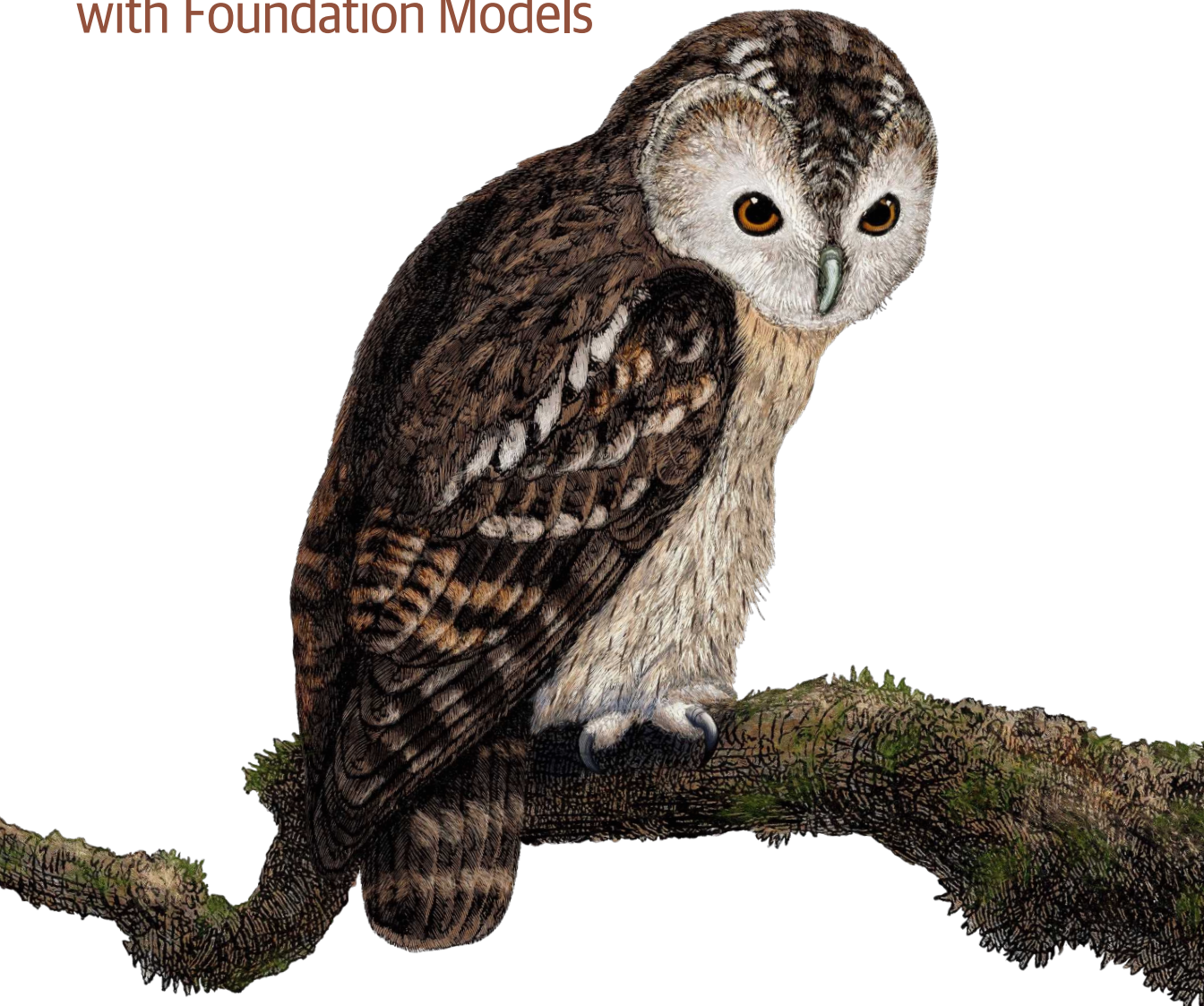


O'REILLY®

AI Engineering

Building Applications
with Foundation Models



Chip Huyen

"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



5 7 9 9 9



O'REILLY®

Praise for *AI Engineering*

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, P&G and Mars

Chip Huyen gets generative AI. On top of that, 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* serves as a comprehensive and holistic guide, masterfully detailing everything required to design and deploy generative AI applications in production.

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

Every AI engineer building real-world applications should read this book. It's a vital guide to end-to-end AI system design, from model development and evaluation to large-scale deployment and operation.

—Andrei Lopatenko, Director Search and AI, Neuron7

This book serves as an essential guide for building AI products that can scale.

Unlike other books that focus on tools or current trends that are constantly changing, Chip delivers timeless foundational knowledge. Whether you're a product manager or an engineer, this book effectively bridges the collaboration gap between cross-functional teams, making it a must-read for anyone involved in AI development.

—Aileen Bui, AI Product Operations Manager, Google

This is the definitive segue into AI engineering from one of the greats of ML engineering! Chip has seen through successful projects and careers at every stage of a company and for the first time ever condensed her expertise for new AI Engineers entering the field.

—swyx, *Curator, AI.Engineer*

AI Engineering is a practical guide that provides the most up-to-date information on AI development, making it approachable for novice and expert leaders alike. This book is an essential resource for anyone looking to build robust and scalable AI systems.

—Vicki Reyzelman, *Chief AI Solutions Architect, Mave Sparks*

AI Engineering is a comprehensive guide that serves as an essential reference for both understanding and implementing AI systems in practice.

—Han Lee, *Director—Data Science, Moody's*

AI Engineering is an essential guide for anyone building software with Generative AI! It demystifies the technology, highlights the importance of evaluation, and shares what should be done to achieve quality before starting with costly fine-tuning.

—Rafal Kawala, *Senior AI Engineering Director, 16 years of experience working in a Fortune 500 company*

AI Engineering

Building Applications with Foundation Models

Chip Huyen

O'REILLY®