

# CHIP HUYEN

LENNY'S PODCAST

BILINGUAL TRANSCRIPT

---

ORIGINAL BY

Lenny Rachitsky

@lennysan • x.com/lennysan

ANALYSIS BY

@Penny777 • x.com/penny777

## Chip Huyen - 双语对照

### Lenny's Podcast: Chip Huyen | Bilingual Transcript

播客: Lenny's Podcast | 嘉宾: Chip Huyen

---

[00:00:00] Chip Huyen

English:

A question that get asked a lot and a lot is, "How do we keep up to date with the latest AI news?" Why do you need to keep up to date with the latest AI news? If you talk to the users who understand what they want or they don't want, look into the feedback, then you can actually improve the application way, way, way more.

中文翻译:

我经常被问到一个问题：“我们该如何紧跟最新的 AI 新闻？”但我想问，你为什么需要紧跟最新的 AI 新闻？如果你去和用户交流，了解他们的需求和痛点，深入研究反馈，你对应用的改进效果其实会比关注新闻要好得多。

---

[00:00:15] Lenny Rachitsky

English:

A lot of companies are building AI products. A lot of companies are not having a good time building AI products.

中文翻译:

现在很多公司都在开发 AI 产品，但也有很多公司在开发过程中举步维艰，进展并不顺利。

---

[00:00:19] Chip Huyen

English:

We are in an ideal crisis. Now, we have all this really cool tools to do everything from scratch and have new design. It can have you write code. You can have new website. So in theory, we should see a lot more, but at the same time, people are somehow stuck. They don't know what to build.

中文翻译:

我们正处于一场“创意危机”中。现在，我们拥有各种酷炫的工具，可以从零开始完成任何事情，实现全新的设计。它可以帮你写代码，帮你建网站。理论上，我们应该看到更多创新，但与此同时，人们却陷入了某种停滞。他们不知道该构建什么。

---

## [00:00:33] Lenny Rachitsky

### English:

All this AI hype, the data is actually showing most companies try it, doesn't do a lot. They stop. What do you think is the gap here?

### 中文翻译:

尽管 AI 炒作得沸沸扬扬，但数据表明，大多数公司尝试之后发现效果不佳，然后就停止了。你认为这其中的差距（Gap）在哪里？

---

## [00:00:38] Chip Huyen

### English:

It's really hard to measure productivity. So, I do ask people to ask their managers, "Would you rather give everyone on the team very expensive coding agent subscriptions or you get an extra head count?" Almost every one, the managers will say head count. But if you ask VP level or someone who manage a lot of teams, they would say, "Want AI assistant." Because as managers, you are still growing, so for you having one HR head count is big. Whereas for executives, maybe you have more business metrics that you care about. So you actually think about what actually drive productivity metrics for you.

### 中文翻译:

衡量生产力确实很难。我会让人们去问他们的经理：“你宁愿给团队里的每个人都买昂贵的 AI 编程助手（Coding Agent）订阅，还是想要一个额外的人头指标（Headcount）？”几乎所有的经理都会选人头。但如果你去问副总裁（VP）级别或者管理很多团队的人，他们会说：“想要 AI 助手。”因为作为基层经理，你还在成长阶段，增加一个正式员工对你来说意义重大；而对于高管来说，他们更关注业务指标，会思考究竟什么才能真正驱动生产力指标。

---

## [00:01:11] Lenny Rachitsky

### English:

Today, my guest is Chip Huyen. Unlike a lot of people who share insights into building great AI products and where things are heading, Chip has built multiple successful AI products, platforms, tools. Chip was a core developer on NVIDIA's NeMo platform, an AI researcher at Netflix. She taught machine learning at Stanford. She's also a two-time founder and the author of two of the most popular books in the world of AI, including her most recent book called *AI Engineering*, which has been the most read book on the O'Reilly platform since its launch.

### 中文翻译:

今天的嘉宾是 Chip Huyen。与许多只分享 AI 见解的人不同，Chip 亲手构建过多个成功的 AI 产品、平台和工具。她曾是 NVIDIA NeMo 平台的内核开发者，也曾是 Netflix 的 AI 研究员。她在斯坦福大学教授过机器学习课程。她还是两次创业的创始人，并撰写了两本 AI 领域最受欢迎的著作，其中包括她最近出版的《AI 工程化》（*AI Engineering*），该书自发布以来一直是 O'Reilly 平台上阅读量最高的书籍。

---

## [00:01:41] Lenny Rachitsky

### English:

She's also gotten to work with a lot of enterprises on their AI strategies, and so she gets to see what's actually happening on the ground inside a lot of different companies. In our conversation, Chip explains a lot of the basics like, what exactly does pre-training and post-training look like? What is RAG? What is reinforcement learning? What is RLHF? We also get into everything she's learned about how to build great AI products, including what people think it takes and what it actually takes.

### 中文翻译:

她还与许多企业合作制定 AI 战略，因此她能洞察到许多公司内部最真实的落地情况。在我们的对话中，Chip 解释了许多基础概念，例如：预训练（Pre-training）和后训练（Post-training）究竟是什么样的？什么是 RAG（检索增强生成）？什么是强化学习（Reinforcement Learning）？什么是 RLHF（基于人类反馈的强化学习）？我们还深入探讨了她在构建优秀 AI 产品方面的所有心得，包括人们“以为”需要什么，以及“实际上”需要什么。

---

## [00:04:40] Lenny Rachitsky

### English:

I want to start with this table/chart that you shared on LinkedIn a while ago that went super viral... It's this very simple table you shared of what people think will improve AI apps and what actually improves AI apps. What people think will improve AI apps: staying up to date with the latest AI news, adopting the newest agentic framework, agonizing about what vector databases to use, constantly evaluating what model is smarter, fine-tuning a model. And then you have what actually improves AI apps: talking to users, building more reliable platforms, preparing better data, optimizing end-to-end workflows, writing better prompts. Why do you think this hit such a nerve with people?

### 中文翻译:

我想从你不久前在 LinkedIn 上分享的一张火爆的图表开始谈起……这张表非常简洁，对比了“人们认为能改进 AI 应用的事”和“实际能改进 AI 应用的事”。人们认为有用的是：紧跟最新 AI 新闻、采用最新的智能体框架（Agentic Framework）、纠结用哪个向量数据库（Vector Database）、不断评估哪个模型更聪明、微调模型（Fine-tuning）。而实际有用的是：与用户交流、构建更可靠的平台、准备更高质量的数据、优化端到端的工作流、编写更好的提示词（Prompts）。你觉得为什么这张图会引起这么多人的共鸣？

---

## [00:05:30] Chip Huyen

### English:

A question that get asked a lot and a lot is that, "How do we keep up to date with the latest AI news?" I'm like, "Why? Why do you need to keep up to date with the latest AI news?" I know it sound very counter-intuitive, but there's just so much news out there... I think it's a question you should ask them is like, "First, how much of the improvement could you get from optimal solutions versus non-optimal solutions?" Right? And sometimes they were like, "Actually, it's not much."

### 中文翻译:

我被问得最多的问题就是：“我们如何紧跟最新的 AI 新闻？”我的反应是：“为什么？你为什么非得紧跟这些新闻？”我知道这听起来很反直觉，但现在的新闻实在太多了。我觉得你应该问他们：“首先，最优方案和次优方案相比，到底能带来多少提升？”有时候他们会发现：“其实差别并不大。”

---

[00:07:34] Chip Huyen

**English:**

So I think it's like one is, supervised fine-tuning when you have demonstration data, and you have a bunch of experts, "Okay, here's a prompt, and here is what the answer should be like." You just train it to emulate what the human expert could be like. That's also what a lot of people would like, so open-source models are doing as they do it by distillation. So instead of having human experts to write really great answers to prompts, they get very popular, famous good models to generate a response to it and getting this train smaller models to emulate.

**中文翻译:**

我认为（后训练）主要分为几种：一种是有监督微调（SFT），当你拥有演示数据和专家时，“给出一个提示词，专家给出标准答案”。你训练模型去模仿人类专家的表现。这也是很多开源模型在做的事情，他们通过“蒸馏”（Distillation）来实现。也就是说，不再让专家去写答案，而是让那些非常流行、强大的模型生成答案，然后训练较小的模型去模仿这些大模型的表现。

---

[00:09:35] Chip Huyen

**English:**

So, I think of language modeling as a way of encoding statistical information about language, right? So, let's say that we both speak English, so we get a sense of what is more statistically likely. If I say my favorite color is, then you would say, "Okay, that should be another color." The word blue would be much more likely to appear than the word like [inaudible], right? Because statistically, blue is more likely to [follow] my favorite color is. So, it's a way of encoding statistical information.

**中文翻译:**

我把语言建模看作是一种对语言统计信息进行编码的方式。假设我们都说英语，我们能感觉到什么词在统计学上更可能出现。如果说“我最喜欢的颜色是”，你会觉得接下来的词应该是一种颜色。单词“蓝色”（blue）出现的概率会比其他无关词汇高得多。因为从统计学上讲，“蓝色”更有可能跟在“我最喜欢的颜色是”后面。所以，这就是一种编码统计信息的方式。

---

[00:14:06] Chip Huyen

**English:**

So the vast majority of time, we don't touch on pre-training model. As users, we don't use it at all... It's very interesting to look at how much of post-training can change the model behavior and I think that's where a lot of time, is a lot of people are spending energy on nowadays, their frontier lab, is on post-training. Because pre-training... needs a lot of data and model size to increase the model capabilities. And at some point, we are actually have kind of maxed out on the internet data.

**中文翻译:**

绝大多数时候，我们不会去碰预训练模型。作为用户，我们根本不直接使用它。观察“后训练”（Post-training）能在多大程度上改变模型行为是非常有趣的，我认为这也是现在许多顶尖实验室投入大量精力的地方。因为预训练需要海量数据和巨大的模型规模来提升能力，而在某种程度上，我们已经快把互联网上的文本数据用光了。

---

[00:15:57] Chip Huyen

## English:

So the idea is that once you... have a model, give the model a prompt and it produce an output. You want to reinforce, encourage the model to produce an output that is better. So now it comes to how do we know that the answer is good or bad? So usually, people relies on signals. One way to get a first one good or bad is human feedback... As humans, we tend to, it's very hard to give a concrete score, but it's easier to do comparisons.

## 中文翻译:

核心思路是：当你有一个模型，给它一个提示词，它会产生输出。你想要强化（Reinforce）并鼓励模型产生更好的输出。那么问题来了，我们怎么知道答案的好坏？通常人们依赖信号。一种方式是人类反馈。作为人类，我们很难给出一个具体的评分，但做比较（Comparisons）却很容易。