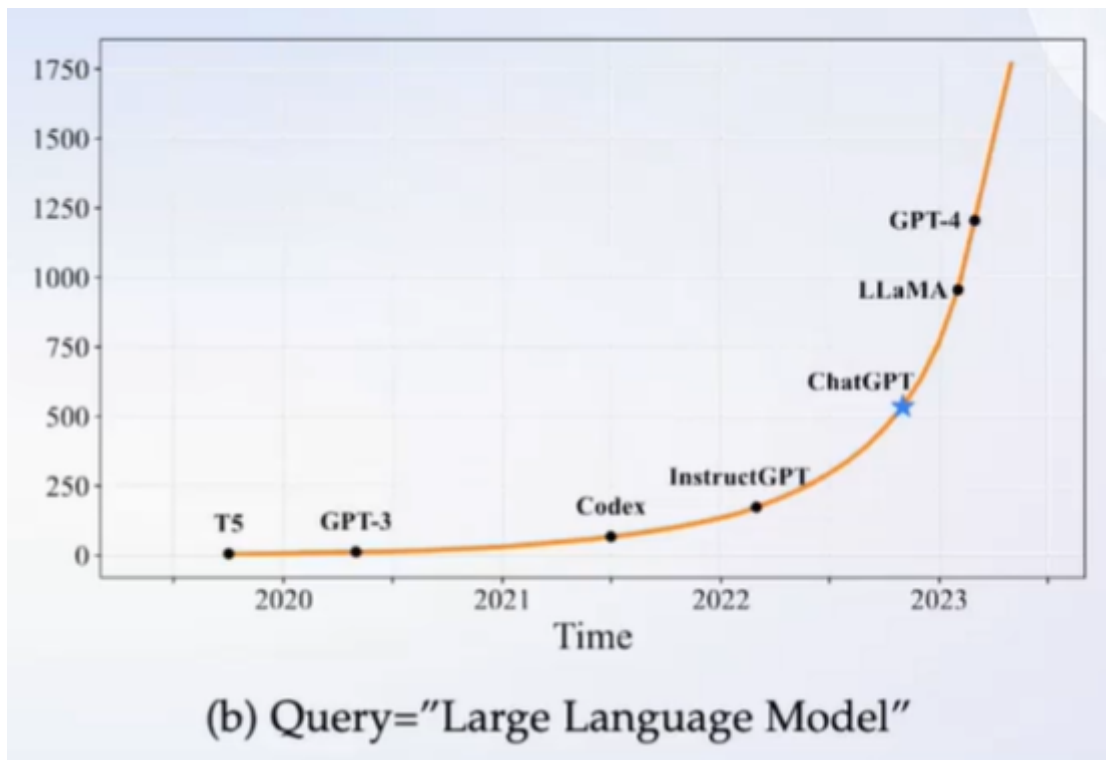


# 书生·浦语大模型全链路开源体系

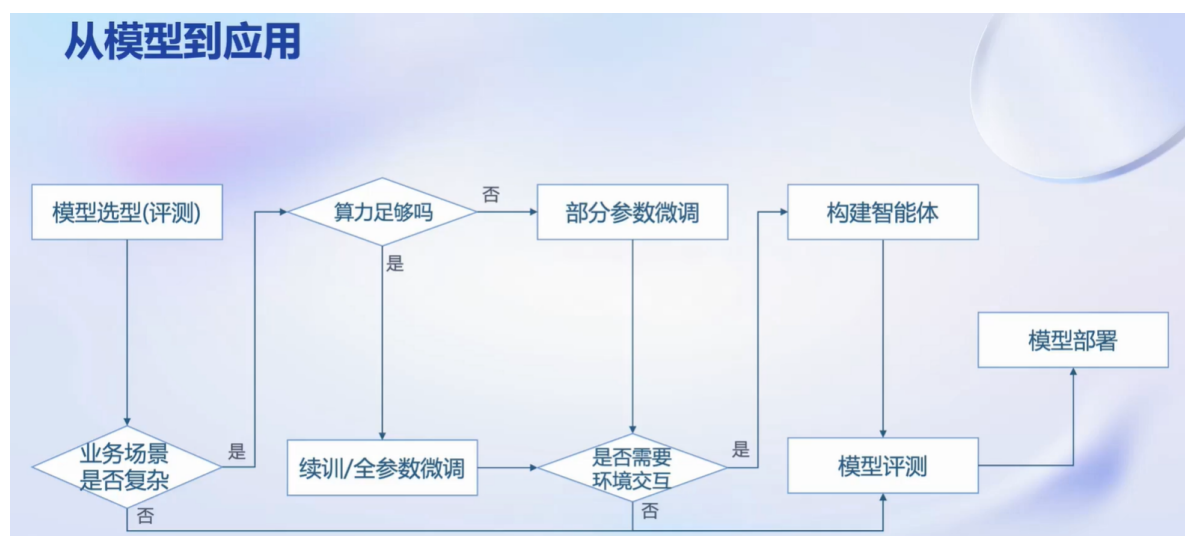
## 大模型发展

- 重要节点：2022年，随着ChatGPT的问世，大模型迎来爆发性增长



- 技术趋势：专用模型→通用模型，发展通用人工智能

## 从模型到应用



- 三个关键点
  - 业务场景
  - 算力
  - 环境交互

# 书生·浦语全链路开源体系

## 1. 数据

- 书生万卷：总数据量2TB，包括文本、图像-文本与三类数据
- OpenDataLab：开放数据平台
- 数据处理
  - 多模态融合
  - 精细化处理
  - 价值观对齐

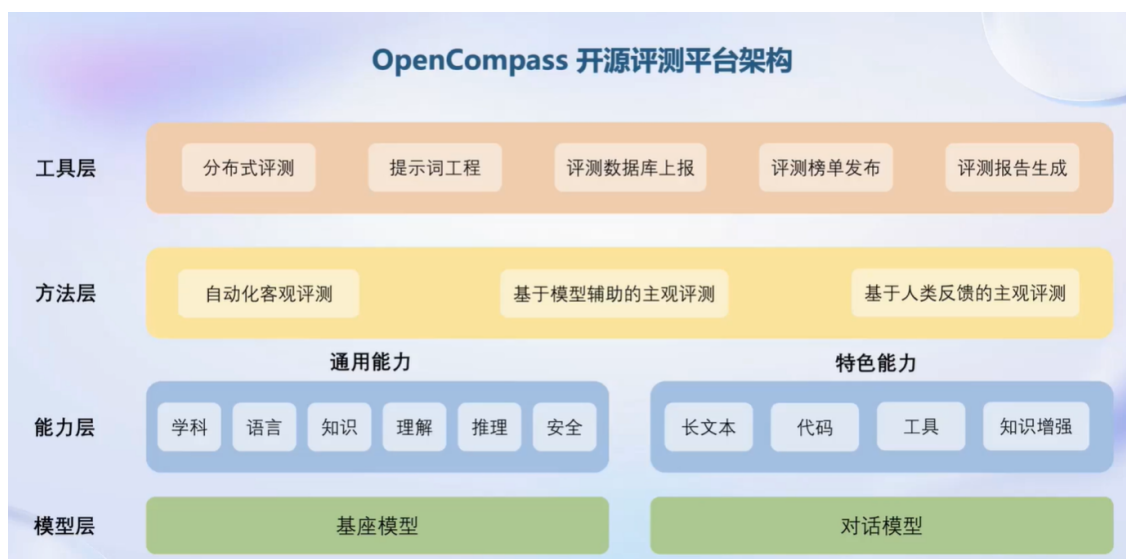
## 2. 预训练

- 高可扩展
- 极致性能优化
- 兼容主流
- 开箱即用

## 3. 微调

- 增量续训
- 有监督微调
  - 全量参数微调
  - 部分参数微调
- 高效微调框架 XTuner
  - 适配多种生态
    - 微调算法
    - 开源生态
    - 自动优化加速
  - 适配多种硬件

## 4. 评测



## 5. 部署



LLM > 智能体

