

选择合适 AI 模型的四个注意事项

随着技术的不断发展，企业组织在利用人工智能（AI）和生成式人工智能（gen AI）实现价值方面面临着越来越大的压力。在这一过程中，关键是要选择合适的生成式 AI 模型，它决定了企业组织如何有效地利用 AI 来实现其战略目标。本检查清单提供了选择 AI 模型时的四个关键注意事项，并概述了[红帽® 企业 Linux® AI](#) 如何为您提供所需的模型，以便针对您的独特用例构建量身定制的解决方案。

1 我的企业组织应该使用哪种规模的模型？

为您的企业组织及特定用例选择合适的模型规模，这是一个关键的起点，因为它会直接影响运行模型的成本、管理模型所需的技能以及最终结果的准确性。

使用具有更多参数的较大模型，有助于提高评估复杂情况的能力。然而，较大的模型（以及为该模型提供支持的数据）需要更多的计算资源和复杂的基础架构，这会显著提高运维成本及模型管理所需的技术门槛。

当您需要使用与您的企业、用例或行业相关的数据来定制较大模型时，这些成本和技能要求会变得更加明显。

采用更小、更专业的模型（如红帽企业 Linux AI 所包含的 [Granite 系列大语言模型（LLM）](#)），这种方法更具成本效益，因为这些模型可以使用您的企业数据进行训练，精准满足特定需求，而无需投入大量成本和时间，且技能方面的要求也会显著降低。

2 我的模型透明度如何？

影响生成式 AI 模型准确性和效率的另一项关键因素是用于训练模型的数据，包括数据的来源、所有权、质量、可访问性、可靠性、相关性、多样性和数据量。

采用训练数据来源不透明的模型，这种做法可能会使企业组织面临重大风险。这可能包括因数据不准确而导致的错误结果，更严重的是，使用基于受版权保护材料训练的模型还可能会引发法律风险。

红帽建议使用具有透明数据源和数据集的模型，如红帽企业 Linux AI 所包含的 Granite 系列模型，它们提供完整的训练数据集和模型内容透明度，并由红帽提供全面的保障和支持。

这样您的企业组织就能够充满信心地推进生成式 AI 计划，因为您可以确信：该模型由可信来源支持，确保构建和训练过程符合伦理规范与监管要求，并经过验证能有效保护敏感数据安全。

3 我应该使用开源模型还是专有模型？

LLM 有多种许可类型，包括开源许可（如 Apache 2.0、通用公共许可等）和商业许可（通常为专有许可或基于订阅的许可）。

您选择的模型类型（开源模型或专有模型）将会对一系列关键指标产生重大影响，包括：

- ▶ 运维效率和成本。
- ▶ 允许的自定义水平。
- ▶ 使用该模型构建的应用和服务的所有权，以及输入到该模型的任何数据的所有权。

专有模型的运行成本可能更高，更难以针对您的具体用例或行业进行训练/调优，可定制性有限，且缺乏较小的开源模型所提供的效率优势，这些小型开源模型受益于社区采用和贡献，更加灵活高效。此外，如果您使用商业许可证，当您想要出售使用该模型构建的服务或应用时，可能需要向模型所有者支付相应费用，或者将输入到模型中的任何数据的所有权交给模型所有者。

借助通过红帽企业 Linux AI 提供的开源许可 Granite LLM，您可以提高运维和财务效率，自由使用、修改和分发模型，同时保留对数据的完整所有权。

4 我该如何训练模型？

通过量身定制模型，您能够以高效、精准和最低成本满足特定用例需求；而如何根据需求调整模型，则会极大地影响模型调优的效率和效果。

红帽企业 Linux AI 通过受支持的、基于生命周期的 [InstructLab](#) 发行版，提供可访问的模型对齐工具。

面向 LLM 的 InstructLab 训练方法可有效解决开发这些模型所面临的挑战，以及传统 LLM 训练所面临的扩展挑战。该方法之所以能够有效解决这些挑战，是因为它采用分类指导的合成数据生成流程和多阶段调优框架，助力提高模型性能，并建立开源社区，使开发人员能够共同为任何模型贡献新技能和新知识。

使用 InstructLab 模型对齐工具，可获享以下优势：

- ▶ 为您提供具有成本效益的解决方案，改善 LLM 的一致性。
- ▶ 通过社区方法推动开源模型的改进。
- ▶ 通过可访问的工具、培训和社区知识，使利用独特数据集定制模型的过程大众化。
- ▶ 允许您保留对训练数据集和模型 IP 的完整所有权。

了解更多

[阅读本电子书](#)，了解红帽如何帮助您开始使用 AI 创新成果。

开始使用

[联系红帽代表](#)，探讨红帽如何帮助您企业组织开始使用 AI。



关于红帽

红帽通过[一流](#)的支持、培训和咨询服务，帮助客户实现整个环境的标准化、开发云原生应用，并实现复杂环境的集成、自动化、安全防护和管理。



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