

Symbols

1-Lipschitz continuous function, 115

A

accuracy, determining, 99, 264, 399, 411

action, in reinforcement learning, 333

activation functions, 32

active inference, 414

Adam (Adaptive Moment Estimation) optimizer, 36

adaptive instance normalization (AdaIN), 279, 282

agent, in reinforcement learning, 333

AI (artificial intelligence), 8, 413

AI ethics, 411-413

approximate density models, 20

artifacts, 103, 281

artificial intelligence (AI), 8, 413

artificial neural networks (ANNs), 25

arXiv, xxii

“Attention Is All You Need” (Vaswani), 236, 394

attention mechanisms

attention equation, 241

attention head, 239

attention scores, 253

attention weights, 241

generating polyphonic music, 313

paper popularizing, 236

self- versus cross-referential, 258

understanding, 238

attribute binding, 376

attributes, entangled, 277

AudioLM, 407

“Auto-Encoding Variational Bayes” (Kingma and Welling), 59

autoencoders (see also variational autoencoders)

architecture of, 63

decoder architecture, 65-67

diagram of process, 61

encoder architecture, 64

Fashion-MNIST dataset, 62

generating new images, 71-74

joining encoder to decoder, 67

reconstructing images, 69

uses for, 64

visualizing latent space, 70

autoregressive models

autoregressive prior of DALL·E 2, 367, 374

bidirectional cells, 153

description of, 130

gated recurrent units (GRUs), 151-153

generative model taxonomy, 19

history of, 394

how LSTMs work, 130

long short-term memory (LSTM) networks, 131-149

masked convolutional layers, 154-156

stacked recurrent networks, 149-151

B

Bach chorale dataset, 317

backpropagation, 26

batch normalization, 46-49, 51, 120

batches, 37

BERT (Bidirectional Encoder Representations from Transformers), 255, 394

- bidirectional cells, 153
- big model era, 395
- BigGAN, 288, 395
- binary cross-entropy loss, 36
- BLOOM, 395, 399
- Boltzmann distribution, 191
- Boltzmann machine, 203
- BookCorpus, 236
- Bricks dataset, 98

C

- Canny edge maps, 402
- categorical cross-entropy loss, 36
- causal masking, 242-244
- CelebFaces Attributes (CelebA) dataset, 85, 275
- CGAN (see conditional GAN)
- challenges, of generative AI, 264, 399, 411-413
- change of variables equation, 173
- change of variables technique, 170-172
- character tokens, 135
- ChatGPT, 260-264, 395, 407
- Chinchilla, 395
- CICERO, 406
- CIFAR-10 dataset, 28
- CLIP (Contrastive Language-Image Pre-training)
 - description of, 362
 - history of, 396
 - key concepts behind, 363
 - training process, 363-367
- CMA-ES (covariance matrix adaptation evolution strategy), 348-353
- CNN (see convolutional neural networks)
- code examples, obtaining and using, xxiii, 20
- codebook, 290
- Codex, 400
- comments and questions, xxv
- compile method, 36
- conditional GAN (CGAN)
 - analysis of, 126
 - architecture of, 123
 - training, 124
- “Conditional Generative Adversarial Nets” (Mirza and Osindero), 122
- context vector, 241
- contrastive divergence, 191, 197-201
- Contrastive Language-Image Pre-training (see CLIP)
- contrastive learning, 362

- ControlNet, 402
- convolutional neural networks (CNNs)
 - batch normalization, 46-49
 - benefits of, 40
 - building, 51-53
 - convolutional layers, 41-46
 - dropout, 49-51
 - masked convolutional layers, 154-156
 - training and evaluating, 53
- convolutional transpose layers, 65
- Copilot, 400
- cosine diffusion schedule, 212
- cosine similarity, 363
- coupling layers, 175-177, 180
- covariate shift, 47
- cross-referential attention, 258
- CycleGAN, 291, 394

D

- DALL·E, 289, 361, 376, 396
- DALL·E 2
 - architecture, 362
 - availability of, 402
 - decoder, 369-373
 - examples generated by, 373-375, 405
 - history of, 361, 396
 - limitations of, 375
 - text encoder, 362-367
 - training the prior model, 367-369
- DCGAN (see deep convolutional GAN)
- DDIM (see Denoising Diffusion Implicit Model)
- DDM (see denoising diffusion models)
- DDPM (see Denoising Diffusion Probabilistic Model)
- decoder Transformers, 244, 255
- deep convolutional GAN (DCGAN)
 - analysis of, 109
 - dataset used, 98
 - discriminator in, 99-101
 - generator in, 101-103
 - history of, 394
 - published paper on, 97
 - training, 104-109
 - training tips and tricks, 110-113
- deep learning
 - deep neural networks, 25-27, 54
 - defined, 23
 - Keras and TensorFlow for, 27

- model creation, 28-40
- model improvement, 40-54
- structured versus unstructured data, 24
- demodulation step, 283
- Denoising Diffusion Implicit Model (DDIM)
 - description of, 226
 - history of, 395
- denoising diffusion models (DDMs)
 - analysis of, 228-230
 - dataset used, 208-209
 - description of, 208
 - diffusion schedules, 211
 - forward diffusion process, 209
 - reparameterization trick, 210
 - reverse diffusion process, 214-216
 - sampling from, 225-227
 - training, 224-225
 - U-Net denoising model, 217-224, 370
- Denoising Diffusion Probabilistic Model (DDPM)
 - development of, 203
 - history of, 395
- denoising models, 64 (see also denoising diffusion models)
- dense layers, 25
- density function, 16, 18
- determinants, 172
- Dhariwal, Prafulla, 395
- diffusion models (see also denoising diffusion models)
 - description of, 205
 - diffusion prior of DALL·E 2, 368
 - generative model taxonomy, 20
 - history of, 395
 - key ideas underpinning, 206-207
 - latent diffusion, 396
- “Diffusion Models Beat GANs on Image Synthesis” (Dhariwal and Nichol), 395
- diffusion schedules, 211
- Dinh, Laurent, 174
- Diplomacy board game, 406
- discrete latent space, 290
- discriminative modeling, 5
- discriminators, 97, 99-101, 110
- Dong, Hae-Wen, 317
- DrawBench, 378
- DreamFusion, 407
- Dreamix, 406
- dropout layers, 49

Du, Yilun, 191

E

- educational applications, 410
- EMA (exponential moving average), 214
- embedding space, 63, 70, 85-93, 219
- embodied large language models, 406
- encoder Transformers, 244, 255
- encoder-decoder Transformers, 255
- encoders, 64
- energy function/energy score, 191, 193
- energy-based models (EBMs)
 - analysis of, 201-202
 - Boltzmann distribution, 191
 - Boltzmann machine, 203
 - dataset used, 192
 - description of, 189
 - energy function, 193
 - generative model taxonomy, 20
 - key concepts behind, 189
 - RBM (restricted Boltzmann machine), 203
 - sampling using Langevin dynamics, 194-196
 - score-based generative models, 206
 - training with contrastive divergence, 197-201
- entangled attributes, 277
- environment, in reinforcement learning, 332
- episode, in reinforcement learning, 333
- epochs, 38
- equalized learning rates, 276
- ethical concerns, of generative AI, 411-413
- evaluate method, 38
- event-based tokenization, 315
- evolutionary strategies, 349
- explicit density models, 19
- exploding gradient problem, 46
- exponential moving average (EMA), 214

F

- facial image generation
 - dataset used, 85
 - generating new faces, 90
 - latent space arithmetic, 91
 - morphing between faces, 92
 - progress in, 7
 - VAE analysis, 89
 - VAE training, 87
- Fashion-MNIST dataset, 62, 70
- fast-track route, 156

- feature engineering, 26
- features
 - descriptions of, 4
 - learning high-level, 26
- FFJORD (Free-Form Continuous Dynamics for Scalable Reversible Generative Models), 187
- filters, 41
- fit method, 37
- Flamingo
 - architecture, 382-387
 - examples generated by, 388-389
 - history of, 396
- Flan-T5, 395
- Flowers dataset, 208
- forward diffusion process, 209
- forward pass, 26
- Free-Form Continuous Dynamics for Scalable Reversible Generative Models (FFJORD), 187
- fully connected layers, 25
- functional API (Keras), 30-35

G

- game state, in reinforcement learning, 333
- GAN (see generative adversarial networks)
- gated recurrent units (GRUs), 132, 151, 394
- Gaussian distribution (normal distribution), 75
- GenAI (see generative AI)
- “Generative Adversarial Nets” (Goodfellow), 95
- generative adversarial networks (GANs)
 - BigGAN, 288
 - challenges of, 110, 113
 - conditional GAN (CGAN), 122-127
 - deep convolutional GANs (DCGANs), 97-113
 - fundamental training concepts, 96
 - generative model taxonomy, 19
 - history of, 95, 394
 - ProGAN, 269-276
 - StyleGAN, 277-281
 - StyleGAN2, 281-286
 - Wasserstein GAN with Gradient Penalty (WGAN-GP), 113-122
 - versus WGAN-GPs, 121
- generative AI (GenAI)
 - current state of, 396-407
 - ethics and challenges related to, 264, 411
 - future of, 407-411, 413
 - history of, 392-396

- generative deep learning (see also models)
 - additional resources, xxii
 - core probability theory, 15-18
 - creating something that is creative, xvii
 - generative modeling framework, 10
 - introduction to, 4-9
 - learning objectives and approach, xviii
 - prerequisites to learning, xix
- generative modeling (see models)
- generative pre-trained transformer (see GPT)
- generators
 - attention-based, 287
 - bar generator, 323
 - DCGAN generator, 101-103, 110
 - in GANs, 97, 101-103
 - MuseGAN generator, 320
 - StyleGAN generator, 277
- GitHub Copilot, 400
- GLIDE (Guided Language-to-Image Diffusion for Generation and Editing), 369-373, 396
- GLOW model, 186
- Goodfellow, Ian, 95
- Gopher, 395
- GPT (generative pre-trained transformer)
 - analysis of, 252-255
 - applications in everyday life, 407
 - attention mechanism, 238
 - causal masking, 242-244
 - dataset used, 237
 - description of, 236
 - evolution of, 259
 - history of, 236, 394
 - improvements to, 237
 - multihead attention, 241
 - positional encoding, 248-250
 - queries, keys, and values, 239-241
 - Transformer block, 245-248
- GPT-2, 237, 394
- GPT-3
 - availability of, 398
 - benefits of, 237
 - evolution of, 259
 - example generated by, 260, 396, 411
 - history of, 395
- GPT-3.5, 237, 262
- GPT-4, 237, 259
- gradient descent using Langevin dynamics, 195
- gradient penalty loss, 117
- Gradient Tape, 83

grid tokenization, 313-315
GRU (see gated recurrent units)
Guided Language-to-Image Diffusion for Generation and Editing (GLIDE), 369-373

H

Ha, David, 331, 337, 394
hallucinations, 264
He initialization, 276
hidden layers, 27
hidden state, 140, 142
Hinton, Geoffrey, 49
Hochreiter, Sepp, 132
Huang, Cheng-Zhi Anna, 313
Hui, Jonathan, 116
hyperparameters, 112

I

image generation (see also facial image generation; PixelCNN)
 benefits of diffusion models for, 205
 BigGAN, 288
 CIFAR-10 dataset for, 28
 DDM analysis, 228-230
 generating new images, 71-74
 generative modeling process, 4
 generative versus discriminative modeling, 5
 history of, 395
 ProGAN, 269-276
 progress in facial image generation, 7
 reconstructing images, 69
 representation learning for, 13
 Self-Attention GAN (SAGAN), 286
 StyleGAN2, 281-286
 visualizing latent space, 70
image-to-image models, 291, 394
Imagen
 architecture, 377
 DrawBench, 378
 examples generated by, 379
 history of, 377, 396
 overview of, 405
Images of LEGO Bricks dataset, 98
implicit density models, 19
“Implicit Generation and Modeling with Energy-Based Models” (Du and Mordatch), 191

“Improving Language Understanding by Generative Pre-Training” (Radford), 236
in-dream training, 353-356
InstructGPT model, 262
isotropic multivariate normal distributions, 76

J

Jacobian determinant, 172
joint token/position encoding, 249

K

Kaggle, 86, 237
Keras (see also models)
 autoencoder creation in, 65
 benefits of, 27
 Conv2DTranspose layer, 66
 creating new layers in, 79
 data loading, 28
 dataset creation, 98
 decoder creation in, 67
 documentation, 36
 GAN discriminator creation in, 100
 model building, 30-35
 model compilation, 35
 model evaluation, 38-40
 model improvement, 40-54
 model training, 37
 MuseGAN generator in, 324
 resources, 20
 StyleGAN tutorial, 278
 VAE creation in, 78
Keras layers
 Activation, 52
 Batch Normalization, 46
 Bidirectional, 153
 Conv2D, 42
 Conv2DTranspose, 66, 103
 Conv3D, 326
 Dense, 32
 Dropout, 49
 Embedding, 138
 Flatten, 32
 GRU, 132
 Input, 32
 LeakyReLU, 52
 LSTM, 140
 MultiHeadAttention, 241
 UpSampling2D, 103
Keras NLP module, 306

- kernels, 41
- key vectors, 239
- Kingma, Diederik, 59
- Kullback–Leibler (KL) divergence, 80

L

- label smoothing, 108
- LAION-5B dataset, 402
- LaMDA, 395
- Langevin dynamics, 191, 194-196
- language modeling, 236
- Large Language Model Meta AI (LLaMA), 398
- large language models (LLMs), 396-400
- Large-scale Scene Understanding (LSUN) data-set, 276
- latent diffusion, 380, 396
- latent space, 63, 70, 85-93
- layer normalization, 245
- layers, 25, 79 (see also Keras layers)
- lazy regularization, 284
- LeakyReLU, 33
- learning rate, 36
- likelihood, 17
- linear diffusion schedule, 211
- Lipschitz constraint, 115, 116
- LLaMA (Large Language Model Meta AI), 398
- LLMs (large language models), 396-400
- logarithm of the variance, 77
- long short-term memory networks (see LSTM networks)
- loss functions, 35, 68, 80, 195
- lower triangle matrix, 178
- LSTM (long short-term memory) networks
 - embedding layer, 138
 - generating datasets, 137
 - generating new text, 146
 - history of, 131, 394
 - LSTM architecture, 138
 - LSTM cell, 142-144
 - LSTM layer, 140-142
 - published paper on, 132
 - tokenizing the text, 134
- LSUN (Large-scale Scene Understanding) data-set, 276

M

- machine learning
 - benefits of, 13
 - data for, 24-25

- dropout principle, 49
- generative modeling and, 4-7
- libraries for, 27
- major branches of, 23, 28, 332
- resources, xxii
- Make-A-Video, 406
- mapping network f , 278
- masked convolutional layers, 154-156
- masking, causal, 242-244
- matrix determinants, 172
- maximum likelihood estimation, 18
- MDN (mixture density network), 337, 346
- mean squared error loss, 35
- Megatron-Turing NLG, 395, 398
- metrics parameter, 36
- MIDI files, 300
- Midjourney, 60, 96, 402, 405
- Mildenhall, Ben, 219
- minibatch standard deviation layer, 275
- Mirza, Mehdi, 122
- mixture distributions, 162-164
- MLP (see multilayer perceptrons)
- MNIST dataset, 192
- mode collapse, 111
- model.summary() method, 34, 45
- models (see also generative deep learning;
Keras)
 - core probability theory, 15-18
 - deep neural networks, 25-28
 - generative model taxonomy, 18
 - generative modeling, 4
 - generative versus discriminative modeling,
5
 - history of, 395
 - improving, 40-54
 - parametric modeling, 16
 - probabilistic versus deterministic, 4
 - variational autoencoders (VAEs), 59
 - World Model architecture, 336-356
- modulation step, 283
- Mordatch
 - Igor, 191
- multihead attention, 241
- multilayer perceptrons (MLPs)
 - data preparation, 28
 - example of, 25
 - model building, 30-35
 - supervised learning and, 28
- multilayer RNNs, 149

- multimodal models
 - challenges of text-to-image generation, 360
 - DALL·E 2, 361-376
 - Flamingo, 381-389
 - history of, 396
 - Imagen, 377-380
 - Stable Diffusion, 380
- multivariate normal distribution, 75
- MUSE, 396, 405
- MuseGAN, 317-329
 - analysis of, 327
 - dataset used, 317
 - MuseGAN critic, 326
 - MuseGAN generator, 320
- “MuseGAN: Multi-Track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment” (Dong), 317
- MuseNet, 394
- MuseScore, 300
- music generation
 - analysis of music generation Transformer, 309-312
 - dataset used, 300
 - generating polyphonic music, 313
 - importing MIDI files, 300
 - inputs and outputs, 307
 - MuseGAN, 317-329
 - music versus text generation, 298
 - prerequisites to, 300
 - sine position encoding, 305-306
 - tokenization, 303
 - training set for, 304
 - Transformers applied to, 394
- Music Transformer, 394
- “Music Transformer: Generating Music with Long-Term Structure” (Huang), 313
- music21 library, 300
- MusicLM, 407

N

- Nain, Aakash Kumar, 113
- natural language processing (NLP), 255
- NCSN (Noise Conditional Score Network), 395
- “NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis” (Mildenhall), 219
- “Neural Discrete Representation Learning” (van den Oord), 289

- neural networks (see also convolutional neural networks; deep learning)
 - deep neural networks, 54
 - defined, 25-27
 - loss functions and, 18
 - role in deep learning, 20
 - using Keras to build, 27, 30-35
- Nichol, Alex, 395
- NLP (natural language processing), 255
- Noise Conditional Score Network (NCSN), 395
- noise, adding to labels, 108
- nontrainable parameters, 48
- normal distribution (Gaussian distribution), 75
- normalizing flow models
 - change of variables equation, 173
 - change of variables technique, 170-172
 - description of, 167
 - FFJORD (Free-Form Continuous Dynamics for Scalable Reversible Generative Models), 187
 - generative model taxonomy, 19
 - GLOW, 186
 - Jacobian determinant, 172
 - key concepts behind, 168
 - motivation of, 169
 - RealNVP model, 174-185

O

- observations, 4
- OPT, 395, 399
- optimizers, 35
- Osindero, Simon, 122
- overfitting, 49
- Oxford 102 Flower dataset, 208

P

- padding, 43
- PaLM-E, 406
- Papers with Code, xxii
- parameters, trainable and nontrainable, 48
- parametric modeling, 16
- Parti, 396, 405
- PatchGAN, 291
- path length regularization, 283
- perceptual loss term, 292
- personal assistants, 407
- piano roll grid, 314
- pix2pix, 291, 394
- PixelCNN

- analysis of, 159-162
- history of, 153, 394
- masked convolutional layers, 154-156
- mixture distributions, 162
- residual blocks, 156-158
- training, 158
- PixelRNN, 394
- pixelwise normalization, 276
- poetry, 397
- Point-E, 407
- positional embedding, 248
- positional encoding, 248-250
- posterior collapse, 289
- prediction, using batch normalization, 48
- probability density function, 16, 18, 75
- probability distributions, 162
- probability theory, 15-18
- ProGAN
 - concept of progressive training, 269-276
 - description of, 269
 - history of, 394
 - outputs, 276
- progressive training, 269-276
- prompt engineering, 405
- prompts, 396

Q

- query, 239
- questions and comments, xxv

R

- Radford, Alec, 97, 236
- random (stochastic) elements, 4
- random noise, 108, 195
- RBM (restricted Boltzmann machine), 203
- RealNVP
 - analysis of, 184
 - coupling layers, 175-177
 - dataset used, 174
 - description of, 174
 - history of, 394
 - passing data through coupling layers, 177-180
 - stacking coupling layers, 180
 - training, 181-183
- Recipes dataset, 132
- recurrent neural networks (see RNNs)
- regularization techniques, 49
- reinforcement learning (RL)

- ChatGPT and, 262
- defined, 332
- key terminology, 332
- process of, 333-335
- Reinforcement Learning from Human Feedback (RLHF), 262
- ReLU (rectified linear unit), 33
- reparameterization trick, 79, 210
- representation learning, 13-14
- residual blocks, 156-158, 221-222
- restricted Boltzmann machine (RBM), 203
- reverse diffusion process, 214-216
- reward modeling, 262
- reward, in reinforcement learning, 333
- RLHF (Reinforcement Learning from Human Feedback), 262
- RMSE (root mean squared error), 68
- RMSProp (Root Mean Squared Propagation)
 - optimizer, 36
- RNNs (recurrent neural networks)
 - bidirectional cells, 153
 - gated recurrent units (GRUs), 151
 - history of, 131
 - LSTM (long short-term memory) networks, 131-149
 - MDN-RNN World Model architecture, 337
 - multilayer, 149
 - stacked recurrent networks, 149
- root mean squared error (RMSE), 68
- Root Mean Squared Propagation (RMSProp)
 - optimizer, 36

S

- SAGAN (self-attention GAN), 286, 395
- sample space, 16
- scaling streams, 177
- Schmidhuber, Jurgen, 132, 331, 394
- score matching technique, 203
- score-based generative models, 206
- Self-Attention GAN (SAGAN), 286, 395
- self-referential layers, 258
- Sequential models (Keras), 30-35
- sigmoid activation, 33
- sine position embedding, 305
- sinusoidal embedding, 219
- skip connections, 156, 217, 245, 284
- softmax activation, 33
- Sparse Transformers, 299
- stabilization phase, 272

- Stable Diffusion
 - advantages of, 402-405
 - architecture, 380
 - examples generated by, 381
 - history of, 380, 396
- stacked recurrent networks, 149
- standard deviation, 75
- standard normal curves, 75
- stemming, 135
- stochastic (random) elements, 4
- stochastic gradient Langevin dynamics, 195
- stochastic variation, 280
- strides parameter (Keras), 43
- structured data, 24
- style mixing, 279
- StyleGAN, 277-281, 395
- StyleGAN-XL, 286, 395
- StyleGAN2, 281-286, 395
- subclassing layers, 79
- summary method, 35
- supervised fine-tuning, 262
- supervised learning, 28, 332
- swish activation, 193
- synthesis network, 279

T

- T5, 256-259, 394
- tape.gradient() method, 83
- taxonomy, 18
- temperature parameter, 146
- temporal networks, 320
- TensorFlow, 27
- text data generation (see also GPT)
 - LSTM (long short-term memory) networks, 131-149
 - RNN (recurrent neural network) extensions, 149-153
 - short story generation example, 130
 - text versus image data, 133
 - text versus music generation, 298
- text-to-3D models, 407
- text-to-code models, 400-400
- text-to-image models, 360, 402-405
- text-to-music models, 407
- text-to-video models, 406
- text-to-X multimodal models, 405
- thermodynamic diffusion, 206
- timeline of AI, 393
- timestep, in reinforcement learning, 333

- token embedding, 248
- tokenization
 - event-based, 315
 - grid, 313-315
 - of notes for music generation, 303
 - process of, 134-137
- Toolformer, 411
- tractable models, 19
- trainable parameters, 48
- training data, 4
- training process, 26
- Transformer block, 245-248
- Transformers (see also GPT; music generation)
 - architectures for, 255
 - BERT (Bidirectional Encoder Representations from Transformers), 255
 - ChatGPT, 260-264
 - decoder versus encoder, 244
 - description of, 236
 - GPT-3 and GPT-4, 259
 - history of, 394
 - Sparse Transformers, 299
 - T5, 256-259
- transition phase, 272
- translation streams, 177
- truncated normal distribution, 288
- truncation trick, 288
- truth, filtering from generated fiction, 99, 264, 399, 410
- two moons dataset, 174

U

- U-Net denoising model, 217-224, 370
- uninformative loss, 112
- unit normal curves, 75
- units, 25, 140
- unstructured data, 24
- unsupervised learning, 332
- “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Network” (Radford), 97
- upsampling, 103, 372

V

- VAE (see variational autoencoders)
- VAE with a GAN discriminator (VAE-GAN), 394
- value vectors, 241
- van den Oord, Aaron, 153, 289

- vanishing gradient problem, 132
- variance, 75
- variational autoencoders (VAEs) (see also autoencoders)
 - analysis of, 84
 - autoencoder architecture, 61-74
 - decoders, 77
 - encoder adjustments, 75-78
 - facial image generation using, 85-93
 - generative model taxonomy, 20
 - history of, 394
 - introduction to, 60, 74
 - published paper on, 59
 - training, 82
 - VAE build in Keras, 78
 - VAE loss function, 80
 - VAE model summary, 80
 - World Model architecture, 336
 - World Model training, 340-344
- Vaswani, Ashish, 219, 236, 394
- Vector Quantized Generative Adversarial Network (VQ-GAN), 289-292, 395
- Vector Quantized VAE (VQ-VAE), 394
- “Vector-quantized Image Modeling with Improved VQGAN” (Yu), 292
- Vision Transformer (ViT), 292, 364, 395
- Visual ChatGPT, 264
- vocabulary, 135
- VQ-GAN (Vector Quantized Generative Adversarial Network), 289-292, 395
- VQ-VAE (Vector Quantized VAE), 394

W

- Wasserstein GAN with Gradient Penalty (WGAN-GP)
 - analysis of, 121

- gradient penalty loss, 117
 - Lipschitz constraint, 115
 - versus standard GANs, 121
 - training, 119
 - tutorial on, 113
 - Wasserstein loss, 114
 - weight clipping, 116
- Wasserstein GANs (WGANs)
 - benefits of, 113
 - history of, 394
- weight clipping, 116
- weight modulation and demodulation, 282-283
- weights, 25
- Welling, Max, 59
- WGAN (see Wasserstein GANs)
- Wine Reviews dataset, 237
- workplace applications, 409
- World Models
 - architecture, 336-338
 - collecting MDN-RNN training data, 346
 - collecting random rollout data, 339
 - published paper on, 331, 337, 394
 - training in-dream, 353-356
 - training process, 338
 - training the controller, 348-353
 - training the MDN-RNN, 346-348
 - training the VAE, 340-344
 - World Model architecture, 336-356

Y

- Yu, Jiahui, 292

Z

- zero-shot prediction, 364