ECS 的自动化申请，并结合 ITSM 审批流。
 - 重构优化 EKS IAC 模块，拆分 Terraform 模块为 EKS 集群、节点和 Add-on 管理。
 - 使用 Python Jinja2 模板替代 Terraform 默认变量文件，结合 GitLab CI 实现更灵活的工作流。

- 为 EKS Terraform IAC 模块新增特性，支持集群的导入迁移、升级及变更操作。
 - 重写 GitLab CI 流水线，添加状态变量机制，实现对集群创建、升级和变更的灵活控制。
2. **先正达种业 DevOps 平台设计项目：**
- 交付基于 AWS EKS 和 GitLab 的 DevOps 平台设计方案，支持自动化应用部署和持续集成。
 - 为平台引入 Prometheus、Grafana、DeepFlow APM 等监控工具，构建全方位的监控体系。
 - 提供基础设施即代码（IAC）流水线模板及应用流水线模板，提升部署自动化水平。
 - 参与编写并审核 Landingzone IAC 模块，涵盖 AWS 账户管理、网络、身份与访问管理（IAM）以及多种云资源（EC2、EKS、S3、RDS、Redis、MongoDB）等。
3. **联合力世纪云平台项目：**
- 为 AWS 和腾讯云的四个账户（测试/生产环境）规划子网，并创建标准化的 SIT、Dev、PRE、PRD 环境。
 - 交付基于 Prometheus、Grafana、DeepFlow APM 和 WireGuard 的跨云监控方案，实现对多云环境的统一监控。
 - 集成 Keycloak 作为统一身份管理服务，对接云账户、GitLab、Harbor 和 Grafana 等多种系统。
 - 提供 Ansible 配置仓库，自动化管理多云环境下的基础设施配置。

江苏博云科技有限公司 高级解决方案架构师 2021.12-2022.01

- **职责：**
 1. 提供技术咨询，帮助客户优化云基础设施并采用 Kubernetes 编排方案。
 2. 开发和部署基于 DevOps 的解决方案，提升 CI/CD 流程的自动化程度。
- **成就：**
 1. 成功帮助客户通过 Kubernetes 提升应用的可扩展性和稳定性。

UCloud 科技有限公司 高级解决方案架构师 2020.07-2021.11

- **技术栈：**云迁移、Kubernetes、混合云、分布式存储、网络安全、日志、监控、CI/CD、Ansible、Python
- **职责：**
 1. 负责推广 UCloud 的云产品及解决方案，提供技术咨询与实施支持，协助销售团队赢得客户项目。
 2. 设计、部署并优化客户的混合云架构，涵盖存储、网络安全、集群管理等关键模块，确保客户基础设施的高可用性和可扩展性。
 3. 参与大规模 Kubernetes 集群的迁移与管理工作，提升自动化运维能力，减少人工操作风险。
 4. 针对客户需求，定制并实施云原生解决方案，推动业务上云及持续集成和交付（CI/CD）流程的自动化。
- **客户项目：**

1. 核桃编程、百望云:

- 为阿里云自建 Kubernetes 集群迁移至 UCloud 制定详细的容器化应用迁移方案，保障数据安全性和业务连续性。
- 通过合理规划迁移过程中的资源调度，确保核心业务在迁移过程中不受影响。

2. Growing IO:

- 负责负载均衡器 (LB) 迁移项目，通过精心设计的流量切换方案，确保用户体验的平滑过渡。
- 利用 Ansible 实现自动化压测脚本，确保系统在高负载条件下的稳定性与可靠性，同时提升了验证效率。

3. 某区块链客户:

- 结合 Python 和 UCloud 云 API，快速自动化批量创建了超过 1000 台云主机，以满足客户大规模弹性计算需求。
- 优化了弹性扩展方案，确保在高并发条件下能够快速响应，提升了客户的业务处理效率。

灵雀云科技有限公司 交付工程师 2018.05-2020.06

• 职责:

1. 升级并管理客户的容器平台，实施 DevOps 流程优化。
2. 为金融行业客户提供 PaaS 平台的技术支持和容器化解决方案。
3. 规划并实施容器平台的监控、自动化运维和持续交付流水线，为客户定制容器化技术方案，提升平台的自动化运维能力。
4. 容器平台的版本升级和功能扩展，确保平台兼容性和未来演进的同时，与客户的 IT 和开发团队紧密合作，提供驻场支持和问题排查，优化性能并减少维护开销，保障业务连续性。

• 成就:

1. 完成光大银行容器平台的 6 次升级，提升了系统的稳定性和可扩展性。
2. 提升了金融行业客户的 PaaS 平台安全性，实施 DevOps 流程以支持快速交付。
3. 驻场运维支持多个大型项目，确保了客户平台的长期可靠运行，提升了客户满意度。

• 客户项目:

1. **中石油/东方地球物理公司:** 实施并运维梦想云 PaaS 平台，采用 DevOps 方法优化平台的持续交付流水线，并提升自动化运维效率。
2. **网上国网:** 提供 Kubernetes 集群的驻场运维支持，确保生产环境稳定运行。
3. **光大银行:** 驻场运维并主导 CPaaS 容器平台的多次升级，实施 DevOps 流程优化，实现了平台的稳定性提升和快速迭代。

深度科技有限公司 软件工程师 2015.05-2018.04

• 技术栈: Linux, Python, Shell 脚本, Docker, Jenkins

• 职责:

1. 参与 Deepin Server OS V15 版本的开发与维护，负责系统打包、编译和部署。
2. 编写自动化脚本，优化服务器版本的打包和编译流程，提高部署效率。
3. 为企业用户提供技术支持，处理远程支持请求，确保生产环境的稳定性。
4. 进行容器技术的前期研究，帮助客户设计定制的容器化解决方案。

• 成就:

1. 成功发布了多个稳定版本的 Deepin Server OS，显著缩短了发版周期。
2. 通过自动化脚本的改进，提升了系统的整体性能和部署效率，发版时间减少了 30%。

- **项目经历:**

- **项目名称:** Deepin Server OS V15 开发项目
- **项目描述:** 开发和维护 Deepin Server OS V15 版本，这是一个为企业环境设计的国产 Linux 操作系统，项目主要涉及系统的自动化打包、编译和部署，以及技术支持和优化。

知道创宇科技有限公司 运维工程师 2013.11-2015.04

- **技术栈:** Linux, Shell, Python, Saltstack, Nginx, Nagios, CDN 技术
- **职责:**
 1. 维护公司的 IT 基础设施，管理 CDN 部署并优化运维平台。
 2. 提供快速响应并解决紧急故障，减少停机时间。
- **成就:**
 1. 提升基础设施的稳定性和性能

浪潮电子信息有限公司 系统软件工程师 2013.05-2013.10

- **技术栈:** Linux, Apache, Bash 脚本, 监控工具
- **职责:**
 1. 维护和优化 Linux 服务器，执行定期更新和安全补丁。
 2. 使用 Bash 编写自动化脚本，提升系统的监控和管理效率。
- **成就:**
 1. 提升了服务器的可用性和响应速度。

中国标准软件有限公司 软件工程师 2011.05-2013.04

- **技术栈:** Linux, RPM 包构建, Koji 构建系统, Shell 脚本
- **职责:**
 1. 将 Linux 操作系统移植到龙芯硬件平台，开发 RPM 包构建和部署自动化工具。
 2. 使用 Koji 构建系统，实现软件包的自动化构建和发布。
- **成就:**
 1. 实施 Koji 自动化系统，提高构建效率