"This book offers a comprehensive, well-structured guide to the essential aspects of building generative AI systems. A must-read for any professional looking to scale AI across the enterprise."

Vittorio Cretella, former global CIO at P&G and Mars

"Chip Huyen gets generative AI. She is a remarkable teacher and writer whose work has been instrumental in helping teams bring AI into production. Drawing on her deep expertise, *AI Engineering* is a comprehensive and holistic guide to building generative AI applications in production."

Luke Metz, cocreator of ChatGPT, former research manager at OpenAI

AI Engineering

Foundation models have enabled many new AI use cases while lowering the barriers to entry for building AI products. This has transformed AI from an esoteric discipline into a powerful development tool that anyone can use—including those with no prior AI experience.

In this accessible guide, author Chip Huyen discusses AI engineering: the process of building applications with readily available foundation models. AI application developers will discover how to navigate the AI landscape, including models, datasets, evaluation benchmarks, and the seemingly infinite number of application patterns. The book also introduces a practical framework for developing an AI application and efficiently deploying it.

- **Understand what AI engineering is and how it differs from traditional machine learning engineering**
- **Learn the process for developing an AI application, the challenges at each step, and approaches to address them**
- **Explore various model adaptation techniques, including prompt engineering, RAG, finetuning, agents, and dataset engineering, and understand how and why they work**
- **Examine the bottlenecks for latency and cost when serving foundation models and learn how to overcome them**
- **Choose the right model, metrics, data, and developmental patterns for your needs**

Chip Huyen works at the intersection of AI, data, and storytelling. Previously, she was with Snorkel AI and NVIDIA, founded an AI infrastructure startup (acquired), and taught machine learning systems design at Stanford. Her book *Designing Machine Learning Systems* (O'Reilly) has been translated into over 10 languages.

DATA

US $79.99 CAN $99.99
ISBN: 978-1-098-16630-4

O'REILLY®