

# TABLE OF CONTENT

Preface.....	3
Chapter: 1. Welcome to the Linguistic Leap .....	8
1.1 Why Language Models Matter Today .....	8
1.2 What “Large” Really Means (and Why It’s Changing).....	12
1.3 From Tokens to Text: A Gentle Intuition .....	16
1.4 What This Book Covers (and What It Skips) .....	22
Chapter: 2. The Building Blocks (Without the Jargon) .....	26
2.1 Tokens, Embeddings, and Vectors Plainly Explained.....	26
2.2 Attention: The “Look Here” Mechanism .....	31
2.3 Encoder vs. Decoder vs. Encoder-Decoder.....	35
2.4 How Size, Data, and Compute Shape Behavior .....	40
Chapter 3 Using LLMs Safely and Effectively.....	45
3.1 Prompting Basics: Roles, Context, Constraints .....	45
3.2 Guardrails: Input/Output Validation & Safety Checks.....	50
3.3 Avoiding Prompt Injection and Data Leaks.....	54
3.4 Human-in-the-Loop: When and How to Intervene .....	60
Chapter: 4. Everyday Applications (Hands-On).....	65
4.1 Summarization and Note-Making.....	65
4.2 Q&A and Retrieval with Your Own Files .....	70
4.3 Drafting, Rewriting, and Style Transfer .....	74
4.4 Brainstorming, Outlining, and Research Assist.....	78
Chapter: 5. Retrieval-Augmented Generation (RAG) .....	84
5.1 Why RAG Beats “Raw” Model Memory .....	84
5.2 Chunking, Indexing, and Embeddings in Practice.....	88
5.3 Query, Rerank, and Response Assembly.....	92
5.4 Pitfalls: Hallucinations, Coverage, and Drift.....	96
Chapter: 6. Evaluation You Can Trust.....	101

6.1 Defining Success: Quality, Safety, and Cost .....	101
6.2 Golden Sets, Rubrics, and Rubric-Guided LLM Judges.....	106
6.3 Measuring Hallucination, Bias, and Toxicity .....	111
6.4 Iterating with A/B Tests and Telemetry.....	116
Chapter: 7. Fine-Tuning and Customization (Lightweight First) .....	121
7.1 When to Choose Prompts, Tools, or RAG over Tuning .....	121
7.2 Instruction Tuning vs. Preference Optimization (Overview) .....	124
7.3 LoRA/Adapters: Practical Customization on a Budget .....	127
7.4 Data Curation: Small, Clean, and Task-True.....	132
Chapter: 8. Behind the Curtain (For the Curious).....	138
8.1 Pretraining: Objective, Data, and Scaling Laws (High-Level).....	138
8.2 Optimization: From Gradient Steps to Checkpoints.....	141
8.3 Inference: Sampling, Temperature, and Decoding .....	144
8.4 Why Bigger Isn't Always Better.....	148
Chapter: 9. Ethics, Privacy, and Compliance .....	151
9.1 Where Data Comes From and Why It Matters .....	151
9.2 Privacy-Preserving Patterns (PII handling, Redaction) .....	154
9.3 Bias, Fairness, and Harms: Practical Mitigations .....	157
9.4 Governance: Policies, Audits, and Documentation .....	161
Chapter: 10. Working with Open and Hosted Models.....	165
10.1 Choosing a Model: Capability, Cost, Constraints.....	165
10.2 Open Models: Pros/Cons, Ecosystem, and Deployment .....	168
10.3 Hosted APIs: Reliability, Rate Limits, and SLAs.....	171
10.4 Portability: Avoiding Vendor Lock-In.....	174
Chapter: 11. Teams, Tools, and Workflows .....	178
11.1 From Prototype to Product: A Minimal Stack .....	178
11.2 Prompt Libraries, Evaluation Loops, and Versioning .....	180
11.3 Observability: Traces, Costs, and Guardrail Logs .....	183
11.4 Collaboration: Roles for PMs, Engineers, and Domain Experts ....	186

Chapter: 12. The Road Ahead (Staying Current, Not Overwhelmed).....	190
12.1 Multimodality: Text + Images + Audio.....	190
12.2 Agents and Tool Use: Promise vs. Reality .....	193
12.3 Model Upgrades Without Breaking Apps.....	196
12.4 A Learning Plan to Keep Your Skills Fresh .....	198
Appendices.....	202