



无锡蔻町智能有限公司基于昇腾AI的原 生大模型解决方案

目录

1 代码大模型技术发展趋势

2 AIGCode昇腾原生基础大模型解决方案

3 AIGCode原生基础大模型优势

4 成功案例和规划

发展趋势

- 持续增长的参数规模: 随着硬件能力的提升和分布式训练技术的成熟, 代码大模型的参数规模不断增加。这种趋势使得模型能够捕捉更复杂的模式和特征。
- 多模态模型: 结合代码和其它数据, 比如自然语言描述、图像、或者其他形式的输入, 打造更强大的通用模型。
- 新的架构创新: 除了Transformer等流行架构, 研究人员还在探索其他神经网络架构, 以提高代码理解和生成的效率与准确性。
- 更高效的训练设施: 更新训练设施, 可以显著提升大模型迭代效率。

需求挑战

- 市面上多数是代码补充产品，缺乏端对端生成解决方案。
- 生成完整工程（End-to-End）的代码大模型在实践中面临许多独特的挑战。这些挑战并不仅仅在于生成单独的代码片段，而是涉及到整个软件生命周期以及复杂的项目管理。
- 长跨度代码的生成需要处理复杂的依赖关系，理解模块之间的依赖并正确引用相关资源和库。需要选择、下载和正确配置第三方库，这是一个复杂且容易出错的过程。

目录

1

代码大模型技术发展趋势

2

AIGCode昇腾原生基础大模型解决方案

3

AIGCode原生基础大模型优势

4

成功案例和规划

AI GCode方案

AI GCode在模型进行大规模的创新。

- 全新的代码transformer架构：
 - 自建数据集
 - 支持多专家
 - 自研共享网络
 - 自研训练优化器

AI GCode方案

打造新一代自动化训练平台

- 完整的训练基础设施管理，包括Ascend NPU，以及cpu，内存，存储资源进行池化弹性管理
- 支持数据集处理流水线
- 支持训练流程自动化
- 支持模型管理
- 支持算子优化

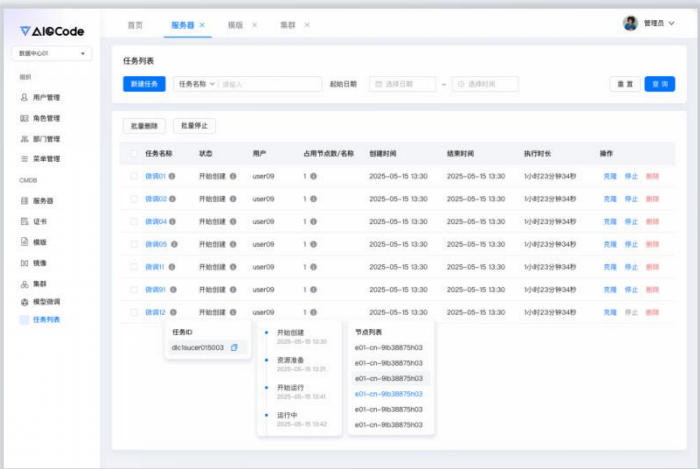
AIGCode平台

平台	数据管理	模型开发	模型训练	模型部署
	数据下载 数据分类 数据合成 数据标注 数据处理 质量管理	交互式建模 可视化建模 零代码建模	词表管理 预训练 指令微调 RLHF 模型优化 模型评估 全周期监测	服务部署 服务调用 推理优化 服务监测
	资产管理（数据集/模型/镜像...）		用户权限	
AI框架	机器学习框架（ Pytorch/Tensorflow/MindSpore/AIGFlow ...）			
	AIGSpeed			
加速库	昇腾CANN算法库		竞品算法库	
IAAS	云原生（容器化/弹性存储/智能调度/持续监测）			
	操作系统（OpenEuler/Ubuntu/Centos...）和基础硬件			

平台功能介绍

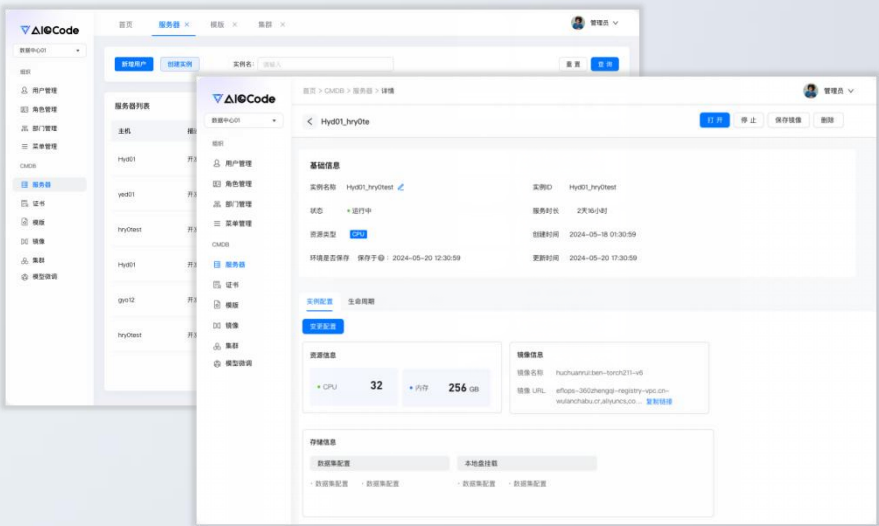
数据管理

支持数据集的下载、生成、标注，同时支持数据预处理流水线和构造新的数据集。支持图像、文本、视频、音频等不同类型数据标注，支持多模态数据标注；提供丰富的标注内容组件和题目组件，用户可以直接使用平台预置的标注模板，也可以自定义模板进行数据标注。



模型开发

平台支持模型低代码构造，同时支持交互式开发和远程开发功能，实现了从数据处理到开发调试、训练模型以及模型部署的AI开发全链路，提供一站式的深度学习开发服务。



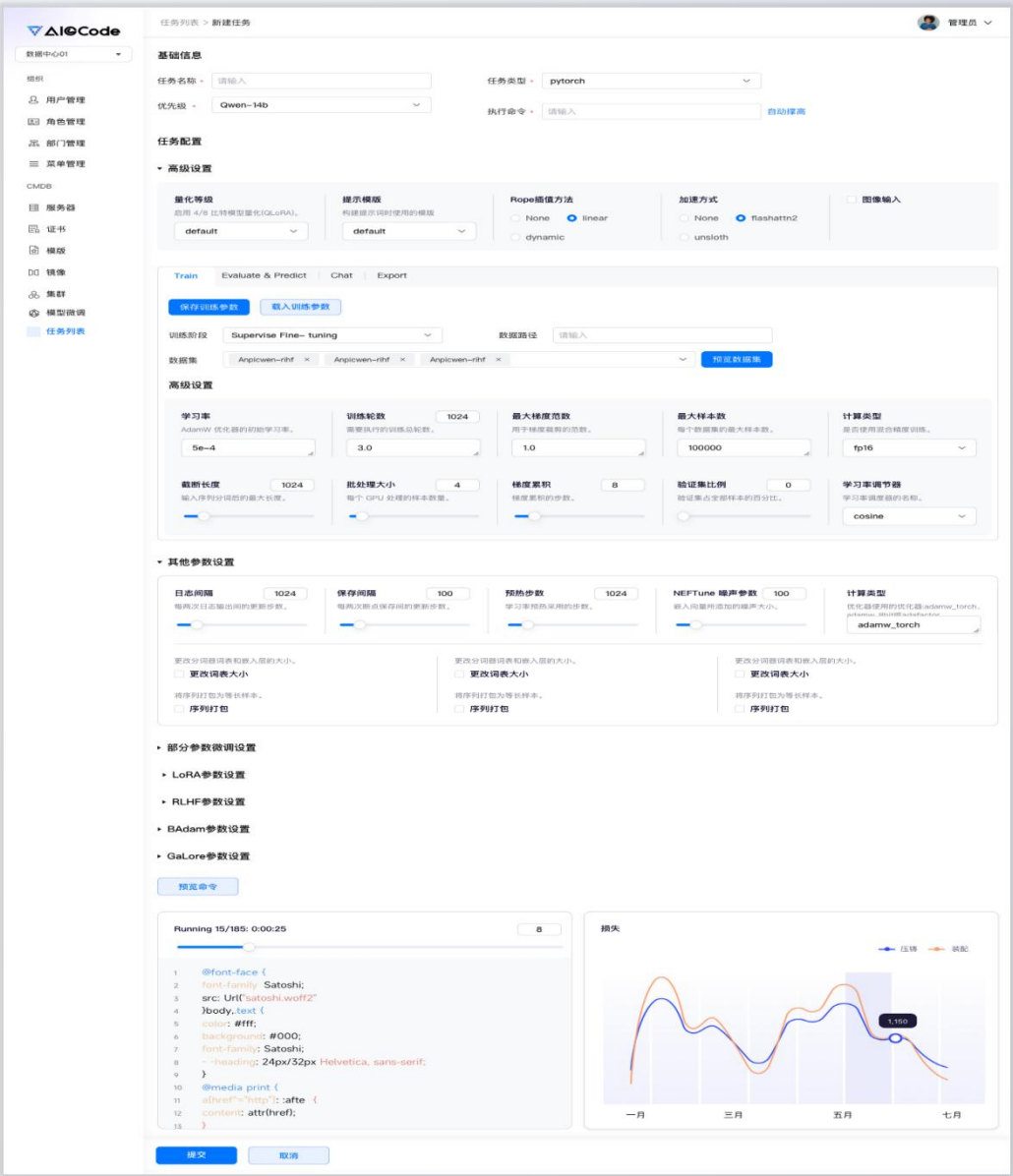
平台功能介绍

模型训练

平台支持用户通过多种方式提交任务。且能清晰的查看任务日志、状态、资源消耗、创建者以及任务配置等。

预置多种运行环境，同时也支持用户指定自定义运行环境以满足特殊的使用场景和需求。

预置多种训练框架和模型，支持创建预训练、模型微调等各类训练任务，且支持多模态模型训练。

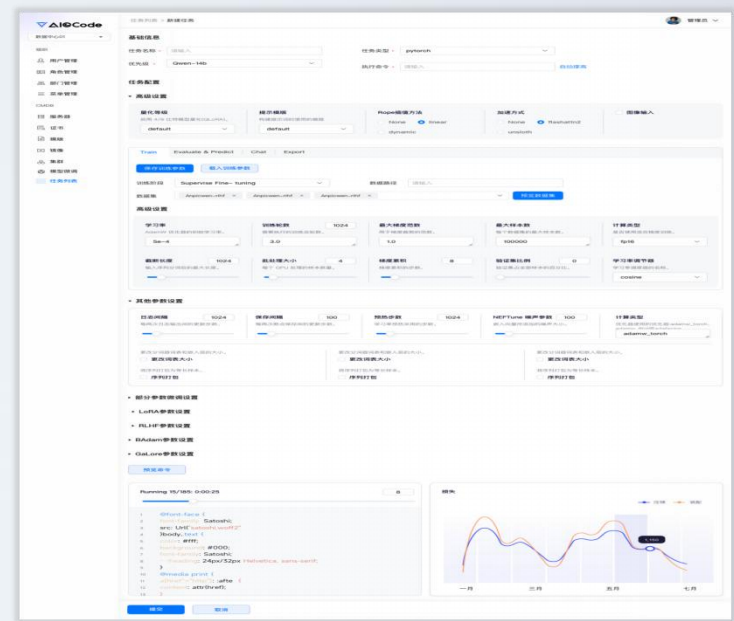


平台功能介绍

模型部署

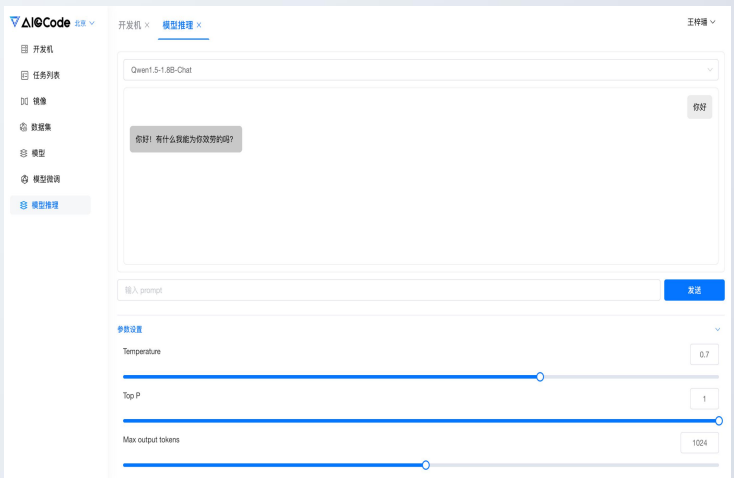
将模型部署成线上服务，支持进行模型推理并提供完整的监测及日志（例如 QPS、响应时长、CPU和NPU利用率等）。用户不仅能充分掌握服务的健康状况，还能通过日志、WebShell 等功能对问题进行详细深入的排查。

高并发高吞吐，服务毫秒级响应时长；服务资源弹性扩缩，保障线上业务的稳定性。



模型推理

平台支持一键部署模型为在线推理服务进行线上推理，用户可实时查看推理结果，同时支持对模型的参数进行在线调整并重新进行模型推理。



目录

1 代码大模型技术发展趋势

2 AIGCode昇腾原生基础大模型解决方案

3 AIGCode原生基础大模型优势

4 成功案例和规划

AIGCode优势

- 完善的训练平台和工具，使用昇腾的torch air框架进行模型编译，训练速度达到竞品的一倍
- 创新的transormer架构，更快的收敛速度
- 完整的端对端代码生成方案
 - 能够端对端生成界面
 - 能够自动化分析和处理数据
 - 能够自动生成后端接口和业务流程

目录

1 代码大模型技术发展趋势

2 AIGCode昇腾原生基础大模型解决方案

3 AIGCode原生基础大模型优势

4 成功案例和规划

成功案例

- 现阶段已经完成1B基座模型研发
- 模型微调在HumanEval测试集上，达到了SOTA水平
- AI数据分析产品上线

aigcode_humaneval

Both Base and Plus results are reported:

Model	HumanEval	HumanEval+
AIGCodeGeek-DS-6.7B	83.5	76.2

原生版本规划

- 昇腾原生7B模型研发

Thank You.