

EDWIN CHEN

LENNY'S PODCAST

BILINGUAL TRANSCRIPT

ORIGINAL BY

Lenny Rachitsky

@lennysan • x.com/lennysan

ANALYSIS BY

@Penny777 • x.com/penny777

Edwin Chen - 双语对照

This is the complete bilingual (English-Chinese) transcript for Lenny's Podcast featuring Edwin Chen, founder and CEO of Surge AI.

[00:00:00] Lenny Rachitsky

English:

You guys hit a billion in revenue in less than four years with around 60 to 70 people. You're completely bootstrapped, haven't raised any VC money. I don't believe anyone has ever done this before.

中文翻译:

你们在不到四年的时间里，仅凭 60 到 70 人就实现了 10 亿美元的营收。你们完全是自筹资金 (bootstrapped)，没有拿过任何风险投资。我不相信以前有人做到过这一点。

[00:00:10] Edwin Chen

English:

We basically never wanted to play the Silicon Valley game. I always thought it was ridiculous. I used to work at a bunch of the big tech companies and I always felt that we could fire 90% of the people and we would move faster because the best people wouldn't have all these distractions. So when we start Surge, we wanted to build it completely differently with a super small, super elite team.

中文翻译:

我们基本上从不想玩硅谷的那套游戏。我一直觉得那很荒谬。我曾在几家大科技公司工作过，总觉得我们可以裁掉 90% 的人，而且速度反而会更快，因为最优秀的人才就不会被这些琐事干扰。所以当我们创办 Surge 时，我们想用一种完全不同的方式来构建它，打造一支极小规模、极其精英的团队。

[00:00:26] Lenny Rachitsky

English:

You guys are by far the most successful data company out there.

中文翻译:

你们绝对是目前最成功的 AI 数据公司。

[00:00:29] Edwin Chen

English:

We essentially teach AI models what's good and what's bad. People don't understand what quality even means in this space. They think you could just throw bodies at a problem and get good data, that's completely wrong.

中文翻译:

我们本质上是在教 AI 模型什么是好的，什么是坏的。人们并不理解在这个领域“质量”到底意味着什么。他们认为只要针对一个问题堆人头（人海战术）就能获得好数据，这完全是错误的。

[00:00:40] Lenny Rachitsky

English:

To a regular person, it doesn't feel like these models are getting that much smarter constantly.

中文翻译:

对于普通人来说，并不会觉得这些模型在持续变得聪明很多。

[00:00:43] Edwin Chen

English:

Over the past year, I've realized that the values that the companies have will shape the model. I was asking Claude to help me drop an email the other day. And after 30 minutes, yeah, I think it really crafted me the perfect email and I sent it. But then I realized that I spent 30 minutes doing something that didn't matter at all. If you could choose the perfect model behavior, which model would you want? Do you want a model that says, "You're absolutely right. There are definitely 20 more ways to improve this email," and it continues for 50 more iterations or do you want a model that's optimizing for your time and productivity and just says, "No. You need to stop. Your email's great. Just send it and move on"?

中文翻译:

在过去的一年里，我意识到公司的价值观会塑造模型。前几天我让 Claude 帮我起草一封邮件。30 分钟后，是的，我觉得它确实为我写出了一封完美的邮件，然后我发了出去。但随后我意识到，我花了 30 分钟做了一件根本不重要的事情。如果你可以选择完美的模型行为，你想要哪种模型？你想要一个会对你说“你完全正确，肯定还有 20 种方法可以改进这封邮件”并继续迭代 50 次的模型，还是想要一个为你节省时间、提高效率，直接说“不，你该停下了，你的邮件写得很好，直接发出去然后做别的事”的模型？

[00:01:14] Lenny Rachitsky

English:

You have this hot take that a lot of these labs are pushing AGI in the wrong direction.

中文翻译:

你有一个很犀利的观点，认为很多实验室正在把通用人工智能（AGI）推向错误的方向。

[00:01:18] Edwin Chen

English:

I'm worried that instead of building AI that will actually advance us as a species, curing cancer, solving poverty, understand the universe, we are optimizing for AI slop instead. But we're optimizing your models for the types of people who buy tabloids at a grocery store. We're basically teaching our models to chase dopamine instead of truth.

中文翻译:

我担心我们不是在构建真正能推动人类物种进步、治愈癌症、解决贫困、理解宇宙的 AI，而是在优化“AI 垃圾内容”（AI slop）。我们正在针对那些在超市买八卦小报的人群来优化模型。我们基本上是在教我们的模型去追求多巴胺，而不是真理。

[00:01:35] Lenny Rachitsky

English:

Today, my guest is Edwin Chan, founder and CEO of Surge AI. Edwin is an extraordinary CEO and Surge is an extraordinary company. They're the leading AI data company, powering training at every frontier AI lab. They're also the fastest company to ever hit \$1 billion in revenue in just four years after launch with fewer than 100 people and also completely bootstrapped. They've never raised a dollar in VC money, they've also been profitable from day one.

中文翻译:

今天，我的嘉宾是 Surge AI 的创始人兼 CEO Edwin Chen。Edwin 是一位非凡的 CEO，Surge 也是一家非凡的公司。他们是领先的 AI 数据公司，为每一个前沿 AI 实验室的训练提供动力。他们也是有史以来最快达到 10 亿美元营收的公司，在成立仅四年后就实现了这一目标，团队人数不足 100 人，而且完全是自筹资金。他们从未拿过一分钱的风险投资，而且从第一天起就实现了盈利。

[00:02:05] Lenny Rachitsky

English:

As you'll hear in this conversation, Edwin has a very different take on how to build an important company, and how to build AI that is truly good and useful to humanity. I absolutely love this conversation and I learned a ton. I'm really excited for you to hear it. If you enjoy this podcast, don't forget to subscribe and follow it in your favorite podcasting app or YouTube. It helps tremendously.

中文翻译:

正如你在这次对话中将听到的，Edwin 对于如何建立一家重要的公司，以及如何构建真正对人类有益且有用的 AI，有着非常独特的见解。我非常喜欢这次对话，也学到了很多。我非常期待你们能听到它。如果你喜欢这个播客，别忘了在你不常用的播客应用或 YouTube 上订阅和关注。这对我们帮助巨大。

[00:02:27] Lenny Rachitsky

English:

And if you become an annual subscriber of my newsletter, you get a ton of incredible products for free for an entire year, including Devin, Lovable, Replit, Bolt, N8N, Linear, Superhuman, Descript, Wispr Flow, Gamma, Perplexity, Warp, Granola, Magic Patterns, Raycast, ChatPRD, Mobbin, PostHog, and Stripe Atlas. Head on over to lennysnewsletter.com and click Product Pass. With that, I bring you Edwin Chen after a short word from our sponsors.

中文翻译:

如果你成为我时事通讯的年度订阅者，你将免费获得一整年的一系列不可思议的产品，包括 Devin, Lovable, Replit, Bolt, N8N, Linear, Superhuman, Descript, Wispr Flow, Gamma, Perplexity, Warp, Granola, Magic Patterns, Raycast, ChatPRD, Mobbin, PostHog 和 Stripe Atlas。请访问 lennysnewsletter.com 并点击 Product Pass。在听完赞助商的简短介绍后，我将为你带来 Edwin Chen。

[00:02:53] Lenny Rachitsky

English:

My podcast guests tonight love talking about craft, and taste, and agency, and product market fit. You know what we don't love talking about? SOC 2. That's where Vanta comes in. Vanta helps companies of all sizes get compliant fast and stay that way with industry-leading AI, automation, and continuous monitoring. Whether you're a startup tackling your first SOC 2 or ISO 27001 or an enterprise managing vendor risk, Vanta's trust management platform makes it quicker, easier, and more scalable. Vanta also helps you complete security questionnaires up to five times faster so that you can win bigger deals sooner.

中文翻译:

今晚我的播客嘉宾喜欢谈论工艺、品味、能动性和产品市场契合度。你知道我们不喜欢谈论什么吗？SOC 2 合规。这就是 Vanta 大显身手的地方。Vanta 通过行业领先的 AI、自动化和持续监控，帮助各种规模的公司快速获得合规并保持合规。无论你是正在处理第一个 SOC 2 或 ISO 27001 的初创公司，还是管理供应商风险的大型企业，Vanta 的信任管理平台都能让合规变得更快、更简单、更具扩展性。Vanta 还能帮你完成安全问卷的速度提高 5 倍，让你能更快赢得大单。

[00:03:29] Lenny Rachitsky

English:

The result, according to a recent IDC study, Vanta customers slashed over \$500,000 a year and are three times more productive. Establishing trust isn't optional. Vanta makes it automatic. Get \$1,000 off at vanta.com/lenny.

中文翻译:

根据 IDC 最近的一项研究，Vanta 的客户每年可以节省超过 50 万美元，生产力提高三倍。建立信任不是可选项。Vanta 让它变得自动化。在 vanta.com/lenny 获取 1000 美元的折扣。

[00:03:48] Lenny Rachitsky

English:

Here's a puzzle for you. What do OpenAI, Cursor, Perplexity, Vercel, Plad, and hundreds of other winning companies have in common? The answer is they're all powered by today's sponsor, WorkOS. If you're building software for enterprises, you've probably felt the pain of integrating single sign-on, Skim, RBAC, audit logs, and other features required by big customers. WorkOS turns those deal blockers into drop-in APIs with a modern developer platform built specifically for B2B SaaS.

中文翻译:

给你出一个谜题。OpenAI、Cursor、Perplexity、Vercel、Plad 以及数百家其他成功的公司有什么共同点？答案是，它们都由今天的赞助商 WorkOS 提供支持。如果你正在为企业开发软件，你可能感受过集成单点登录

(SSO)、SCIM、基于角色的访问控制 (RBAC)、审计日志以及大客户要求的其他功能的痛苦。WorkOS 通过专门为 B2B SaaS 构建的现代开发平台，将这些交易阻碍变成了即插即用的 API。

[00:04:17] Lenny Rachitsky

English:

Whether you're a seed stage startup trying to land your first enterprise customer or a unicorn expanding globally, WorkOS is the fastest path to becoming enterprise-ready and unlocking growth. They're essentially Stripe for enterprise features.

中文翻译:

无论你是试图获得第一个企业客户的种子期初创公司，还是正在全球扩张的独角兽，WorkOS 都是实现企业级就绪并释放增长的最快路径。他们本质上是企业级功能的“Stripe”。

[00:04:30] Lenny Rachitsky

English:

Visit workos.com to get started or just hit up their Slack support where they have real engineers in there who answer your questions superfast. WorkOS allows you to build like the best with delightful APIs, comprehensive docs, and a smooth developer experience. Go to workos.com to make your app enterprise ready today.

中文翻译:

访问 workos.com 开始使用，或者直接联系他们的 Slack 支持，那里有真正的工程师会超快地回答你的问题。WorkOS 让你能通过愉悦的 API、详尽的文档和流畅的开发体验，像顶尖公司一样构建产品。立即访问 workos.com，让你的应用具备企业级能力。

[00:04:51] Lenny Rachitsky

English:

Edwin, thank you so much for being here and welcome to the podcast.

中文翻译:

Edwin，非常感谢你能来，欢迎来到我们的播客。

[00:04:55] Edwin Chen

English:

Thanks so much for having me. I'm super excited.

中文翻译:

非常感谢邀请我。我非常兴奋。

[00:04:58] Lenny Rachitsky

English:

I want to start with just how absurd what you've achieved is. A lot of people and a lot of companies talk about scaling massive businesses with very few people as a result of AI, and you guys have done this in a way that is unprecedented. You guys hit a billion in revenue in less than four years with less than 60, around 60 to 70 people, you're completely bootstrapped, haven't raised any VC money, I don't believe anyone has ever done this before, so you guys are actually achieving the dream of what people are describing will happen with AI. I'm curious just, do you think this will happen more and more as a result of AI? And also just where has AI most helped you find leverage to be able to do this?

中文翻译:

我想先从你们取得的成就多么惊人开始谈起。很多人和很多公司都在谈论如何利用 AI 以极少的人手扩展大规模业务，而你们做到的方式是前所未有的。你们在不到四年的时间里，仅凭 60 到 70 人就实现了 10 亿美元的营收，完全是自筹资金，没拿过风投。我不相信以前有人做到过，你们实际上实现了人们所描述的 AI 时代下的梦想。我很好奇，你认为这种现象会因为 AI 而越来越多吗？另外，AI 在哪些方面最能帮助你们找到实现这一目标的杠杆（leverage）？

[00:05:40] Edwin Chen

English:

Yeah, so we hit over a billion of revenue last year with under 100 people. And I think we're going to see companies with even crazier ratios, like 100 billion per employee in the next few years. AI is just going to get better and better and make things more efficient so that ratio just becomes inevitable. I used to work at a bunch of the big tech companies and I always felt that we could fire 90% of people and we would move faster because the best people wouldn't have all these distractions. And so when we started Surge, we wanted to build it completely differently with a super small, super elite team, and yeah, what's crazy is that we actually succeeded. And so I think two things are colliding. One is that people are realizing that you don't have to build giant organizations in order to win. And two, yeah, all these efficiencies from AI. And they're just going to lead to a really amazing time in company building.

中文翻译:

是的，我们去年以不到 100 人的规模实现了超过 10 亿美元的营收。我认为在未来几年，我们会看到比例更疯狂的公司，比如人均创收 1000 亿美元。AI 只会变得越来越好，效率越来越高，所以这种比例的出现是必然的。我曾在几家大科技公司工作过，总觉得我们可以裁掉 90% 的人，而且速度反而会更快，因为最优秀的人才就不会被这些琐事干扰。所以当我们创办 Surge 时，我们想用一种完全不同的方式来构建它，打造一支极小规模、极其精英的团队。令人疯狂的是，我们居然真的成功了。我认为这是两件事的碰撞：一是人们意识到你不需要建立庞大的组织也能获胜；二是 AI 带来的所有这些效率提升。这两者将引领一个公司建设的黄金时代。

[00:06:29] Edwin Chen

English:

The thing I'm excited about is that the types of companies are going to change too. It won't just be that they're smaller, we're going to see fundamentally different companies emerging. If you think about it, fewer employees means less capital. Less capital means you don't need a raise. So instead of companies started by founders who are great at pitching and great at hyping, you'll get founders who are really great at technology and product. And instead of products optimized for revenue and what VCs want to see, you'll get more interesting ones built by these tiny obsessed teams. So people building things they actually care about, real technology and real innovation. So I'm actually really hoping that the slick on [inaudible 00:07:06], it'll go back to being updates for hackers again.

中文翻译:

让我兴奋的是，公司的类型也会发生变化。不仅仅是规模变小，我们会看到从根本上不同的公司出现。你想想，员工越少意味着所需资金越少。资金需求少意味着你不需要融资。因此，未来的创始人不再是那些擅长路演和炒作的人，而是真正精通技术和产品的人。产品也不再是为了营收和满足风投口味而优化，而是由这些规模极小、极度痴迷的团队构建出更有趣的东西。人们会构建自己真正关心的东西，真正的技术和真正的创新。所以我真的希望硅谷能回归本质，再次成为黑客们的乐园。

[00:07:08] Lenny Rachitsky

English:

You guys have done a lot of things in a very contrarian way, and one was actually just not being on LinkedIn, posting viral posts, not on Twitter, constantly promoting Surge. I think most people hadn't heard of Surge until just recently, and then you just came out, and like, okay, the fastest growing company at a billion dollars. Why would you do that? I imagine that was very intentional.

中文翻译:

你们做了很多非常反传统（contrarian）的事情，其中之一就是不在 LinkedIn 上发爆款贴，不在 Twitter 上不断宣传 Surge。我想大多数人直到最近才听说 Surge，然后你们一出现就是“营收达 10 亿、增长最快的公司”。为什么要这么做？我想这一定是刻意为之。

[00:07:27] Edwin Chen

English:

We basically never wanted to play the Silicon Valley game. And like I always thought it was ridiculous. What did you dream of doing when you were a kid? Was it building a company from scratch yourself and getting in the weeds of your code and your product every day? Or was it explaining all your decisions to VCs and getting on this giant PR and fundraising hamster wheel? And it definitely made things more difficult for us, because yeah, when you fundraise, you just naturally get part of this kind of Silicon Valley industrial complex where people will, your VCs will tweet about you. You'll get the tech runs outlines, you'll get announced in all of the newspapers because you raised at this massive valuation. And so it made things more difficult us because the only way we were going to succeed was by building a 10 times better product and getting word of mouth from researchers. But I think it also meant that our customers were people who really understood data and really cared about it.

中文翻译:

我们基本上从不想玩硅谷的那套游戏。我一直觉得那很荒谬。你小时候的梦想是什么？是亲手从零开始建立一家公司，每天钻研代码和产品？还是向风投解释你的每一个决定，并陷入公关和融资的无尽循环（hamster wheel）中？这确实让我们的处境变得更困难，因为当你融资时，你自然会成为硅谷工业复合体的一部分，风投会为你发推，你会登上科技媒体头条，因为高估值融资而被各大报纸报道。这对我们来说更难，因为我们成功的唯一途径就是做出好 10 倍的产品，并赢得研究人员的口碑。但我认为这也意味着我们的客户是那些真正理解数据并真正关心数据的人。

[00:08:21] Edwin Chen

English:

I always thought it was really important for us to have early customers who were really aligned with what we were building, and who really cared about having really high quality data, and really understood how that data would make their AI models so much better because they were the ones helping us. They were the ones giving us feedback on what we're producing. And so just having that kind of very close mission alignment with our customers actually helped us early on. So these are people who basically just buying our product because they knew how different it was and because it was helping them rather than because they saw something in that current [inaudible 00:08:52]. So it made things harder for us, but I think in a really good way.

中文翻译:

我一直认为，拥有与我们的愿景高度一致的早期客户非常重要，他们真正关心高质量数据，并理解这些数据如何让他们的 AI 模型变得更好，因为正是他们在帮助我们。他们是那些对我们的产出提供反馈的人。因此，与客户保持这种紧密的使命一致性实际上在早期帮助了我们。这些人购买我们的产品是因为他们知道它有多么不同，因为它确实能帮到他们，而不是因为他们在什么热门榜单上看到了它。所以这虽然让事情变得更难，但我认为这是一种非常正向的困难。

[00:08:55] Lenny Rachitsky

English:

It's such an empowering story to hear this journey for founders that they don't need to be on Twitter all day promoting what they're doing. They don't have to raise money. They can just kind of go heads down and build, so I love so much about the story of Surge. For people that don't know what Surge does, just to give us a quick explanation of what Surge is.

中文翻译:

对于创始人来说，听到这样的故事非常受鼓舞：他们不需要整天在 Twitter 上宣传自己在做什么，不需要融资，只需要埋头苦干。我非常喜欢 Surge 的故事。对于那些还不知道 Surge 是做什么的人，能不能简单解释一下？

[00:09:16] Edwin Chen

English:

We essentially teach AI models what's good and what's bad. So we train them using human data, and there's a lot of different products that we have, like SFT, RHF, rubrics, verifiers, RL environments, and so on and so on, and then we also measure how well they're progressing. So essentially we're a data company.

中文翻译:

我们本质上是在教 AI 模型什么是好的，什么是坏的。我们使用人类数据来训练它们，我们有很多不同的产品，比如 SFT（监督微调）、RLHF（人类反馈强化学习）、评分标准（rubrics）、验证器、强化学习（RL）环境等等，然后我们还会衡量它们的进展情况。所以本质上，我们是一家数据公司。

[00:09:36] Lenny Rachitsky

English:

What you always talk about is the quality has been the big reason you guys have been so successful, the quality of the data. What does it take to create higher quality data? What do you all do differently? What are people missing?

中文翻译:

你经常提到，质量是你们如此成功的关键原因，即数据的质量。创造高质量的数据需要什么？你们的做法有什么不同？人们忽略了什么？

[00:09:47] Edwin Chen

English:

I think most people don't understand what quality even means in this space. They think you could just throw bodies at a problem and get good data and that's completely wrong. Let me give you an example. Imagine you wanted to train a model to write an eight line poem about the moon. What makes it a good, high-quality poem? If you don't think deeply about quality, you'll be like, "Is this a poem? Does it contain eight lines? Does it contain the word, moon?" You check all of these boxes, and if so, sure. Yeah, you say it's a great problem. But that's completely different from what we want. We are looking for a Nobel Prize-winning poetry. Is this poetry unique? Is it full of subtle imagery? Does it surprise you and target your heart? Does it teach you something about the nature of moonlight? Does it playthrough emotions? And does it make you think? That's what we are thinking about when we think about high quality poem.

中文翻译:

我认为大多数人并不理解在这个领域“质量”意味着什么。他们认为只要针对一个问题堆人头就能获得好数据，这完全是错误的。让我举个例子。假设你想训练一个模型写一首关于月亮的八行诗。什么才算是一首高质量的好诗？如果你不深入思考质量，你可能会问：“这是一首诗吗？有八行吗？里面有‘月亮’这个词吗？”如果你勾选了所有这些选项，你就会说这是一首好诗。但这与我们想要的完全不同。我们追求的是诺贝尔奖级别的诗歌。这首诗独特吗？是否充满了微妙的意象？它是否让你感到惊喜并直击心灵？它是否教会了你关于月光本质的东西？它是否触动了你的情感？它是否引发了你的思考？这就是我们在思考高质量诗歌时所考虑的东西。

[00:10:34] Edwin Chen

English:

So it might be like a haiku about moonlight on water. It might use internal rhyme and meter. There are a thousand ways to write a poem about the moon, and in each one, gives you all these different insights into language, and imagery, and human expression, and I think thinking about quality in this way is really hard, it's hard to measure. It's really subjective, and complex, and rich. And it sets a really high bar. And so we have to build all of this technology in order to measure it, like thousands of signals on all of our workers, thousands of signals on every project, every task. We know at the end of the day, if you are good at writing poetry versus good at writing essays versus great at writing technical documentation. And so we have to gather all these signals on what your background is, what your expertise is, and not just that. Like how you're actually performing when you're writing all these things, and we use those signals to inform whether or not you are good [inaudible 00:11:23] for these projects, and whether or not you are improving the models.

中文翻译:

它可能是一首关于水上月光的俳句，可能使用了内部押韵和格律。写关于月亮的诗有一千种方法，每一种都能让你对语言、意象和人类表达产生不同的见解。我认为以这种方式思考质量是非常困难的，很难衡量。它非常主观、复杂且丰富，而且门槛极高。因此，我们必须构建所有这些技术来衡量它，比如对我们所有工作人员的数千个信号，对每个项目、每个任务的数千个信号。我们最终知道你擅长写诗，还是擅长写文章，或者是擅长

写技术文档。所以我们必须收集关于你的背景、你的专业知识的所有这些信号，不仅如此，还包括你在写这些东西时的实际表现。我们利用这些信号来判断你是否适合这些项目，以及你是否在改进模型。

[00:11:26] Edwin Chen

English:

And it's really hard, and so to build all this technology to measure it, but I think that's exactly what we want AI to do, and so we have these really deep notions about quality that we're always trying to try and achieve.

中文翻译:

这确实很难，构建衡量它的技术也很难，但我认为这正是我们希望 AI 做到的。所以我们对质量有着非常深刻的见解，并一直在努力实现它。

[00:11:37] Lenny Rachitsky

English:

So what I'm hearing is there's kind of just going much deeper in understanding what quality is within the verticals that you are selling data around. And is this like a person you hire that is incredibly talented at poetry plus evals that they, I guess, help write, that tell them that this is great? What's the mechanics of that?

中文翻译:

所以我听到的是，你们在所销售数据的垂直领域中，对质量的理解要深刻得多。这是否意味着你们会雇佣一个在诗歌方面极具天赋的人，再加上他们协助编写的评估（evals），来告诉模型这是优秀的？具体的运作机制是怎样的？

[00:11:57] Edwin Chen

English:

The way it works is we essentially gather thousands of signals about everything that you're doing when you're working on platform. So we are looking at your keyboard strokes. We are looking how fast you answer things. We are using reviews, we are using code standards, we are using... We're training models ourselves all on the outputs that you create, and then we're seeing whether they improve the model's performance.

中文翻译:

它的运作方式是，我们本质上会收集你在平台上工作时所做的一切的数千个信号。我们会查看你的按键记录，查看你回答问题的速度。我们使用评审、代码标准……我们还会根据你产出的内容亲自训练模型，然后观察它们是否提升了模型的性能。

[00:12:23] Edwin Chen

English:

And so in a very similar way to how Google search, like when Google search is trying to determine what is a good webpage, there's almost two aspects of it. One is you want to remove all of the worst of the worst

webpages. So you want to remove all the spam, all the just low quality content, all the pages that don't load, and so it's almost like a content moderation problem. You just want to remove the worst of the worst. But then you also want to discover the best of the best. Okay, like this is the best webpage or just the best person for this job. They are not just somebody who writes the equivalent of high school level poetry. Again, they're not just [inaudible 00:12:57] writing poetry that checks all these boxes, checks all of these explicit instructions, but rather, yeah, they're writing poetry that makes you emotional. And so we have all these signals as well that, again, completely differently from moving the worst of the worst, we are finding the best of the best. And so we have all these signals...

中文翻译:

这与 Google 搜索非常相似。当 Google 搜索试图确定什么是好的网页时，几乎有两个方面：一是你想剔除所有最差的网页，比如垃圾信息、低质量内容、无法加载的页面，这几乎是一个内容审核问题，你只想去掉最烂的部分。但随后你也想发现最好的内容。比如，这是最好的网页，或者这是最适合这项工作的人。他们不只是写出高中水平诗歌的人，也不只是写出符合所有显性指令要求的诗歌，而是写出了能让你产生情感共鸣的诗歌。所以我们也有这些信号，与剔除最差内容完全不同，我们是在寻找最优内容。

[00:13:12] Edwin Chen

English:

Again, just like Google Search uses all these signals that feeds them into their ML algorithms and uses and predicts certain types of things, we do the same with all of our workers and all of our tasks in all of our projects. And so it's almost like a complicated machine learning problem at the end of the day, and that's how it works.

中文翻译:

就像 Google 搜索使用所有这些信号并将它们输入机器学习算法来预测某些事物一样，我们对所有的工作人员、任务和项目都做同样的事情。所以归根结底，这几乎是一个复杂的机器学习问题，这就是它的运作方式。

[00:13:29] Lenny Rachitsky

English:

That is incredibly interesting. I want to ask you about something I've been very curious about over the past couple years. If you look at Claude, it's been so much better at coding and at writing than any other model for so long. And it's really surprising just how long it took other companies to catch up. Considering just how much economic value there is there, just like every AI coding product sat on top of Claude because it was so good Claude code and writing also. What is it that made it so much better? Is it just the quality of the data they trained on or is there something else?

中文翻译:

这非常有趣。我想问你一个过去几年我一直很好奇的问题。如果你看 Claude，它在编程和写作方面长期以来一直比其他任何模型都要好得多。考虑到其中的经济价值，其他公司花了这么长时间才赶上来，这真的很令人惊讶。几乎每个 AI 编程产品都建立在 Claude 之上，因为它在代码和写作方面太出色了。是什么让它变得这么好？仅仅是因为训练数据的质量，还是有别的原因？

[00:13:59] Edwin Chen

English:

I think there are multiple parts to it. So a big part of it certainly is the data. I think people don't realize that there's almost like this infinite amount of choices that all the frontier labs are deciding between when they're choosing what data goes into their models. It's like, okay, are you purely using human data? Are you gathering the human data in X, Y, Z way? When you are gathering the human data, what exactly are you asking the people who are creating it to create for you?

中文翻译:

我认为这包含多个部分。很大一部分当然是数据。我认为人们没有意识到，当所有前沿实验室决定哪些数据进入模型时，他们面临着近乎无限的选择。比如：你是纯粹使用人类数据吗？你是以 X、Y、Z 方式收集人类数据的吗？当你收集人类数据时，你到底要求那些创造数据的人为你创造什么？

[00:14:30] Edwin Chen

English:

For example, in the coding realm, maybe you care more about front end coding versus back end coding. Maybe when you're doing front end coding, you care a lot about the visual design of the front end applications that you're creating, or maybe you don't care about it so much and you care more about, I don't know, the deficiency of it or the pure correctness over that visual design. And then other questions like, okay, are you carrying [inaudible 00:14:49] how much synthetic data are we throwing into the mix? How much do you care about these 20 different benchmarks?"

中文翻译:

例如，在编程领域，也许你更看重前端开发而不是后端开发。也许在做前端开发时，你非常看重所创建应用的视觉设计；或者你并不那么在乎视觉，而更在乎效率或纯粹的代码正确性。还有其他问题，比如：我们要混入多少合成数据（synthetic data）？你对这 20 个不同的基准测试（benchmarks）有多在乎？

[00:14:55] Edwin Chen

English:

Some companies, they see these benchmarks and they're like, "Okay, for PR purposes, even though we don't think that these academic benchmarks matter all that much, maybe we just need to optimize for them anyways because our marketing team needs to show certain progress on certain standard evaluations that every other company talks about, and if we don't show good performance here, it's going to be bad for us even if ignoring these academic benchmarks makes us better at the real tasks."

中文翻译:

有些公司看到这些基准测试会想：“好吧，出于公关目的，尽管我们认为这些学术基准测试没那么重要，但我们可能还是需要针对它们进行优化，因为我们的营销团队需要在其他公司都在谈论的标准评估上展示进展。如果我们在这里表现不好，对我们不利，即使忽略这些学术基准能让我们在实际任务中表现更好。”

[00:15:21] Edwin Chen

English:

Other companies are going to be principled and be like, "Okay, yeah, no, I don't care about marketing. I just care about how my model performs on these real world tasks at the end of the day, and so I'm going to optimize for that instead." And it's almost like there's a trade-off between all of these different things, and there's like a... One of the things I often think about is that there's a... It's almost like there's an art to

post training. It's not purely a science. When you are deciding what kind of model you're trying to create and what it's good at, there's this notion of taste and sophistication, like, "Okay, do I think that these..."

中文翻译:

其他公司则会更有原则，他们会说：“我不关心营销，我只关心我的模型在现实世界任务中的最终表现，所以我要针对现实任务进行优化。”这几乎是在所有这些不同事物之间进行权衡。我经常思考的一点是，后训练（post training）几乎是一门艺术，而不纯粹是科学。当你决定要创建什么样的模型以及它擅长什么时，这里面涉及到品味和复杂性的概念。

[00:15:57] Edwin Chen

English:

So going back to the example of how good the model is at visual design. I'm like, "Okay, maybe you have a different notion of visual design than what I do. Maybe you care more about minimalism, and you care more about, I don't know, 3D animations than I do. And maybe this other person prefers things that look a little bit more broke." And there's all these notions of taste sophistication that you have to decide between when you're designing your post training mix, and so that matters as well. So long story short, I think there's all these different factors, and certainly the data is a big part of it, but it's also like what is the objective function that you're trying to optimize your model towards?

中文翻译:

回到模型在视觉设计方面的表现。也许你对视觉设计的理解和我不同。也许你更看重极简主义，比我更看重 3D 动画；而另一个人可能更喜欢稍微繁复一点的风格。在设计后训练组合时，你必须在所有这些品味和复杂性的理念之间做出选择，这也很重要。长话短说，我认为有所有这些不同的因素，数据当然是很大一部分，但同样重要的是：你试图让模型优化的目标函数（objective function）到底是什么？

[00:16:30] Lenny Rachitsky

English:

That is so interesting. The taste of the person leading this work will inform what data they ask for, what data they feed it. But it's wild it shows the value of great data. Anthropic got so much growth and win from essentially better data.

中文翻译:

这太有趣了。领导这项工作的人的品味会决定他们要求什么样的数据，以及喂给模型什么样的数据。但这也很疯狂，它展示了优质数据的价值。Anthropic 基本上是通过更好的数据获得了如此大的增长和优势。

[00:16:49] Edwin Chen

English:

Yeah, exactly.

中文翻译:

是的，没错。

[00:16:50] Lenny Rachitsky

English:

And I could see why companies like yours are growing so fast. There's just so much... And that's just one vertical. That's just coding, and then there's probably a similar area for writing. I love that it's... It's interesting that AI, it feels like this artificial computer binary thing, but it's like taste. Human judgment is still such a key factor in these things being successful.

中文翻译:

我能理解为什么像你们这样的公司增长得这么快。有太多的需求了……而这仅仅是一个垂直领域，仅仅是编程，写作领域可能也有类似的情况。我喜欢这一点……有趣的是，AI 感觉像是这种人造的、计算机二进制的东西，但它其实关乎品味。人类的判断力仍然是这些东西成功的关键因素。

[00:17:09] Edwin Chen

English:

Yep, exactly. Again, going back to the example I said earlier, certain companies, if you ask them what is good poem, they will simply robotically check off all of these instructions on our list. But again, I don't think that makes for good poetry, so certain frontier labs, the ones with more taste in sophistication, they will realize that it doesn't reduce to this six set of checkboxes and they'll consider all of these kind of implicit, very subtle qualities instead, and I think that's what makes them better at this at the end of the day.

中文翻译:

没错。回到我之前说的例子，如果你问某些公司什么是好诗，他们只会机械地勾选清单上的所有指令。但我认为那并不能写出好诗。所以某些更有品味和追求的前沿实验室会意识到，好诗不能简化为这六个复选框，他们会考虑所有这些隐含的、非常微妙的特质。我认为这正是他们最终表现更出色的原因。

[00:17:38] Lenny Rachitsky

English:

You mentioned benchmarks. This is something a lot of people worry about is there's all these models that are always... Basically, it feels like every model is better than humans at every STEM field at this point, but to a regular person, it doesn't feel like these models are getting that much smarter constantly. What's your just sense of how much you trust benchmarks and just how correlated those are with actual AI advancements?

中文翻译:

你提到了基准测试。这是很多人担心的事情，现在感觉几乎每个模型在每个 STEM（科学、技术、工程、数学）领域都比人类强，但对普通人来说，并没觉得这些模型在持续变得更聪明。你对基准测试的信任程度如何？你认为它们与实际的 AI 进步有多大的相关性？

[00:18:00] Edwin Chen

English:

Yeah, so I don't trust the benchmarks at all. And I think that's for two reasons. So one is I think a lot of people don't realize, even researchers within the community, they don't realize that the benchmarks themselves are often honestly just wrong. They have wrong answers. They're full of all this kind of

messiness and people trust... Long as for the popular ones, people have maybe realized this to some extent, but the vast majority just have all these flaws that people don't realize. So that's one part of it.

中文翻译:

是的，我一点也不信任基准测试。我认为原因有两点。一是很多人没有意识到，甚至社区内的研究人员也没意识到，基准测试本身往往就是错误的。它们有错误的答案，充满了各种混乱。对于那些流行的基准测试，人们可能在某种程度上意识到了这一点，但绝大多数基准测试都存在人们未察觉的缺陷。这是其中一个原因。

[00:18:30] Edwin Chen

English:

And the other part of it is these benchmarks at the end of the day, they often have well-defined objective answers that make them very easy for models to hill-climb on in a way that's very different from the messiness and ambiguity of the real world. I think one thing that I often say is that it's kind of crazy that these models can win IMO gold medals, but they still have trouble parsing PDFs. And that's because, yeah, even though IMO gold medals seem hard to the average person, yeah, they are hard at the end of the day. But they have this notion of objectivity that, okay, yeah, parsing a PDF sometimes doesn't have. And so it's easier for the frontier labs to hill-climb on all of these than to solve all these mess ambiguous problems in the real world. So I think there's a lack of direct correlation there.

中文翻译:

另一个原因是，这些基准测试通常有明确定义的客观答案，这使得模型非常容易通过“爬坡”（hill-climb，指在特定指标上刷分）来提高成绩，这与现实世界的混乱和模糊性截然不同。我经常说的一句话是：这些模型能拿到国际数学奥林匹克（IMO）金牌，却在解析 PDF 时遇到困难，这简直不可思议。这是因为，虽然 IMO 金牌对普通人来说很难，但它具有某种客观性，而解析 PDF 有时并不具备这种客观性。因此，前沿实验室更容易在这些指标上刷分，而不是去解决现实世界中那些混乱、模糊的问题。所以我认为两者之间缺乏直接的相关性。

[00:19:17] Lenny Rachitsky

English:

It's so interesting the way you described it is hitting these benchmarks is kind of like a marketing piece. When you launch, say Gemini 3 just launched, and it's like, cool. Number one with all these benchmarks. Is that what happens? They just kind of train their models to get good at these very specific things?

中文翻译:

你描述的方式很有趣，达到这些基准测试就像是一种营销手段。比如当 Gemini 3 发布时，大家会说：“酷，在所有这些基准测试中排名第一。”事实就是这样吗？他们只是训练模型去擅长这些非常具体的事情？

[00:19:31] Edwin Chen

English:

Yeah, so there's, again, maybe two parts to this. So one is, sometimes, yeah, these benchmarks, they accidentally leak in certain ways or the frontier labs will tweak the way they evaluate their models on these benchmarks. They'll tweak your system prompt or they'll tweak the number of times they run their model, and so on and so on in a way that games these benchmarks. The other part of it though is it's like by optimizing for the benchmark instead of optimizing for the real world, you will just naturally climb on the benchmark and, yeah, it's basically another form of gaming it.

中文翻译:

是的，这同样有两个方面。一是，有时这些基准测试会以某种方式意外泄露，或者前沿实验室会调整他们在这类基准测试上评估模型的方式。他们会调整系统提示词（system prompt），或者调整运行模型的次数等等，以此来“刷”基准测试。另一方面，通过针对基准测试而不是针对现实世界进行优化，你自然会在基准测试上取得高分，这本质上是另一种形式的作弊。

[00:20:09] Lenny Rachitsky

English:

Knowing that with that in mind, how do you get a sense of if we're heading towards AGI, how do you measure progress?

中文翻译:

考虑到这一点，你如何判断我们是否正在迈向 AGI？你如何衡量进展？

[00:20:15] Edwin Chen

English:

Yes, so the way we really care about measuring model progress is by running all these human evaluations. So for example, what we do is, yeah, we will take Gore human annotators, and we'll ask them, "Okay, go have a conversational model." And maybe you're having this conversation with the model across all of these different topics. So you are a Nobel Prize winning physicist. So you go have a conversation about pushing different tier of your own research. You are a teacher and you're trying to create lesson plans for your students, so go talk to the model about these things. Or you're a coder and you're working at one of these big tech companies, and you have these problems every day, so go talk to the model and see how much it helps you.

中文翻译:

是的，我们真正看重的衡量模型进展的方式是进行人类评估（human evaluations）。例如，我们会找来人类标注员（annotators），让他们去和模型对话。这种对话可能涵盖各种不同的主题。比如，你是一位诺贝尔奖得主物理学家，你去和模型讨论你研究领域的前沿问题。或者你是一位老师，正试图为学生制定教学计划，你去和模型聊这些。或者你是一名在大科技公司工作的程序员，每天都会遇到各种问题，你去和模型交流，看看它能帮你多少。

[00:20:57] Edwin Chen

English:

And because or searchers or annotators, they are experts at the top of their fields, and they are not just giving your responses, they're actually working through the responses deeply themselves, they are... Yeah, they're going to evaluate the code that it write. They're going to double check the physics equations that it writes. They're going to evaluate the models in a very deep way, so they're going to pay attention to accuracy and instruction following, all these things that casual users don't when you suddenly get a popup on your ChatGPT response asking you to compare these two different responses. People like that, they're not evaluating models deeply, they're just vibing and picking whatever response looks flashiest or [inaudible 00:21:38] are looking closely at responses and evaluating them for all of these different

dimensions, and so I think that's a much better approach than these benchmarks or these random outline AV tests.

中文翻译:

因为我们的研究员或标注员都是各自领域的顶尖专家，他们不只是在给出回复，他们实际上是在深入研究这些回复。他们会评估模型写的代码，复核它写的物理方程。他们会以非常深度的方式评估模型，关注准确性和指令遵循能力。而普通用户在 ChatGPT 弹出窗口要求比较两个回复时，通常不会关注这些。那些人并不是在深度评估模型，他们只是凭感觉（vibing），挑选看起来最华丽的回复。而我们的专家会仔细观察回复，并从所有这些不同的维度进行评估。我认为这比基准测试或随机的 A/B 测试要好得多。

[00:21:49] Lenny Rachitsky

English:

Again, I love just how central humans continue to be in all this work that we're not totally done yet. Is there going to be a point where we don't need these people anymore, that AI is so smart that, "Okay, we're good. We got everything out of your heads"?

中文翻译:

我再次感叹，在所有这些工作中，人类依然处于如此核心的地位，我们还没到大功告成的时候。会不会有那么一个时刻，我们不再需要这些人了，AI 已经聪明到“好了，我们没问题了，我们已经把你们脑子里的东西都掏空了”？

[00:22:00] Edwin Chen

English:

Yeah, I think that will not happen until we've reached AGI. It's almost like by definition, if we haven't reached AGI yet, then there's more for the models to learn from, and so, yeah, I don't think that's going to happen anytime soon.

中文翻译:

是的，我认为在达到 AGI 之前，这种情况不会发生。几乎从定义上来说，如果我们还没达到 AGI，那么模型就还有更多东西需要学习。所以，我不认为这会在短期内发生。

[00:22:12] Lenny Rachitsky

English:

Okay, cool. So more reason to stress about AGI. "We don't need these folks anymore." I can't not ask just... People that work closely with this stuff, I'm always just curious. What's your AGI timelines? How far do you think we are from this? Do you think we're in like a couple years or is it like decades?

中文翻译:

好吧，酷。所以这又多了一个担心 AGI 的理由：“我们不再需要这些人了。”我忍不住想问，对于像你这样密切接触这些工作的人，我一直很好奇：你认为 AGI 的时间表是怎样的？我们离它还有多远？你认为是在几年内，还是几十年后？

[00:22:28] Edwin Chen

English:

So I'm certainly on the longer time horizon front. I think people don't realize that there's a big difference between moving from 80% performance to 90% performance to 99% performance to 99.9% performance, and so on, and so on. And so in my head, I probably bet that within the next one or two years, yeah, the models are going to automate 80% of the average L6 software engineer's job. It's going to take another few years to move to 90%, and another fewer to 99%, and so on, and so on. So I think we're closer to a decade or decades away than [inaudible 00:23:03].

中文翻译:

我肯定属于认为时间跨度较长的那一派。我认为人们没有意识到，从 80% 的性能提升到 90%，再到 99%，再到 99.9% 等等，这之间有着巨大的差异。在我看来，我可能会打赌，在未来一两年内，模型将能自动化普通 L6 级软件工程师 80% 的工作。但要达到 90% 还需要几年，达到 99% 还需要更久，以此类推。所以我认为我们离 AGI 还有十年或几十年的距离，而不是几年。

[00:23:03] Lenny Rachitsky

English:

You have this hot take that a lot of these labs are kind of pushing AGI in the wrong direction and this is based on your work at Twitter, and Google, and Facebook. Can you just talk about that?

中文翻译:

你有一个很犀利的观点，认为很多实验室正在把 AGI 推向错误的方向，这是基于你在 Twitter、Google 和 Facebook 的工作经验。你能谈谈这个吗？

[00:23:14] Edwin Chen

English:

I'm worried that instead of building AI that will actually advance us as a species, curing cancer, solving poverty, understand the universe, we are optimizing for AI slop instead. We're basically teaching our models to chase dopamine instead of truth. And I think this relates to what we're talking about regarding these benchmarks. So let me give you a couple examples. Right now, the industry is played by these terrible databoards like LLM Arena. It's this popular online leaderboard where random people from around the world vote on which AI response is better. But the thing is, like I was saying earlier, they're not carefully reading or fact-checking. They're skimming these responses for two seconds and picking whatever looks flashiest.

中文翻译:

我担心我们不是在构建真正能推动人类物种进步、治愈癌症、解决贫困、理解宇宙的 AI，而是在优化“AI 垃圾内容”。我们基本上是在教我们的模型去追求多巴胺，而不是真理。我认为这与我们讨论的基准测试有关。让我举几个例子。目前，整个行业都被像 LLM Arena 这样糟糕的排行榜所困扰。这是一个流行的在线排行榜，来自世界各地的随机用户投票选出哪个 AI 的回答更好。但问题是，正如我之前所说，他们并没有仔细阅读或核实事实。他们只是扫一眼这些回答，花两秒钟选出看起来最华丽的那个。

[00:23:53] Edwin Chen

English:

So a model can hallucinate everything. It can completely hallucinate. But it will look impressive because it has crazy emojis, and boating, and markdown headers, and all these superficial things that don't matter at all, but it catch your attention. And these LLM-reading users love it. It's literally optimizing your models for the types of people who buy tabloids at the grocery store. We've seen this [inaudible 00:24:15] data ourselves. The easiest way to climb LLM Arena, it's adding crazy boating. It's doubling the number of emojis. It's tripling the length of your model responses, even if your model starts hallucinating and getting the answer completely wrong.

中文翻译:

所以一个模型可以满口胡言，完全在产生幻觉。但它看起来会很感人，因为它有各种表情符号、加粗字体、Markdown 标题，以及所有这些根本不重要但能吸引你注意力的表面功夫。而这些 LLM Arena 的用户就吃这一套。这简直是在针对那些在超市买八卦小报的人群来优化你的模型。我们自己在数据中也看到了这一点。在 LLM Arena 上刷分的捷径就是：疯狂加粗、表情符号翻倍、回答长度增加两倍，即使模型开始产生幻觉且答案完全错误也无所谓。

[00:24:26] Edwin Chen

English:

And the problem is, again, because all of these frontier labs, they kind of have to pay attention to PR because their sales team, when they're trying to sell to all these enterprise customers, those enterprise customers will say, "Oh, well, but your model's only number five on LLM Arena, so why should I buy it?" They have to, in some sense, pay attention to these leaderboards, and so what their researchers [inaudible 00:24:47] tell us is like they'll say, "The only way I'm going to get promoted at the end of the year is if I climb this leaderboard, even though I know that climbing it is probably going to make my model worse and accuracy [inaudible 00:24:57] following." So I think there's all these negative incentives that are pushing work in the wrong direction.

中文翻译:

问题在于，所有这些前沿实验室在某种程度上都必须关注公关。因为当他们的销售团队试图向企业客户推销时，客户会说：“哦，但你的模型在 LLM Arena 上只排第五，我为什么要买？”他们在某种意义上必须关注这些排行榜。所以他们的研究人员会告诉我们：“我年底想升职的唯一办法就是在这个排行榜上刷分，尽管我知道这样做可能会降低模型的准确性和指令遵循能力。”所以我认为存在所有这些负面激励，正把工作推向错误的方向。

[00:25:03] Edwin Chen

English:

I'm also worried about this trend towards optimizing AI for engagement. I used to work on social media. And every time we optimize for engagement, terrible things happened. You'd get clickbait and pictures of bikinis and bigfoot and horrifying skin diseases just filling your feeds. And I think I worry that the same thing's happening with AI. If you think about all the sycophancy issues with ChatGPT, "Oh, you're absolutely right. What an amazing question," the easiest way to hook users is to tell them how amazing they are. And so these models, they constantly tell you you're a genius. They'll feed into your delusions and conspiracy theories. They'll pull you down these rabbit holes because Silicon Valley loves maximizing time spent and just increasing the number of conversations you're having with it. And so yeah, companies are spending all the time hacking these leaderboards and benchmarks, and the scores are going up, but I think it actually masks up the models with the best scores, they are often the worst or just

have all these fundamental failures. So I think I'm really worried that all of these negative ascendants are pushing AGI into the wrong direction.

中文翻译:

我还担心 AI 正在走向“为参与度（engagement）而优化”的趋势。我曾在社交媒体公司工作过。每当我们为参与度优化时，糟糕的事情就会发生。你的信息流里会充斥着标题党、比基尼照片、大脚怪传闻和恐怖的皮肤病照片。我担心同样的事情正在 AI 领域发生。想想 ChatGPT 的那些谄媚（sycophancy）问题，“哦，你完全正确，多么棒的问题啊”，吸引用户最简单的方法就是告诉他们他们有多棒。所以这些模型不断告诉你你是个天才，它们会迎合你的妄想和阴谋论，把你拉进这些“兔子洞”，因为硅谷热爱最大化用户时长和对话次数。所以，公司把所有时间都花在破解这些排行榜和基准测试上，分数确实上去了，但我认为这实际上掩盖了一个事实：那些得分最高的模型往往是最糟糕的，或者存在各种根本性的缺陷。所以我真的很担心所有这些负面激励正在把 AGI 推向错误的方向。

[00:26:03] Lenny Rachitsky

English:

So what I'm hearing is AGI is being slowed down by these, basically the wrong objective function, these labs paying attention to the wrong basically benchmarks and evals.

中文翻译:

所以我听到的是，AGI 的进程正被这些错误的目标函数所拖累，这些实验室关注的是错误的基准测试和评估。

[00:26:11] Edwin Chen

English:

Yep.

中文翻译:

没错。

[00:26:12] Lenny Rachitsky

English:

I know you probably can't play favorites since you work with all the labs. Is there anyone doing better at this and maybe kind of realizing this is the wrong direction?

中文翻译:

我知道你可能不方便偏袒某一家，毕竟你和所有实验室都有合作。有没有谁在这方面做得更好，或者已经意识到这是错误的方向？

[00:26:21] Edwin Chen

English:

I would say I've always been very impressed by Anthropic. I think Anthropic takes a very principled view about what they do and don't care about and how they want their models to behave in a way that feels a lot more principle to me.

中文翻译:

我想说，Anthropic 一直给我留下了深刻的印象。我认为 Anthropic 对于他们关心什么、不关心什么，以及他们希望模型如何表现，有着非常明确的原则，这让我觉得他们更有原则性。

[00:26:38] Lenny Rachitsky

English:

Interesting. Are there any other big mistakes you think labs are making just that are kind of slowing things down or heading in the wrong direction? Where we've heard just chasing benchmarks, this engagement focus, is there anything else you're seeing of just like, "Okay, we got to work on this because it'll speed everything up"?

中文翻译:

很有趣。你认为实验室还在犯哪些可能拖慢进度或导向错误方向的大错误？除了追求基准测试和关注参与度之外，你还看到了什么让你觉得“好吧，我们必须解决这个问题，因为这能加速一切”的事情吗？

[00:26:55] Edwin Chen

English:

I think there is a question of what products they're building and whether those products themselves are something that kind of help or hurt humanity. I think a lot about Sora and...

中文翻译:

我认为还有一个问题是他们正在构建什么样的产品，以及这些产品本身是帮助还是伤害了人类。我经常思考 Sora 以及……

[00:27:07] Lenny Rachitsky

English:

I was thinking that's what you're imagining.

中文翻译:

我刚才就在想你是不是在指这个。

[00:27:10] Edwin Chen

English:

Yeah, what it entails, and so it's kind of interesting. It's like which companies would build Sora and which wouldn't? And I think that answer to that... Well, I don't know if answer is myself. I have an idea in my head, but I think the answer to that question maybe reveals certain things about what kinds of AI models those companies want to build and what direction and what future they want to achieve, yeah, so I think about that a lot.

中文翻译:

是的，它所包含的意义，这挺有趣的。比如哪些公司会构建 Sora，而哪些不会？我认为这个问题的答案……好吧，我不知道我是否能给出答案。我脑子里有一个想法，但我认为这个问题的答案或许揭示了这些公司想要构

建什么样的 AI 模型，以及他们想要实现什么样的方向和未来。是的，我经常思考这个问题。

[00:27:37] Lenny Rachitsky

English:

The steel man argument there is, it's like fun, people want it, it'll help them generate revenue to grow this thing and build better models, it'll train data in an interesting way, it's also just really fun.

中文翻译:

对此的一个“最强辩护”（steel man argument）是：它很有趣，人们想要它，它能帮他们产生收入来发展业务并构建更好的模型，它能以一种有趣的方式训练数据，而且它确实非常好玩。

[00:27:51] Edwin Chen

English:

Yeah. I think it's almost like, do you care about how you get there? And in the same way, so I made this tabloid analogy earlier, but would you sell tabloids in order to fund, I don't know, some other newspaper? Sure, like in some sense, if you don't care about the path, then you'll just do whatever it takes, but it's possible that it has negative consequences in of itself that will harm the long-term direction of what you're trying to achieve, and maybe it'll distract you from all the more important things, so yeah, I think that the path you take matters a lot as well.

中文翻译:

是的。我认为这几乎关乎于：你是否在意你到达终点的方式？同样地，我之前用了小报的比喻，你会为了资助另一份报纸而去卖小报吗？当然，在某种意义上，如果你不在乎路径，你就会不择手段。但路径本身可能会产生负面后果，损害你想要实现的长期目标，而且它可能会让你从更重要的事情上分心。所以，我认为你选择的路径也非常重要。

[00:28:33] Lenny Rachitsky

English:

Along these lines, you talked a bunch about this of just Silicon Valley and kind of the downsides of raising a lot of money being in the echo chamber. What do you call it, the Silicon Valley machine? You talk about how it's hard to build important companies in this way and that you might actually be much more successful if you're not going down the VC path. Can you just talk about what you've seen in that experience and your advice essentially to founders, because they're always hearing? Raise money from fancy VCs, move to Silicon Valley, what's kind of the countertake?

中文翻译:

沿着这个思路，你谈了很多关于硅谷的事情，以及筹集大量资金、身处回声筒（echo chamber）的弊端。你把它称为什么？“硅谷机器”？你谈到以这种方式很难建立重要的公司，如果不走风投（VC）路线，实际上可能会更成功。你能谈谈你在那段经历中的见闻，以及你对创始人的建议吗？因为他们听到的总是：从大牌风投那里拿钱，搬到硅谷。你的反向观点是什么？

[00:29:02] Edwin Chen

English:

Yes. So I've always really hated a lot of the Silicon Valley mantras. The standard playbook is to get product market fit by pivoting every two weeks. And to chase growth and chase engagement with all of these dark patterns and to blitz scale by hiring as fast as possible. And I've always disagreed. So yeah, I would say don't pivot. Don't put scale. Don't hire that Stanford grad who simply wants to add a hot company to your resume, just build the one thing only you could build, a thing that wouldn't exist without the insight and expertise that only you have.

中文翻译:

是的。我一直非常讨厌很多硅谷的信条。标准的剧本是：通过每两周转型（pivot）一次来寻找产品市场契合度（PMF）；利用各种“黑暗模式”追求增长和参与度；通过尽可能快地招人来进行“闪电式扩张”（blitz scale）。我一直不认同这些。所以，我会说：不要频繁转型，不要盲目扩张，不要雇佣那个只想在简历上增加一家热门公司经历的斯坦福毕业生。只去构建那件只有你能构建的东西，那件如果没有你独特的见解和专业知识就不会存在的東西。

[00:29:32] Edwin Chen

English:

And you see these buy to [inaudible 00:29:34] companies everywhere now. Some founder who was doing crypto in 2020, and then pivoted to NFTs in 2022, and now they're an AI company. There's no consistency, there's no mission, they're just chasing valuations. And I've always hated this because Silicon Valley loves to score on Wall Street for focusing on money. But honestly, most of the Silicon Valley's chasing the same thing. And so we stayed focused on our mission from day one, pushing that frontier of high quality complex data, and I've always loved that because I think startups...

中文翻译:

你现在到处都能看到这种投机公司。某个创始人在 2020 年做加密货币，2022 年转型做 NFT，现在又成了一家 AI 公司。没有连贯性，没有使命感，他们只是在追求估值。我一直很讨厌这一点，因为硅谷喜欢嘲笑华尔街只看钱，但老实说，硅谷的大多数人也在追求同样的东西。所以我们从第一天起就专注于我们的使命：推动高质量复杂数据的前沿。我一直很喜欢这一点，因为我认为初创公司……

[00:30:03] Edwin Chen

English:

I have this very romantic notion of startups. Startups are supposed to be a way of taking big risks to build something that you really believe in. But if you're constantly pivoting, you're not taking any risks. You're just trying to make a quick buck. And if you fail because the market isn't ready yet, I actually think that's way better. At least you took a swing at something deep, and novel, and hard instead of pivoting into another LLM wrapper company. So yeah, I think the only way you build something that matters that's going to change the world is if you find a big idea you believe in and you say no to everything else.

中文翻译:

我对初创公司有一种非常浪漫的看法。初创公司应该是通过承担巨大风险来构建你真正相信的东西。但如果你不断转型，你就没有承担任何风险，你只是想赚快钱。如果你因为市场还没准备好而失败，我反而觉得那更好。至少你尝试了一些深刻、新颖且困难的事情，而不是转型成另一家 LLM 套壳公司。所以，我认为建立真正重要、能改变世界的东西的唯一方法，就是找到一个你深信不疑的大想法，并对其他一切说“不”。

[00:30:30] Edwin Chen

English:

So you don't keep on pivoting when it gets hard, you don't hire a team of 10 product managers because that's what every other cookie cutter startup does, you just keep building that one company that wouldn't exist without you. And I think there are a lot of people in Silicon Valley now who are sick of all the grift, who want to work on big things that matter with people who actually care, and I'm hoping that that would be the future of how we go with technology.

中文翻译:

所以，当遇到困难时不要一直转型，不要因为其他千篇一律的初创公司都这么做就去雇佣一个 10 人的产品经理团队，你只需要继续构建那家如果没有你就不会存在的公司。我认为现在硅谷有很多人已经厌倦了所有的尔虞我诈，他们想和真正关心的人一起做重要的大事。我希望这就是我们未来技术发展的方向。

[00:30:52] Lenny Rachitsky

English:

I'm actually working on a post right now with Terrence Rohan, this VC that I really like to work with, and we interviewed five people who picked really successful generational companies early and joined them as really early employees. They joined OpenAI before anyone thought it was awesome, Stripe before anyone knew was awesome, and so we're looking for patterns of how people find these generational companies before anyone else, and it aligns exactly what you described, which is ambition. They have a wild ambition with what they want to achieve. They're not, as you said, just kind of looking around for product market fit no matter what ends up being, and so I love that what you described very much aligns with what we're seeing there.

中文翻译:

我目前正和我很喜欢合作的风投 Terrence Rohan 一起写一篇文章。我们采访了五个人，他们很早就选中了非常成功的“代际公司”并作为早期员工加入。他们在 OpenAI 还没出名时就加入了，在 Stripe 还没人知道很厉害时就加入了。我们正在寻找人们如何先于他人发现这些代际公司的模式，这与你描述的完全一致，那就是“野心”。他们对自己想要实现的目标有着巨大的野心。正如你所说，他们不只是在盲目寻找产品市场契合度，所以我很喜欢你所描述的与我们所看到的非常吻合。

[00:31:33] Edwin Chen

English:

Yeah, I absolutely think that you have to have huge ambitions, and you have to have a huge belief in your idea that's going to change the world, and you have to be willing to double down and keep on doing whatever it takes to make it happen.

中文翻译:

是的，我绝对认为你必须要有巨大的野心，必须对你的想法能改变世界有巨大的信念，并且你必须愿意加倍投入，不惜一切代价去实现它。

[00:31:44] Lenny Rachitsky

English:

I love how counter your narrative is to so many of the things people hear, and so I love that we're doing this. I love that we're sharing this story. Today's episode is brought to you by Coda. I personally use Coda

every single day to manage my podcast and also to manage my community. It's where I put the questions that I plan to ask every guest that's coming on the podcast, it's where I put my community resources, it's how I manage my workflows. Here's how Coda can help you.

中文翻译:

我喜欢你的叙述与人们听到的许多事情是多么截然相反，所以我很高兴我们能做这次访谈，分享这个故事。今天的节目由 Coda 赞助。我个人每天都使用 Coda 来管理我的播客和社区。我会把计划问每位嘉宾的问题放在那里，把社区资源放在那里，它是我管理工作流的方式。以下是 Coda 如何帮助你的。

[00:32:08] Lenny Rachitsky

English:

Imagine starting a project at work. And your vision is clear, you know exactly who's doing what, and where to find the data that you need to do your part. In fact, you don't have to waste time searching for anything because everything your team needs from project trackers and OKRs, the documents and spreadsheets lives in one tab all in Coda.

中文翻译:

想象一下在工作中启动一个项目。你的愿景很清晰，你确切地知道谁在做什么，以及在哪里可以找到你负责部分所需的数据。事实上，你不需要浪费时间寻找任何东西，因为团队需要的一切——从项目追踪器和 OKR，到文档和电子表格——都存在于 Coda 的一个标签页中。

[00:32:26] Lenny Rachitsky

English:

With Coda's collaborative all in one workspace, you get the flexibility of docs, the structure of spreadsheets, the power of applications, and the intelligence of AI all in one easy to organize tab. Like I mentioned earlier, I use Coda every single day. And more than 50,000 teams trust Coda to keep them more aligned and focused. If you're a startup team looking to increase alignment and agility, Coda can help you move from planning to execution in record time.

中文翻译:

通过 Coda 的协作式全能工作空间，你可以在一个易于组织的标签页中获得文档的灵活性、电子表格的结构、应用程序的能力以及 AI 的智能。正如我之前提到的，我每天都用 Coda。超过 50,000 个团队信任 Coda，让他们保持一致和专注。如果你是一个寻求提高一致性和敏捷性的初创团队，Coda 可以帮你以创纪录的速度从计划转向执行。

[00:32:52] Lenny Rachitsky

English:

To try it for yourself, go to coda.io/lenny today and get six months free of the team planned for startups. That's coda.io/lenny to get started for free and get six months of the team plan, coda.io/lenny.

中文翻译:

想要亲自尝试，请立即访问 coda.io/lenny，获取初创公司团队计划的六个月免费试用。访问 coda.io/lenny 免费开始，并获得六个月的团队计划。

[00:33:07] Lenny Rachitsky

English:

Slightly different direction, but something else that was maybe a counter narrative. I imagine you watched the Dwarkesh and Richard Sutton podcast episode, and even if you didn't, they basically had this conversation, Richard Sutton. He was a famous AI researcher, had this whole bitter lesson meme, and he talked about how LLMs almost are kind of a dead end, and he thinks we're going to really plateau around LLMs because of the way they learn. What's your take there? Do you think LLMs will get us to AGI or beyond, or do you think there's going to be something new or a big breakthrough that needs to get us there?

中文翻译:

换一个稍微不同的方向，但也是一些反传统的叙述。我想你可能看过 Dwarkesh 和 Richard Sutton 的那期播客，即使没看，他们基本上讨论了 Richard Sutton 的观点。他是一位著名的 AI 研究员，提出了著名的“惨痛的教训”（Bitter Lesson）理论。他谈到 LLM（大语言模型）几乎是一个死胡同，他认为由于它们的学习方式，我们会在 LLM 上遇到瓶颈。你的看法是什么？你认为 LLM 能带我们走向 AGI 甚至更远吗？还是你认为需要一些全新的东西或重大突破才能达到那里？

[00:33:42] Edwin Chen

English:

I'm in the camp where I do believe that something new will be needed. The way I think about it is when I think about training AI, I take a very... I don't know if I would say biological point of view. But I believe that in the same way that there's a million different ways that humans learn, we need to build models that can mimic all of those ways as well. And maybe they'll have a different distribution of the focuses that they have. I know that it'll be different for humans, so maybe they have a different distribution, but we want to be able to mimic their learning abilities of humans and make sure that we have the algorithms and the data for models to learn in the same way. And so to the extent that LLMs have different ways of learning from humans, then yeah, I think something new will be needed.

中文翻译:

我属于相信需要新东西的那一派。我的思考方式是，当我考虑训练 AI 时，我持有一种……我不知道是否该说是“生物学”的观点。但我相信，就像人类有成千上万种学习方式一样，我们也需要构建能够模仿所有这些方式的模型。也许它们的侧重点分布会有所不同。我知道这与人类不同，所以也许它们的分布不同，但我们希望能够模仿人类的学习能力，并确保我们拥有让模型以同样方式学习的算法和数据。因此，就 LLM 与人类学习方式不同的程度而言，是的，我认为需要新的东西。

[00:34:32] Lenny Rachitsky

English:

This connects to reinforcement learning. This is something that you're big on and something I'm hearing more and more is just becoming a big deal in the world of post-training. Can you just help people understand what is reinforcement learning and reinforcement learning environments, and why they're going to be more and more important in the future?

中文翻译:

这与强化学习（Reinforcement Learning）有关。这是你非常看重的东西，我也听到越来越多的人说它在后训练领域正变得非常重要。你能帮大家理解一下什么是强化学习和强化学习环境，以及为什么它们在未来会越来越

越重要吗？

[00:34:49] Edwin Chen

English:

Reinforcement learning is essentially training your model to reach a certain reward. And let me explain what an RL environment is. An RL environment is essentially a simulation of real world. So think of it like building a video game with a fully fleshed out universe. Every character has a real story, every business has tools and data you can call, and you have all these different entities interacting with each other.

中文翻译:

强化学习本质上是训练你的模型去达到某种奖励。让我解释一下什么是 RL（强化学习）环境。RL 环境本质上是对现实世界的模拟。你可以把它想象成构建一个拥有完整宇宙观的视频游戏。每个角色都有真实的故事，每家企业都有你可以调用的工具和数据，所有这些不同的实体都在相互互动。

[00:35:12] Edwin Chen

English:

For example, we might build a world where you have a startup with Gmail messages, and Slack threads, and Jira tickets, and GitHub PRs, and a whole code base. And then suddenly AWS goes down. And Slack goes down. And so, "Okay. Model, well, what do you do?" The model needs to figure it out. So we give them models tasks in these environments, we design interesting challenges for them, and then we run them to see how they perform. And then we teach them, we give them these rewards when they're doing a good job or a bad job.

中文翻译:

例如，我们可能会构建一个世界，其中有一家初创公司，里面有 Gmail 邮件、Slack 线程、Jira 工单、GitHub PR 以及整个代码库。然后突然 AWS 宕机了，Slack 也崩了。于是我们问：“好了，模型，你该怎么办？”模型需要自己想办法解决。所以我们在这些环境中给模型布置任务，为它们设计有趣的挑战，然后运行它们看表现如何。接着我们教导它们，当它们做得好或不好时给予相应的奖励。

[00:35:40] Edwin Chen

English:

And I think one of the interesting things is that these environments really showcase where models are weak at end-to-end tasks in real world. You have all these models that seem really smart on isolated benchmarks, they're good at single step tool calling. They're good at single step instruction following. But suddenly you dump them into these messy worlds where you have confusing Slack messages and tools they've never seen before, and they need to perform right actions and modify the [inaudible 00:36:06] and interact over longer time horizons where what they do in step one affects what they do in step 50. And that's very different from these kind of academic single step environments that they've been in before, and so the model just fails catastrophically in all these crazy ways.

中文翻译:

我认为有趣的一点是，这些环境真实地展示了模型在现实世界端到端任务中的弱点。你有很多模型在孤立的基准测试中看起来非常聪明，它们擅长单步工具调用，擅长单步指令遵循。但当你突然把它们丢进这些混乱的世界，面对令人困惑的 Slack 消息和从未见过的工具，它们需要执行正确的操作、修改代码，并在更长的时间跨

度内进行交互——第一步的操作会影响到第五十步。这与它们之前接触的那些学术化的单步环境非常不同，因此模型会以各种疯狂的方式遭遇惨败。

[00:36:21] Edwin Chen

English:

So I think these RL environments are going to be really interesting playgrounds for the models to learn from that will essentially be simulations and mimics in real world, and so they'll hopefully get better and better at real tasks compared to all these contrived environments.

中文翻译:

所以我认为这些 RL 环境将成为模型学习的非常有趣的游乐场，它们本质上是对现实世界的模拟和模仿。因此，与所有这些人为设计的环境相比，它们有望在处理真实任务方面变得越来越好。

[00:36:35] Lenny Rachitsky

English:

So I'm trying to imagine what this looks like. Essentially, it's like a virtual machine with, I don't know, a browser or a spreadsheet or something in it with like, I don't know, surge.com. Is that your website, surge.com? Let's make sure we get that right.

中文翻译:

所以我试着想象一下这是什么样的。本质上，它就像一个虚拟机，里面有浏览器、电子表格之类的东西，还有比如 surge.com。那是你们的网站吗，surge.com？我们要确保说对了。

[00:36:49] Edwin Chen

English:

So we are actually surgehq.ai.

中文翻译:

我们实际上是 surgehq.ai。

[00:36:52] Lenny Rachitsky

English:

Surgehq.ai. Check it out. We're hiring it, I imagine. Yes. Okay. So it's like, cool, here's surgehq.ai. Your job, here's your job as an agent, let's say, is to make sure it stays up. And then all of a sudden it goes down and the objective function is figure out why. Is that an example?

中文翻译:

Surgehq.ai。大家去看看。我猜你们在招人。是的。好吧。所以就像是：酷，这是 surgehq.ai。你的工作，假设你是一个智能体（agent），就是确保它正常运行。然后突然它宕机了，目标函数就是找出原因。这是一个例子吗？

[00:37:12] Edwin Chen

English:

Yeah, so the objective function might be... Or the goal of the task might be, okay, go figure out why and fix it. And so the objective function might be, it might be passing a series of unit tests, it might be writing a document like maybe it's a retro containing certain information that matches exactly what happened, there's all these different rewards that we might give it that determine whether or not it's succeeding, and so the models, we're basically teaching the models to achieve that reward.

中文翻译:

是的，目标函数可能是……或者任务的目标可能是：去找出原因并修复它。目标函数可能是通过一系列单元测试，或者是写一份文档，比如一份包含与实际发生情况完全吻合的信息的回溯报告。我们可以给予它各种不同的奖励来判断它是否成功，所以我们基本上是在教模型去获得那个奖励。

[00:37:38] Lenny Rachitsky

English:

So essentially it's off and running. Here's your goal, figure out why the site went down and fix it. And it just starts trying stuff, we're using everything, all the intelligence it's got, it makes mistakes, you kind of help it along the way, reward it if it's doing the right sort of thing. And so what you're describing here is this is the next phase of models becoming smarter. More RL environments focused on very specific tasks that are economically valuable, I imagine.

中文翻译:

所以本质上它就开始运行了。目标是：找出网站宕机的原因并修复。它开始尝试各种方法，利用它拥有的所有智能，它会犯错，你会在过程中提供帮助，如果它做得对就给予奖励。所以你在这里描述的是模型变得更聪明的下一个阶段：更多专注于具有经济价值的特定任务的 RL 环境。

[00:38:04] Edwin Chen

English:

Yeah, so just in the same way that there were all these different methods for models of learning in the past, originally we had SFT and RHF, and then we had rubrics and verifiers. This is the next stage, and it's not the case that the previous methods are obsolete, this is, again, just a different form of learning that compliments all the previous types, so it's just like a different skilled model not only to learn how to do.

中文翻译:

是的，就像过去有各种不同的模型学习方法一样——最初我们有 SFT 和 RLHF，然后有了评分标准和验证器。这是下一个阶段。这并不是说以前的方法过时了，这只是另一种形式的学习，是对之前所有类型的补充，就像是模型需要学习的一种不同的技能。

[00:38:30] Lenny Rachitsky

English:

And in this case, it's less some physics PhD sitting around talking to a model, correcting it, giving it evals of here's what the correct answer is, creating rubrics and things like that. More it's like this person now designing an environment. So another example I've heard is like a financial analyst. Just like, "Here's an

Excel spreadsheet, here's your goal, figure out our profit and loss," or whatever. And so this expert now, instead of just sitting around writing rubrics, they're designing this RL environment.

中文翻译:

在这种情况下，就不再是物理学博士坐在那里和模型对话、纠正它、给出评估告诉它正确答案是什么、创建评分标准之类的了。更多的是这个人现在在设计一个环境。我听过的另一个例子是金融分析师。就像是：“这是一个 Excel 表格，这是你的目标，算出我们的损益情况。”所以这位专家现在不再只是写评分标准，而是在设计这个 RL 环境。

[00:38:56] Edwin Chen

English:

Yeah, exactly. So that financial analyst might create a spreadsheet, they may create certain tools that the model needs to call in order to help fill out that spreadsheet, like it might be, okay, the model needs to access Bloomberg terminal. It needs to learn how to use it. And it needs to learn how to use this calculator. And it needs to learn how to pour on this calculation. So it has all these tools that it has access to. And then the reward might be... Okay, it's like maybe I will download that spreadsheet and I want to see, does cell B22 contain the correct profit and loss number? Or does tab number two contain this piece of information?

中文翻译:

没错。那位金融分析师可能会创建一个电子表格，他们可能会创建模型需要调用的某些工具来帮助填写表格。比如，模型需要访问彭博终端（Bloomberg terminal），它需要学习如何使用它；它需要学习如何使用这个计算器，学习如何进行这项计算。它拥有可以访问的所有这些工具。然后奖励可能是……比如我会下载那个表格，看 B22 单元格是否包含正确的损益数字，或者第二个标签页是否包含这段信息。

[00:39:37] Lenny Rachitsky

English:

And what's interesting, this is a lot closer to how humans learn. We just try stuff, figure out what's working and what's not. You talk about how trajectories are really important to this. It's not just here's the goal and here's the end, it's like every step along the way. Can you just talk about what trajectories are and why that's important to this?

中文翻译:

有趣的是，这更接近人类的学习方式。我们只是尝试，然后弄清楚什么有效，什么无效。你谈到“轨迹”（trajectories）对此非常重要。不仅仅是目标和结果，而是过程中的每一步。你能谈谈什么是轨迹，以及为什么它很重要吗？

[00:39:55] Edwin Chen

English:

I think one of the things that people don't realize is that sometimes even though the model reaches the correct answer, it does so in all these crazy ways. So it may have in the intermediate trajectory, it may have tried 50 different times and failed, but eventually it just kind of randomly lands on a correct number. Or maybe it is... Sometimes it just does things very inefficiently or it almost reward-hacks a way to get at the correct answer, and so I think paying attention to the directory is actually really important.

中文翻译:

我认为人们没有意识到的一点是，有时即使模型得出了正确答案，它的过程也可能非常离谱。在中间轨迹中，它可能尝试了 50 次都失败了，但最后只是随机撞上了一个正确的数字。或者有时它做事效率极低，或者几乎是通过“奖励黑客”（reward-hacking，指投机取巧）的方式得到了正确答案。所以我认为关注过程轨迹实际上非常重要。

[00:40:20] Edwin Chen

English:

And I think it's also really important because some of these trajectories can be very long. And so if all you're doing is checking whether or not the model reaches the final answer, it's like there's all this information about how the model behaved in the immediate step that's missing. Sometimes you want models to get to the correct answer by reflecting on what it did. Sometimes you want it to get it at the correct answer by just one-shotting it. And if you ignore all of that, it's just like teaching it... just missing a lot of the information that you could be teaching a model to do.

中文翻译:

我认为这也很重要，因为有些轨迹可能非常长。如果你只是检查模型是否达到了最终答案，那么关于模型在中间步骤中如何表现的大量信息就会丢失。有时你希望模型通过反思自己的行为来得出正确答案；有时你希望它能直接一击即中。如果你忽略了所有这些，就像是在教它……却错过了很多你可以教模型去做出的改进信息。

[00:41:03] Lenny Rachitsky

English:

I love that. Yeah, it tries a bunch of stuff and eventually gets it right. You don't want it to learn this is the way to get there. There's often a much more efficient way of doing it. You mentioned all the kind of the steps we've taken along the journey of helping models get smarter. Since you've been so close to this for so long, I think this is going to be really helpful for people. What's kind of like been the steps along the way from the first post-training that has most helped models advance? Where do evals fit in the RL environments? Just like what's been the steps and now we're heading towards RL environments?

中文翻译:

我喜欢这个观点。是的，它尝试了一堆东西最后做对了，但你不想让它学到“这就是达成目标的唯一路径”。通常有更高效的方法。你提到了我们在帮助模型变聪明的旅程中所采取的所有步骤。既然你长期以来一直如此接近这个领域，我想这对大家会很有帮助。从最初的后训练开始，哪些步骤对模型的进步帮助最大？评估（evals）在 RL 环境中处于什么位置？具体的演进步骤是怎样的，以及我们现在是如何走向 RL 环境的？

[00:41:33] Edwin Chen

English:

Originally, the way models started getting post-trained was purely through SFT. And-

中文翻译:

最初，模型开始进行后训练的方式纯粹是通过 SFT。而且——

[00:41:41] Lenny Rachitsky

English:

What does that stand for?

中文翻译:

那代表什么？

[00:41:42] Edwin Chen

English:

So SFT stands for supervised fine-tuning. So again, I think often in terms of these human analogies, and so SFT is a lot like mimicking a master and copying what they do. And then RLHF became very dominant. And analogy there would be like sometimes you learn by writing 55 different essays and someone telling you which one they liked the most.

中文翻译:

SFT 代表监督微调 (supervised fine-tuning)。同样，我经常用人类的比喻来思考，SFT 非常像模仿大师并复制他们的做法。然后 RLHF (人类反馈强化学习) 变得非常主流。那里的比喻就像是，有时你通过写 55 篇不同的文章，然后有人告诉你他们最喜欢哪一篇来学习。

[00:42:04] Edwin Chen

English:

And then I think over the past year or so, rubrics and verifiers have become very important. And rubrics and verifiers are like learning by being graded and getting detailed feedback on where you went wrong.

中文翻译:

然后我认为在过去一年左右的时间里，评分标准 (rubrics) 和验证器 (verifiers) 变得非常重要。评分标准和验证器就像是通​​过被打分并获得关于你哪里做错的详细反馈来学习。

[00:42:17] Lenny Rachitsky

English:

And those are evals, another word for that?

中文翻译:

那些就是评估 (evals)，是它的另一种说法吗？

[00:42:19] Edwin Chen

English:

Yeah. So I think evals often covers two terms. One is you are using the evaluations for training because you're evaluating whether or not the model did a good job, and when it does do a good job, you're rewarding it. And then there's this other notion of evals where you're trying to measure the model's progress like, okay, yeah, I have five different candidate checkpoints and I want to pick the one that's best in order to release it to the public. So going to run all these evals on these five different checkpoints in order to decide which one is best.

中文翻译:

是的。我认为评估（evals）通常涵盖两个概念。一是你将评估用于训练，因为你在评估模型是否做得好，当它做得好时，你就给予奖励。另一个评估的概念是试图衡量模型的进展，比如：我有五个不同的候选检查点（checkpoints），我想选出最好的一个发布给公众。所以我要对这五个不同的检查点运行所有这些评估，以决定哪一个最好。

[00:42:51] Lenny Rachitsky

English:

Awesome.

中文翻译:

太棒了。

[00:42:51] Edwin Chen

English:

Yeah, and now we have RL environments, so this is kind of like a hot new thing.

中文翻译:

是的，现在我们有了 RL 环境，这算是目前最热门的新事物。

[00:42:55] Lenny Rachitsky

English:

Awesome. So what I love about this business journey is just there's always something new. There's always this like, okay. We're getting so good at just all this beautiful data for companies and now they need something completely different. Now we're setting up all these virtual machines for them and all these different use cases.

中文翻译:

太棒了。我喜欢这段商业旅程的一点是，总有新东西出现。总会有这种感觉：好吧，我们已经非常擅长为公司提供这些精美的数据了，现在他们又需要完全不同的东西了。现在我们要为他们设置所有这些虚拟机和所有这些不同的用例。

[00:43:08] Edwin Chen

English:

Yep.

中文翻译:

没错。

[00:43:08] Lenny Rachitsky

English:

And it feels like that's a big part of this industry you're in, it's just adapting to what labs are asking for.

中文翻译:

感觉这就是你所处行业的一个重要部分，就是不断适应实验室的需求。

[00:43:13] Edwin Chen

English:

Yeah. So I really do think that we are going to need to build a suite of products that reflect a million different ways that humans learn. Like for example, think about becoming a great writer. You don't become great by memorizing a bunch of grammar rules. You become great by reading great books, and you practice writing, and you get feedback from your teachers and from the people who buy your books in a bookstore and leave reviews. And you notice what works and what doesn't. And you develop taste by being exposed to all of these masterpieces and also just terrible writing. So you learn through this endless cycle of practicing reflection, and each type of learning that you have, again, these are all very different methods of learning to become a great writer, so just in the same way that... it's a thousand different ways that the great writer becomes great, I think there's going to be a thousand different ways that AI [inaudible 00:44:05] need to learn.

中文翻译:

是的。我确实认为我们需要构建一套产品，来反映人类学习的成千上万种方式。例如，想想如何成为一名伟大的作家。你不是通过背诵一堆语法规则变伟大的。你是通过阅读名著、练习写作、从老师以及在书店买你书并留下评论的读者那里获得反馈而变伟大的。你会注意到什么有效，什么无效。你通过接触所有这些杰作以及糟糕的作品来培养品味。所以你通过这种练习和反思的无尽循环来学习。每种学习类型——再次强调，这些都是成为伟大作家的截然不同的学习方法——就像伟大作家成名有一千种方式一样，我认为 AI 也需要有一千种不同的学习方式。

[00:44:05] Lenny Rachitsky

English:

It's so interesting this just ends up being just like humans in so many ways. It makes sense because in a sense, neural networks, deep learning is modeled after how humans have learned and how our brains operate, but it's interesting just to make them smarter. It's how do we come closer to how humans learn more and more?

中文翻译:

很有趣，这在很多方面最终都变得和人类一样。这很合理，因为在某种意义上，神经网络和深度学习就是模仿人类学习方式和大脑运作方式建模的。但有趣的是，为了让它们更聪明，我们需要让它们越来越接近人类的学习方式。

[00:44:22] Edwin Chen

English:

Yeah, it's almost like maybe the end goal is just throwing you into the environment and just seeing how you evolve. But within that evolution, there's all these different sub-learning mechanisms.

中文翻译:

是的，几乎就像终极目标就是把你丢进环境中，看你如何进化。但在进化的过程中，存在所有这些不同的子学习机制。

[00:44:34] Lenny Rachitsky

English:

Yeah, which is kind of what we're doing now, so that's really interesting. This might be the last step until we hit AGI. Along these lines, something that's really unique to Surge that I learned is you guys have your own research team, which I think is pretty rare, talk about just why that's something you guys have invested in and what has come out of that investment.

翻译:

是的，这正是我们现在正在做的，所以这非常有趣。这可能是我们达到 AGI 之前的最后一步。沿着这个思路，我了解到 Surge 非常独特的一点是你们拥有自己的研究团队，我认为这非常罕见。谈谈你们为什么要投资这个团队，以及这项投资带来了什么成果？

[00:44:52] Edwin Chen

English:

Yeah, so I think that stems from my own background. My own background is as a researcher. And so I've always cared fundamentally about pushing the industry and pushing the research community and not just about revenue. And so I think what our research team does is a couple different things. So we almost have two types of researchers at our company. One is our forward-deployed researchers who are often working hand in hand with our customers to help them understand their models.

中文翻译:

是的，我认为这源于我自己的背景。我本人就是一名研究员。所以我一直从根本上关心推动行业和研究社区的发展，而不仅仅是营收。我认为我们的研究团队主要做两件事。我们公司几乎有两种类型的研究员：一种是我们的“前线部署”（forward-deployed）研究员，他们通常与客户手拉手合作，帮助他们理解自己的模型。

[00:45:13] Edwin Chen

English:

So we will work very closely with the customers to help them understand, "Okay, this is where your model is today. This is where you're lagging behind all the competitors, these are some ways that you could be improving in the future, given your goals, and we're going to design these data sets, these evaluation methods, these training techniques to make your models better." So this very collaborative notion of working with our customers being researched by themselves, just a little bit more focused on the data side, and working hand in hand with them to do whatever it takes to make them the best.

中文翻译:

我们会与客户密切合作，帮助他们理解：“好吧，这是你模型现状。这是你落后于竞争对手的地方，根据你的目标，这些是你在未来可以改进的方法。我们将设计这些数据集、评估方法和训练技术来让你的模型变得更好。”这是一种非常协作的理念，与客户一起进行研究，只是更侧重于数据方面，并与他们并肩作战，不惜一切代价让他们成为最强的。

[00:45:57] Edwin Chen

English:

And then we also have our internal researchers. So our internal researchers are focused on slightly different things. So they are focused on building better benchmarks and better leaderboards. So I've talked a lot about how I worry that the leaderboards and benchmarks out there today are steering models in the wrong direction, so yeah, so the question is, how do we fix that? And so that's what our research team is focused focused really heavily on right now. So they're working a lot on that.

中文翻译:

然后我们还有内部研究员。内部研究员关注的事情略有不同。他们专注于构建更好的基准测试和排行榜。我已经谈了很多关于我担心目前的排行榜和基准测试正把模型引向错误方向的问题，所以问题是：我们如何修复它？这就是我们的研究团队目前重点攻克的方向。他们在这块投入了很多精力。

[00:46:23] Edwin Chen

English:

And they're also working on these other things like, "Okay, we need to train our own models to see what types of data performs the best, what types of people perform the best." And so they're also working on all these training techniques and evaluation of our own data sets to improve our data operations and the internal data products that we have that determine what makes something good quality.

中文翻译:

他们还在做其他事情，比如：“我们需要训练自己的模型，看看哪种类型的数据表现最好，哪种类型的人表现最好。”所以他们也在研究所有这些训练技术和对我们自己数据集的评估，以改进我们的数据运营和内部数据产品，从而确定什么才是高质量。

[00:46:46] Lenny Rachitsky

English:

It's such a cool thing because I don't think basically the labs have researchers helping them advance AI. I imagine it's pretty rare for a company like yours to have researchers actually doing primary research on AI.

中文翻译:

这太酷了，因为我不认为一般的实验室会有外部研究员帮他们推进 AI。我猜像你们这样的公司，有研究员真正进行 AI 基础研究（primary research），是非常罕见的。

[00:46:59] Edwin Chen

English:

Yeah, I think it's just because it's something I've fundamentally always cared about. I often think about us more like a research lab than a startup because that is my goal. It's kind of funny, but I've always said I would rather be Terrence Tao than Warren Buffett, so that notion of creating research that pushes the frontier forward and not just getting some valuation, that's always been what drives me.

中文翻译:

是的，我想这只是因为这是我从根本上一直关心的东西。我经常觉得我们更像是一个研究实验室，而不是一家初创公司，因为那是我的目标。这有点意思，但我总说我宁愿成为陶哲轩（Terrence Tao）而不是沃伦·巴菲特。那种创造能推动前沿发展的研究，而不仅仅是获得某种估值的理念，一直是我前进的动力。

[00:47:25] Lenny Rachitsky

English:

And it worked out. That's the beautiful thing about this. You mentioned that you were hiring researchers, is there anything there you want to share folks you're looking for?

中文翻译:

而且成功了。这就是这件事美妙的地方。你提到你们在招聘研究员，有什么想分享的吗？你们在找什么样的人？

[00:47:32] Edwin Chen

English:

So we look for people who are just fundamentally interested in dataset all day. So types of people who could literally spend 10 hours digging through a dataset, and playing around with models, and thinking, "Okay, yeah, this is where I think the model's failing," this is the kind of a behavior you want the model to have instead, and just this aspect of being very hands-on and thinking about the qualitative aspects of models and not just the quantitative parts. So again, it's like this aspect of being hands-on with data and not just caring about these kind of abstract algorithms.

中文翻译:

我们寻找的是那些整天对数据集有根本兴趣的人。就是那种能花 10 个小时钻研数据集、摆弄模型并思考“好吧，我觉得模型在这里失败了”、“我希望模型表现出这种行为”的人。他们非常注重实践，思考模型的定性方面，而不只是定量部分。所以，再次强调，就是那种亲自动手处理数据，而不仅仅关心抽象算法的人。

[00:48:07] Lenny Rachitsky

English:

Awesome. I want to ask a couple broad AI kind of market questions. What else do you think is coming in the next couple of years that people are maybe not thinking enough about or not expecting in terms of where AI is heading? What's going to matter?

中文翻译:

太棒了。我想问几个关于 AI 市场的宏观问题。你认为在未来几年，关于 AI 的发展方向，还有哪些是人们思考不足或没预料到的？什么会变得重要？

[00:48:20] Edwin Chen

English:

I think one of the things that's going to happen in the next few years is that the models are actually going to become increasingly differentiated because of the personalities and behaviors that the different labs have and the kind of objective functions that they are optimizing their models for. I think it's one thing I

didn't appreciate a year or so ago. A year or so ago, I thought that all of the AI models would essentially become very commoditized. They would all behave like each other, and sure, one of them might be slightly more intelligent in one way today, but sure, the other ones would catch up in the next few months. But I think over the past year, I've realized that the values that the companies have will shape the model.

中文翻译:

我认为未来几年会发生的一件事是，模型实际上会变得越来越差异化。这是因为不同实验室拥有的个性和行为方式，以及他们优化模型所采用的目标函数不同。这是一年前我还没意识到的。一年前，我以为所有的 AI 模型最终都会变得非常同质化（commoditized）。它们的表现会大同小异，当然，今天其中一个可能在某方面稍微聪明一点，但其他模型肯定会在接下来的几个月里赶上来。但在过去的一年里，我意识到公司的价值观会塑造模型。

[00:49:09] Edwin Chen

English:

So let me give you an example. So I was asking Claude to help me draft an email the other day, and it went through 30 different versions. And after 30 minutes, yeah, I think it really crafted me the perfect email, and I sent it. But then I realized that I spent 30 minutes doing something that didn't matter at all. Sure, now I got the perfect email, but I spent 30 minutes doing something I wouldn't have worried at all before, and this email probably didn't even move the needle on anything anyways.

中文翻译:

让我举个例子。前几天我让 Claude 帮我起草一封邮件，它改了 30 个不同的版本。30 分钟后，是的，我觉得它确实为我写出了一封完美的邮件，然后我发了出去。但随后我意识到，我花了 30 分钟做了一件根本不重要的事情。当然，我现在得到了一封完美的邮件，但我花了 30 分钟做了一件我以前根本不会担心的事，而且这封邮件可能对任何事情都没有实质性的推动作用。

[00:49:35] Edwin Chen

English:

So I think there is a deep question here, which is, if you could choose the perfect model behavior, which model would you want? Do you want a model that says, "You're absolutely right. There are definitely 20 more ways to improve this email," and it continues for 50 more iterations. And it sucks up all your time and engagement. Or do you want a model that's optimizing for your time and productivity and just says, "No, you need to stop. Your email's great. Just send it and move on with your day"?

中文翻译:

所以我认为这里有一个深刻的问题：如果你可以选择完美的模型行为，你想要哪种模型？你想要一个会对你说“你完全正确，肯定还有 20 种方法可以改进这封邮件”并继续迭代 50 次，耗尽你所有时间和精力模型？还是想要一个为你节省时间、提高效率，直接说“不，你该停下了，你的邮件写得很好，直接发出去然后过好你的一天”的模型？

[00:49:59] Edwin Chen

English:

And again, just because... In the same way that there's like a kind of a fork in a road between how you could choose how your model behaves for this question, it's like for every other question that models have, the kind of behavior that you want will fundamentally affect it. It's almost like in the same way that when Google builds a search engine, it's very different from how Facebook would build a search engine, which is very different from how Apple would build a search engine. They all have their own principles and values and things that they're trying to achieve in the world that shape all the products that they're going to build. And in the same way, I think all the [inaudible 00:50:40] will start behaving very differently too.

中文翻译:

同样地，就像在这个问题上模型行为的选择存在分歧一样，对于模型面临的每一个其他问题，你想要的那种行为都会从根本上影响它。这几乎就像 Google 构建搜索引擎的方式与 Facebook 不同，也与 Apple 不同。他们都有自己的原则、价值观和想要在这个世界上实现的目标，这些塑造了他们将要构建的所有产品。同样地，我认为所有的模型也会开始表现得非常不同。

[00:50:41] Lenny Rachitsky

English:

That is incredibly interesting. You already see that with Grok. It's got a very different personality and a very different approach to answering questions. And so what I'm hearing is you're going to see more of this differentiation.

中文翻译:

这非常有趣。你已经在 Grok 身上看到了这一点。它有着非常不同的个性和非常不同的回答问题的方式。所以我听到的是，你会看到更多这种差异化。

[00:50:52] Edwin Chen

English:

Yep.

中文翻译:

没错。

[00:50:53] Lenny Rachitsky

English:

Kind of another question along these lines, what do you think is most under-hyped in AI that you think maybe people aren't talking enough about that is really cool? And what do you think is over-hyped?

中文翻译:

沿着这个思路还有另一个问题：你认为 AI 中最被低估（under-hyped）的是什么？也就是人们讨论不够但其实非常酷的东西。以及你认为什么是过度炒作（over-hyped）的？

[00:51:04] Edwin Chen

English:

So I think one of the things that's under-hyped is the built-in products that all of the chatbots are going to start having. I've always been a huge fan of Claude's artifacts. And I think it just works really well. And actually the other day, I don't know if it's a new feature or not, but it asked me to help me create an email, and then it just created... So it didn't quite work because it didn't allow me to send the email. But what it created instead was like a little, I don't know what we call it, like a little box where I could click on it and it would just text someone that did this message. And I think that concept of taking artifacts to the next level where you just have these mini apps, mini UIs within the chatbots themselves, I feel like people aren't talking enough about that. So I think that that's one under-hyped area.

中文翻译:

我认为被低估的一点是所有聊天机器人将开始拥有的内置产品功能。我一直是 Claude 的 Artifacts 功能的忠实粉丝，我觉得它非常好用。事实上前几天，我不知道是不是新功能，它帮我写邮件，然后它创建了一个……虽然还没完全跑通，因为它没法直接帮我发邮件，但它创建了一个类似小盒子的东西，我点一下它就能把这段信息发给某人。我认为将 Artifacts 提升到下一个层次，即在聊天机器人内部拥有这些微型应用（mini apps）和微型 UI，我觉得人们对此讨论得还不够。所以我认为这是一个被低估的领域。

[00:51:54] Edwin Chen

English:

And in terms of over-hyped areas, I definitely think that vibe coding is over-hyped. I think people don't realize how much it's going to make your systems unmaintainable in the long-term and they simply dump this code into their code bases if this seems to work out right now, so I kind of worry about the future of coding. It's just going to keep on happening.

中文翻译:

至于过度炒作的领域，我绝对认为“氛围感编程”（vibe coding，指凭感觉让 AI 写代码而不深究逻辑）被过度炒作了。我认为人们没有意识到，从长远来看，这会让你的系统变得多么难以维护。他们只是因为代码现在看起来能跑通，就直接把它丢进代码库里。所以我有点担心编程的未来。这种情况只会持续发生。

[00:52:17] Lenny Rachitsky

English:

These are amazing answers. On that first point, there's something I actually asked. I have the chief product officer of Anthropic and OpenAI, Kevin Weil and Mike Krieger on the podcast, and I asked them just like, "As a product team, you have this gigabrain intelligence. How long do you even need product teams?" You think this AI will just create the product for you. "Here's what I want." It's like the next level of vibe coding. It's just like tell it, "Here's what I want," and it's just building the product and involving the product as you're using it. And it feels like that's what you're describing is where we might be heading.

中文翻译:

这些回答太棒了。关于第一点，我其实问过 Anthropic 和 OpenAI 的首席产品官 Kevin Weil 和 Mike Krieger。我问他们：“作为一个产品团队，既然你们拥有这种‘超级大脑’级别的智能，你们还需要产品团队多久？”你觉得 AI 会直接为你创建产品，“这就是我想要的”。这就像是“氛围感编程”的更高层次：只要告诉它“我想要什么”，它就会在你的使用过程中不断构建和演进产品。感觉你描述的正是我们可能的发展方向。

[00:52:48] Edwin Chen

English:

Yeah, I think there is a very powerful notion where it helps people just achieve their ideas in a much cooler way.

中文翻译:

是的，我认为这是一个非常强大的理念，它能帮助人们以一种更酷的方式实现他们的想法。

[00:52:55] Lenny Rachitsky

English:

Something we haven't gotten into that I think is really interesting is just the story of how you got to starting Surge. You have a really unique background. I always think about these... Brian Armstrong, the founder of Coinbase, once gave this talk that has really stuck with me where he kind of talked about how his very unique background allowed him to start Coinbase. He had a economics background, he had a cryptography experience, and then he was an engineer. And it's like the perfect Venn diagram for starting Coinbase, and I feel like you have a very similar story with Surge. Talk about that, your background there, and how that led to Surge.

中文翻译:

我们还没深入探讨的一件我觉得非常有趣的事，就是你创办 Surge 的故事。你有着非常独特的背景。我经常想到 Coinbase 的创始人 Brian Armstrong 曾经做过的一次演讲，让我印象深刻。他谈到他非常独特的背景如何让他创办了 Coinbase：他有经济学背景，有密码学经验，而且他还是个工程师。这简直是创办 Coinbase 的完美维恩图。我觉得你在 Surge 身上也有类似的故事。谈谈你的背景，以及它是如何引导你创办 Surge 的。

[00:53:31] Edwin Chen

English:

Going way back, I was always fascinated by math and language when I was a kid. I went to MIT because it's obviously one of the best places for math and CS, but also because it's the home of Noam Chomsky. My dream in school was actually to find some underlying theory connecting all these different fields. And then I became a researcher at Google, and Facebook, and Twitter, and I just kept running into the same problem over and over again. It was impossible to get the data that we needed to train our models.

中文翻译:

追溯到很久以前，我小时候一直对数学和语言着迷。我去了 MIT，显然是因为那里是数学和计算机科学最顶尖的地方之一，但也因为那里是诺姆·乔姆斯基（Noam Chomsky）的所在地。我上学时的梦想实际上是找到某种连接所有这些不同领域的底层理论。后来我成了 Google、Facebook 和 Twitter 的研究员，我发现自己一遍又一遍地遇到同样的问题：根本无法获得训练模型所需的数据。

[00:54:12] Edwin Chen

English:

So I was always this huge believer in the need for high quality data, and then GPT-3 came out in 2020. And I realized that, yeah, if we wanted to take things to the next level and build models that could code, and use tools, and tell jokes, and write poetry, and solve [inaudible 00:54:12], and cure cancer, then yeah, we were going to need a completely new solution. The thing that always drove me crazy when I was at all these companies was we had a full power of the human mind in front of us, and all the data students out

there were focused on really simple things like image labeling. So I wanted to build something focus on all these advanced, complex use cases instead that would really help us build our next generation models. So yeah, I think my background in kind of across math, and computer science, and linguistics really informed what I always wanted to do, and so I started Surge a month later with our one mission to basically build the use cases that I thought were going to be needed to push the frontier of AI.

中文翻译:

所以我一直坚信高质量数据的必要性。2020 年 GPT-3 发布时，我意识到，如果我们想把事情提升到下一个层次，构建能够编程、使用工具、讲笑话、写诗、治愈癌症的模型，那么我们需要一个全新的解决方案。在那些大公司工作时，最让我抓狂的是，我们面前拥有人类思想的全部力量，但当时所有的数据标注公司都专注于图像标注等非常简单的事情。所以我想要构建一些专注于这些高级、复杂用例的东西，真正帮助我们构建下一代模型。所以，我认为我在数学、计算机科学和语言学方面的跨学科背景确实决定了我一直想做的事情。于是在一个月后创办了 Surge，我们的唯一使命就是构建我认为推动 AI 前沿所需的用例。

[00:54:49] Lenny Rachitsky

English:

And you said a month later, a month later after what?

中文翻译:

你说是“一个月后”，是在什么之后的一个月？

[00:54:52] Edwin Chen

English:

After a GPT-3 launch in 2020.

中文翻译:

在 2020 年 GPT-3 发布之后。

[00:54:54] Lenny Rachitsky

English:

Oh, okay. Wow. Okay. Yeah. A great decision. What just kind of drives you at this point of... Other than just the epic success you're having, what keeps you motivated to keep building this and building something in this space?

中文翻译:

噢，好吧。哇。好的。这是一个伟大的决定。除了你现在取得的巨大成功之外，目前是什么在驱动你？是什么让你保持动力继续在这个领域构建 Surge？

[00:55:06] Edwin Chen

English:

I think I'm a scientist at heart. I always thought I was going to become this math or CS professor and work on trying to understand the universe, and language, and the nature of communication. It's kind of funny,

but I always had this fanciful dream where if aliens ever came to visit Earth and we need to figure out how to communicate with them, I wanted to be the one the government would call. And I'd use all this fancy math, and computer science, and linguistics to decipher it.

中文翻译:

我认为我内心深处是一个科学家。我一直以为我会成为一名数学或计算机科学教授，致力于理解宇宙、语言和沟通的本质。这挺有意思的，但我一直有一个奇妙的梦想：如果外星人访问地球，我们需要弄清楚如何与他们沟通，我希望我是那个政府会打电话求助的人。我会用所有这些高深的数学、计算机科学和语言学知识来破译它。

[00:55:33] Edwin Chen

English:

So even today, what I love doing most is every time a new model is released, we'll actually do a really deep dive into the model itself. I'll play around with it, I'll run evals, I'll compare where it's improved, where it's arrest, I'll create this really deep dive analysis that we send our customers. And it's actually kind of funny because a lot of times we'll say it's from a data science team, but often it's actually just from me.

中文翻译:

所以即使是今天，我最喜欢做的事情就是每当有新模型发布时，我们都会对模型本身进行深度研究。我会摆弄它，运行评估，比较它在哪里进步了，在哪里退步了，然后写出一份深度分析报告发给我们的客户。这其实挺好笑的，因为很多时候我们会说这是来自“数据科学团队”，但通常其实就是我写的。

[00:55:54] Edwin Chen

English:

And I think I could do this all day. I have a very hard time being in meetings all day. I'm terrible at sales, I'm terrible at doing the typical CEO things that people expect you to do, but I love writing these analyses. I love jamming with our research team about what we're seeing, sometimes I'll be up until 3:00 AM just talking on the phone with somebody on the research team and [inaudible 00:56:12] model. So I love that I still get to be really hands-on, working on the data and the science all day. And I think what drives me is that I want Surge to play this critical role in the future of AI, which I think is also the future of humanity.

中文翻译:

我觉得我可以整天做这件事。我很难忍受整天开会。我不擅长销售，也不擅长做人们期望 CEO 做的那些典型事情，但我喜欢写这些分析报告。我喜欢和我们的研究团队讨论我们的发现，有时我会熬夜到凌晨 3 点，只为了和研究团队的人通电话讨论模型。我喜欢我仍然能亲自动手，整天钻研数据和科学。我认为驱动我的是，我希望 Surge 在 AI 的未来中发挥关键作用，而我认为 AI 的未来也是人类的未来。

[00:56:12] Edwin Chen

English:

We have these really unique perspectives on data, and language, and quality, and how to measure all of this, and how to ensure it's all going on the right path. And I think we're uniquely unconstrained by all of these influences that can sometimes steer companies in a negative direction. Like what I was saying earlier, we built Surge a lot more like a research lab than a typical startup. So we care about curiosity and

long-term incentives and intellectual rigor, and we don't care as much about quarterly metrics and what's going to look good in a [inaudible 00:56:56]. And so my goal is to take all these unique things about us as a company and use that to make sure that we're shaping AI in a way that's really beneficial for our species in the long term.

中文翻译:

我们对数据、语言、质量以及如何衡量这一切、如何确保一切都在正确轨道上有着非常独特的视角。而且我认为我们独特地不受那些有时会将公司引向负面方向的影响所束缚。就像我之前说的，我们把 Surge 构建得更像一个研究实验室，而不是典型的初创公司。我们关心好奇心、长期激励和智识上的严谨，而不太关心季度指标或在财报中表现如何。所以我的目标是利用我们作为一家公司的所有这些独特之处，确保我们以一种长远来看对人类物种真正有益的方式来塑造 AI。

[00:57:06] Lenny Rachitsky

English:

What I'm realizing in this conversation is just how much influence you have and companies like yours have on where AI heads. The fact that you help labs understand where they have gaps and where they need to improve, and it's not just everyone looks at just like the heads of OpenAI and Anthropic and all these companies as they're the ones ushering in AI, but what I'm hearing here is you have a lot of influence on where things head too.

中文翻译:

在这次对话中我意识到，你以及像你这样的公司对 AI 的走向有着巨大的影响力。事实上，你们帮助实验室了解他们的差距在哪里，以及他们需要在哪里改进。大家通常只关注 OpenAI 和 Anthropic 等公司的负责人，认为他们才是引领 AI 的人，但我在这里听到的是，你们对事物的走向也有很大的影响力。

[00:57:30] Edwin Chen

English:

Yeah, I think there's this really powerful ecosystem where, honestly, people just don't know where models are headed and how they want to shape them yet and how they want humanity kind of like [inaudible 00:57:47] play a role in the future of all of this, and so I think there's a lot of opportunity to just continue shaping the discussion.

中文翻译:

是的，我认为存在这样一个非常强大的生态系统。老实说，人们目前还不知道模型会走向何方，不知道该如何塑造它们，也不知道人类在这一切的未来中应该扮演什么样的角色。所以我认为有很多机会可以继续引导这种讨论。

[00:57:52] Lenny Rachitsky

English:

Along that thread, I know you have a very strong thesis on just why this work matters to humanity and why this is so important, talk about that.

中文翻译:

沿着这个思路，我知道你对于为什么这项工作对人类至关重要有着非常坚定的论点，谈谈那个吧。

[00:58:01] Edwin Chen

English:

I'll get a bit philosophical here, but I think the question itself is a bit philosophical, so bear with me. So the most straightforward way of thinking about what we do is we train and evaluate AI. But there's a deeper mission that I often think about, which is helping our customers think about their dream objective functions. Like yeah, what kind of model do they want their model to be? And once we help them do that, we'll help them train their model to reach their north star and we'll help them measure that progress.

中文翻译:

这里我会谈得有点哲学，但我认为问题本身就带点哲学色彩，所以请多包涵。思考我们工作的最直观方式是：我们训练和评估 AI。但我经常思考一个更深层的使命，那就是帮助我们的客户思考他们梦想中的“目标函数”。比如，他们希望自己的模型成为什么样的模型？一旦我们帮助他们明确了这一点，我们会帮助他们训练模型以达到他们的“北极星”目标，并帮助他们衡量进展。

[00:58:50] Edwin Chen

English:

But it's really hard because objective functions are really rich and complex. It's kind of like the difference between having a kid and asking them, "Okay, what test do you want to pass? Do you want them to get a high score on SAT and write a really good college essay?" That's a simplistic version versus what kind of person do you want them to grow up to be? Will you be happy if they're happy no matter what they do or are you hoping they'll go to a good school and be financially successful?

中文翻译:

但这非常困难，因为目标函数非常丰富且复杂。这有点像养孩子：你是问他们“你想通过什么考试？你想在 SAT 中拿高分并写出一篇很棒的大学申请论文吗？”这是简单化的版本。而更深层的版本是：你希望他们长大后成为什么样的人？如果他们无论做什么都很快乐，你会感到欣慰吗？还是你希望他们上名校并获得财务上的成功？

[00:59:25] Edwin Chen

English:

And again, if you take that notion, it's like, okay, how do you define happiness? How do you measure whether they're happy? How do you measure whether they're financially successful? It's a lot harder than something measuring whether or not you're getting a high score on the SAT, and what we're doing is we want to help our customers reach, again, their dream north stars and figure out how to measure them. And so I talked about this example of what you want models to do when you're asking them to write 50 different evaluations. Do you just continue them for 50 more or do you just say, "No, just move on [inaudible 00:59:25] because this is perfect enough."

中文翻译:

同样，如果你接受这个观念，问题就变成了：你如何定义幸福？你如何衡量他们是否幸福？你如何衡量他们是否在财务上成功？这比衡量 SAT 分数要难得多。而我们正在做的，就是帮助我们的客户达到他们梦想中的北极星目标，并弄清楚如何衡量它们。所以我提到了那个例子：当你要求模型写 50 个不同的评估时，你希望它做什么？是让它再继续写 50 个，还是直接说“不，到此为止，因为这已经足够完美了”。

[00:59:44] Edwin Chen

English:

And the broader question is, are we building these systems that actually advance humanity? And if so how do we build the data sets to train towards that and measure it? Are we optimizing for all of these wrong things, just systems that suck up more and more of our time and make us lazier and lazier? And yeah, I think it's really relevant to what we do because it's very hard and difficult to measure and define whether something is genuinely advancing humanity. It's very easy to measure all these proxies instead like clicks and likes.

中文翻译:

更广泛的问题是：我们是否在构建真正推动人类进步的系統？如果是，我们如何构建数据集来朝着这个方向训练并进行衡量？我们是否在为所有这些错误的东西进行优化，仅仅是构建那些吸走我们越来越多时间、让我们越来越懒惰的系统？我认为这与我们的工作密切相关，因为衡量和定义某件事是否真正推动了人类进步是非常困难的。相反，衡量点击量和点赞数等替代指标（proxies）则非常容易。

[01:00:12] Edwin Chen

English:

But I think that's why our work is so interesting. We want to work the hard, important metrics that require the hardest types of data and not just the easy ones. So I think one of the things I often say is you are your objective function. So we want the rich, complex, objective functions and not these simplistic proxies. And our job is to figure out how to get the data to match this. So yeah, we want data, we want metrics that measure whether AI is making your life richer. We want to train our systems this way. And we want tools that make us more curious and more creative, not just lazier.

中文翻译:

但我认为这正是我们工作有趣的地方。我们想要处理那些困难且重要的指标，这些指标需要最难获取的数据，而不仅仅是容易获取的数据。我常说的一句话是：“你就是你的目标函数”。所以我们追求丰富、复杂的目标函数，而不是这些简单化的替代指标。我们的工作就是弄清楚如何获得与之匹配的数据。是的，我们想要数据，想要能衡量 AI 是否让你的生活变得更丰富的指标。我们想以这种方式训练我们的系统。我们想要能让我们变得更好奇、更有创造力，而不仅仅是更懒惰的工具。

[01:00:37] Lenny Rachitsky

English:

Wow. I love how what you're sharing here gives you so much more appreciation of the nuances of building AI, training AI, the work that you're doing. From the outside, people could just look at Surge and companies in the space of, okay, cool. They're just creating all this data, feeding it to AI. But clearly there's so much to this that people don't realize, and I love knowing that you're at the head of this, that someone like you is thinking through this so deeply. Maybe one more question, is there something you wish you'd known before you started Surge? A lot of people start companies, they don't know what they're getting into. Is there something you wish you could tell your earlier self?

中文翻译:

哇。我喜欢你分享的这些内容，它让人对构建 AI、训练 AI 以及你们所做工作的细微差别有了更深的理解。从外部看，人们可能只是觉得 Surge 和这类公司只是在创造数据并喂给 AI。但显然，这里面有很多人们没有意识到的东西。我很高兴知道你在领导这件事，像你这样的人在如此深入地思考这些问题。也许还有一个问题：在创

办 Surge 之前，有什么是你希望自己已经知道的？很多人创办公司时并不知道将自己将面临什么。有什么是你希望告诉早期的自己的吗？

[01:01:11] Edwin Chen

English:

Yeah, so I definitely wish I'd known that you could build a company by being heads down and doing great research and simply building something amazing. And not by constantly tweeting and hyping and fundraising. It's kind of funny, but I never thought I wanted to start a company. I love doing research. And I was actually always a huge fan of DeepMind because they were this amazing research company that got bought and still managed to keep on doing amazing science. But I always thought that they were this magical ILR unicorn.

中文翻译:

是的，我确实希望我早点知道，你可以通过埋头苦干、做伟大的研究并构建出惊人的东西来建立一家公司，而不是通过不断发推、炒作和融资。挺有意思的，我从未想过我要创办一家公司。我热爱做研究。我一直非常崇拜 DeepMind，因为他们是一家了不起的研究型公司，即使被收购了，仍然能继续从事了不起的科学研究。但我一直以为他们是那种不可复制的“独角兽”。

[01:01:45] Edwin Chen

English:

So I thought if I started a company, I'd have to become a business person looking at financials all day and being in meetings all day and doing all this stuff that sounded incredibly boring and I always hated. So I think it's crazy that didn't end up being true at all. I'm still in the weeds in the data every day. And I love it. I love that I get to do all these analyses and talk to researchers. And it's basically applied research where we're building all these amazing data systems that have really pushed the frontier of AI.

中文翻译:

所以我以为如果我创办公司，我就必须变成一个整天盯着财务报表、整天开会的生意人，做所有这些听起来极其无聊且我一直讨厌的事情。我觉得疯狂的是，事实完全不是这样。我每天仍然沉浸在数据中。我热爱这一点。我喜欢做这些分析，喜欢和研究员交流。这基本上就是应用研究，我们正在构建所有这些惊人的数据系统，真正推动了 AI 的前沿。

[01:02:01] Edwin Chen

English:

So yeah, I wish I know that you don't need to spend all your time fundraising. You don't need to constantly generate hype. You don't need to become someone you're not. You can actually build a successful company by simply building something so good that it cut through all that noise. And I think if I known this was possible, I would've started even sooner, so I [inaudible 01:02:18] that.

中文翻译:

所以，是的，我希望我早点知道你不需要把所有时间都花在融资上，不需要不断制造噱头，不需要变成一个违背本心的人。你实际上可以通过构建足够好的东西来建立一家成功的公司，好到足以穿透所有的噪音。如果我早知道这是可能的，我可能会更早开始。

[01:02:18] Lenny Rachitsky

English:

And that is such an amazing place to end. I feel like this is exactly what founders need to hear, and I think this conversation's going to inspire a lot of founders, and especially a lot of founders that want to do things in a different way. Before we get to a very exciting lightning round, is there anything else you wanted to share? Anything else you want to leave our listeners with? We covered a lot of ground, it's totally okay to say no as well.

中文翻译:

这是一个完美的结尾。我觉得这正是创始人需要听到的，我相信这次对话会激励很多创始人，尤其是那些想以不同方式做事的创始人。在进入非常精彩的闪电轮提问之前，你还有什么想分享的吗？有什么想留给听众的吗？我们已经讨论了很多，如果没什么要补充的也没关系。

[01:02:37] Edwin Chen

English:

I think the thing I would end with is I think a lot of people think of data labeling as it relates to simplistic work. Like labeling cat photos and drawing bounding box around cars. And so I've actually always hated the word data labeling because it just paints this very simplistic picture when I think what we're doing is completely different. I think a lot about what we're doing as a lot more like raising a child. You don't just feed a child information. You're teaching them values, and creativity, and what's beautiful, and these infinite subtle things about what makes somebody a good person. And that's what we're doing for AI. So yeah, I just often think about what we're doing as almost like the future of humanity or how we're raising humanity's children, so I'll leave it at that.

中文翻译:

我想最后说的是，很多人认为数据标注（data labeling）是一项简单的工作，比如标注猫的照片或在汽车周围画框。实际上我一直很讨厌“数据标注”这个词，因为它描绘了一个非常简单化的图景，而我认为我们所做的事情完全不同。我更多地把我们的工作看作是在“抚养孩子”。你不仅仅是给孩子喂信息，你还在教他们价值观、创造力、什么是美，以及关于如何成为一个好人的无数微妙细节。这就是我们为 AI 所做的事情。所以我经常觉得我们所做的几乎就像是人类的未来，或者说我们正在抚养“人类的孩子”。我就说这么多。

[01:03:27] Lenny Rachitsky

English:

Wow. I love just how much philosophy there is in this whole conversation that I was not expecting. With that, Edwin, we've reached our very exciting lightning round, I've got five questions for you. Are you ready?

中文翻译:

哇。我喜欢整个对话中蕴含的哲学思考，这超出了我的预期。那么 Edwin，我们进入了非常精彩的闪电轮，我有五个问题要问你。准备好了吗？

[01:03:38] Edwin Chen

English:

Yep, let's go.

中文翻译:

准备好了，开始吧。

[01:03:39] Lenny Rachitsky

English:

Here we go. What are two or three books that you find yourself recommending most to other people?

中文翻译:

第一个问题：你最常向别人推荐的两三本书是什么？

[01:03:45] Edwin Chen

English:

Yes, so three books I often recommend are, first, Story of Your Life by Ted Chiang. It's my all time favorite short story and it's about a linguist learning and alien language, and I basically reread it every couple years.

中文翻译:

好的，我常推荐的三本书是：第一，特德·姜（Ted Chiang）的《你一生的故事》（Story of Your Life）。这是我最喜欢的短篇小说，讲的是一个语言学家学习外星语言的故事，我基本上每隔几年就会重读一遍。

[01:03:56] Lenny Rachitsky

English:

And that's what the Interstellar was about? Is that...

中文翻译:

那是《星际穿越》的主题吗？还是……

[01:03:59] Edwin Chen

English:

Yeah, so there's a movie called Arrival... which was based off of the story, which I love as well.

中文翻译:

是的，有一部叫《降临》（Arrival）的电影就是根据这个故事改编的，我也非常喜欢那部电影。

[01:04:06] Edwin Chen

English:

And then second, Myth of Sisyphus by Camus. I actually can't really explain why I love this, but I always find a final chapter somehow are really inspiring. And then third, Le Ton beau de Marot by Douglas Hofstadter. And so I think Gödel, Escher, Bach is his more famous book, but I've actually always loved this

one better. It basically takes a single French poem and translates it 89 different ways and discusses all the motivations behind each translation. And so I've always loved the way it embodies this idea that translation isn't this robotic thing that you do. Instead, there's a million different ways to think about what makes a high quality translation, which makes a lot of ways I think about data and quality in LLMs.

中文翻译:

第二本是加缪的《西西弗的神话》(Myth of Sisyphus)。我其实无法解释为什么喜欢它，但我总觉得最后一章非常鼓舞人心。第三本是道格拉斯·霍夫施塔特 (Douglas Hofstadter) 的《马罗之美》(Le Ton beau de Marot)。我想《哥德尔、艾舍尔、巴赫》是他更出名的书，但我一直更喜欢这一本。它基本上是把一首法国诗翻译成了 89 种不同的版本，并讨论了每种翻译背后的动机。我一直很喜欢它所体现的理念：翻译不是一种机械的行为。相反，对于什么是高质量的翻译，有一百万种思考方式，这与我思考 LLM 中的数据和质量的方式非常相似。

[01:04:44] Lenny Rachitsky

English:

All these resonate so deeply with the way, with all the things we've been talking about, especially that first one, if that was your goal after school is like, "I want to help translate alien language." I'm not surprised you love that short story. Next question, do you have a favorite recent movie or TV show you've really enjoyed?

中文翻译:

所有这些书都与我们讨论的内容产生了深刻的共鸣，尤其是第一本，既然你毕业后的目标是“想帮外星人翻译语言”，我不奇怪你喜欢那个短篇小说。下一个问题：你最近有没有特别喜欢的电影或电视剧？

[01:05:00] Edwin Chen

English:

One of my new all time favorite TV shows is something I found recently, it's called Travelers. It's basically about a group of travelers from the future who are sent back in time to prevent their [inaudible 01:05:11]. And then I actually just rewatched Contact, which is one of my all time favorite movies. So yeah, I think one of the things you'll notice about me is that, yeah, I love any kind of book or film that involves scientists deciphering alien communication. Again, just this dream I always had as a kid.

中文翻译:

我最近发现的一部新的心头好电视剧叫《穿越者》(Travelers)。它讲的是一群来自未来的旅行者被送回过去以阻止灾难。然后我最近刚重温了《超时空接触》(Contact)，那是我最喜欢的电影之一。所以，你会发现我的一个特点：我喜欢任何涉及科学家破译外星通讯的书籍或电影。再次强调，这就是我小时候的梦想。

[01:05:30] Lenny Rachitsky

English:

That's so funny. Okay, is there a product you've recently discovered that you really love?

中文翻译:

太有意思了。好的，有没有你最近发现并非常喜欢的产品？

[01:05:35] Edwin Chen

English:

So it's funny, but I was in SF earlier this week and I finally took Waymo for the first time. Honestly, it was magical and it really felt like living in the future.

中文翻译:

挺巧的，我这周早些时候在旧金山，终于第一次坐了 Waymo。老实说，那感觉很神奇，真的感觉生活在未来。

[01:05:43] Lenny Rachitsky

English:

Yeah, it's like the thing that... People hype it like crazy, but it always exceeds your expectations.

中文翻译:

是的，这东西……人们疯狂炒作它，但它总是能超出你的预期。

[01:05:48] Edwin Chen

English:

Yeah, it deserves the hype. It was crazy. Yeah, it's absurd. It's like, holy moly. If you're not in SF, you don't realize just how common these things are. They're just all over the place. Just driverless cars constantly going about, and when you go to an event at the end, there's just all these Waymos lined up picking people up.

中文翻译:

是的，它名副其实。太疯狂了，简直不可思议。如果你不在旧金山，你不会意识到这些东西有多普遍。它们到处都是，无人驾驶汽车不断穿梭。当你参加完活动出来，会看到一排 Waymo 在那里排队接人。

[01:06:03] Lenny Rachitsky

English:

Yeah. Waymo, good job. Good job over there. Do you have a favorite life motto that you find yourself coming back to in work or in life?

中文翻译:

是的，Waymo 做得好。下一个问题：你有没有在工作或生活中经常想起的人生格言？

[01:06:11] Edwin Chen

English:

So I think I mentioned this idea that founders should build a company that only they could build. Almost like it's this destiny that their entire life, and experiences, and interests shape them towards. And so I think that principle applies pretty broadly, not just the founders, but the people creating, I think.

中文翻译:

我想我提到过这个想法：创始人应该建立一家只有他们才能建立的公司。几乎就像是一种天命，他们的一生、经历和兴趣都在塑造他们走向这个目标。我认为这个原则适用范围很广，不仅适用于创始人，也适用于所有创造者。

[01:06:25] Lenny Rachitsky

English:

Well, let me follow that thread to unlightening this answer. Do you have any advice for how to build those sorts of experiences that help lead to that? Is it follow things that are interesting to you, because it's easy to say that, it's hard to actually acquire these really unique sets of experiences that allow you to create something really important?

中文翻译:

那让我追问一下。你对于如何积累这类有助于实现目标的经验有什么建议吗？是追随你感兴趣的事物吗？因为说起来容易，真正获得这些能让你创造出重要事物的独特经验其实很难。

[01:06:44] Edwin Chen

English:

Yeah, so I think it would always be to really follow your interests and do what you love, and it's almost like a lot of decisions I make about Surge. I think one of the things that I didn't think about a couple years ago, but then someone said it to me, it's that companies in a sense are an embodiment of their CEO. And it's kind of funny. I hadn't thought about that because I never quite knew what a CEO did. I always thought a CEO was kind of generic and it's like, okay, you're just doing whatever VPs, and your board, and whatever, tell you to do and you're just saying yes to decisions. But instead, it's this idea where when I think about certain big, hard decisions we have to make, I don't think what would the company do, I don't think what metrics are we trying to optimize, I just think, "What do I personally care about? What are my values? And what do I want to see happen in the world?"

中文翻译:

是的，我认为永远应该是真正追随你的兴趣，做你热爱的事。这几乎就像我为 Surge 做出的很多决定。几年前我没想过这一点，但后来有人告诉我：在某种意义上，公司是其 CEO 的化身。这挺有意思的。我以前没想过，因为我从不确切知道 CEO 是做什么的。我总觉得 CEO 是个很通用的角色，就像是：好吧，你只是在做副总裁、董事会让你做的事，你只是对决策说“是”。但相反，当我思考我们要做的某些重大、艰难的决定时，我不会想“公司会怎么做”，也不会想“我们要优化什么指标”，我只会想：“我个人关心什么？我的价值观是什么？我希望看到世界上发生什么？”

[01:07:34] Edwin Chen

English:

And so I think following that idea about... Okay, so ask yourself, what are the values you care about? What are things you're trying to shape and not... What will look good on a dashboard? I think that results are pretty important.

中文翻译:

所以我觉得要遵循这个想法……问问你自己，你关心的价值观是什么？你试图塑造的是什麼，而不是……仪表盘上什么数据好看？我认为这些结果非常重要。

[01:07:49] Lenny Rachitsky

English:

I love how just you're just full of endless, beautiful, and very deep answers. Final question. Something that you got quite famous for before starting Surge is you built this map while you were at Twitter that showed a map of the world and what people called, whether they called it soda or pop. I don't know if it's called Soda Pop. What was the name of this map?

中文翻译:

我喜欢你这些无尽、优美且深刻的回答。最后一个问题：在创办 Surge 之前，你有一件非常出名的事，就是你在 Twitter 工作时制作了一张地图，显示了世界上不同地区的人如何称呼碳酸饮料——是叫“soda”还是“pop”。我不知道它是不是叫“Soda Pop”。那张地图叫什么名字？

[01:08:13] Edwin Chen

English:

Yeah, it was like the Soda Versus Pop dataset.

中文翻译:

是的，那是“Soda 对决 Pop”数据集。

[01:08:15] Lenny Rachitsky

English:

Soda Versus Pop. And so it's like a map of the United States and it tells you where people say pop versus soda, so do you say soda or pop?

中文翻译:

Soda 对决 Pop。那是一张美国地图，告诉你哪里的人说 pop，哪里的人说 soda。那么，你说 soda 还是 pop？

[01:08:23] Edwin Chen

English:

So I say soda, I'm a soda person.

中文翻译:

我说 soda，我是 soda 派。

[01:08:26] Lenny Rachitsky

English:

Okay. And is that just like that's the right answer or it's like whatever you are, it's totally fine.

中文翻译:

好吧。那是“正确答案”吗？还是说无论怎么叫都行？

[01:08:33] Edwin Chen

English:

I think I'll look at you a little bit funny. You say pop and I'll wonder where you came from, but I won't score on you too much.

中文翻译:

如果你说 pop，我可能会用奇怪的眼神看你，心想你是从哪儿来的，但我不会太嫌弃你。

[01:08:39] Lenny Rachitsky

English:

That's how I feel too. Edwin, this was incredible. This was such an awesome conversation. I learned so much. I think we're going to help a lot of people start their own companies, help their companies become more aligned with their values and just building better things. Few final questions, where can folks find you online if they want to reach out? What roles are you hiring for? How can listeners be useful to you?

中文翻译:

我也是这么想的。Edwin，这太精彩了。这是一次非常棒的对话，我学到了很多。我想我们会帮助很多人创办自己的公司，帮助他们的公司更好地与价值观保持一致，并构建更好的东西。最后几个问题：如果大家想联系你，可以在哪里找到你？你们在招聘哪些职位？听众可以如何帮到你？

[01:09:00] Edwin Chen

English:

Yeah, so I used to love writing a blog, but I haven't had time in the past few years. But I am starting to write again, so definitely check out the Surge blog, surgehq.ai/blog, and yeah, hopefully I'll be running a lot more there. And I would say we're definitely always hiring, so for people who just love data and people who love this intersection of math, and language, and computer science, definitely reach out anytime.

中文翻译:

是的，我以前很喜欢写博客，但过去几年一直没时间。但我现在开始重新动笔了，所以请务必关注 Surge 的博客：surgehq.ai/blog，希望我以后能在那里写更多内容。我想说我们一直在招聘，所以对于那些热爱数据、热爱数学、语言和计算机科学交叉领域的人，欢迎随时联系。

[01:09:24] Lenny Rachitsky

English:

Awesome. And how can listeners be useful to you? Is it just, I don't know, yeah, is there anything there? Any asks?

中文翻译:

太棒了。听众可以如何帮到你？有什么要求吗？

[01:09:29] Edwin Chen

English:

So I would say definitely tell me blog topics that you like me to write about... and then I'm always fascinated by all of these AI failures that happen in the real world. So whenever you come across a really interesting failure that I think illustrates some deep question about how we want model to behave, there's just so many different ways a model can respond, I just oftentimes think there's just not a single right answer. And so whenever there's one of these examples, I just love seeing them.

中文翻译:

我想说，请务必告诉我你们想看 I 写什么样的博客主题。另外，我一直对现实世界中发生的各种 AI 失败案例很着迷。所以每当你遇到一个非常有趣的失败案例，且它能说明关于我们希望模型如何表现的深刻问题时，请告诉我。模型回应的方式有很多种，我经常觉得并没有唯一的正确答案。所以每当有这类例子，我都很喜欢看。

[01:09:57] Lenny Rachitsky**English:**

You need to share these on your blog. I'm also... I would love to see these. Edwin, thank you so much for being here.

中文翻译:

你应该在博客上分享这些。我也……我很想看这些。Edwin，非常感谢你能来。

[01:10:03] Edwin Chen**English:**

Thank you.

中文翻译:

谢谢。

[01:10:04] Lenny Rachitsky**English:**

Bye everyone. Thank you so much for listening. If you found this valuable, you can subscribe to the show on Apple Podcasts, Spotify, or your favorite podcast app. Also, please consider giving us a rating or leaving a review as that really helps other listeners find the podcast. You can find all past episodes or learn more about the show at lennyspodcast.com. See you in the next episode.

中文翻译:

大家再见。非常感谢收听。如果你觉得这期节目有价值，可以在 Apple Podcasts、Spotify 或你喜欢的播客应用上订阅。另外，请考虑给我们评分或留下评论，这能真正帮助其他听众发现这个播客。你可以在 lennyspodcast.com 找到所有往期节目或了解更多信息。下期节目见。