**核心内容：介绍全年度开源体系**

**（一）大模型：通往通用人工智能的关键途径**

专用模型（2000-2021）：针对特定任务使用特定模型，如AlphaGo

通用大模型（2022-至今）：ChatGPT

大模型本质上在做的是语言建模：通过给定的context预测token；其关键是高质量语料

**（二）书生·浦语大模型**

**历程**



**体系：**



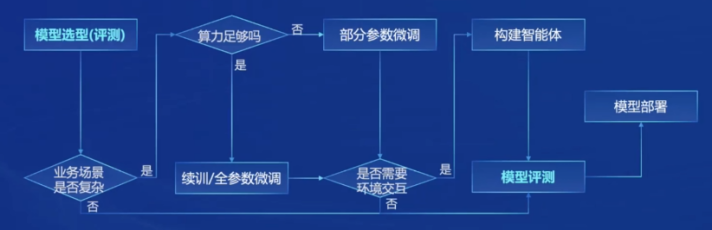
**InternLM2优势：**

20万token；

推理、数学、代码能力强；

内生计算能力强，加入代码解释器后能比肩GPT4

**从模型到应用：**



**（三）全链条开放开源体系：**



**1.数据集获取：**



1. **预训练：**

高可扩展、极致性能优化、兼容主流技术生态、支持多规格语言模型

**3.微调：**

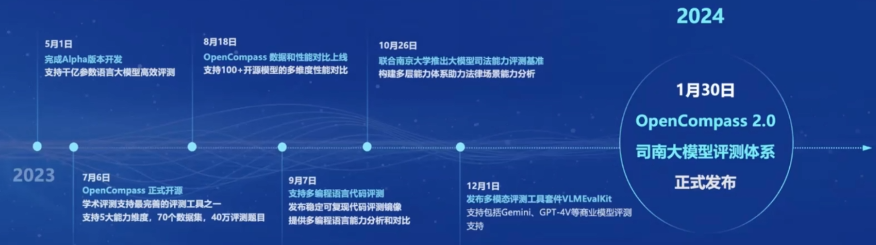
大语言模型下游应用常用方式：

**1.增量续训**：使用文章、书籍、代码等数据来学习垂类领域知识，其过程类似预训练。

**2.有监督微调**：使用高质量的对话和问答数据，让模型学会理解何种指令进行对话，或注入少量领域知识。分为全量参数微调和部分参数微调（降低微调成本）。

高效微调框架XTuner：适配多生态；适配多硬件（覆盖20系及以上显卡）

1. **评测：**

OpenCompass（最完善的国产大模型评测体系，已有广泛应用）

1.提供中立全面的大模型榜单

2.开源评测的各类工具：数据污染检查、推理模型接入、长文本能力评测、中英文双语主观评测

3.评测基准社区：CompassHub

**特点：**采用循环评测（选项轮换），相比于单纯的选择题，减少猜测与运气因素。

**客观评测结论：**

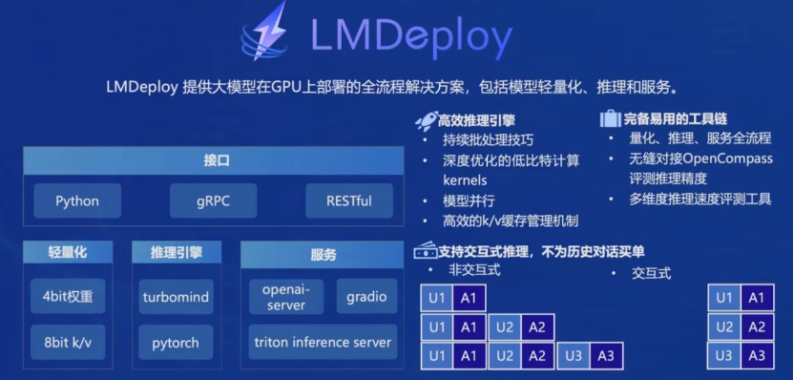
1. 大模型整体能力提升比较大，即使是GPT4-turbo，在百分制客观评测基准中也只有及格。
2. 复杂推理仍然是短板。模型间的复杂推理能力差距大；复杂推理能力与模型尺寸关系较大。
3. 理科维度能力与模型尺寸关联性高，如数学、代码、推理等。（语言和知识等文科维度，中轻量级模型和重量级/闭源商业模型差距较小）
4. 单个模型的主客观性能差距可能较大。主观性能：满足用户偏好、对话体验。

**主观评测结论：**

1. 国内闭源大模型接近GPT-4水平
2. 国内模型在中文环境下有很强竞争力，有时甚至更加优异。
3. 开源模型性能越来越优秀。

**5.部署：**

LMDeploy



交互式推理：对话

非交互式推理：发送整个对话历史

**6.智能体：**

Lagent（轻量级）



通过Lagent框架：代码解题；零样本泛化

多模态智能体工具箱AgentLego：工具集合

**课程目标：**

基于大模型做应用开发和微调实现，在InternLM里都可以找到相关工具。