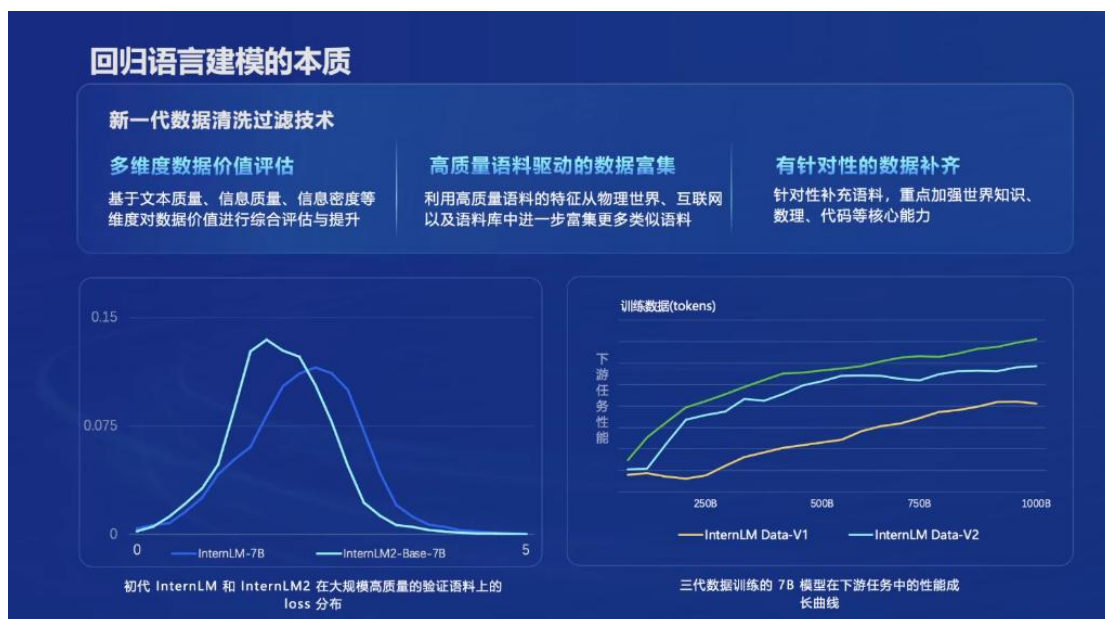


# About InternLM2



提供轻量级（7B）和综合模型（20B）两种，同时在不同专业领域中进行强化用以对接商业应用场景。



通过高质量的语料加强模型建模能力，相比上一代有优势（左图中 loss 的横轴是什么？）

## 书生·浦语 2.0 (InternLM2) 的主要亮点



超长上下文

模型在 20 万 token 上下文中，几乎完美实现“大海捞针”



综合性能全面提升

推理、数学、代码提升显著 InternLM2-Chat-20B 在重点评测上比肩 ChatGPT



优秀的对话和创作体验

精准指令跟随，丰富的结构化创作，在 AlpacaEval2 超越 GPT-3.5 和 Gemini Pro



工具调用能力整体升级

可靠支持工具多轮调用，复杂智能体搭建



突出的数理能力和实用的数据分析功能

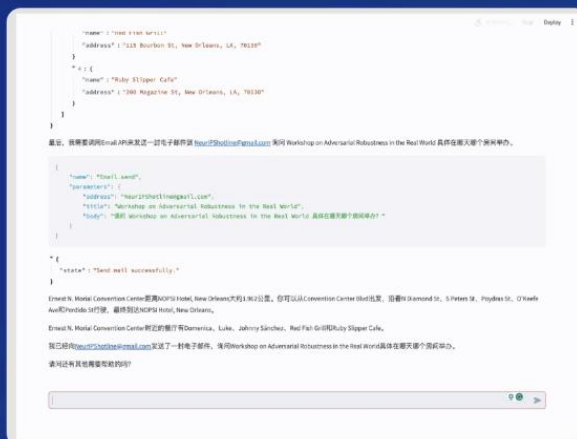
强大的内生计算能力，加入代码解释后，在 GSM8K 和 MATH 达到和 GPT-4 相仿水平

相比其他竞品和上代产品，InternLM2 具有如下优势：

1. 对文本位置信息敏感，上下文检索能力强
2. 代码，数学等部分传统 LLM 薄弱能力上有所加强
3. 多模型封装

## 工具调用能力升级

工具调用能够极大地拓展大语言模型的能力边界，使得大语言模型能够通过搜索、计算、代码解释器等获取最新的知识并处理更加复杂的问题。InternLM2 进一步升级了模型的工具调用能力，能够更稳定地进行工具筛选和多步规划，完成复杂任务



## 强大的内生计算能力

### 高准确率

InternLM2 针对性提高了模型的计算能力，在不依靠计算器等外部工具的情况下，在 100 以内的简单数学运算上能够做到接近 100% 的准确率，在 1000 以内达到 80% 左右的运算准确率

1123 + 456 - 789 + 234 - 1 = 1023

9 \* cos(0°) + (-8) - 4 + 3 + 10 = 10

9 \* cos(0°) + (-8) - 4 + 3 + 10 = 9 \* 1 + (-8) - 4 + 3 + 10 = 9 - 8 - 4 + 3 + 10 = 10

### 复杂运算和求解

依赖模型优秀的内生能力，InternLM2 不借助外部工具就能够进行部分复杂数学题的运算和求解

动点P在直线 $x + y - 4 = 0$ , O为原点, 则 $|OP|$ 的最小值为\_\_\_\_

由题意, 动点P在直线 $x + y - 4 = 0$ 上, 因此P的坐标为 $(a, b)$ , 其中 $a + b = 4$ .  
设O到直线 $x + y - 4 = 0$ 的距离为d, 则O到P的最小距离为O到直线的距离d.  
根据点到直线的距离公式, 有:  
$$d = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}}$$
  
其中, A和B是直线方程 $Ax + By + C = 0$ 中的系数,  $(x_0, y_0)$ 是点O的坐标.  
而直线方程 $x + y - 4 = 0$ 的系数代入上述公式, 得到:  
$$d = \frac{|0 + 0 - 4|}{\sqrt{1^2 + 1^2}} = \frac{4}{\sqrt{2}} = 2\sqrt{2}$$
  
因此,  $|OP|$ 的最小值为 $2\sqrt{2}$ .  
最终答案:  
 $[2\sqrt{2}]$ .

## 代码解释器：更上一层楼

在典型的数学评测集 GSM8K 和 MATH 上，配合代码解释器，InternLM2 都能够在本身已经较高的分数上，进一步获得提升。其中对于难度更高的 MATH 数据集，借助代码解释器，精度从 32.5 大幅提升到 51.2，甚至超过了 GPT-4 的表现

	GSM8K	MATH
InternLM2 单纯依靠模型能力	79.6	32.5
InternLM2 配合代码解释器	84.5	51.2
ChatGPT (GPT-3.5)	78.2	28.0
GPT-4	91.4	45.8

配合代码解释器，20B 模型已经能够完成一些例如积分求解等大学级别的数学题目

以L为底取对数 $y^2 = 2x$ , 计算面积积分:  $I = \int_0^1 x dy = \_\_\_\_\_\_$

首先, 我们需要将给定的方程转换为函数形式。由于L是底数, 我们可以写成 $y^2 = 2x$ , 即 $y = \sqrt{2x}$ 。这里我们假设L为自然对数的底数e, 即 $y = \sqrt{2x}$ 。这样我们就可以将方程转换为 $y = \sqrt{2x}$ 。接下来, 我们需要计算积分 $I = \int_0^1 x dy$ 。为了计算这个积分, 我们可以使用换元法。令 $u = \sqrt{2x}$ , 则 $x = \frac{u^2}{2}$ 。当 $y = 0$ 时,  $x = 0$ ; 当 $y = 1$ 时,  $x = \frac{1}{2}$ 。因此, 积分可以转换为 $I = \int_0^1 \frac{u^2}{2} du$ 。计算这个积分, 我们得到 $I = \frac{1}{6}$ 。因此, 最终答案是 $\frac{1}{6}$ 。

```
from sympy import symbols, cos, sin, sqrt, pi, integrate\n\n# 定义变量\ntheta = symbols('theta')\n\n# 函数方程\nx = 1 / cos(theta)\ny = sin(theta)\n\n# 计算 dx\ndx = sqrt(1 + sin(theta)**2) + (y * diff(theta)**2) * theta.diff(theta)\n\n# 计算积分 I\nI = integrate(x * dx, (theta, 0, pi/2))\n\n# 输出结果\nI.simplify()
```

最终答案:  $\frac{1}{6}$

## 从模型到应用



书生·浦语



智能客服

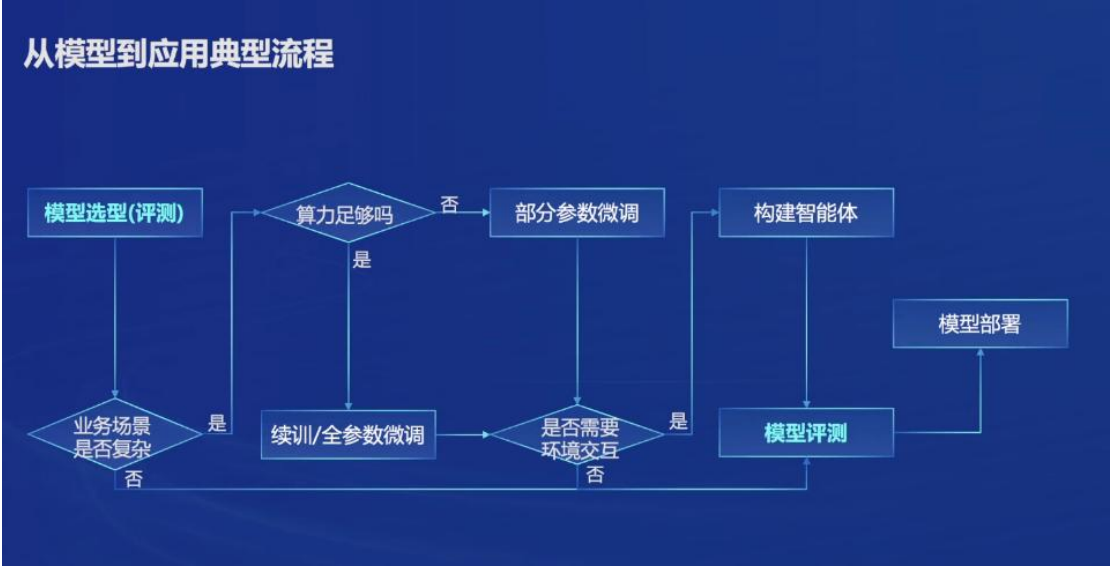


个人助手



行业应用

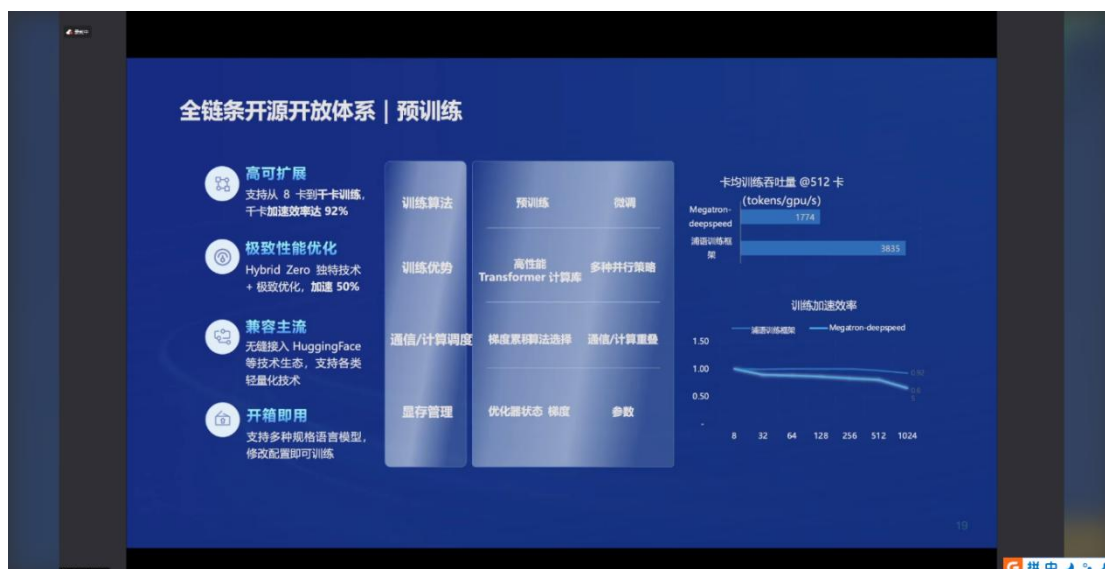
目前大模型的商业产品化仍然存在一定距离。



环境交互：业务中是否需要与外界 api 进行交互







## 全链条开源开放体系 | 微调

大语言模型的下游应用中，增量续训和有监督微调是经常会用到两种方式。

### 增量续训

使用场景：让基座模型学习到一些新知识，如某个垂类领域知识

训练数据：文章、书籍、代码等

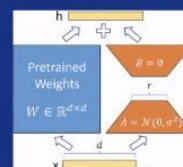
全量参数微调

部分参数微调

### 有监督微调

使用场景：让模型学会理解各种指令进行对话，或者注入少量领域知识

训练数据：高质量的对话、问答数据



## 全链条开源开放体系 | 微调

### 高效微调框架 XTuner



### 适配多种生态

- 多种微调算法  
多种微调策略与算法，覆盖各类 SFT 场景
- 适配多种开源生态  
支持加载 HuggingFace、ModelScope 模型或数据集
- 自动优化加速  
开发者无需关注复杂的显存优化与计算加速细节

### 适配多种硬件

- 训练方案覆盖 NVIDIA 20 系以上所有显卡
- 最低只需 8GB 显存即可微调 7B 模型

## CompassRank: 中立全面的性能榜单

数据社区 评测榜单 评测工具 文档中 | EN加入评测贡献数据集登录

### CompassRank 将未来的可能性定位在今天

致力于探索最先进的语言与视觉模型，为工业界和研究社区提供全面、客观、中立的评测参考

大语言模型总榜

全部 24-01 23-12

多模态模型总榜

全部 24-01 23-12

#### NLP

大语言模型评测

100+ 大语言模型已加入评测

模型	评测结果
OpenAI	GPT-4o / GPT-4o / GPT-4o / GPT-4o
Google	Gemini-1.5
Anthropic	Claude-3.5
Meta	Llama-3.3
Alibaba	Qwen-2.5
Microsoft	GPT-4o
OpenAI	GPT-4o
Google	Gemini-1.5
Anthropic	Claude-3.5
Meta	Llama-3.3
Alibaba	Qwen-2.5
Microsoft	GPT-4o

#### MULTI MODAL

多模态模型评测

100+ 多模态大模型已加入评测

模型	评测结果
OpenAI	GPT-4o / GPT-4o / GPT-4o / GPT-4o
Google	Gemini-1.5
Anthropic	Claude-3.5
Meta	Llama-3.3
Alibaba	Qwen-2.5
Microsoft	GPT-4o
OpenAI	GPT-4o
Google	Gemini-1.5
Anthropic	Claude-3.5
Meta	Llama-3.3
Alibaba	Qwen-2.5
Microsoft	GPT-4o

## CompassKit: 大模型评测全栈工具链

### OpenCompass 核心代码库功能全面升级

#### 数据污染检查



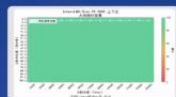
- ✓ 提供多种数据污染检测方法
- ✓ 支持包括GSM-8K, MMLU等主流数据集上的污染检测

#### 更丰富的模型推理接入



- ✓ 支持近20个商业模型API
- ✓ 支持LMDeploy, vLLM, Lighthouse LLM等推理后端

#### 长文本能力评测



- ✓ 支持200K大海捞针测试
- ✓ 支持多个主流长文本评测基准

#### 中英文双语主观评测



- ✓ 支持基于大模型评价的主观评测
- ✓ 提供模型打分、模型对战等多种能力
- ✓ 灵活切换上百种评价模型

## OpenCompass 助力大模型产业发展和学术研究

### 广泛应用于头部大模型企业和科研机构

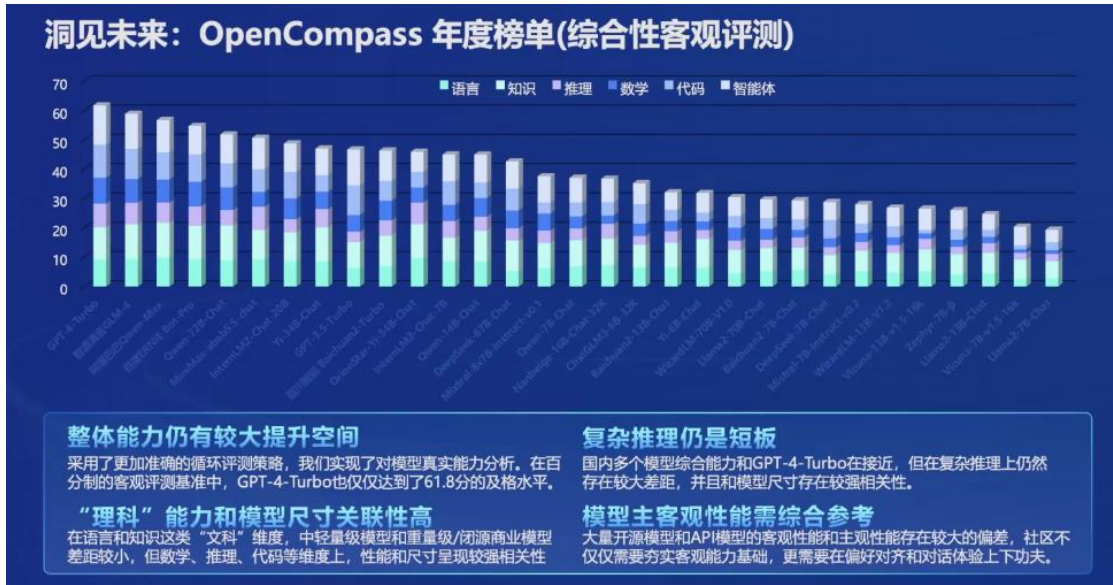


#### 获得 Meta 官方推荐 唯一国产大模型评测体系

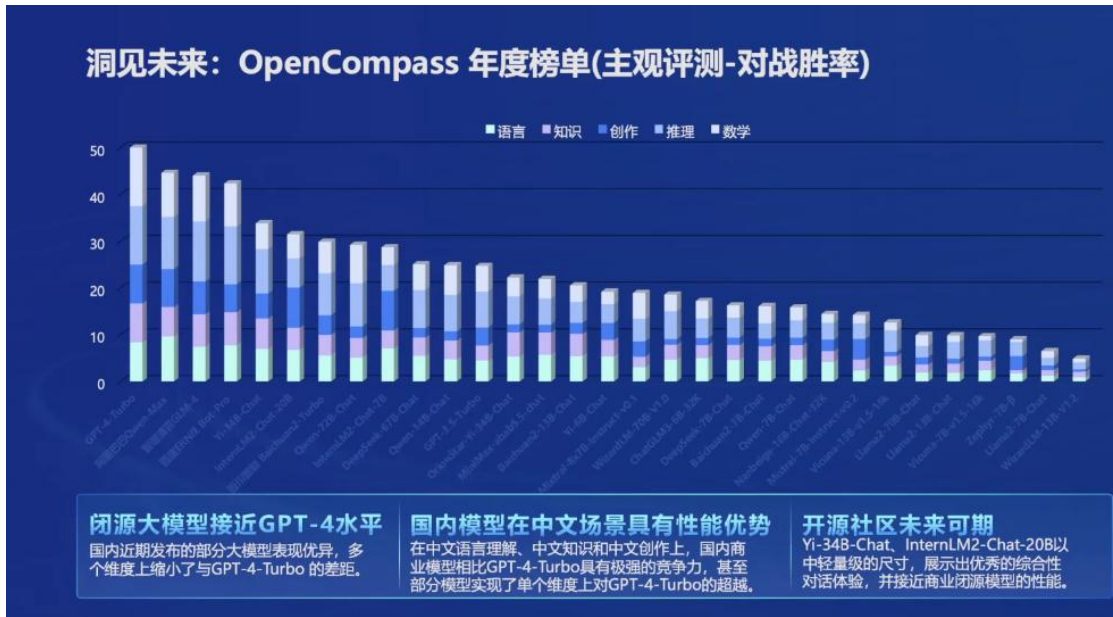
These types of projects provide a quantitative way of looking at the models performance in simulated real world examples. Some of these projects include the [LM Evaluation Harness](#) (used to create the [HF leaderboards](#)), [HELM](#), [BIG-bench](#) and [OpenCompass](#).

#### 社区支持最完善的评测体系之一 100+ 评测集 50万+ 题目





循环评测：将选项进行轮换，总是答对才算对  
部分理科建模能力（数学，代码）与模型尺寸强相关



中文场景下国内模型部分维度优于 CHATGPT4



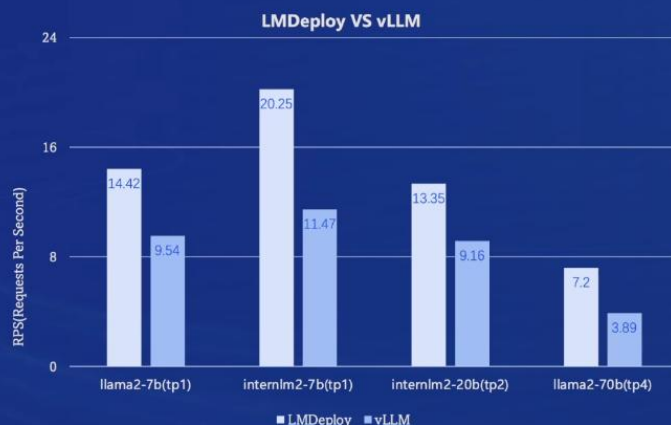
## 全链条开源开放体系 | 部署



LMDeploy 提供大模型在GPU上部署的全流程解决方案，包括模型轻量化、推理和服务。



## 全链条开源开放体系 | 部署



领先的推理性能

备注：同一 80G A100 机器上进行的测试

31

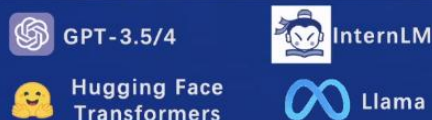
## 全链条开源开放体系 | 智能体

### 轻量级智能体框架 Lagent

#### 支持多种类型的智能体能力



#### 灵活支持多种大语言模型



#### 简单易拓展，支持丰富的工具

AI 工具	能力拓展	Rapid API
文生图	搜索	出行 API
文生语音	计算器	财经 API
图片描述	代码解释器	体育资讯 API



## 全链条开源开放体系 | 智能体

### 多模态智能体工具箱 AgentLego

- 丰富的工具集合，尤其是提供了大量视觉、多模态相关领域的前沿算法功能
- 支持多个主流智能体系统，如 LangChain, Transformers Agent, lagent 等
- 灵活的多模态工具调用接口，可以轻松支持各类输入输出格式的工具函数
- 一键式远程工具部署，轻松使用和调试大模型智能体



封装其他模型使用