老师从书生浦语大模型的全年度开源体系开始讲起，以及书生浦语大模型的发展历程和特点，其中包括轻量级和重量级模型以及不同能力的模型。

介绍到了通用大模型成为人工智能发展趋势，书生浦语大模型开源历程，书生浦语大模型2.0提供不同尺寸和类型的模型，支持多语言和多模态任务

二、书生浦语模型在各种能力评测中的表现，包括语言知识、推理、数学、代码等方面，同时还介绍了模型的应用和数据分析功能。20B模型在推理数学代码等方面的性能优于GP3.5和germini pro模型在综合性能方面达到同量级的开源模型领先水平，模型内生的计算能力和数据分析功能能够处理复杂的任务和数据分析

三、从模型选型到应用的整个流程，以及各个环节需要做的事情，并介绍了书生葡语的全链条工具体系和开源数据集。

模型选型是第一步，需要考虑模型的复杂度和算力

书生浦语的全链条工具体系开源，包括数据、预训练、微调、部署、评测、应用等环节

书生万卷cc数据集开源，包括2013年至2023年的互联网公开内容，并进行精细化的清洗和处理

四、open compass 2.0思南大模型评测体系，包括评测框架的开发和开源、评测基准社区的建立以及对大模型能力提升的分析。，可以把更多精力投入到数据准备和优化上，已经·发布open compass 2.0思南大模型评测体系，open compass已经适配超过100个评测集，是国内最完善的评测体系之一

五、书生浦语模型推理和部署工具的评测和发展趋势，以及智能体框架和多媒体多模态智能体工具箱的使用和开发。

开源社区发展趋势，中轻量级模型性能接近商业闭源模型

Mdepot提供全链条部署解决方案，支持模型轻量化、推理引擎、服务模块等

智能体框架Legend支持多种智能体能力，提供多模态AI工具箱AgentLego和多媒体算法功能

这份技术报告介绍了InternLM2，InternLM2是由上海人工智能实验室、SenseTime集团、香港中文大学以及复旦大学共同开发的开源大型语言模型（LLM）。InternLM2在六个维度和30个基准测试中的全面评估中超越了其前身，并且在长文本建模和开放式主观评估方面表现出色。报告详细介绍了InternLM2的预训练过程，包括多样化数据类型的准备，如文本、代码和长文本数据。InternLM2通过创新的预训练和优化技术，有效地捕捉长期依赖关系，并在预训练和微调阶段从4k标记开始，逐步提高到32k标记，表现出在200k“针堆”测试中的卓越性能。

报告介绍了InternLM2的对齐策略，包括监督式微调（SFT）和新颖的条件在线强化学习（COOL RLHF）策略，该策略解决了冲突的人类偏好和奖励黑客问题。通过发布不同训练阶段和模型大小的InternLM2模型，为社区提供了模型演变的洞察。

报告的内容包括：

引言：介绍了大型语言模型（LLMs）的流行和对人工通用智能（AGI）的讨论。

基础设施：介绍了用于模型训练的InternEvo框架，包括模型结构和训练过程。

预训练：详细介绍了预训练数据、设置和阶段。

对齐：介绍了监督式微调和COOL RLHF策略。

评估和分析：提供了模型在各种下游任务上的性能分析。

结论：总结了InternLM2的主要贡献和性能。

附录包括致谢和用于评估的提示示例。

报告中讨论了数据污染问题，并提供了对齐能力的评估，包括英语和中文的主观评估、指令遵循评估以及条件奖励模型的消融研究。此外，报告还提供了关于InternLM2训练和评估的详细信息，包括使用的数据处理管道、质量过滤策略、依赖排序、长文本数据准备以及工具增强型LLMs的实践。

最后，报告强调了InternLM2在长文本理解、推理、数学问题解决和编码方面的能力，并通过多个基准数据集对其进行了评估。InternLM2在各种任务和基准测试中都取得了优异的性能，证明了其作为一种多功能和高效的语言模型的潜力。