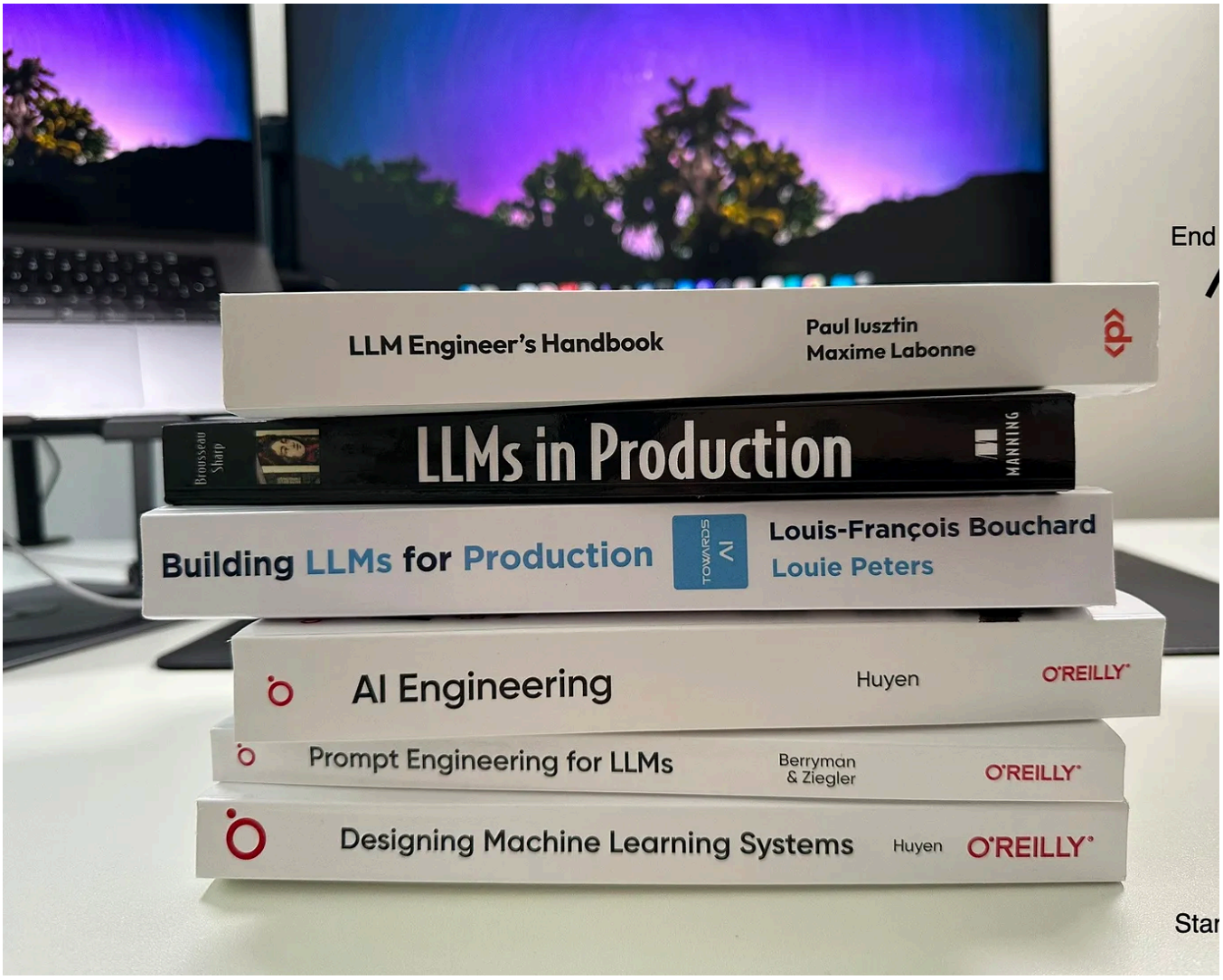
5 Books to Ship AI Products in 2025

That helped me build LLM and agentic systems

PAUL IUSZTIN
JUL 03, 2025

In today's article, I want you to present my top 5 favorite books (that I actually read getting into AI Engineering in 2025, plus a bonus at the end.



I will start with more general books on fundamentals and gradually add more niche content on AI Engineering as we progress through the list. Thus, if you are unsure

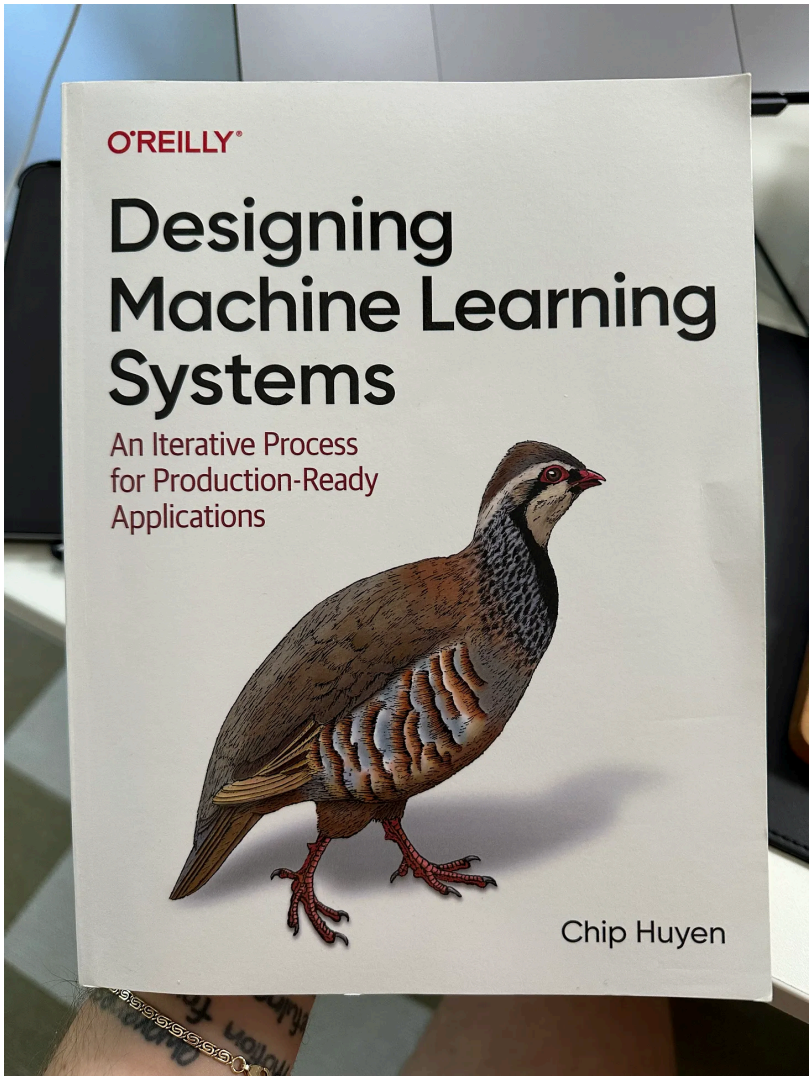
the order in which to read them. You can follow my list from 1 to 5, in that order.

1. Designing Machine Learning Systems

Designing Machine Learning Systems by Chip Huyen is the best book I have read far, which is both beginner-friendly and provides a strong intuition into what an en to-end ML system looks like.

To me, it was an eye-opener. It made me realize that an ML system is more than ju code. It provided me with a bird's-eye view of what it takes to build a production- grade ML system, touching aspects such as:

- Architecting ML pipelines
- Managing models and datasets
- MLOps & Infrastructure



But wait, this book is about ML Systems, not AI Systems. Well... It doesn't matter. book was written before the AI hype, when “ML” was the broad umbrella of the domain. Thus, most of the principles still apply, even if it doesn’t talk about LLMs.

2. Prompt Engineering for LLMs

Prompt Engineering for LLMs by John Berryman and Albert Ziegler (one of the co engineers who developed GitHub Copilot) is the best book on prompt engineering have read so far. As I am obsessed with seeing the big picture and thinking in syst I was fascinated by how they presented prompt engineering in this book.

They weren’t limited to concrete examples of how to write prompts for X, Y, Z, or s on well-known techniques such as thinking step-by-step or few-shot learning.

Instead, they presented how to effectively engineer your prompts by considering t context you add to the prompt as packets of knowledge and discussing how to manipulate it to make it flexible, scalable, and squeeze the most out of your mode

P.S. The book is written in such a funny way that it's entertaining to read. 100% recommend it.

3. AI Engineering

The 3rd book on fundamentals that I recommend is *AI Engineering* by (yet again) Chip Huyen. Similar to her first book, this is a fantastic resource on providing a bir eye view of what AI Engineering is all about.

In my opinion, it has one of the best and most clear explanations on:

- AI vs. ML Engineering

- RAG (from classic text search to more complex, semantic-based systems)
- Building agentic systems (guardrails, gateways, caches, storages, memory... a the good stuff interconnected into a single system)
- LLMOps (Especially observability and user feedback)

This book is a no-brainer if you want to level up in AI Engineering.

4. Building LLMs for Production

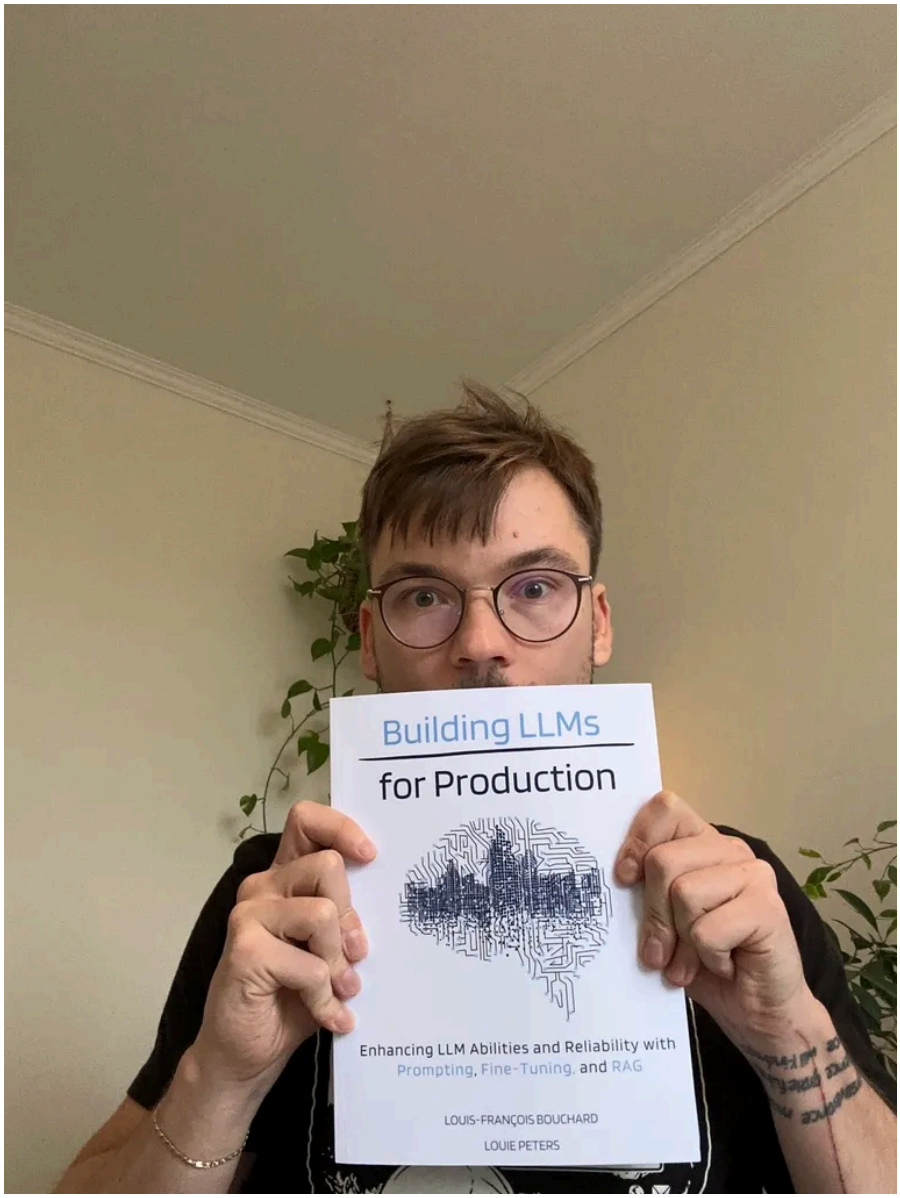
Now, let's get into more hands-on books that actually teach you how to build stuff

An excellent entry-to-mid-level book is *Building LLMs for Production* by Louis-François Bouchard and Louie Peters (founders of Towards AI). Throughout t

book, they will walk you through how to implement all the core algorithms required build LLM applications.

Using LangChain and LlamaIndex, they show you how to implement all kinds of RA techniques, starting from the vanilla architecture to more advanced techniques su as GraphRAG.

However, there is more to it than just RAG; they tackle everything, from prompt engineering and LLMs to agents, fine-tuning, and the core of deploying LLM apps



Suppose you're overwhelmed by all the frameworks and tools out there. This book great way to start!

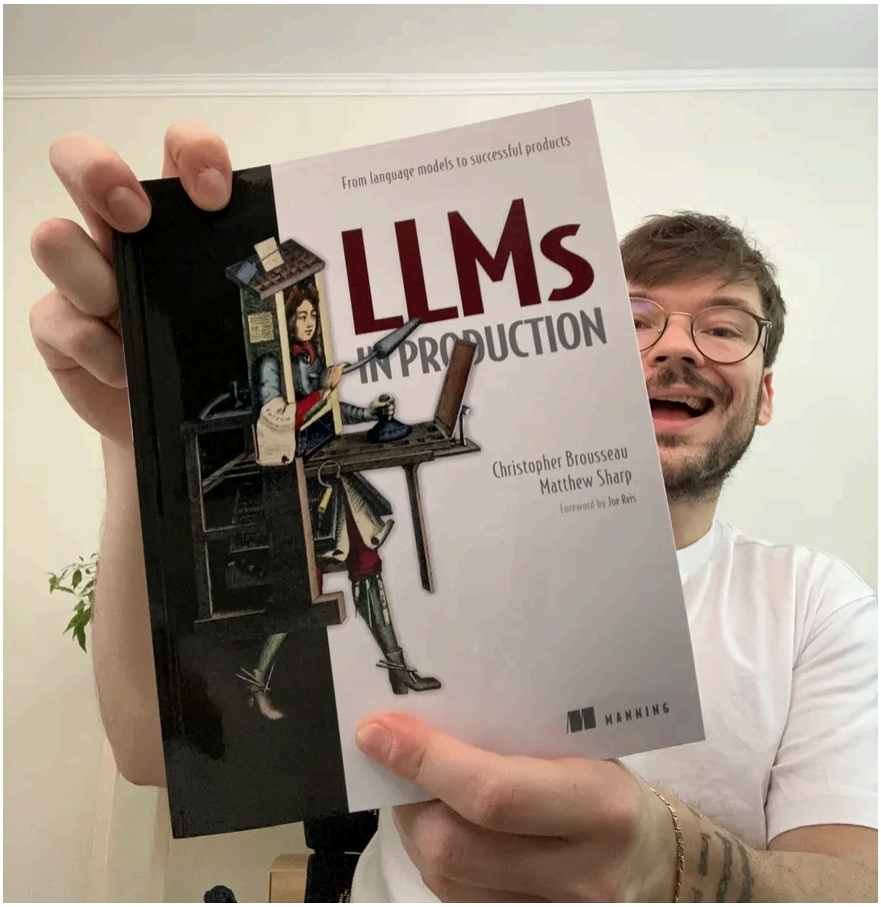
> As I've partnered with the Towards AI team (on our future agents course and m
> I can offer you on this one **15% off** using my affiliate code Paul_15.

5. LLMs in Production

The last book on my list is *LLMs in Production* by Christopher Brousseau and Matt
Sharp.

This is a fantastic book focused entirely on shipping LLMs, well... to production! T
do a great job of explaining:

- How to optimize your LLMs using different tools such as ONNX, TensorRT and
  llama.cpp.
- What's data engineering for building an LLM app.
- Training and serving LLMs at scale.
- How to think about infrastructure when serving LLMs (disk, cold-starts, memo
  etc.).
- A fun capstone project where they deploy an LLM on a Raspberry Pi and hook
  a chat UI, having your own self-hosted LLM to chat with.



This is a highly underrated book!

Bonus. LLM Engineer's Handbook

I couldn't help but add my book to this list. After all, it's my little baby. Hehe 🥹

Jokes aside, the *LLM Engineer's Handbook* by Paul Iusztin and Maxime Labonne would be on this list anyway.

Because of my obsession with thinking in end-to-end systems and Maxime's for fi tuning, during the book, we walk you through building a single, but complex, LLM RAG application, walking you through all the steps you will find in the industry:

- Architecting the LLM & RAG system
- Collecting data
- Building fine-tuning and RAG pipelines, and datasets
- Fine-tuning and evaluating LLMs

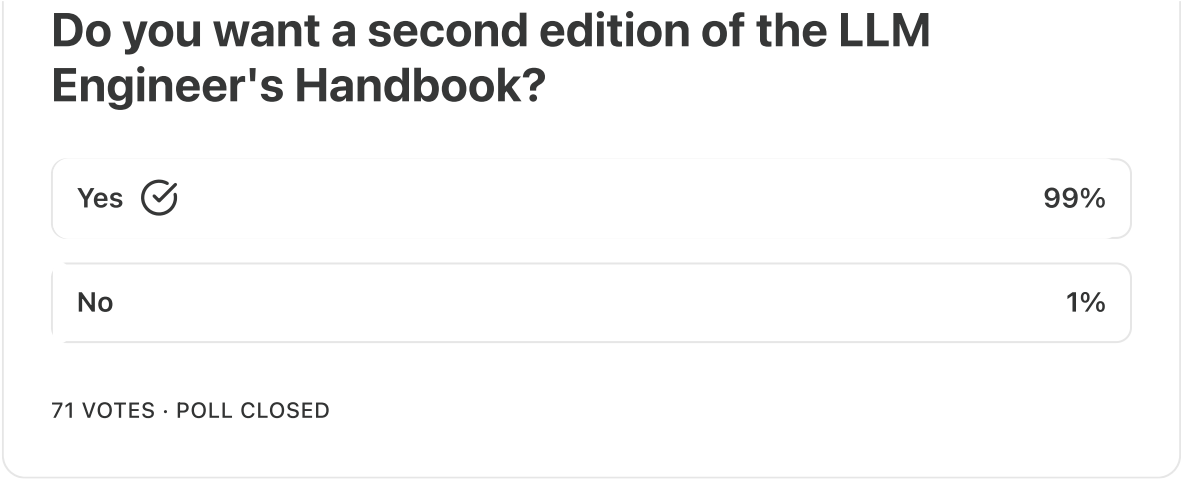
- Deploying and scaling LLMs as REST APIs
- LLMOps (orchestration, data versioning, CI/CD pipelines, Docker, monitoring,

Even if we don't touch anything on agents, I still think this is a relevant book on teaching you how to think about when implementing end-to-end AI systems.

P.S. Maxime and I are considering a second edition of this book, where we aim to update it with the latest fine-tuning and agent techniques.

👉 The same fundamental principle: implementing an end-to-end project. But nov with reasoning models and agents. Would that interest you?

POLL



Final Thoughts

These are my favorite books and recommendations to get into AI Engineering in 2

Let me know in the comments which one you read, plan to read, liked or disliked. Would love to hear your thoughts ✌️ ↓

Whenever you're ready, there are 3 ways we can help you:

1. **Perks:** Exclusive discounts on our recommended learning resources (books, live courses, self-paced courses and learning platforms).
2. **The LLM Engineer's Handbook:** Our bestseller book on teaching you an enc end framework for building production-ready LLM and RAG applications, from data collection to deployment (get up to 20% off using our discount code).
3. **Free open-source courses:** Master production AI with our end-to-end open- source courses, which reflect real-world AI projects and cover everything fror system architecture to data collection, training and deployment.

Images