**火山引擎-豆包同传介绍**

**发言部分**

我是火山引擎豆包同传模型产品负责人殷悦，非常荣幸能与大家分享我们在 AI 同传方向的进展。

豆包 AI 同传方案，结合了豆包同声传译模型、声音复刻大模型、实时音视频技术。具有高翻译质量、低延时输出的产品优势。为本次大会提供了现场同传和线上双语字幕的支持。现场有很多外宾都提到本次会议的 AI 同传，上下文语义更连贯通顺，表达自然，翻译质量特别高，想要了解是怎么实现的。  
  
首先，传统的 AI 同传采用的是级联方案，也就是先识别语音到文本，再进行文本翻译，最后进行语音合成，这个过程会不可避免的会传递一些错误，比如很多中文同音字识别错了，那么翻译的结果必然会错。而且整个链路的实时性也比较难保证。

豆包同传模型，采用端到端大模型架构，避免了错误累积。在结合特定语境下更“聪明”，比如会挑选最合适的表达习惯，比如遇到发言人有一些支支吾吾的冗余词会智能精简，比如在中文表达上缺少主语和时态的时候在英语表达上补充上。这些「大模型的智能」让我们翻译更地道流畅，大幅超越了传统机器学习的翻译方案。  
  
另外，我们专门有一套在同传领域的「有效字段占比」的评估方案作为指导，这个指标越高，发言者的意图和信息就能更加完整的表达。人类同传也是在短时间内进行翻译表达，所以这个指标并不是 100%，大部分在 70%-80% 我们的同传模型在中英英中的评测表现突出，可以媲美高水平人类同传，这也使得豆包同传大模型能够满足高级别会议和专业场合的需求。  
  
最后在声音复刻模型上，我们仅需 5 秒原声采样，就能跨语种完整复刻发言人音色，发言人说中文，AI 同传频道听到的是地道的母语级别的英语和同音色的流畅表达，这给现场的听众更还原更沉浸的体验。  
  
字节跳动内部本身也有很多跨国沟通的场景，在产品正式对外前，我们在公司内部的会议软件，和全员大会直播上都应用了该同传模型的能力，来帮助各方高效合作。今年我们也在逐步推进，希望未来为更多企业提供高质量的 AI 同传服务。

**问答：**

**豆包同传模型后续的迭代计划是怎么样的？**

我们会继续迭代模型、提升翻译质量的同时进一步压低延时，并且提升跨语种语言表现力：

* 在翻译上，不同语种之间其实表达的长短不同。我们都知道中文会比较精炼，那么翻译到其他语种进行播报，就会要求同传模型具有高级同传译员「精简的智慧」，我们叫如何处理「膨胀语向」的问题
* 在断句决策上，我们也会持续性的用一些新的方案，更好激发模型潜能。比如会场上会有人反应说我们的大模型延时不稳定，时快时慢，实际上这个是我们模型的一个特点，当说长难句的时候，和说简单句的时候，模型对合适决定翻译的实际不同，是一个动态的延时，所以后续我们会继续研究如何模型更智能决策断句，动态降低平均输出延时
* 在生成上，我们会更进一步的提升生成语音的表现力，例如演讲人的的语气，停顿，情感等等如何进行迁移，也会积极探索更多语言