
Table of Contents

Preface.....	xi
--------------	----

Part I. Leveraging Open Models

1. An Introduction to Generative Media.....	3
Generating Images	4
Generating Text	7
Generating Sound Clips	8
Ethical and Societal Implications	8
Where We've Been and Where Things Stand	9
How Are Generative AI Models Created?	10
Summary	12
2. Transformers.....	13
A Language Model in Action	14
Tokenizing Text	14
Predicting Probabilities	17
Generating Text	20
Zero-Shot Generalization	28
Few-Shot Generalization	30
A Transformer Block	33
Transformer Model Genealogy	35
Sequence-to-Sequence Tasks	35
Encoder-Only Models	37
The Power of Pretraining	40
Transformers Recap	43
Limitations	45

Beyond Text	46
Project Time: Using LMs to Generate Text	50
Summary	50
Exercises	52
Challenges	52
References	53
3. Compressing and Representing Information.....	55
AutoEncoders	57
Preparing the Data	57
Modeling the Encoder	61
Decoder	64
Training	65
Exploring the Latent Space	70
Visualizing the Latent Space	74
Variational AutoEncoders	78
VAE Encoders and Decoders	79
Sampling from the Encoder Distribution	80
Training the VAE	83
VAEs for Generative Modeling	93
CLIP	93
Contrastive Loss	94
Using CLIP, Step-by-Step	95
Zero-Shot Image Classification with CLIP	101
Zero-Shot Image-Classification Pipeline	102
CLIP Use Cases	103
Alternatives to CLIP	104
Project Time: Semantic Image Search	105
Summary	106
Exercises	108
Challenges	108
References	109
4. Diffusion Models.....	111
The Key Insight: Iterative Refinement	112
Training a Diffusion Model	115
The Data	116
Adding Noise	118
The UNet	119
Training	121
Sampling	123
Evaluation	124

In Depth: Noise Schedules	126
Why Add Noise?	126
Starting Simple	128
The Math	130
Effect of Input Resolution and Scaling	135
In Depth: UNets and Alternatives	137
A Simple UNet	138
Improving the UNet	141
Alternative Architectures	142
In Depth: Diffusion Objectives	143
Project Time: Train Your Diffusion Model	145
Summary	145
Exercises	146
Challenges	147
References	148
5. Stable Diffusion and Conditional Generation.....	149
Adding Control: Conditional Diffusion Models	149
Preparing the Data	150
Creating a Class-Conditioned Model	152
Training the Model	153
Sampling	157
Improving Efficiency: Latent Diffusion	159
Stable Diffusion: Components in Depth	160
The Text Encoder	161
The Variational AutoEncoder	164
The UNet	167
Stable Diffusion XL	169
FLUX, SD3, and Video	171
Classifier-Free Guidance	172
Putting It All Together: Annotated Sampling Loop	173
Open Data, Open Models	176
Challenges and the Sunset of LAION-5B	177
Alternatives	178
Fair and Commercial Use	178
Project Time: Build an Interactive ML Demo with Gradio	179
Summary	180
Exercises	181
Challenge	181
References	182

Part II. Transfer Learning for Generative Models

6. Fine-Tuning Language Models.....	185
Classifying Text	186
Identify a Dataset	187
Define Which Model Type to Use	189
Select a Good Base Model	190
Preprocess the Dataset	190
Define Evaluation Metrics	192
Train the Model	194
Still Relevant?	202
Generating Text	203
Picking the Right Generative Model	204
Training a Generative Model	208
Instructions	212
A Quick Introduction to Adapters	216
A Light Introduction to Quantization	221
Putting It All Together	225
A Deeper Dive into Evaluation	230
Project Time: Retrieval-Augmented Generation	233
Summary	235
Exercises	236
Challenge	236
References	237
 7. Fine-Tuning Stable Diffusion.....	 239
Full Stable Diffusion Fine-Tuning	239
Preparing the Dataset	240
Fine-Tuning the Model	242
Inference	246
DreamBooth	248
Preparing the Dataset	250
Prior Preservation	250
DreamBooth the Model	251
Inference	252
Training LoRAs	253
Giving Stable Diffusion New Capabilities	256
Inpainting	256
Additional Inputs for Special Conditionings	256
Project Time: Train an SDXL DreamBooth LoRA by Yourself	257
Summary	258

Exercises	259
Challenge	259
References	260

Part III. Going Further

8. Creative Applications of Text-to-Image Models.....	263
Image to Image	263
Inpainting	265
Prompt Weighting and Image Editing	267
Prompt Weighting and Merging	268
Editing Diffusion Images with Semantic Guidance	270
Real Image Editing via Inversion	274
Editing with LEDITS++	276
Real Image Editing via Instruction Fine-Tuning	277
ControlNet	279
Image Prompting and Image Variations	283
Image Variations	283
Image Prompting	285
Project Time: Your Creative Canvas	288
Summary	289
Exercises	289
References	290
9. Generating Audio.....	291
Audio Data	293
Waveforms	297
Spectrograms	298
Speech to Text with Transformer-Based Architectures	307
Encoder-Based Techniques	308
Encoder-Decoder Techniques	312
From Model to Pipeline	315
Evaluation	318
From Text to Speech to Generative Audio	324
Generating Audio with Sequence-to-Sequence Models	324
Going Beyond Speech with Bark	329
AudioLM and MusicLM	332
AudioGen and MusicGen	335
Audio Diffusion and Riffusion	336
Dance Diffusion	339
More on Diffusion Models for Generative Audio	340

Evaluating Audio-Generation Systems	340
What's Next?	341
Project Time: End-to-End Conversational System	342
Summary	342
Exercises	345
Challenges	345
References	346
10. Rapidly Advancing Areas in Generative AI	349
Preference Optimization	349
Long Contexts	352
Mixture of Experts	354
Optimizations and Quantizations	356
Data	358
One Model to Rule Them All	359
Computer Vision	360
3D Computer Vision	362
Video Generation	363
Multimodality	365
Community	367
A. Open Source Tools	369
B. LLM Memory Requirements	373
C. End-to-End Retrieval-Augmented Generation	377
Index	385