

书生·浦语大模型全链路开源体系

一、大模型

1. 大模型

在计算机科学和人工智能领域，"大模型"通常指的是参数数量庞大、拥有大量隐藏层和神经元的深度学习模型。这些模型通常使用大规模的数据集进行训练，并且在多个任务上表现出色。以下是大模型的一些特征和定义要点：

- 大量参数**：大模型通常具有数十亿至数千亿个参数。这些参数用于表示模型的权重和偏置，以便模型能够对输入数据进行复杂的映射和抽象。
- 深层结构**：大模型通常具有深层的神经网络结构，包含多个隐藏层。通过增加隐藏层的数量，模型可以学习到更加复杂和抽象的特征表示。
- 大规模数据集**：为了训练大模型，通常需要使用大规模的数据集。这些数据集可能包含数百万甚至数十亿条数据样本，以确保模型可以学习到足够泛化的特征。
- 高计算资源需求**：由于大模型拥有庞大的参数量和复杂的结构，因此训练和使用这些模型需要大量的计算资源，包括高性能的GPU或TPU加速器。
- 广泛应用性**：大模型通常能够在多个任务和领域中表现出色，例如自然语言处理、计算机视觉、语音识别等。它们可以被用于各种任务，包括语言生成、文本分类、图像识别、语音生成等。

2. 专用模型

针对特定任务，一个模型解决一个问题。

3. 通用大模型

一个模型应对多种任务、多种模态。

4. 通用人工智能的发展方向

从单一任务模型转向通用大模型，解决多种任务和模态。

二、书生·浦语大模型

1. 书生·浦语大模型

2023年7月6日，2023世界人工智能大会（WAIC 2023）隆重开幕。在开幕式和科学前沿全体会议上，上海人工智能实验室（上海AI实验室）与商汤科技联合香港中文大学、复旦大学、上海交通大学及清华大学发布全新升级的“书生通用大模型体系”（以下简称“书生大模型”），包括书生·多模态、书生·浦语和书生·天际等三大基础模型，以及首个面向大模型研发与应用的全链条开源体系。

2. 开源历程




三、书生·浦语2.0（InternLM2）的体系

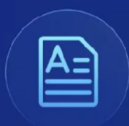


四、书生·浦语2.0（InternLM2）的主要亮点

长上下文理解、对话与创作、数学能力等，例如通过模型进行行程规划和情感对话。

书生·浦语 2.0 (InternLM2) 的主要亮点

OpenMMLab 



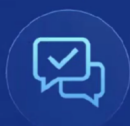
超长上下文

模型在 20 万 token 上下文中，几乎完美实现“大海捞针”



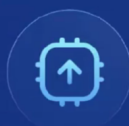
综合性能全面提升

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




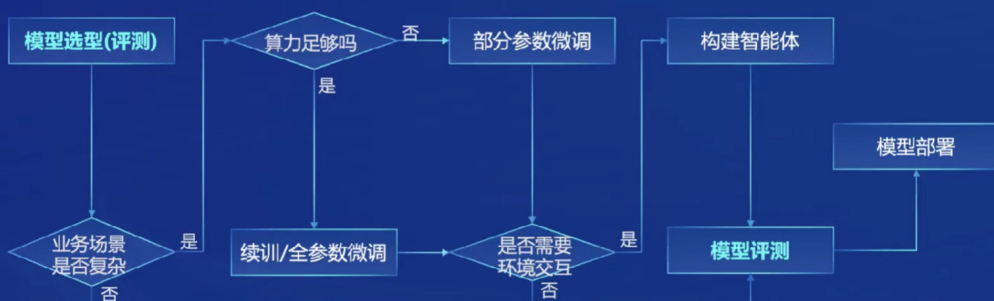
突出的数理能力和实用的数据分析功能

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

五、从模型到应用的典型流程

从模型到应用典型流程

OpenMMLab 



六、书生·浦语全链条开源开放体系



1. 数据：书生·万卷

数据是训练大型模型的基础。这些数据可以是文本、图像、音频、视频等形式的信息。数据通常需要大规模收集、清洗和标注，以用于训练模型。

书生·万卷开放高质量语料数据，提供丰富多样的数据，支持数据清洗、安全处理和公开使用。

2. 预训练：InternLM-Train

预训练是指在大规模数据上使用无监督或半监督学习方法对模型进行初始化训练。预训练的目标是让模型学会对自然语言、图像等领域的基本特征进行抽象表示，为后续任务提供良好的初始化参数。

3. 微调：XTuner

微调是在预训练模型的基础上，利用特定任务的有标注数据进行进一步的训练。微调的目的是让模型适应于特定任务的需求，提高模型在该任务上的性能表现。

4. 部署：LMDeploy

部署是指将训练好的模型应用到实际场景中，让模型能够处理真实的数据并产生相应的预测或输出。部署涉及到模型的集成、优化、部署到特定的硬件或软件环境中等一系列工作。

LMDeploy支持模型轻量化、量化和推理服务，与评测工具无缝对接。

5. 评测：OpenCompass

评测是对训练好的模型在特定任务上进行性能评估的过程。评测可以包括定量指标（如准确率、召回率、F1值等）和定性评估（如模型生成文本的流畅度、图像生成的质量等）。

6. 应用：Lagent AgentLego

应用是指将训练好的模型应用到具体的实际问题中，解决实际的任务或提供特定的服务。应用可以涉及到多个领域，如自然语言处理、计算机视觉、语音识别、推荐系统等。

