

[[github-release-link]  
[[github-releasedate-link]  
[[github-contributors-link]

[[github-forks-link]  
[[github-stars-link]  
[[github-issues-link]  
[[github-license-link]

[🌐官方网站](#) |  
[📖数据集社区](#) |  
[📊性能榜单](#) |  
[📖文档教程](#) |  
[🔧安装](#) |  
[😬报告问题](#)

[English](#) | [简体中文](#)

[[github-trending-url]

👋 加入我们的 [Discord](#) 和 [微信社区](#)

[!IMPORTANT]

收藏项目，你将能第一时间获取 OpenCompass 的最新动态～🌟

## 📢 OpenCompass 2.0

我们很高兴发布 OpenCompass 司南 2.0 大模型评测体系，它主要由三大核心模块构建而成：[CompassKit](#)、[CompassHub](#)以及[CompassRank](#)。

**CompassRank** 系统进行了重大革新与提升，现已成为一个兼容并蓄的排行榜体系，不仅囊括了开源基准测试项目，还包含了私有基准测试。此番升级极大地拓宽了对行业内各类模型进行全面而深入测评的可能性。

**CompassHub** 创新性地推出了一个基准测试资源导航平台，其设计初衷旨在简化和加快研究人员及行业从业者在多样化的基准测试库中进行搜索与利用的过程。为了让更多独具特色的基准测试成果得以在业内广泛传播和应用，我们热忱欢迎各位将自定义的基准数据贡献至CompassHub平台。只需轻点鼠标，通过访问[这里](#)，即可启动提交流程。

**CompassKit** 是一系列专为大型语言模型和大型视觉-语言模型打造的强大评估工具合集，它所提供的全面评测工具集能够有效地对这些复杂模型的功能性能进行精准测量和科学评估。在此，我们诚挚邀请您在学术研究或产品研发过程中积极尝试运用我们的工具包，以助您取得更加丰硕的研究成果和产品优化效果。

► [Star History](#)

 **欢迎**

来到OpenCompass!

就像指南针在我们的旅程中为我们导航一样，我们希望OpenCompass能够帮助你穿越评估大型语言模型的重重迷雾。OpenCompass提供丰富的算法和功能支持，期待OpenCompass能够帮助社区更便捷地对NLP模型的性能进行公平全面的评估。

▶▶▶ 欢迎加入 OpenCompass！我们目前招聘全职研究人员/工程师和实习生。如果您对 LLM 和 OpenCompass 充满热情，请随时通过[电子邮件](#)与我们联系。我们非常期待与您交流！

🔥🔥🔥 祝贺 **OpenCompass** 作为大模型标准测试工具被Meta AI官方推荐, 点击 Llama 的 [入门文档](#) 获取更多信息.

注意

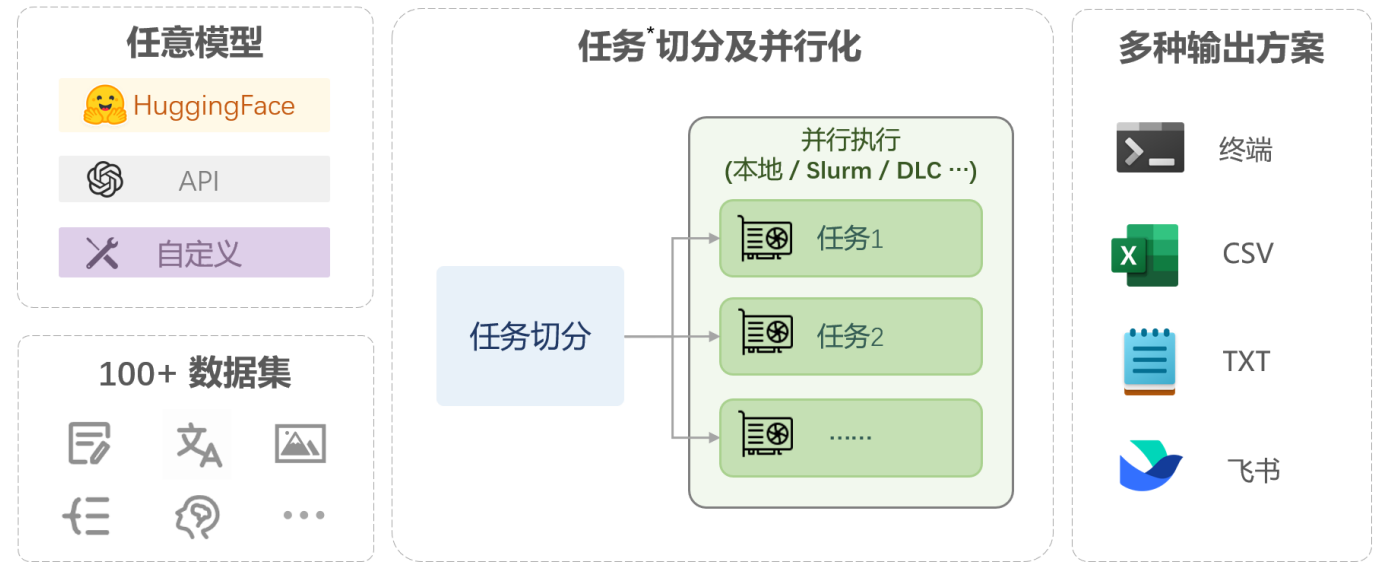
我们正式启动 OpenCompass 共建计划，诚邀社区用户为 OpenCompass 提供更具代表性和可信度的客观评测数据集！  
点击 [Issue](#) 获取更多数据集。  
让我们携手共进，打造功能强大易用的大模型评测平台！

🚀 最新进展 NEW!

- [2024.04.22] 我们支持了 [LLaMA3](#) 和 [LLaMA3-Instruct](#) 的评测，欢迎试用！🔥🔥🔥.
- [2024.02.29] 我们支持了MT-Bench、AlpacaEval和AlignBench，更多信息可以在[这里](#)找到。
- [2024.01.30] 我们发布了OpenCompass 2.0。更多信息，请访问[CompassKit](#)、[CompassHub](#)和[CompassRank](#)。

[更多](#)

✨ 介绍



任务\*：OpenCompass 会将评测请求切分为多个独立执行的任务，从而最大化利用计算资源。

OpenCompass 是面向大模型评测的一站式平台。其主要特点如下：

- **开源可复现**：提供公平、公开、可复现的大模型评测方案
- **全面的能力维度**：五大维度设计，提供 70+ 个数据集约 40 万题的模型评测方案，全面评估模型能力
- **丰富的模型支持**：已支持 20+ HuggingFace 及 API 模型

- **分布式高效评测**：一行命令实现任务分割和分布式评测，数小时即可完成千亿模型全量评测
- **多样化评测范式**：支持零样本、小样本及思维链评测，结合标准型或对话型提示词模板，轻松激发各种模型最大性能
- **灵活化拓展**：想增加新模型或数据集？想要自定义更高级的任务分割策略，甚至接入新的集群管理系统？OpenCompass 的一切均可轻松扩展！

## 性能榜单

---

我们将陆续提供开源模型和 API 模型的具体性能榜单，请见 [OpenCompass Leaderboard](#) 。如需加入评测，请提供模型仓库地址或标准的 API 接口至邮箱 `opencompass@pjlab.org.cn` 。

 [返回顶部](#)