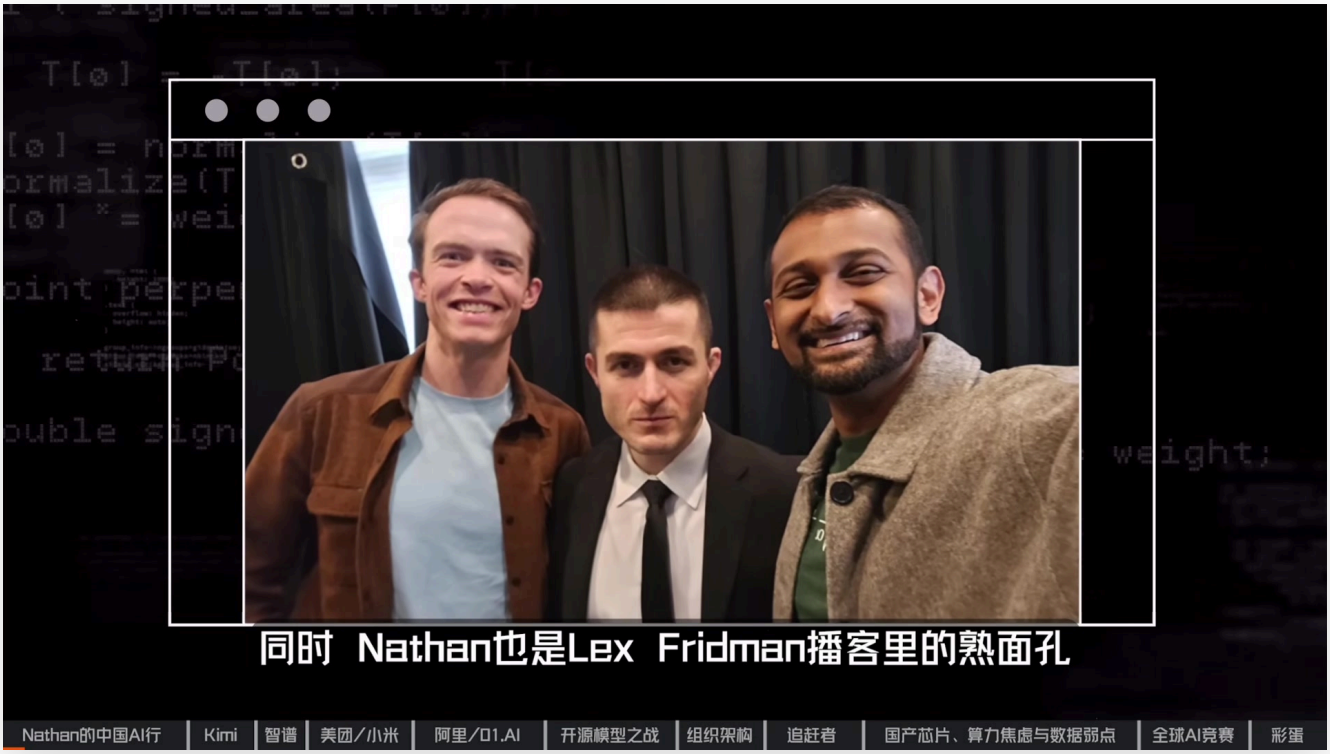
美国AI研究员的中国之旅:年轻人,追赶者,算力焦虑与“AGI展示厅”|专访Nathan Lambert【101视频播客】 – 图文解说

> 一张关键画面对应一段完整解说，按原视频时间推进。

文章摘要

本期围绕《美国AI研究员的中国之旅:年轻人,追赶者,算力焦虑与“AGI展示厅”|专访Nathan Lambert【101视频播客】》展开，按时间顺序讨论scene_001、scene_002、scene_003、scene_004。**以下内容保留完整问答、例子、数字、画面信息和限定条件，适合先读这段摘要建立框架，再结合每个关键帧逐段阅读。**

01. scene_001（00:00–02:34）



场景 1

本节主旨 scene_001

完整内容

陈茜：美国 AI 研究员怎么看中国 AI 模型的发展？**2026 年 4 月**，Nathan Lambert 走访了北京、杭州和上海的 AI 实验室，拜访阿里巴巴、月之暗面、智谱、清华、美团、小米、蚂蚁和零一万物等团队。Nathan 长期参与 Hugging Face 的 RLHF 研究，后来负责 AI2 的大模型后训练，参与 Omni、Tulu 等开源模型，也多次参加 Lex Fridman 的节目。

Nathan：如果一定要回答“谁有最好的中国开源模型”，截至目前他会在 Kimi 和 ZAI 之间选择。中国实验室最大的劣势仍然是算力，Meta 或 OpenAI 可能拥有比这些中国公司合计更多的 GPU。关于华为芯片，他听到的普遍说法是更适合**推理**，还不太适合训练；如果能获得更多华为芯片，一些公司愿意用它来扩大客户。**关于 Anthropic CEO Dario Amodei 对中国 AI 和开放权重模型风险的判断，他认为短期内有些过于激进。**

陈茜：**美国 SOTA 模型领先中国模型多少个月？**这个差距会缩小还是扩大？随后正式从西雅图开始这场全英文访谈。

画面说明

画面是 Nathan 与 Lex Fridman 等人的资料合影，字幕说明他长期参与开放模型研究和播客讨论；这一帧承担嘉宾背景信息，不是普通人物截图。

02. scene_002（02:34–12:11）

场景 2

本节主旨 scene_002

完整内容

Nathan：回来以后，大家总问他最大的结论是什么。AI 行业里谁在做模型，大家其实已经知道，所以他没有一个特别想制造的“热辣观点”。这次旅行对他最重要的价值，是认识真正做这些模型的人，理解他们为什么这样做，而不是只根据美国人的猜测解释中国实验室。

他看到一个正在形成的生态：**美国有很多围绕真正开放模型搭建产品和基础设施的公司，而许多模型在中国被训练出来。**过去 Meta、Google 等公司的政策甚至会让带着工作电脑去中国变得不可能，两边像隔着一堵墙；但现在两个技术社区正在逐渐接触。他希望通过认识具体的人、把他们人性化，降低 AI 议题里的社会和地缘政治误读，因为研究者并不等于做出政治决定的人。

这次大约去了六七晚，主要在北京和杭州。北京住在奥运中心附近，附近清华一带聚集了很多 AI 初创公司，走路就能从一家实验室到另一家，感觉很像湾区。杭州去了阿里巴巴，也和 DeepSeek 相关人员交流，但没有进入 DeepSeek 办公室。访问名单还包括 Kimi（月之暗面）、ZAI、美团、清华、小米、蚂蚁、ModelScope，以及上海的 MiniMax 和可能的 StepFun。**DeepSeek 的神秘感和字节跳动强大的资源、豆包产品，是各家公司都会谈到的话题。**

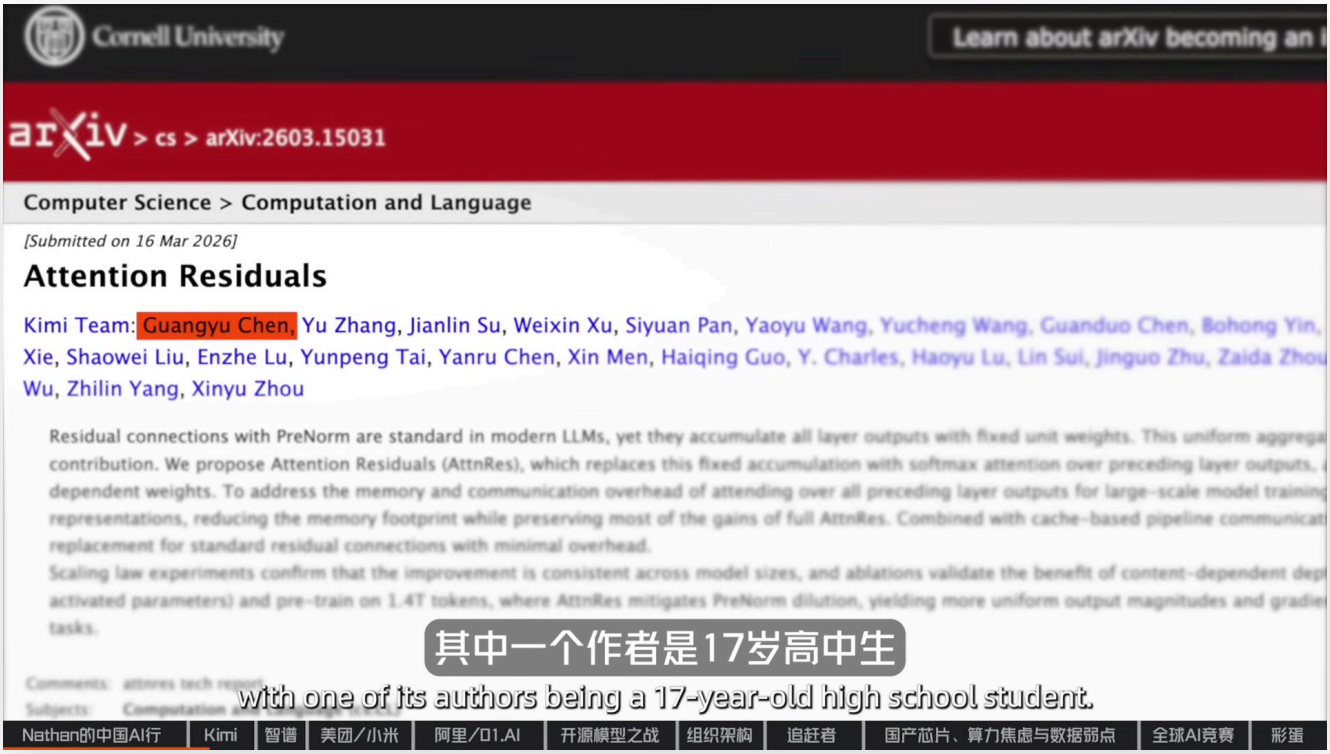
他说自己“带着谦逊回来”，因为只看了两座城市，而且行程像被面包车快速传送，无法凭短暂访问形成对中国社会、教育和研究者直觉的完整理解。**正式访问通常会有公司公关人员在场，领导者更熟悉对外表达，可能少一些细节；最有意思的对话反而是通过微信约一位研究员在办公室外聊天。**

他问得最多的问题包括：为什么要自己训练模型、为什么不直接使用 DeepSeek、目前的瓶颈是什么、AI 将如何影响劳动市场。中国公司普遍很务实：模型要服务自己的用户，最好掌握完整技术栈；即便开放发布，也能获得反馈。很多人说华为芯片可以用于**推理**，但训练仍然依赖更强的 GPU。和美国相比，中国研究者更少在播客生态里公开谈宏观影响，更多回答“我负责把东西做出来”。

画面说明

选中的是访谈双人全景，陈茜与 Nathan 坐在植物和落地窗前，字幕正在讨论“一个在这里成长的人会怎么想”；画面用于承载长段对话关系，不额外虚构地点。

03. scene_003（12:12–16:50）



场景 3

本节主旨 scene_003

完整内容

陈茜：月之暗面给人的感觉和其他公司有什么不同？**她提到 Moonshot 的 Attention Residuals 论文，其中一位作者还是 17 岁高中生。**

Nathan：Kimi 的人似乎很紧密，既享受把有趣的东西做出来，也在同步建立业务。他把月之暗面称为这次访问中“氛围最好”的公司之一：不太像大公司，更像一群人一起投入很酷的事情。**相比之下，阿里和蚂蚁更有大公司和云业务的企业感，ZAI 则非常强调 AGI，带着明显的使命感。**这样的判断只是几小时接触的直觉，不能据此断言一家公司的长期轨迹，但人的表达方式和办公室氛围仍然会透露一些信息。

Attention Residuals 论文被 Elon Musk 转发后，年轻作者在中国迅速受到关注。**Nathan 认为这不只是月之暗面的个案：他在中国实验室连续见到很多非常年轻、深度嵌入前沿模型工作的研究者，有些人同时做研究、产品和对外沟通。**清华周边高校和创业公司的联系也比他预想得更紧密，学生更容易进入实验室。

年轻人没有太多过去的技术路线需要维护，也没有家庭和职业之外的大量分心事项。他们可以把预训练、后训练、数据配方和实验瓶颈全部塞进自己的工作上下文，并直接追着最新范式走。Nathan 说，科学往往“一代新人推动一代变化”，新人没有旧观点和既有声望要保护，反而更愿意接受 MoE、RL、Agent 等新方向。

画面说明

画面是 Cornell arXiv 页面上的 Attention Residuals 论文，能够看到论文标题、作者列表和正文；它比人物近景更直接地证明了本段谈论的论文和年轻作者。

04. scene_004（16:50–20:06）

场景 4

本节主旨 scene_004

完整内容

陈茜：接下来谈智谱。Nathan：智谱让他第一次强烈感受到中国公司的“展示厅文化”。接待来访者时，公司会把发展历史和技术能力做成一套可观看、可演示的空间。智谱的展板和屏幕上不断出现 AGI 功能、AGI loading，以及“AGI 已经走到 **42%**”一类带有《银河系漫游手册》意味的表达。

他没有像在 Kimi 那样充分认识团队，但智谱对自己的工作很自豪，也已经因为上市和被美国列入实体清单等因素，迅速获得“大公司”的外部身份。**Kimi 仍是私人公司，智谱和 MiniMax 已经上市，市场会给上市公司更多注意力和融资机会；另一方面，上市也意味着要面对短期资本市场的要求。**

Nathan 不认为上市天然决定模型质量。Kimi 和智谱的模型各有取舍，未来可能因为资本结构不同而出现不同的文化和发展路径。**智谱更公开、更像一家需要向市场解释自己的公司，Kimi 则更像一支专注把模型和产品继续做下去的团队。**

画面说明

选中画面是 Z.AI 舞台上的“AGI: LOADING…”展示页，字幕提到这是《银河系漫游手册》的隐喻。**这里保留了公司对外展示的视觉证据，而不是用 Nathan 的普通近景代替。**

05. scene_005（20:07–25:20）

场景 5

本节主旨 scene_005

完整内容

陈茜：美国的 Uber、Apple 并没有普遍自己训练开放模型，但在中国，美团、小米等大型公司更倾向于自己建模。为什么？

Nathan：以美团为例，公司希望有能替用户完成任务的 **Agent**，例如帮用户确定路线、叫车或下单。美国消费者可能直接拿一个 API key 来做这件事；中国公司的第一反应往往是，既然有算力和人才，为什么不自己做，而且这样可能更便宜。它们可以先发布通用模型，让更多人使用并获得反馈，再在内部把模型微调成专门的 **Agent** 和产品模型，这些专业模型不一定对外发布。

小米则体现了大公司全栈整合的倾向：语言模型、机器人模型、机器人硬件、汽车和云端服务都想自己连接起来。AI 最终会和所有数字技术交织在一起，可能成为汽车功能、手机功能以及超级 App 的底层能力。**Nathan 还提到，小米在很短时间内从自有模型走到较受关注的模型产品；对于一家已经做过汽车等复杂硬件的公司，训练模型不一定比造车更难。**

中国公司的优势是大平台带来的反馈飞轮：资源越多，模型越能改善产品；产品做得越好，平台就产生更多回报，继续投入训练和**推理**。蚂蚁也在把模型能力扩展到健康等更大范围。**风险是大平台可能压缩小公司的空间；有公司做了能在字节应用里工作的 Agent，却因为平台不允许访问数据而被禁用。**小模型公司仍在寻找自己的接口和细分位置，Nathan 对中型公司最终会独立成长还是被整合保持谨慎，但他看好 AI 作为生意的需求。

画面说明

双人全景展示采访关系和“让 AI 进入业务”的讨论，不额外把字幕中的公司推断成画面中看不见的产品。

06. scene_006（25:21–27:04）

场景 6

本节主旨 scene_006

完整内容

陈茜：你说中国公司有一种“技术所有权心态”，具体是什么意思？

Nathan：这和超级 App 的思路有关。公司相信自己可以把东西做出来，做成以后继续投入，让它更便宜、更高效，并且针对自己的业务场景调优。**中国很多产业就是这样逐渐发展成全球最大的行业之一，所以公司更希望自己建设，而不是长期购买一个关键外部服务；只有把技术和服务掌握在自己手里，才能控制长期路线、利润和竞争力。**

他把这种心态和中国更激烈、更长期的商业竞争联系起来：如果竞争者依赖别人的服务，就很难在所有环节做到最好。训练大模型目前不是一项“不可能负担”的投入，他估计一些实验室一年可能花费 5 亿到数十亿美元。**能力和投入大致呈对数关系，继续花钱仍会变强，但边际回报会下降，因此市场最**

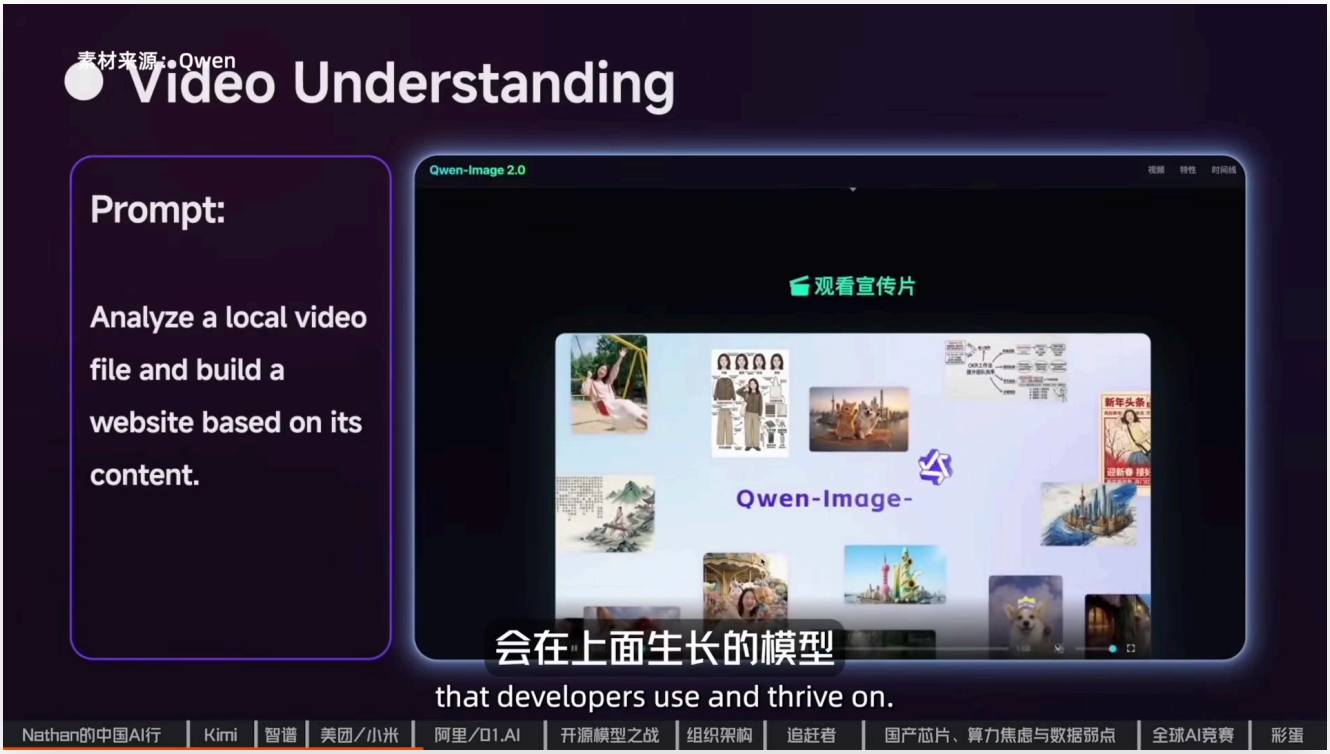
终可能整合。

目前仍处在一个有趣阶段：人才和技术已经存在，公司可以迅速把团队重新组织起来训练大模型。**Nathan 举例说，Richard 曾提到组建大型模型团队只花了大约六个月，因为基础人才和技术设施本来就在公司里。**短期内大家都愿意尝试，但他不确定这种投入会持续多久。

画面说明

选中的是双人全景，Nathan 正在回答“长期拥有技术”这一概念；画面没有额外图表，因此说明保持为访谈构图和字幕中的问题。

07. scene_007（27:04–30:32）



场景 7

本节主旨 scene_007

完整内容

陈茜：阿里是大公司，和 Kimi 等小实验室有什么不同?

Nathan：千问通过持续发布不同尺寸、性能很强的小模型，赢得了开发者的认可。它们在 Llama 4 和 Meta 暂停推进的时期进一步扩大了影响，但千问最大的模型 Qwen Plus、Qwen Max 已经保持闭源，这和阿里云的商业逻辑有关。**阿里在中国云市场的位置类似 Google Cloud 在美国的位置，公司知道最大的模型可以直接变成云收入，所以不会轻易把所有能力开放出去。**

他强调，千问的一到十亿、三十亿参数等小模型并不是偶然做出来的，而是有成熟配方和明确策略。**发布模型需要长期支持：要处理 VLM、开发者生态、兼容性和社区反馈，模型团队不能只负责“训练完成”。**因此千问一方面不想发布过多难以维护的模型，另一方面也不公开最大的模型，以保留云端

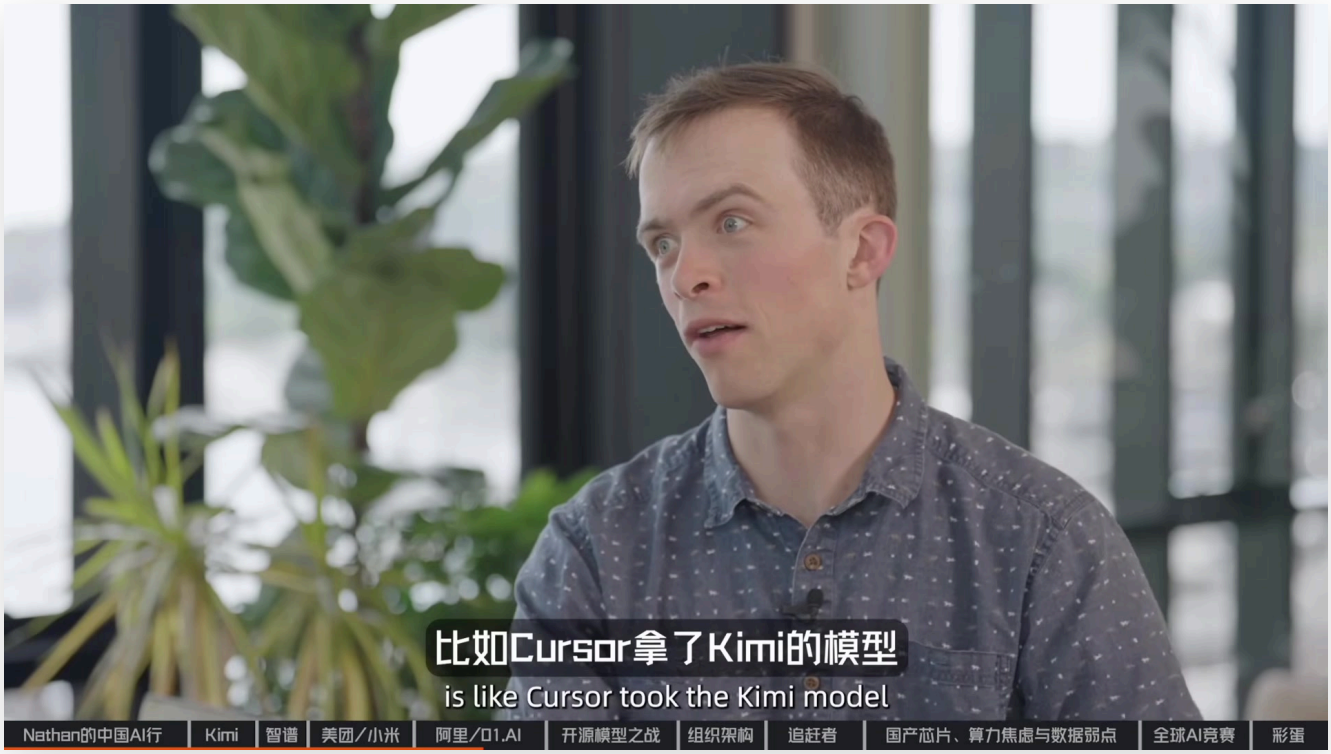
变现空间。

中国过去被认为不愿意像美国 SaaS 那样按席位、按月付软件费，但所有现代公司都在使用云。**Nathan 预计中国会逐渐把 AI 支出理解为一种必要的云支出，而不是普通办公软件订阅。**陈茜补充说，机场、高铁和城市里都能看到阿里云广告，说明云已经成为基础设施品牌。

画面说明

选中画面是“Video Understanding”演示界面，屏幕中有输入提示和视频理解结果。它把本段的模型产品和开发者使用场景具体呈现出来。

08. scene_008（30:32–32:48）



场景 8

本节主旨 scene_008

完整内容

陈茜：阿里云和 AWS 的一个差别是，AWS Bedrock 同时承载很多大模型、自己的模型和小模型，而中国云往往更强调自己的模型。这意味着什么？

Nathan：美国出现了一个很大的中间层，企业可以把别人的模型重新作为 Token 服务，也可以提供 GPU 托管、**推理**、后训练和模型部署。CoreWeave、Nebula、Fireworks、Lambda、OpenRouter 等公司让模型实验室、基础设施和最终产品之间形成多方协作。英伟达也希望生态足够多元，不让 AI 收入只集中到一家模型公司。

中国相对缺少这一层基础设施和优化服务。Nathan 以 Cursor 使用 Kimi 模型为例：Cursor 可以和 Kimi 合作，在 Fireworks 的强化学习沙盒中后训练，再通过 Fireworks 提供服务。这种多方协作能形成强模型和强产品，但在中国可能需要更垂直、更一体化地完成，因为很多公司仍然希望把 GPU 托管、**推理**和产品都握在自己手里。

这不是说中国实验室不会合作。Kimi 等新实验室可能比传统大公司更愿意和 Cursor 这样的产品方交流，但“把一个重要环节交给外部服务商”的商业习惯还没有完全形成。**Nathan 认为，所有权和整合能力是优势，也可能让中国生态少掉一层专业分工。**

画面说明

选中的是双人全景，字幕讨论“美国市场里很大的一部分”和中国缺少的基础设施层。**没有强证据画面时保留全景，避免用重复近景制造虚假的信息密度。**

09. scene_009（32:48–35:47）

场景 9

本节主旨 scene_009

完整内容

陈茜：零一万物由李开复创办，正在经历从模型实验室转向更完整商业模式的变化。中国 To B 市场在 AI 时代会怎样?

Nathan：未来关键在于这些公司能否建立合作关系，把 AI 能力嵌入缺少 AI 专长的企业。美国正在流行的 Forward Deployed Engineer，或者 FDE，本质上是把工程师带进客户组织里解决真实问题。它在美国可能会快速增长，但在中国，企业是否愿意让外部团队进入内部流程、同时接受自己不是技术所有者，仍然取决于本地商业文化。他和李开复见过面，对方已经在重新思考如何做下一阶段的业务；至于为什么零一万物没有成为中国领先模型实验室，最好直接问李开复。

Nathan 还把“训练模型”和“做 AI 生意”分开。训练大模型像不断转动曲柄、持续增加资源，模型不会做完一次就停止。Databricks 的思路是，如果别人能把模型做好，自己就不必继续支付训练成本，专注自己的业务。**相比之下，很多公司还在追逐一万亿参数的 MoE，投入非常大。**对不可能领先的公司来说，尽早停止训练、转向小模型、应用或专门场景，可能比晚几年退出更理性。

结论不是每家公司都不该训练模型，而是要根据资源、产品和客户决定：模型是核心产品，就做；模型只是业务能力，就可能购买、合作或使用更小的模型。

画面说明

选中的是产品演示界面，画面中可以看到聊天/应用操作状态；它对应“把 AI 能力放进业务”的讨论，比单独人物近景更能承载零一万物和 To B 转型的语境。

10. scene_010（35:48–38:57）

场景 10

本节主旨 scene_010

完整内容

陈茜：**美国 SOTA 模型领先中国模型多少？**差距会缩小还是扩大？

Nathan：传统 benchmark 上，差距一段时间以来大约是 6 到 9 个月。但在真实使用中，尤其**推理**模型刚出现时，差距非常紧。新范式的进步速度很快，中国实验室常常在最后一轮 RL 完成后几天内就发布模型，美国实验室则往往需要数周的内部流程。**即使中国模型在技术曲线上落后，只要更快发布，用户真正能用到的性能就会靠近。**

他认为，流行 benchmark 上的六到九个月可能会稳定存在，但美国模型在知识工作领域还有难以量化的稳健性和易用性优势。Claude Coding、Codex 等产品带来的变化，不只是代码质量，而是让普通用户几乎不用学习就能使用。若开放权重模型能以每月 **1 美元的成本**提供同样体验，而不是

OpenAI 每月 **200 美元的**产品，用户会迅速迁移；但截至录制时，他还没有看到这样清晰的分水岭。

中国模型的代码质量会继续提升，公开数据和 GitHub 让训练具备基础。**美国的优势在于真实用户数据：Cursor、ChatGPT、Codex、Claude 都能从用户实际使用中获得反馈；开放权重实验室发布模型后，很多使用并不会回流到开发者。**中国实验室要追上编码产品，不只是解决代码题，还要解决产品易用性和真实使用数据闭环。

画面说明

双人全景展示关于模型差距的完整问答，画面没有可读的图表；说明重点放在字幕中明确出现的“6—9 个月”、发布速度和真实用户数据。

11. scene_011（38:57–44:26）

场景 11

本节主旨 scene_011

完整内容

陈茜：为什么中国实验室发布这么多强大的开源模型？**如果 Meta 没能成为最好的开放模型，是否意味着美国把开放模型领导权输给了中国?**

Nathan：发布开放模型是中国实验室获得相关性的现实路径。即使一家公司的 API 不被知道，公开发布模型也会让开发者试用，帮助公司进入全球前沿 AI 的讨论。**它们知道自己在追赶，因此会继续用开放模型建立客户、反馈和品牌，而不是把这件事包装成纯粹的理想主义。**

他判断，美国在 **2025 年**夏天之后已经把开放模型领导权让给了中国，GLM 4.5、Kimi K2 等模型成为转折点。但美国仍有机会，因为英伟达有强烈的经济动机，不希望 AI 被任何一家模型公司完全控制。Meta、Microsoft 等拥有其他产品的公司也能发布开放权重模型，只是它们可能担心分散注意

力、市场营销和安全风险。

英伟达重新发布 Nemotron 系列，标志着它再次尝试开放模型。大 MoE 模型可以生成合成数据、蒸馏出更小的模型，并继续扩展尺寸。与此同时，开放生态不应该只追逐最大的模型：小模型便宜，如果能稳定完成一个任务，成本会远低于 Claude 或 Codex。AI2 更适合做有研究价值的小模型和公共资源，而不是花费无法承受的资金去追逐一万亿参数模型。Nathan 认为当前 Kimi 和 ZAI 的开放模型很强，DeepSeek 的 V3、R1 曾在 **2025 年**占据明显优势，竞争仍然会以很小的差距快速变化。

画面说明

英伟达 Jensen Huang 的资料画面出现在讨论“英伟达为什么可能成为美国开放模型的希望”时。**它是本段的关键证据帧，替代了普通人物近景。**

12. scene_012（44:26–56:46）

场景 12

本节主旨 scene_012

完整内容

陈茜：你在文章里谈到很多中国研究者。**他们和美国 AI 实验室有什么不同?**

Nathan：他在中国实验室看到的研究者整体更年轻、更同质化，很多人来自中国本土高校；美国实验室则更国际化，成员来自欧洲、中东、中国和亚洲其他地区。**中国新一代研究者普遍英语很好，关注 Twitter、英文博客和美国 AI 生态；反方向却不成立，很多美国研究者并不了解中国技术生态，信息流动并不对称。**

美国研究界更热衷讨论模型性格、未来哲学、递归自我改进和职业焦虑。中国研究者更务实、少一些集体性的 AI 社会焦虑，但并不代表没有压力。他们更关心模型能否追上、算力够不够、下一轮训练是否成功。**美国限制高技能移民让 Nathan 担心，因为中国研究者在美国几乎每家前沿实验室都承担**

关键工作；如果地缘政治压力让人才回流，美国公司会受到严重影响。

训练一个模型不只是“有一个天才带队”。它更像果园问题：要把很多人才组织到同一个模型上，把大问题拆成大量可协作的小问题。**Nathan 猜测，中国工程师的训练方式可能更适合画出这样的组织结构，让每个人把自己的部分磨好并服务于更大的模型，但他强调这只是推测，因为前沿实验室的真实组织方式仍缺乏公开资料。**

年轻人也更愿意拥抱新范式。**MoE 不是 DeepSeek 发明的，但 DeepSeek 把它优化得非常有效；没有旧路线和既有成功记录要维护的研究者，更容易接受新方法。**标准化考试和长期竞争又可能让人更擅长执行与优化，却不一定自动培养出最具原创性的“从零到一”型研究者。最后，更多中国研究者进入美国和全球讨论，对双方理解和全球平衡都有好处。

画面说明

选中的是 Nathan 近景，字幕正在讨论“中国研究者”与“美国实验室”的差异。此处没有资料画面，保留说话人近景并把组织、人才和信息流动完整交给文字。

13. scene_013（00:56:46–01:05:45）

场景 13

本节主旨 scene_013

完整内容

陈茜：和中国研究者交流时，你看到更多的是自信、焦虑还是好奇？

Nathan：最明显的焦虑是跟不上，而跟不上主要表现为算力不足。**中国实验室都希望拿到更多 NVIDIA GPU 来训练最新模型，研究者最直接地感受到资源限制。**相比之下，美国社会里关于 AI 取代工作的焦虑更强：美国 CEO 反复谈论失业，甚至顶尖研究者也会担心自己的工作被替代；中国开发者也知道软件工作会改变，但这还没有形成同样强烈的集体恐慌。他们更像是先把东西做出来，再考虑后果。

中国经历了很多连续的技术和经济变化，超级 App、电动车等已经改变了日常生活，所以 AI 被许多人看成下一轮变化，而不是一个突然打破稳定的外来力量。安全问题在中国实验室通常被放在“安全”和发布风险的框架下，团队会避免开放模型权重带来明显的新风险，但较少出现美国那种围绕 AI 末日风险的意识形态式争论。Nathan 支持开放模型，但不是绝对主义者：并不是每个模型都必须第一天就把权重上传到 Hugging Face，技术扩散可以有时间差。

中国开放模型的一个优势是长期积累了真实的开发者反馈。**Kimi、ZAI 等模型经过多轮发布，逐渐知道人们如何使用它们，这种“什么让开放模型真正好用”的知识很难从 benchmark 直接得到。**美国公司也在加速发布 Nemotron 等模型，以获得更多反馈。

但中国实验室最根本的限制仍是 GPU。Meta 或 OpenAI 很可能拥有比中国公司总和更多的 GPU；一家中国实验室说最新预训练要花六个月，如果这次失败，整条模型路线都可能不存在。算力不足会限制实验空间，迫使实验室更谨慎地优化现有路线。

画面说明

画面是双人全景，主题从研究者的情绪转向算力约束；截图不把字幕中提到的 GPU 数量伪装成画面中真实出现的图表。

14. scene_014（01:05:46–01:11:43）

场景 14

本节主旨 scene_014

完整内容

陈茜：算力稀缺会怎样改变模型策略？训练和**推理**分别会发生什么？

Nathan：模型设计取决于下游怎么使用，也取决于训练集群的形状。集群规模会影响哪种参数规模更容易训练；之后还要考虑如何部署和服务用户。小实验室的预训练周期更长，可能更少改变基础架构。若希望在华为芯片上做**推理**，模型还要同时兼顾另一套推理栈。训练集群、推理系统、用户需求和参数规模，是一个多目标优化问题。

谈到华为芯片，他听到的共识是：可以用于推理，但暂时不适合大规模训练。很多语言模型公司愿意用华为芯片做推理，因为这能增加服务能力和客户数量；但在训练端，目前更现实的是训练小模型。**他预计这种差距还会持续若干年，因为 NVIDIA 仍然在把全球供应链集中到顶级产品上，而华为是在**

“需求推动创新”的环境里追赶。

关于 Jensen Huang 和 Dario Amodei 对中国芯片与 AI 风险的不同立场，Nathan 认为自己处在中间但更偏向 Amodei 一侧。芯片出口既有商业市场，也有技术和地缘政治风险；限制销售可能无法完全阻止中国技术进步，却会影响美国企业、政策和供应链的平衡。他没有足够信息给出精确的市场均衡判断。

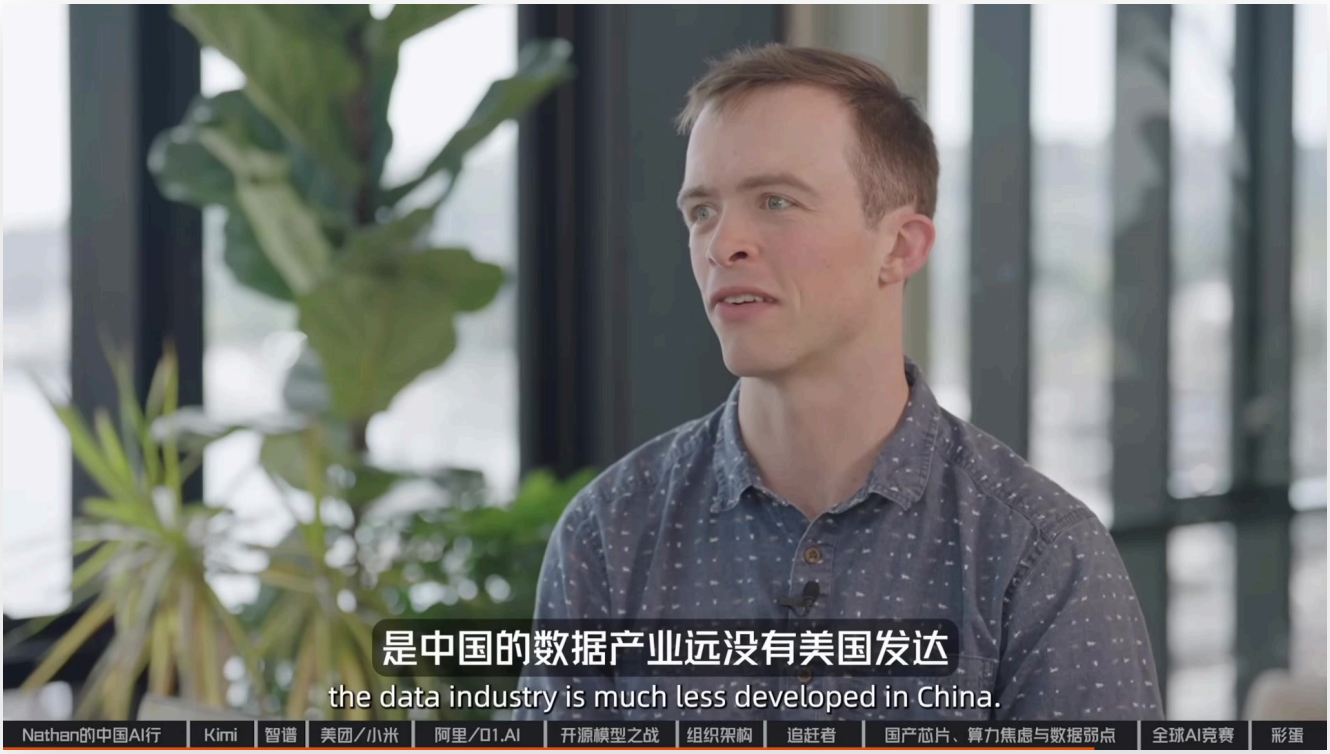
他更确定的是需求会推动本地 AI 栈。中国开发者已经在使用 Claude 等工具，需求一旦被国内云和应用承接，就会汇聚到华为能生产多少芯片、软件生态有多容易使用。这个过程可能以月为单位发生，而不是几年。**瓶颈不仅是芯片数量，也包括类似 CUDA 的软件和模型适配层。**

最后谈到 Dario Amodei 的风险表达，Nathan 认为他对开放权重模型的短期风险判断略显激进，应该区分中国政府、个体研究者和模型本身的风险。

画面说明

画面是访谈插入的一段技术/办公资料画面，展示真实场景中的 AI 研发与设备语境；它比人物近景更适合承接国产芯片和**推理**基础设施的讨论。

15. scene_015（01:11:43–01:14:50）



场景 15

本节主旨 scene_015

完整内容

陈茜：你说中国 AI 的数据产业弱点比很多人意识到的更重要，为什么？

Nathan：他最惊讶的是，中国的数据产业没有美国发达。**美国的 Anthropic、OpenAI 会花数十亿美元从外部供应商购买环境和高质量训练数据；他问到中国公司时，很多人的回答是没有类似规模的外部采购，主要还是内部建设。**字节跳动和阿里有能力建立内部数据团队，但“模型哪里不好，就去外面购买针对性数据”的市场在中国不常见。

这会影响知识工作模型。美国前沿模型正在进入医疗、法律、咨询、金融等工作场景，很多训练数据由专业人士通过 Scale AI、Mercor、Turing 等公司生成。市场也在把数据生产从低成本外包推向更高技能的专业数据。**每一轮数据都应该让模型在具体任务上多提升一点，这种增量可能很小，却是模**

型变好所需要的长期闭环。

如果中国实验室没有成熟的外部数据供应商，它们要么依赖内部构造环境，要么等待国内 Agent 和知识工作市场成长。**只有当中国专业用户真正用本土模型完成工作，模型公司才有理由花钱购买国内专业数据。**否则，研究员自己编一个“请制作金融 Excel 表格”的环境，未必能像真实从业者那样构造有价值的任务。

因此这不是简单的“数据多或少”，而是数据公司、专业用户、模型部署和反馈之间有没有形成产业闭环。Nathan 认为这可能是中国模型要补上的结构性短板。

画面说明

选中的是 Nathan 近景，字幕明确出现“data industry is much less developed in China”。**本段没有可读的数据图表，因此不添加虚构的市场数字。**

16. scene_016（01:14:51–01:16:34）

场景 16

本节主旨 scene_016

完整内容

陈茜：移动互联网时代，美团、微信、阿里巴巴改变了日常生活。**模型准备好以后，中国是不是会在 AI 应用上更快?**

Nathan：美国和中国都会有很强的应用创新。**一个差别是，美国用户更习惯下载新 App、试用它、为它付费；中国用户已经被超级 App 组织起来，因此问题变成：超级 App 能否承载同样多样的新场景。**

他以 Sora 为例：它曾经成为第一名应用，有数百万人尝试，但后来又像一次失败的实验。**美国的软件生态允许一个应用在相对独立的沙盒里快速试错；中国的超级 App 生态可能更适合把 AI 直接嵌入已有入口，但也可能让新的独立产品更难获得空间。**两种模式都能产生应用，只是试错和分发方式不

同。

关于商业化，Nathan 没有观察到中国前沿实验室公开地表现出特别强的变现焦虑。**OpenAI、Anthropic 需要上市，产品经理会关心收入和算力预算，但许多模型研究者的直接目标仍是把模型做得最好。**如果模型持续变强、确实有用，融资不会是唯一决定因素；产品经理、实验室管理层和研究者可能有不同的优先级。

画面说明

双人全景把问题和回答放在同一画面中，适合表现“超级 App 与独立应用”的对比讨论；画面没有展示具体 App，因此文字不把画面误称为产品演示。

17. scene_017（01:16:34–01:22:01）

场景 17

本节主旨 scene_017

完整内容

陈茜：**美国实验室是否应该更多访问中国实验室？**一年后会发生什么？

Nathan：首先会有一批美国中小公司主动去中国。它们不一定训练基础模型，而是围绕开放模型做微调 API、专门部署、基础设施和产品。**中国实验室比较愿意分享技术和专业知识，双方可以从“访问”发展到真正的跨境合作。**对美国公司来说，目前去中国的签证和过境安排相对容易；中国研究者去美国则要面对更复杂的签证流程。

未来 12 个月，他预计会出现新的 **Agent** 产品，可能是覆盖整个工作流程的“复杂 **Agent**”，不一定由今天最知名的公司推出。模型能力大概率继续变好，OpenAI、Anthropic 会继续增长，中型公司 Fireworks、Baseten 以及 Kimi、ZAI、MiniMax 等也会靠销售 Token 和模型能力获得收入。美国

与中国前沿模型的可见差距可能从几个月扩大到六到九个月甚至一年，但中国不会因此停止，会继续出现 DeepSeek、Kimi、GLM 等新版本。

更长的三到五年，部分实验室可能因为融资和市场整合而退出。**Nathan 对全球 AI 竞赛总体中性，真正担心的是美国公众对 AI 失去信心，从而拖慢进展。**地缘政治上，美国仍领先，中国在追赶，芯片销售和限制会改变生态，但不能把研究者和用户等同于最高层的政策决策者。

他和陈茜都认为，亲自走进另一边的实验室，是理解这片影响全球技术的人群的最好方式。**Nathan 希望以后再次来中国，也希望两边更多人能看到对方的真实工作方式，让竞赛少一些极端化叙事。**

画面说明

双人全景是主画面，字幕从跨境合作转到“新一代”的讨论。**这里保留人物关系，因为主要信息来自预测、限定条件和对双方动机的解释，而非外部图表。**

18. scene_018（01:22:04–01:26:57）

场景 18

本节主旨 scene_018

完整内容

陈茜：最后进入 AI2 的办公室。Nathan：AI2 做开放源代码 AI 研发，过去更偏学术，后来转向“可复用产物”开发：模型、代码库和数据集都让社区继续使用。Olmo 等开放语言模型很受欢迎，另有多模态方向和 Semantic Scholar。AI2 也在做帮助研究人员理解科学文献的 **Agent** 工具。

AI2 的核心文化是科学和公共服务。它大约有 200 人，约一半是技术研究与工程人才，另一部分负责集群和产品支持。它是完整的非营利机构，资金来自慈善基金会和 NSF 等资助，不以销售产品为主要目标，而是把成果免费交给别人学习和使用。

和大学相比，AI2 有更多资源，也有更强的压力去做出真正有效的模型；和追逐 AGI 的商业前沿实验室相比，它拥有更多研究自由。它处在“封闭实验室到学术机构”的中间位置：需要让模型足够好，但不必被商业产品和上市节奏完全牵引。

聊到西雅图和湾区，Nathan 说西雅图更安静，没有每天不断发生的行业活动，更适合安定下来组建家庭；湾区的 AI 几乎进入每一次对话。陈茜看到窗外的 Lake Union，称这是自己梦想中的办公室。**谈到中国实验室的办公室，Nathan 说普通科技公司办公空间其实很相似，真正独特的是接待来访者时把企业历史和能力做成展示厅的文化。**

陈茜：感谢 Nathan，也感谢大家对硅谷101的留言、转发和点赞。她用中文完成收尾。

画面说明

画面从猫和多模态 AI2 资料切换到 Nathan 与陈茜在 AI2 办公室走廊、公共区域和落地窗边行走。选中的画面同时保留两位人物和办公室空间，明确这是 office tour，而不是普通访谈镜头。