

This is f*cking insane. Former OpenAI CTO, Mira Murati: "The models ...

Soham Mehta 引述前 OpenAI CTO 米拉·穆拉蒂关于新一代「交互模型」的发言，介绍其创办的 Thinking Machines 实验室对实时人机协作的构想。

视频号 小红书 公众号 englishspring.com B站

扫码获取更多学习资料

原文对照

第1章 从实验室愿景到交互模型

The reason why we even started Thinking Machines is because we wanted to build the frontier AI lab that's really focused on really human AI collaboration piece, and that means a lot of things. And there are specific research bets that go into that.

我们之所以创办 Thinking Machines，是因为我们想打造一家真正专注于「人机协作」这一环节的前沿 AI 实验室，而这意味着一系列具体的事情，其中也包含一些特定的研究押注方向。

frontier n.
/ˈfrʌntɪə(r)/
前沿，边界

collaboration n.
/kəˌlæbəˈreɪʃn/
协作，合作

bet n.
/bet/
押注，赌注（此处指研究方向上的「押宝」）

But we wanted to showcase our work in one of the first bets through the interaction models.

但我们希望通过「交互模型」来展示我们在其中最初几个押注方向上所做的工作。

interaction n.
/ˌɪntərˈækʃn/
交互，互动

showcase v.
/ˈʃəʊkeɪs/
展示，亮相

第2章 轮次式模型的局限

And the interaction models are a new kind of model. If you consider the types of models that we work with today, they're very turn-based, and so you talk, they talk, then they go off and think, um, once you've given sort of a prompt on what it is that you want to do, and while they're thinking, it's almost like they're deaf and blind.

交互模型是一种全新的模型。如果你看看我们如今使用的这些模型，它们是非常「轮次式」的：你说完，它们说；然后一旦你给出了某种提示、说明你想做什么，它们就跑开去「思考」，而在思考的时候，它们几乎又聋又瞎。

turn-based adj.
/ˈtɜːn beɪst/
回合制的，轮次式的

prompt n.
/prɒmpt/
提示词，提示

deaf adj.
/def/
聋的

blind adj.
/blaɪnd/
瞎的

They cannot perceive anything else about what's going on. And then it's your turn. And while you're talking,

它们无法感知正在发生的其他任何事情。然后就轮到你了。而当你说话时，

perceive v.
/pəˈsiːv/
察觉，感知

they really cannot perceive anything about how you're talking.

它们其实完全无法感知你是怎么说话的。

It's not happening in real time. And, by contrast,

这一切并非实时发生。相比之下，

by contrast
/baɪ ˈkɒntrɑːst/
短语. 相比之下

第3章 实时交互模型与高带宽协作

our interactions with each other are very rich. There is a lot of information in our interactions when we are silent, when we're thinking, when we're interrupting one another, and so interaction models are able to capture all of these nuances.

我们人与人之间的互动是非常丰富的。在我们的互动中蕴含着大量信息——比如我们沉默的时候、思考的时候、互相打断的时候，因此交互模型能够捕捉到所有这些细微之处。

rich adj.
/rɪtʃ/
丰富的

nuance n.
/ˈnjuːɑːns/
细微差别，微妙之处

interrupt v.
/ˌɪntəˈrʌpt/
打断，插话

capture v.
/ˈkæptʃə(r)/
捕捉，记录

They're not turn-based, they're more like time-based interaction, where they're continuously taking in audio, text, video, continuously providing output.

它们不是轮次式的，而更接近基于时间的交互：它们持续接收音频、文本、视频，并持续给出输出。

Now we cut this up in chunks of 20 milliseconds, and this enables you to actually catch these things like interruptions and simultaneous speech and really create a rich, high bandwidth interaction between humans and machines, which we think is critical to actually enable agency and enable people to be more in this loop as we advance AI further and further.

现在我们把这切成每块 20 毫秒，这就让你能够真正捕捉到打断、同时说话这类情况，并在人与机器之间创造出丰富的、高带宽的交互——我们认为，随着我们把 AI 推进得越来越远，这对于真正实现「主体性」、让人们更深度地留在这一循环之中至关重要。

continuously adv.
/kən'tɪnjuəsli/
持续地，不断地

output n.
/'aʊtpʊt/
输出

chunk n.
/tʃʌŋk/
块，片段

simultaneous adj.
/ˌsɪml'teɪniəs/
同时的

bandwidth n.
/'bændwɪdθ/
带宽

agency n.
/'eɪdʒənsi/
能动性，主体性

critical adj.
/'krɪtɪkl/
至关重要的

句型解析

1. The reason why we even started Thinking Machines is because we wanted to build the frontier AI lab that's really focused on really human AI collaboration piece. 出自【c1p1】

   The reason why we even started Thinking Machines　is because

   我们创办 Thinking Machines 的原因　是因为

   we wanted to build the frontier AI lab

   我们想打造一家前沿 AI 实验室

   that's really focused on really human AI collaboration piece

   真正专注于人机协作这一环节的

   我们之所以创办 Thinking Machines，是因为我们想打造一家真正专注于人机协作这一环节的前沿 AI 实验室。

公式 The reason why + 从句 + is because + 从句 = 「……的原因是……」
仿写 The reason why she moved to the countryside is because she wanted a quieter life.
译文 她搬到乡下的原因是她想要更安静的生活。

2. they're more like time-based interaction, where they're continuously taking in audio, text, video, continuously providing output. 出自【c3p2】

they're more like time-based interaction | where | they're continuously taking in audio | text

它们更接近基于时间的交互 | 在其中

video | continuously providing output

它们持续接收音频、文本、视频 | 并持续给出输出

它们更接近基于时间的交互：持续接收音频、文本、视频，并持续给出输出。

公式 名词 + where + 从句（补充说明该名词的特征） = 「……的……，在其中……」

仿写 They built a platform where users can share ideas and get instant feedback.

译文 他们搭建了一个平台，用户可以在其中分享想法并获得即时反馈。

---

3. Now we cut this up in chunks of 20 milliseconds, and this enables you to actually catch these things like interruptions and simultaneous speech. 出自【c3p3】

Now we cut this up in chunks of 20 milliseconds | and this enables you

现在我们把它切成 20 毫秒一块 | 而这让你能够

to actually catch these things | like interruptions and simultaneous speech

真正捕捉到这些东西 | 比如打断和同时说话

现在我们把它切成每块 20 毫秒，这就让你能够真正捕捉到打断、同时说话这类情况。

公式 cut sth up in chunks of + 数量 = 「把某物切成……大小的块」；enable sb to do = 「使某人能够做」

仿写 Cut the task up in chunks of thirty minutes, and it becomes much easier to finish.

译文 把任务切成每三十分钟一块，完成起来就容易多了。

背景知识

人物

1. 米拉·穆拉蒂（Mira Murati） 人物

曾任 OpenAI 首席技术官，主导 ChatGPT 等产品的技术落地，2024 年离开 OpenAI。本文引述的是她离开后关于下一代 AI 模型形态的观点，她随后创办了 Thinking Machines 实验室。

文化与术语

1. Thinking Machines 术语

米拉·穆拉蒂创办的 AI 研究实验室，定位为「前沿 AI 实验室」，核心方向是人机协作。「Thinking Machines」一名也呼应了早期人工智能史上同名公司，本文介绍的是其「交互模型」这一研究方向。

2. 交互模型（Interaction Models） 术语

指不采用「你问我答」轮次结构、而是持续同步接收多模态输入并实时输出的模型形态。本文将其与当今常见的轮次式模型对比，强调对沉默、打断、语气等副语言信息的捕捉能力。