

Table of Contents

Foreword.....	xv
---------------	----

Preface.....	xvii
--------------	------

Part I. Introduction to Generative Deep Learning

1. Generative Modeling.....	3
------------------------------------	----------

What Is Generative Modeling?	4
------------------------------	---

Generative Versus Discriminative Modeling	5
---	---

The Rise of Generative Modeling	6
---------------------------------	---

Generative Modeling and AI	8
----------------------------	---

Our First Generative Model	9
----------------------------	---

Hello World!	9
--------------	---

The Generative Modeling Framework	10
-----------------------------------	----

Representation Learning	12
-------------------------	----

Core Probability Theory	15
-------------------------	----

Generative Model Taxonomy	18
---------------------------	----

The Generative Deep Learning Codebase	20
---------------------------------------	----

Cloning the Repository	20
------------------------	----

Using Docker	21
--------------	----

Running on a GPU	21
------------------	----

Summary	21
---------	----

2. Deep Learning.....	23
------------------------------	-----------

Data for Deep Learning	24
------------------------	----

Deep Neural Networks	25
----------------------	----

What Is a Neural Network?	25
Learning High-Level Features	26
TensorFlow and Keras	27
Multilayer Perceptron (MLP)	28
Preparing the Data	28
Building the Model	30
Compiling the Model	35
Training the Model	37
Evaluating the Model	38
Convolutional Neural Network (CNN)	40
Convolutional Layers	41
Batch Normalization	46
Dropout	49
Building the CNN	51
Training and Evaluating the CNN	53
Summary	54

Part II. Methods

3. Variational Autoencoders.....	59
Introduction	60
Autoencoders	61
The Fashion-MNIST Dataset	62
The Autoencoder Architecture	63
The Encoder	64
The Decoder	65
Joining the Encoder to the Decoder	67
Reconstructing Images	69
Visualizing the Latent Space	70
Generating New Images	71
Variational Autoencoders	74
The Encoder	75
The Loss Function	80
Training the Variational Autoencoder	82
Analysis of the Variational Autoencoder	84
Exploring the Latent Space	85
The CelebA Dataset	85
Training the Variational Autoencoder	87
Analysis of the Variational Autoencoder	89
Generating New Faces	90

Latent Space Arithmetic	91
Morphing Between Faces	92
Summary	93
4. Generative Adversarial Networks.....	95
Introduction	96
Deep Convolutional GAN (DCGAN)	97
The Bricks Dataset	98
The Discriminator	99
The Generator	101
Training the DCGAN	104
Analysis of the DCGAN	109
GAN Training: Tips and Tricks	110
Wasserstein GAN with Gradient Penalty (WGAN-GP)	113
Wasserstein Loss	114
The Lipschitz Constraint	115
Enforcing the Lipschitz Constraint	116
The Gradient Penalty Loss	117
Training the WGAN-GP	119
Analysis of the WGAN-GP	121
Conditional GAN (CGAN)	122
CGAN Architecture	123
Training the CGAN	124
Analysis of the CGAN	126
Summary	127
5. Autoregressive Models.....	129
Introduction	130
Long Short-Term Memory Network (LSTM)	131
The Recipes Dataset	132
Working with Text Data	133
Tokenization	134
Creating the Training Set	137
The LSTM Architecture	138
The Embedding Layer	138
The LSTM Layer	140
The LSTM Cell	142
Training the LSTM	144
Analysis of the LSTM	146
Recurrent Neural Network (RNN) Extensions	149
Stacked Recurrent Networks	149

Gated Recurrent Units	151
Bidirectional Cells	153
PixelCNN	153
Masked Convolutional Layers	154
Residual Blocks	156
Training the PixelCNN	158
Analysis of the PixelCNN	159
Mixture Distributions	162
Summary	164
6. Normalizing Flow Models.....	167
Introduction	168
Normalizing Flows	169
Change of Variables	170
The Jacobian Determinant	172
The Change of Variables Equation	173
RealNVP	174
The Two Moons Dataset	174
Coupling Layers	175
Training the RealNVP Model	181
Analysis of the RealNVP Model	184
Other Normalizing Flow Models	186
GLOW	186
FFJORD	187
Summary	188
7. Energy-Based Models.....	189
Introduction	189
Energy-Based Models	191
The MNIST Dataset	192
The Energy Function	193
Sampling Using Langevin Dynamics	194
Training with Contrastive Divergence	197
Analysis of the Energy-Based Model	201
Other Energy-Based Models	202
Summary	203
8. Diffusion Models.....	205
Introduction	206
Denoising Diffusion Models (DDM)	208
The Flowers Dataset	208

The Forward Diffusion Process	209
The Reparameterization Trick	210
Diffusion Schedules	211
The Reverse Diffusion Process	214
The U-Net Denoising Model	217
Training the Diffusion Model	224
Sampling from the Denoising Diffusion Model	225
Analysis of the Diffusion Model	228
Summary	231

Part III. Applications

9. Transformers.....	235
Introduction	236
GPT	236
The Wine Reviews Dataset	237
Attention	238
Queries, Keys, and Values	239
Multihead Attention	241
Causal Masking	242
The Transformer Block	245
Positional Encoding	248
Training GPT	250
Analysis of GPT	252
Other Transformers	255
T5	256
GPT-3 and GPT-4	259
ChatGPT	260
Summary	264
10. Advanced GANs.....	267
Introduction	268
ProGAN	269
Progressive Training	269
Outputs	276
StyleGAN	277
The Mapping Network	278
The Synthesis Network	279
Outputs from StyleGAN	280
StyleGAN2	281

Weight Modulation and Demodulation	282
Path Length Regularization	283
No Progressive Growing	284
Outputs from StyleGAN2	286
Other Important GANs	286
Self-Attention GAN (SAGAN)	286
BigGAN	288
VQ-GAN	289
ViT VQ-GAN	292
Summary	294
11. Music Generation.....	297
Introduction	298
Transformers for Music Generation	299
The Bach Cello Suite Dataset	300
Parsing MIDI Files	300
Tokenization	303
Creating the Training Set	304
Sine Position Encoding	305
Multiple Inputs and Outputs	307
Analysis of the Music-Generating Transformer	309
Tokenization of Polyphonic Music	313
MuseGAN	317
The Bach Chorale Dataset	317
The MuseGAN Generator	320
The MuseGAN Critic	326
Analysis of the MuseGAN	327
Summary	329
12. World Models.....	331
Introduction	331
Reinforcement Learning	332
The CarRacing Environment	334
World Model Overview	336
Architecture	336
Training	338
Collecting Random Rollout Data	339
Training the VAE	341
The VAE Architecture	341
Exploring the VAE	343
Collecting Data to Train the MDN-RNN	346

Training the MDN-RNN	346
The MDN-RNN Architecture	347
Sampling from the MDN-RNN	348
Training the Controller	348
The Controller Architecture	349
CMA-ES	349
Parallelizing CMA-ES	351
In-Dream Training	353
Summary	356
13. Multimodal Models.....	359
Introduction	360
DALL.E 2	361
Architecture	362
The Text Encoder	362
CLIP	362
The Prior	367
The Decoder	369
Examples from DALL.E 2	373
Imagen	377
Architecture	377
DrawBench	378
Examples from Imagen	379
Stable Diffusion	380
Architecture	380
Examples from Stable Diffusion	381
Flamingo	381
Architecture	382
The Vision Encoder	382
The Perceiver Resampler	383
The Language Model	385
Examples from Flamingo	388
Summary	389
14. Conclusion.....	391
Timeline of Generative AI	392
2014–2017: The VAE and GAN Era	394
2018–2019: The Transformer Era	394
2020–2022: The Big Model Era	395
The Current State of Generative AI	396
Large Language Models	396

Text-to-Code Models	400
Text-to-Image Models	402
Other Applications	405
The Future of Generative AI	407
Generative AI in Everyday Life	407
Generative AI in the Workplace	409
Generative AI in Education	410
Generative AI Ethics and Challenges	411
Final Thoughts	413

Index..... 417

Introduction	1
What is Generative AI?	1
Why is it important?	2
How it works	3
Applications	4
The Future	5
Conclusion	6
Chapter 1: The Basics of Generative AI	7
1.1 Introduction to Generative AI	7
1.2 The History of Generative AI	8
1.3 The Architecture of Generative AI	9
1.4 The Role of Data	10
1.5 The Role of Algorithms	11
1.6 The Role of Hardware	12
1.7 The Role of Software	13
1.8 The Role of the User	14
1.9 The Role of the Environment	15
1.10 The Role of the Future	16
1.11 Conclusion	17
Chapter 2: Text-to-Text Models	18
2.1 Introduction to Text-to-Text Models	18
2.2 The History of Text-to-Text Models	19
2.3 The Architecture of Text-to-Text Models	20
2.4 The Role of Data	21
2.5 The Role of Algorithms	22
2.6 The Role of Hardware	23
2.7 The Role of Software	24
2.8 The Role of the User	25
2.9 The Role of the Environment	26
2.10 The Role of the Future	27
2.11 Conclusion	28
Chapter 3: Text-to-Image Models	29
3.1 Introduction to Text-to-Image Models	29
3.2 The History of Text-to-Image Models	30
3.3 The Architecture of Text-to-Image Models	31
3.4 The Role of Data	32
3.5 The Role of Algorithms	33
3.6 The Role of Hardware	34
3.7 The Role of Software	35
3.8 The Role of the User	36
3.9 The Role of the Environment	37
3.10 The Role of the Future	38
3.11 Conclusion	39
Chapter 4: Image-to-Image Models	40
4.1 Introduction to Image-to-Image Models	40
4.2 The History of Image-to-Image Models	41
4.3 The Architecture of Image-to-Image Models	42
4.4 The Role of Data	43
4.5 The Role of Algorithms	44
4.6 The Role of Hardware	45
4.7 The Role of Software	46
4.8 The Role of the User	47
4.9 The Role of the Environment	48
4.10 The Role of the Future	49
4.11 Conclusion	50
Chapter 5: Video-to-Video Models	51
5.1 Introduction to Video-to-Video Models	51
5.2 The History of Video-to-Video Models	52
5.3 The Architecture of Video-to-Video Models	53
5.4 The Role of Data	54
5.5 The Role of Algorithms	55
5.6 The Role of Hardware	56
5.7 The Role of Software	57
5.8 The Role of the User	58
5.9 The Role of the Environment	59
5.10 The Role of the Future	60
5.11 Conclusion	61
Chapter 6: Audio-to-Audio Models	62
6.1 Introduction to Audio-to-Audio Models	62
6.2 The History of Audio-to-Audio Models	63
6.3 The Architecture of Audio-to-Audio Models	64
6.4 The Role of Data	65
6.5 The Role of Algorithms	66
6.6 The Role of Hardware	67
6.7 The Role of Software	68
6.8 The Role of the User	69
6.9 The Role of the Environment	70
6.10 The Role of the Future	71
6.11 Conclusion	72
Chapter 7: Code-to-Code Models	73
7.1 Introduction to Code-to-Code Models	73
7.2 The History of Code-to-Code Models	74
7.3 The Architecture of Code-to-Code Models	75
7.4 The Role of Data	76
7.5 The Role of Algorithms	77
7.6 The Role of Hardware	78
7.7 The Role of Software	79
7.8 The Role of the User	80
7.9 The Role of the Environment	81
7.10 The Role of the Future	82
7.11 Conclusion	83
Chapter 8: Music-to-Music Models	84
8.1 Introduction to Music-to-Music Models	84
8.2 The History of Music-to-Music Models	85
8.3 The Architecture of Music-to-Music Models	86
8.4 The Role of Data	87
8.5 The Role of Algorithms	88
8.6 The Role of Hardware	89
8.7 The Role of Software	90
8.8 The Role of the User	91
8.9 The Role of the Environment	92
8.10 The Role of the Future	93
8.11 Conclusion	94
Chapter 9: Game-to-Game Models	95
9.1 Introduction to Game-to-Game Models	95
9.2 The History of Game-to-Game Models	96
9.3 The Architecture of Game-to-Game Models	97
9.4 The Role of Data	98
9.5 The Role of Algorithms	99
9.6 The Role of Hardware	100
9.7 The Role of Software	101
9.8 The Role of the User	102
9.9 The Role of the Environment	103
9.10 The Role of the Future	104
9.11 Conclusion	105
Chapter 10: The Future of Generative AI	106
10.1 Introduction to the Future of Generative AI	106
10.2 The Current State of Generative AI	107
10.3 The Challenges of Generative AI	108
10.4 The Opportunities of Generative AI	109
10.5 The Role of the Future	110
10.6 Conclusion	111
Chapter 11: Large Language Models	112
11.1 Introduction to Large Language Models	112
11.2 The History of Large Language Models	113
11.3 The Architecture of Large Language Models	114
11.4 The Role of Data	115
11.5 The Role of Algorithms	116
11.6 The Role of Hardware	117
11.7 The Role of Software	118
11.8 The Role of the User	119
11.9 The Role of the Environment	120
11.10 The Role of the Future	121
11.11 Conclusion	122
Chapter 12: The Conclusion	123
12.1 Introduction to the Conclusion	123
12.2 The Summary	124
12.3 The Future	125
12.4 Conclusion	126