

生成式 AI 基础模型平台

关键优势

- ▶ 用户能够使用 InstructLab 来更新和增强大型语言模型 (LLM)
- ▶ 将 LLM 与专有数据安全稳妥地对齐，以根据您的业务需求定制模型
- ▶ 借助值得信赖且安全至上的红帽® 企业 Linux® 平台，快速开始使用生成式人工智能 (生成式 AI) 并交付结果
- ▶ 打包为可引导的红帽企业 Linux 容器镜像，以用于安装和更新

产品概述

红帽企业 Linux AI 是一个企业级的生成式 AI 基础模型平台，可用于开发、测试和部署生成式 AI 业务用例的 LLM。

红帽企业 Linux AI 汇集了：

- ▶ Granite 系列的开源 LLM。
- ▶ InstructLab 模型对齐工具，提供社区推动的 LLM 微调方法。
- ▶ 可引导的红帽企业 Linux 镜像，以及用于 NVIDIA、Intel 和 AMD 的 AI 库和依赖项，如 PyTorch 和 AI 加速器驱动程序软件。
- ▶ 红帽提供的企业级技术支持和模型知识产权保障。
- ▶ 红帽企业 Linux AI 为您提供值得信赖的红帽企业 Linux 平台并添加必要的组件，让您开启新一代 AI 之旅并收获成果。

开放和透明是 AI 未来发展的关键

红帽企业 Linux AI 包括开源 Granite 语言和代码模型的一个子集，由红帽提供全面保障。开源 Granite 模型为企业组织提供了在成本和性能方面得以优化的模型，与各式各样的 AI 用例相一致。Granite 模型是依据 Apache 2.0 许可证发布的。除了模型是开源的之外，用于训练模型的数据集也保持透明且开放。

表 1. 红帽企业 Linux AI 中的 Granite 模型

IBM Granite 语言模型	Granite-7B-Starter Granite-7B-RedHat-Lab
IBM 代码模型	Granite-8B-Code-Instruct Granite-8B-Code-Base

易于访问的生成式 AI 模型训练，加快价值实现时间

除了开源 Granite 模型外，红帽企业 Linux AI 还包括基于聊天机器人大规模对齐 (LAB) 技术的 InstructLab 模型对齐工具。使用 InstructLab 时，企业内的团队能够高效地为 LLM 贡献技能和知识，并根据业务的特定需求来定制模型。

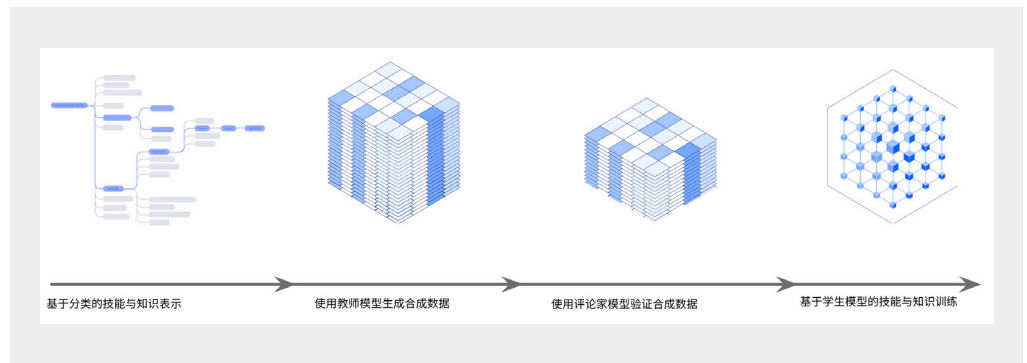


红帽官方微博



红帽官方微信

- ▶ 技能：是指一种能力领域，旨在教导模型如何完成某一任务。技能划分成两个类别。
 - ▶ 复合技能
 - ▶ 让 AI 模型执行特定的任务或功能。
 - ▶ 可以有根基的（包含上下文）或无根基的（不含上下文）
 - ▶ 有根基示例：添加技能，使模型能够读取 markdown 格式的表。
 - ▶ 无根基示例：添加技能，教导模型如何押韵。
 - ▶ 基础技能
 - ▶ 诸如数学、推理和编程等技能。
- ▶ 知识：是指数据和事实，为模型提供额外是数据和信息，从而更准确地回答问题。



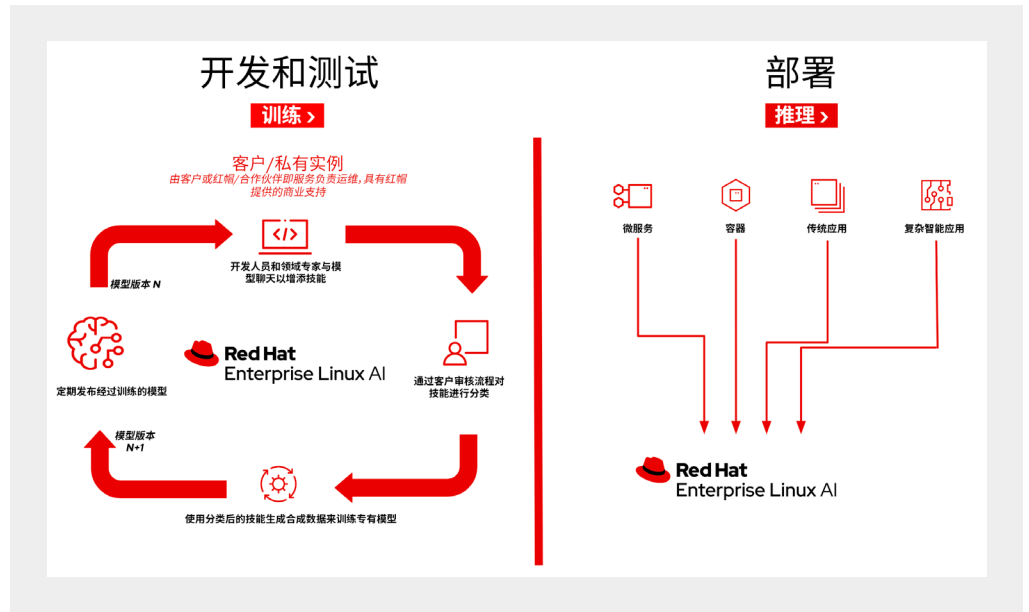
上图概述了 InstructLab 模型微调的工作流。

1. 对技能和知识的贡献被放入基于分类的数据存储库中。
2. 使用分类数据来生成大量的合成数据，以便产生足够大的数据集来成功地更新和更改 LLM。
3. 由评论家模型对合成数据输出进行审核、验证和修剪。
4. 使用基于人类生成的人工输入的合成数据，对模型进行训练。

InstructLab 可供开发人员和领域专家使用，这些人员可能缺乏微调 LLM 所需的必要数据科学专业知识。借助 InstructLab 方法，团队可以将特别契合业务用例要求的数据或技能添加到所选的模型中，进而以协作方式对模型进行训练，从而更快实现价值。

在任何位置进行训练和部署

红帽企业 Linux AI 为企业组织提供所需的各种工具，使之能够在混合云中的任何位置训练、调优和部署模型，进而协助企业组织加快从概念验证迈向基于生产服务器的部署过程。然后，部署的模型可以由公司内的各种服务和应用使用。



当企业组织准备就绪时，红帽企业 Linux AI 还可为红帽 OpenShift®AI 提供助力，以便在分布式集群环境中使用与红帽企业 AI 部署中所用相同的 Granite 模型和 InstructLab 方法来大规模训练、调优和提供这些模型。

功能和优势

功能	优势
享有全面支持的 Granite 语言和代码模型，依据 Apache 2.0 许可证进行开源	开源和透明的 LLM 以及可公开访问的训练数据，提高了数据透明度，解决了数据内容和来源方面的道德伦理问题，最终降低了整体业务风险。

功能	优势
Granite 模型的模型知识产权保障	对红帽企业 Linux AI 中 Granite 模型的保障反映了红帽和 IBM 对这些模型的严格开发和测试的坚定信心。这种保障能让客户更加放心，在红帽的承诺下，以更坚定的信任和信心去使用 Granite 模型。
InstructLab LLM 对齐工具，实现可扩展和易访问的模型微调	InstructLab 提供了一种便于访问的方法来微调 LLM，降低了对深度数据科学专业知识的需求，使得企业组织内的各类角色都能够做出贡献。如此一来，您的企业可以采用生成式人工智能，进而加快价值实现的时间，并最大限度地提高投资回报率。
经过优化的可引导模型运行时实例	红帽企业 Linux AI 能以可引导容器镜像的形式交付，这是一种被称为红帽企业 Linux 镜像模式的部署方法。这种技术降低了安装、配置和更新的复杂性，实现了便捷的设置与变更管理过程。
适用于 AI 硬件的生成式 AI 软件包依赖项和软件驱动程序	运用一套完备的工具，诸如 PyTorch、vLLM 和 NVIDIA 驱动程序等必需的软件包与驱动程序，即刻开启生成式 AI，从而保证您从一开始就具备应对生成式 AI 业务用例的实力。



关于红帽

红帽是世界领先的企业开源软件解决方案供应商，依托强大的社区支持，为客户提供稳定可靠且高性能的 Linux、混合云、容器和 Kubernetes 技术。红帽致力于帮助客户开发云原生应用，集成现有和新的 IT 应用，并实现复杂环境的自动化和管理。作为深受《财富》500 强公司信赖的技术顾问，红帽旨在提供一流的支持、培训和咨询服务，努力将开放创新的优势赋能于各行各业。红帽作为全球企业、合作伙伴和社区网络的互连枢纽，致力于帮助企业发展和转型，并拥抱数字化未来。



红帽官方微博



红帽官方微信

销售及技术支持

800 810 2100
400 890 2100

红帽北京办公地址

北京市朝阳区东大桥路 9 号侨福芳草地大厦 A 座 8 层 邮编:100020
8610 6533 9300