
Index

- $(\mu + \lambda)$ -ES, 2041
- (μ, λ) -ES, 2041
- (ϵ, δ) -differential privacy, 4495
- L^p space, 5213
- Q-Wiener process, 3270
- $[A][B] \cdots [A] \rightarrow [B]$ pattern, 4022
- β -VAE, 2754
- ℓ_1 regularisation, 2749
- x_0 -prediction, 2764
- \$100 device target, 4901
 - architecture, 4920
- 1×1 convolution, 3532
- 1-bit SGD, 4476
- 1/5-th success rule, 2280
- 10–20 electrode system, 3791
 - definition, 3791
 - figure, 3792
- 10-metre wind components, 4225
- 1F1B
 - bubble fraction, 5052
- 1F1B schedule, 4994–4999, 5051–5060, 5137
 - bubble fraction, 5141, 5173
 - bubble fraction exercise, 4999
 - definition, 4994
- 2-Poisson model, 2352–2353
- 2-layer transformer
 - example, 4010
- 2.5D approach, 3507
- 2.5D segmentation, 3537
- 21 CFR Part 882, 3939
- 2D segmentation, 3506
- 2D torus, 4945
- 2D vs. 3D segmentation
 - trade-off, 3507
- 2D+time segmentation, 3507–3508
- 3-D Gaussian Splatting
 - (lip synthesis), 4738
 - face animation exercise, 4744
- 3D Earth-Specific Transformer
 - architecture, 4255
- 3D Swin Transformer
 - weather forecasting, 4254–4264
- 3D U-Net
 - MRI-to-PET, 3858
- 3D convolution, 563–566
 - causal, 564
 - definition, 563
 - parameter count, 563
- 3D molecule generation, 2891
- 3D parallelism, 5058–5060, 5135–5144
 - configuration, 5059
 - data-tensor-pipeline, 5136
 - definition, 4932, 5058
- 3D printing
 - assistive devices, 4869
- 3D segmentation, 3506–3507, 3537–3538
 - anisotropy, 3507
 - memory constraints, 3507
- 3D torus, 4945
 - topology, 4939
- 3D+time segmentation, 3727
- 3D-Var, 4216
 - exercise, 4411
- 3DReasonKnee, 3688
- 4D CT
 - respiratory gating, 3508
- 4D MRI
 - reconstruction, 3585
- 4D cardiac CT synthesis, 4614

- 4D medical volumes, [3727](#)
- 4D parallelism, [4932](#), [5137](#)
- 4D segmentation, [3508](#)
- 4D-Var, [4191](#), [4216](#)
 - as special case, [4383](#)
- 510(k)
 - AI device, [3939](#)
- 510(k) pathway, [3939](#)
- 5D parallelism, [5138](#)
- 5G/6G
 - federated generation, [4623–4624](#)
- 7–10 day threshold
 - predictability, [4398](#)
- 70B LLM
 - all-reduce example, [5039](#)
 - gradient communication, [5039](#)
- 70B model
 - memory budget, [5077](#)
- 7B language model
 - training recipe, [5112](#)
- 7B model
 - memory requirements, [4930](#)
 - training time calculation, [4940](#)
- 850-hPa temperature, [4225](#)
- A-GEM, [1196](#)
 - definition, [1196](#)
- A/B testing
 - evolutionary heuristics, [2212](#)
- A100
 - BF16 training, [4935](#)
 - HBM2e bandwidth, [4934](#)
 - NVLink bandwidth, [4948](#)
 - specification, [4935](#)
 - training speed, [4926](#)
- AAC
 - emotional expression, [4760](#)
 - intent anticipation, [4767](#)
 - keystroke efficiency analysis, [4768](#)
 - limitations, [4759](#)
 - linguistic limitations, [4759](#)
 - LLM-powered, [4760](#)
 - operational complexity, [4759](#)
 - practice opportunities, [4759](#)
 - predictive intent, [4767](#)
 - predictive text, [4917](#)
 - social participation, [4760](#)
 - strategic communication, [4760](#)
 - symbol sets, [4759](#)
 - text generation, [4694](#)
 - voice identity, [4699](#)
- AANet, [4369](#)
 - gradient loss, [4370](#)
- Aardvark
 - computational efficiency, [4377](#)
 - end-to-end forecasting, [4375](#)
- abc conjecture, [2941](#)
- AbdomenCT-1K dataset, [3512](#)
- abdominal segmentation, [3512](#)
 - PAN, [3578](#)
- ablation methods
 - comparison, [3980](#)
- ablation study
 - LLM components, [2108](#)
- abstract syntax trees, [3036](#)
- abstraction layers
 - mechanistic interpretability, [3972](#)
- AC⁰, [1036](#)
 - parity lower bound, [1036](#)
- ACC, *see* anomaly correlation coefficient
- acceleration
 - audio generation, [873–880](#)
 - consistency distillation, [876](#)
- accelerator
 - landscape, [4933](#), [4942](#)
- accelerator comparison, [4939](#)
 - table, [4939](#)
- access problem, [2687](#)
 - examples, [2687–2688](#)
- AccessEval corpus, [4889–4890](#)

- accessibility
 - automated segmentation, 3488
 - bias, 4841–4853
 - blind professional workflow, 4862
 - color vision, 4769–4779
 - consumer ecosystem, 4854–4864
 - day-in-life scenario, 4862
 - definition, 4691
 - DHH
 - lip synthesis, 4735–4745
 - economic impact, 4696
 - ecosystem
 - 2026 landscape, 4854
 - equity, 4860
 - ethics, 4744
 - faithfulness evaluation, 4838
 - formal definition, 4691
 - formal model exercise, 4697
 - generative AI, 4689
 - generative shift, 4697–4706
 - generative vs. discriminative, 4693
 - hallucination, 4829–4841
 - imperative, 4690
 - language coverage, 4860
 - market failure, 4742
 - motor impairment, 4864–4875
 - neurodiversity, 4789
 - safety stakes, 4829
 - scale statistics, 4696
 - sign language, 4725–4735
 - tools
 - comparison table, 4857
 - performance benchmarks, 4856
 - trust formation, 4860
- accessibility AI
 - chapter summary, 4921
 - evaluation, 4885–4897
 - framework, 4891
 - measurement problem, 4885
 - exercises, 4911–4922
 - open problems, 4897–4910
 - real-time perception, 4897
 - real-world evaluation, 4893
 - research challenges, 4919–4921
 - accessibility evaluation
 - standard protocol, 4895
 - accessibility metrics, 4888–4889
 - multi-dimensional evaluation, 4891
 - organised by disability type, 4891
 - accessibility technology
 - evolution timeline, 4701
 - history, 4693–4694
 - accessible navigation, 4716–4725
 - integration exercise, 4724
 - accumulated improvement signal, 2172
 - accuracy
 - comparability, 3918
 - ACDC algorithm, 3975
 - complexity, 3976
 - description, 3975
 - Ackermann, Wilhelm, 3016
 - ACL2, 3004–3005
 - key features, 3004
 - acquisition function, 2227
 - acquisition functions
 - segmentation, 3734
 - ACT, *see* Adaptive Computation Time
 - action potential, 1857
 - action-outcome circuit, 4118
 - definition, 4118
 - exercise, 4123
 - activation caching, 4087
 - Prisma, 4087
 - activation checkpointing, 5070–5081
 - algorithm, 5071
 - formal definition, 5071
 - optimal strategy, 5072
 - activation compression, 5077
 - activation drift, 1183
 - activation engineering, 4139

- activation function, 92
 - in generative models, 93
 - Snake, 793, 958
- activation functions, 1501–1505
 - dying ReLU, 1502
 - GELU, 1503
 - Mish, 1504
 - SELU, 1502
 - self-normalising property, 1503
 - SiLU, 1503
 - StarReLU, 1504
 - wavelet, 3261
- activation memory
 - anatomy, 5070
 - exercise, 5080
 - quadratic scaling, 5071
 - single device, 5098
- activation patching, 3974–3981, 4014, 4032–4035, 4118
 - bias auditing, 4137–4138
 - concept-level, 3973
 - cross-modal, 4079
 - deceptive alignment, 4135
 - definition, 3974, 4032
 - exercise, 4036
 - implementation, 3977
 - toy transformer, 4165
- activation recomputation
 - boundaries, 5101
 - boundary
 - recommendation, 5101
- activation resampling, 3975
- activation steering, 4073
- active learning
 - pool-based, 2224
 - stream-based, 2224
- active contours, 3487, 3525
 - energy functional, 3525
 - definition, 3526
 - Euler–Lagrange equation, 3742
 - external energy, 3526
 - internal energy, 3526
 - level set formulation, 3527
 - numerical solution, 3527
- active learning, 2224–2237, 2881–2882
 - acquisition functions, 3734
 - definition, 2224
 - endoscopy, 3514
 - evolutionary connection, 2225
 - fundamentals, 2224–2226
 - generative amplification, 3736
 - medical segmentation, 3734
 - optimal stopping, 3736
 - query synthesis, 2224
 - synthesis loop, 2881
 - synthetic data, 2881
 - with evolutionary proposal, 2226
- actor model
 - for agents, 1335
- AdaEvolve, 2170–2180, 2286
 - accumulated improvement signal, 2171
 - algorithm, 2174
 - convergence rate, 2177
 - empirical results, 2179
 - exercises, 2180
 - formal definition, 2174
 - Level 1, 2173
 - Level 2, 2173
 - Level 3, 2173
 - motivation, 2170
 - stability, 2176
 - theoretical analysis, 2176
 - three-level hierarchy, 2172
- AdaIN, 353–355
 - exercise, 360
- AdaLN
 - weather, 4333
- adaLN-Zero, *see* adaptive layer normalisation

- Adam optimiser, 95
 - checkpoint size, 5147
 - memory cost, 4930
- Adam optimizer
 - connection to AdaEvolve, 2172
 - distributed, 5111
 - memory cost, 5060
 - warmup, 5027
- adapter composition, 693
- adapter modules, 1962
- adapters
 - medical segmentation, 3716
- adaptive agents, 1284–1289
- adaptive AI
 - FDA, 3939
- adaptive algorithm
 - FDA regulatory, 3940
- adaptive allocation
 - compute savings, 2199
- adaptive architecture
 - comparison, 1890
 - motivation, 1889
- adaptive attacks, 1686–1688
 - against DP-SGD, 1687
 - lower bound, 1687
- adaptive computation, 1955
- Adaptive Computation Time, 1058
- adaptive depth, 1955
- Adaptive FNO, 3223
- adaptive foundation models, 1967–1968
 - algorithm, 1967
- Adaptive Fourier Neural Operator, 4244–4254
 - definition, 4246
 - formal definition, 4247
- adaptive interface generation
 - definition, 4866
- adaptive landscape, 2022
- adaptive layer normalisation
 - conditioning, 614
- adaptive layer normalisation, 593–598
 - adaLN-Zero, 594
 - affine modulation, 594
 - definition, 593
 - exercise, 598
 - expressiveness, 596
 - FGN, 4333
 - mathematical analysis, 594
- adaptive learning
 - neurodivergent, 4791
- adaptive network
 - definition, 1871
- adaptive network hypothesis, 1868–1880
- adaptive networks, 1845–1880, 2005
 - central question, 1871
 - connections, 1969–1981
 - exercises, 1992–2003
 - future directions, 1981–1989
 - network growth, 1869–1871
- adaptive retrieval, 2527
- adaptive UI generation, 4866
- adaptive width, 1956
- ADAS, 2181–2190, 2285
 - algorithm, 2183
 - benchmark results, 2186
 - boundary dissolution, 2187
 - cross-model transfer, 2186
 - explanation, 2186
 - emergent topologies, 2184
 - exercises, 2189
 - Meta Agent Search, 2183
 - search space, 2182
- ADASYN, 3805
 - description, 3806
 - limitations, 3806
- ADHD, 4690, 4789
 - dynamic difficulty adjustment, 4796
- adjacency matrix
 - electrode graph, 3886
 - learnable, 2748

- learned, 2747
- learning, 2747–2749
- adjoint method, 3400
- Adobe Firefly, 11
- Adult dataset, 2854
- adult neurogenesis, 1853
- advantage function, 3054, 3102
- advection equation, 3416
 - neural network embedding, 4366
- adversarial autoregressive network, 4369
- adversarial coevolution, 2154–2170
 - general framework, 2165
- adversarial curriculum, 2393
- adversarial interpretability, 4140–4141
- adversarial loss
 - audio, 785
 - intensity constraint, 4356
 - negative examples, 3573
 - video, 748
- adversarial perturbation
 - cultural bias, 4142
- adversarial prompts
 - coevolution, 2166
- adversarial robustness
 - agents, 1407–1414
 - analogy, 1687
 - coevolution, 2166
- adversarial training
 - audio, 942
 - domain adaptation, 3578
 - pathologies, 2758
 - SegAN, 3573
 - segmentation, 3572–3580
- aesop tactic, 2990
- affine coupling layers, 3855
- affine function, 5244
- Africa
 - weather prediction, 4227
- Agda, 3002–3003
- key features, 3002
- agent alignment, 1294–1301
 - constrained POMDP, 1297
- agent applications, 1418–1425
- agent architecture
 - performance ceiling, 2181
- agent architecture search space, 2182
- agent evaluation, 1414–1418
- agent evolution, 1390–1400
- agent loop, 1279–1284
- agent planning, 1289–1294
 - coding example, 1294
- agent platforms, 1334–1344
 - scaling, 1336
- agent reasoning, 1289–1294
- agent safety, 1294–1301
- agent security, 1407–1414
- agent self-improvement, 1287–1289
 - algorithm, 1288
- agent topologies, 1320–1334
 - chain, 1322
 - definition, 1321
 - dynamic adaptation, 1330
 - formal definition, 1321
 - graph, 1322
 - graph attention, 1332
 - message complexity, 1323
 - performance bounds, 1332
 - selection heuristics, 1329
 - star, 1322
 - LongAgent, 1327
 - task structure, 1329
 - tree, 1322
 - variants, 1322
- AgentEvol, 1287
 - definition, 1287
 - variational bound, 1447
- agentic AI, 1267–1463
 - attack surface, 1823
 - attack taxonomy, 1827

- data exfiltration, 1822
- defense architecture, 1828
- EchoLeak, 1823
- exercises, 1445–1462
- future directions, 1440–1445
- multi-agent systems, 1826
- privacy, 1822–1828
- prompt injection, 1823
- research pipeline, 1827
- self-modifying architecture, 1976
- Slack AI vulnerability, 1822
- video world model, 749
- warning, 1823
- agentic retrieval, 2631
- agentic systems
 - evolution, 2285
- aging population
 - digital access, 4696
- AGMC hypothesis
 - exercise, 4124
- agriculture
 - weather dependence, 4182
- aha moment, 3052
- AI
 - and theorem provers, 3013
- AI accelerator
 - performance trajectory, 4816
- AI accelerators
 - wearable devices, 4815
- AI alignment
 - evolutionary perspective, 2167
- AI assistant
 - mathematician, 3133
- AI competition
 - weather forecasting, 4283
- AI for Math Initiative, 3132–3133
- AI for mathematics
 - approach comparison, 3135–3137
 - DeepMind ecosystem, 3131–3146
 - landscape, 2953–2964
 - open challenges, 3141–3142
- AI safety, 4133–4152
 - alignment as arms race, 2167
 - dual-use concern, 2167
 - interpretability motivation, 4133
 - recursive self-improvement, 1988
 - scalable oversight, 4146
 - self-modifying architectures, 1968, 1977
- AI Scientist, 2285
- AI scientist, 2215–2223, 2277
 - limitations, 2221–2223
 - research loop, 2216
- AI weather forecasting
 - economics, 4391
- AI weather models
 - failure modes, 4386
 - exercise, 4393
- AI4AD initiative, 4800
- AIC, 155
- AIFS-CRPS, 4304
- AIGEFs, *see* NOAA Project EAGLE
- AIGFS, *see* NOAA Project EAGLE
- air quality forecasting
 - Aurora, 4379
- airfoil
 - flow prediction, 3247
 - Transolver, 3247
- AirfRANS dataset, 3247
- AISTATS 2017, 4450
- aleatoric uncertainty, 2878, 3410
 - disentangled, 3568
 - segmentation, 3490
- Alemi, Alex, 3034
- Aletheia, 3173–3178
 - Erdos problems, 3163
 - Feng26 paper, 3174
- Aletheia project, 3133–3135
 - methodology, 3133
- AlexNet

- 2012 revolution, [4236](#)
- GTX 580, [4931](#)
- single GPU training, [4931](#)
- ALGA-Ed, [4791](#), [4914](#)
- algebraic hierarchy, [2973](#)
- algebraic reasoning (AR), [3122](#)
- algorithm discovery
 - mathematical implications, [3132](#)
- algorithm understanding, [3996](#)
- algorithmic ableism, [4841–4853](#)
 - definition, [4842](#)
 - disparity computation exercise, [4852](#)
 - empirical evidence, [4842](#)
 - formal definition, [4842](#)
 - legal context, [4848](#)
 - mitigation, [4849](#)
 - summary, [4852](#)
- algorithmic audit
 - disability, [4851](#)
- algorithmic discovery, [2108–2118](#)
- algorithmic fairness, [2690](#)
 - definition, [3724](#)
 - dermatology, [3725](#)
 - epilepsy, [3941–3942](#)
 - medical segmentation, [3724–3725](#)
 - Pareto analysis, [3738](#)
 - sources, [3724](#)
- algorithmic transparency, [3723](#)
- algorithms
 - macroscopic, [3972](#), [3973](#)
 - neural network, [3971](#)
- aliasing, [770](#)
 - in generation, [771](#)
 - temporal, [527](#)
- ALiBi, *see* positional encoding, ALiBi, [2389](#)
 - vs RoPE, [1553](#)
- ALIGN, [2411–2412](#)
 - architecture, [2411](#)
- aligned agent
 - definition, [1297](#)
- alignment, [1123–1167](#)
 - architectural ceiling, [1975](#)
 - architecture adaptation, [1885](#)
 - architecture design, [1974](#)
 - as continual learning, [1223](#), [1241](#)
 - as ongoing process, [2263](#)
 - challenge exercise, [2296](#)
 - collective, [1433](#)
 - composability, [1457](#)
 - cross-system, [3009](#)
 - drift rate, [2273](#)
 - emergent misalignment, [1461](#)
 - evolutionary systems, [2256–2267](#), [2272–2274](#)
 - exercise, [2293](#)
 - game-theoretic exercise, [2266](#)
 - instability conjecture, [2272](#)
 - interpretability-first, [4133–4152](#)
 - multi-agent
 - open problem, [1440](#)
 - NAS for, [1976](#)
 - open problems, [1166](#)
 - pipeline design, [1166](#)
 - preservation mechanisms, [2273](#)
 - retrieval-based, [2614](#)
 - synthesis, [1159](#)
 - unified framework, [1159](#)
- Alignment Forum, [4137](#)
- alignment problem, [4133](#)
 - evolutionary, [2257–2258](#)
 - for evolutionary systems, [2257](#)
 - formal definition, [2257](#)
 - unique challenges, [2258](#)
- all-gather, [4972](#)
 - FSDP
 - latency hiding, [4988](#)
 - sequence parallelism, [5138](#)
 - ZeRO-3, [5064](#)

- all-reduce, 5031–5041
 - 7B model example, 4957
 - algorithm comparison exercise, 5040
 - beyond gradient averaging, 5031
 - binary tree, 5032
 - bucket granularity, 4991
 - communication volume, 4930
 - cost, 5135
 - diffusion training, 5090
 - fat-tree example, 4949
 - FP16 bandwidth, 5076
 - gradient, 5022
 - latency bound, 4973
 - layer-wise pipelining, 4984
 - LLM training example, 4968
 - lower bound, 5033
 - naive reduce-broadcast, 5032
 - NVSwitch efficiency, 4936
 - profiling, 5039
 - recursive halving-doubling, 5035
 - ring algorithm, 5033
 - time analysis, 5169
 - topology dependence, 4942
 - TPU torus, 4939
 - training heartbeat, 5040
 - transformer layer, 5047
- all-to-all
 - NVSwitch, 4936
 - transformer training, 4977
- all-to-all communication, 5138
 - attention, 5157
- Allen–Cahn equation, 3357–3364
 - formulation, 3357
 - PINN implementation, 3363
- alpha peak frequency
 - identifiable, 3827
- alpha waves, 3793
- alpha-beta model, 4951–4959
 - all-reduce time, 5169
 - definition, 4953
 - formal definition, 4953
 - latency vs bandwidth diagnosis, 4958
 - timeline, 4956
 - typical values, 4953
- AlphaEvolve, 2108–2118, 2221, 2286, 3131–3132
 - algorithm, 2110
 - architecture, 2108–2111
 - Borg scheduler, 2114
 - combinatorial optimisation, 2114
 - evaluation, 2116–2118
 - evolutionary database, 2109
 - FlashAttention, 2114
 - formal definition, 2109
 - index optimization, 2611
 - mathematical discovery, 3175
 - multi-block coevolution, 2111–2113
 - NAS connection, 2254
 - prompt sampler, 2109
 - quality-diversity connection, 2089
 - results, 2113–2116
 - summary of results, 2115
 - tiered models, 2110
 - TPU Verilog, 2115
- AlphaFold, 14, 2893
 - comparison with mathematical AI, 3183
 - connection to AlphaProof, 3137–3139
- AlphaGeometry, 3112, 3120–3123
 - architecture, 3121–3123
- AlphaGeometry 2, 3120–3123
 - improvements, 3123
- AlphaGo, 3042
 - architecture, 3114
 - lineage to AlphaProof, 3113–3116
- AlphaProof, 15, 2207–2208, 2535, 2961, 3034, 3112–3131

- architecture, [3116–3120](#)
- pre-training, [3118](#)
- reinforcement learning, [3118–3119](#)
- test-time compute, [3119–3120](#)
- vs AlphaGeometry, [3123–3124](#)
- AlphaTensor, [3132](#)
- AlphaZero
 - applied to proofs, [3099](#)
 - architecture, [3114–3115](#)
 - lineage to AlphaProof, [3113–3116](#)
 - training loop for proofs, [3101](#)
- ALS
 - speech loss, [4690](#)
 - voice banking case study, [4756](#)
 - voice cloning, [4746](#)
 - voice preservation, [4900](#)
- alt text
 - missing, [4692](#)
- alternating projection, [373](#)
- AltGDA, *see* gradient descent ascent, alternating
- Alzheimer’s disease
 - AI support, [4799](#)
 - segmentation, [3479](#)
- Amazon Transcribe, [4728](#)
- Ambiguity Controlling Network (ACN), [3607](#)
- ambiguity modeling
 - diffusion, [3606–3615](#)
- Ambiguity Modeling Network (AMN), [3607](#)
- ambiguity modelling
 - segmentation, [3493](#)
 - spatiotemporal, [3727](#)
- AMD
 - verification, [3005](#)
- American Community Survey, [2689](#)
- American Meteorological Society
 - award, [4231](#)
- amnesia
 - anterograde, [2340](#)
- amortised inference
 - operator learning, [3211](#)
 - OT, [3937](#)
- amortization gap, [256](#)
- amortization gap, [231–268](#)
 - bound, [232](#)
 - definition, [231](#)
- amortized inference, [230–236](#), [391–392](#), [2871](#)
 - definition, [230](#), [392](#)
 - example, [234](#)
 - VAE, [256](#)
- AMOS dataset, [3512](#)
- amplitude modulation artefact, [2793](#)
- amyotrophic lateral sclerosis, [4746](#)
- anamnesis, [2333](#)
- ANCE, [2393](#)
- ancestral sampling, [147](#)
- Anderson acceleration, [1054](#)
- Anderson’s theorem, [2771](#), [3269](#)
- Anderson–Darling test, [2851](#)
- Anemoi
 - framework, [4387](#)
- AnimateAnyone, [681](#)
- AnimateDiff, [692](#)
- AnimateLCM, [647](#)
- ANN review
 - meteorology 1999, [4234](#)
- ANN search, *see* approximate nearest neighbour search, [2375](#)
 - recall-latency trade-off, [2383](#)
- annotation
 - bias, [4135](#)
 - cost, [3483](#)
 - expert time, [3484](#)
 - formats, [3484](#)
 - inter-rater variability, [3795](#)
 - noise vs ambiguity, [3485](#)
 - variability, [3549](#)

- annotation ambiguity, 3478
 - connection to interface, 3521
 - distribution perspective, 3491
- annotation amplifier, 3520
- annotation averaging
 - suboptimality, 3498
- annotation bottleneck
 - federated generative solution, 4463
- annotation cost, 3519
- annotation crisis, 3483
- annotation interface, 3518–3521
 - historical evolution, 3521
- annotator quality, 3568
- annotator weighting, 3760
- annotator-specific predictions, 3568
- annotator-specific styles, 3568
- Annoy, 2377
- anomaly correlation coefficient, 4190–4191
- anomaly detection, 158
 - data needs, 2788
 - EEG diffusion, 3839
 - epilepsy, 3783
 - generative model foundations, 3846
 - GMM, 162
 - in-training, 5152
 - protocol, 3859
 - reconstruction principle, 3840
 - seizure prediction, 3888
 - VAE
 - MRI, 3786
- answer relevance, 2533
- ant colony optimisation, 1270–1271
 - as multi-agent system, 1273
- antenna design, 3350
 - generative models, 3350
- Anthony, Thomas, 3042
- Anthropic
 - interpretability research, 4150
- anti-seizure medication
 - failure rate, 3777
- antisymmetry
 - wavefunctions, 3365
- any-to-any retrieval, 2471
 - formal definition, 2471
- APAC-Net, 3873
 - algorithm, 3873
- aphasia, 4690
 - post-stroke, 4746, 4747
- APOLLO, 3090–3091
 - agents, 3090
 - convergence, 3095
 - results, 3090
- Appa
 - model, 4343–4344
 - storm reconstruction, 4344
- appearance flow, 746
 - definition, 746
- Appel, Kenneth, 3029
- applications of generative models, 10
- applied behaviour analysis (ABA)
 - AI, 4806
- approximate density, 7
- approximate nearest neighbour
 - exercise, 2633
- approximate nearest neighbour search, 2375–2385
- AR smart glasses
 - CVD correction, 4783
- ARC benchmark, 2186
- Archimedean property, 5192
- architectural constraints, 2703
- architectural limitations
 - vs. data limitations, 3048
- architectural plasticity
 - Mamba, 4128–4129
- architectural safety, 1975
- architecture
 - as first-class variable, 1980
 - fixed, 2025

- input-dependent, 1955
- tokenisation, 1946
- architecture capacity
 - bound, 1864
 - information-theoretic limits, 1863–1868
- architecture change
 - forgetting bound, 1883
- architecture consolidation, 1964–1969
 - definition, 1964
 - efficiency bound, 1965
- architecture convergence conjecture, 1963
- architecture DAG, 1896
- architecture design
 - exercise, 3264
 - LLM prompt example, 1951
 - LLM-based, 1951
- architecture diffusion
 - convergence, 1950
- architecture evaluation
 - LLM-based, 1952
- architecture evolution
 - weather forecasting, 4241
- architecture generation
 - diffusion models, 1949
 - problem formulation, 1950
- architecture heterogeneity, 4579–4589
- architecture parameters, 1900
- architecture plasticity, 1882
- architecture prior, 1865
- architecture selection
 - epilepsy, 3937
 - PDE Transformers, 3263
- architecture space
 - cardinality, 1898
 - mathematical structure, 1897–1910
- architecture suboptimality, 1936
- architecture-based methods, 1198–1204
- architecture-memory connection, 1968
- architecture-performance landscape, 1983
- Archive of Formal Proofs, 3001, 3025
- arcuate fasciculus, 3890
- area weighting
 - spherical grid, 4311
- argument from design, 2020
- Aristotle
 - association, 2334
- Aristotle (Harmonic), 3152
- arithmetic
 - LLM failures, 3047
- arithmetic intensity
 - definition, 4936
 - exercise, 4941
 - formal definition, 4936
 - GenAI operations, 4938
- ARM processors
 - accessibility, 4815
- arms race
 - evolutionary, 2058
 - in Tierra, 2240
- array theory, 3023
- artificial life, 2195, 2237–2249
 - connection to meta-evolution, 2195
 - formal definition, 2241
 - foundations, 2239
 - history, 2239
- artificial neural network
 - meteorology, 4232–4243
- ASD
 - gamified learning, 4806
 - social mediation, 4797
- ASL
 - datasets, 4727
 - gloss translation, 4912
 - vocabulary coverage, 4732
- ASL avatar, 4880–4882
- ASL gloss, 4728
 - limitations, 4729

-
- ASLLVD, [4727–4728](#)
 - description, [4727](#)
 - aspect ratio
 - training, [656](#)
 - ASR
 - error compounding, [2541](#)
 - assignment problem, [367–369](#)
 - definition, [368](#)
 - assistive AI ecosystem, [4857](#)
 - definition, [4857](#)
 - assistive technology
 - definition, [4692](#)
 - generative models, [4689](#)
 - low-income countries, [4696](#)
 - market size, [4696](#)
 - associative memory, [1077](#), [2334](#)
 - capacity, [2346](#)
 - associative recall, [4170](#)
 - exercise, [4132](#)
 - associative recall (AR) task, [4126–4127](#)
 - definition, [4126](#)
 - formal definition, [4126](#)
 - associative treecall (ATR) task, [4127](#)
 - formal definition, [4127](#)
 - associativity
 - of addition, [2968](#)
 - asymmetric distance computation, [2378](#)
 - asymmetric embeddings, [2462](#)
 - asymmetric fusion, [4081](#)
 - asymmetric representation, [4135–4136](#)
 - asymmetry loss
 - PET synthesis, [3858](#)
 - async checkpointing, [5148](#)
 - asynchronous All-Reduce, [5132](#)
 - asynchronous federated learning, [4451](#), [4481](#)
 - asynchronous FL, [4581](#)
 - definition, [4581](#)
 - asynchronous SGD, [4931](#)
 - asynchronous training, [5022](#)
 - Atari games, [2054](#)
 - atlas-based segmentation, [3487](#), [3528–3530](#)
 - multi-atlas label fusion, [3529](#)
 - atmospheric blocking
 - superensemble, [4322](#)
 - atmospheric equations, [4202–4221](#)
 - coordinate systems, [4203](#)
 - atmospheric layers, [4208](#)
 - atmospheric phase space
 - diffusion prior, [4321](#)
 - atmospheric prediction, [4182–4192](#)
 - atmospheric prediction pipeline
 - diffusion, [4318](#)
 - atmospheric river
 - FuXi forecast, [4282](#)
 - atmospheric rivers
 - machine learning, [4401](#)
 - atmospheric science, [4181–4202](#)
 - atmospheric stability
 - bifurcation, [4230](#)
 - atmospheric state vector, [4188–4192](#)
 - definition, [4188](#)
 - exercise, [4192](#)
 - ATP invocation, [3025](#)
 - AttenMIA, [1754–1762](#)
 - attention entropy, [1756](#)
 - attention visualization, [1759](#)
 - concentration score, [1757](#)
 - entropy bound, [1757](#)
 - full score, [1758](#)
 - score function, [1758](#)
 - attention
 - as explanation, [3988](#), [3989](#)
 - as kernel integral, [3236](#)
 - as memory, [1076–1081](#)
 - computational cost
 - video, [548](#)
 - deep content-focused, [4107–4113](#)
 - definition, [101](#)

- diagnostic utility, 3989
- distributed partitioning, 4927
- distributed variant, 5178
- efficient, 1511–1516
- entropy
 - exercises, 4113
 - proof, 4109
- entropy comparison, 4109
- faithfulness, 3988
- federated unlearning, 4604
- Flash Attention
 - extensions, 1515
- gated, 111
- global workspace analogy, 4149
- grouped-query, 107
- head parallelism, 5045
- heterogeneous, 3242
- information routing, 4012
- IO-aware, 109
- linear, 1514
 - definition, 1514
- local window, 1512
- masked diffusion models,
 - 4104–4113
- memory bottleneck, 4937
- memory scaling, 4932
- multi-head, 101
- multi-query, 107
- not explanation warning, 3989
- quadratic complexity, 3237
- ring, 1513
 - communication, 1514
- ring attention, 5011
- roofline analysis, 4937
- scaled dot-product, 101, 3234
- scaling, 101
- shallow structure-aware, 4107–4113
- sliding window, 1512
- sparse, 1511
 - complexity, 1512
 - strided, 1512
- attention decomposition, 4012–4020
- attention floating, 4104–4113
 - definition, 4107
 - formal characterisation
 - exercises, 4113
 - intuition, 4107
 - key insight, 4105
 - MDM, 4106–4113
- attention flooding, 4169
- attention gates, 3539–3541
 - analysis, 3548
 - mechanism, 3539
 - spatial attention, 3539
- Attention Gathers MLPs Compose
 - hypothesis, 4117–4118
 - definition, 4117
- Attention Gathers, MLPs Compose
 - hypothesis, 4016–4017
 - key idea, 4017
- attention head
 - additive logit contributions, 4015
 - additive unit, 4003
 - importance score, 1944
 - independent, 4003
 - interpretability key idea, 4016
 - logit space additivity, 4014
 - previous token, 4021
 - spatiotemporal specialisation, 4118
- attention heads
 - self-explanation, 4145
- attention injection, 714
 - definition, 715
- attention maps
 - epilepsy interpretation, 3942
- attention matrix
 - forget direction, 4605
- attention mechanism, 2629
 - brain networks, 3869
 - brain regions, 3882

- integral operators, 3452
 - PDE solving, 3233–3264
- attention pattern
 - interpretation, 4013
 - QK circuit, 4004
 - softmax, 4013
- attention pattern visualisation, 4119
- attention patterns
 - broadcast, 4149
- attention sink, 1487
- attention sinks, 4105–4113, 4169–4170
 - definition, 4105
 - lost-in-the-middle, 4106
 - origin, 4106
- Attention U-Net, 3487, 3539–3541
 - brain tumour segmentation, 3541
 - motivation, 3539
- attention visualisation, 3988–3989
- attention window
 - consistency, 708
 - consistency guarantee, 708
- attribute inference, 2705, 2844–2846
- attribution, 2532–2533
 - formal definition, 2532
 - limitations, 2533
 - signal patches
 - EEG, 3943
- attribution accuracy, 2534
- attribution chain problem, 3993
- attribution methods
 - path-based, 3987
- attribution problem
 - generative models, 3991–3992
 - generative models key idea, 3992
- atypicality
 - exercises, 1669
 - membership statistic, 1664
- AUC
 - membership inference, 4597
 - seizure detection, 3923
- AUC-ROC, 1585, 3809
 - definition, 3810
- audio autoencoder
 - variational perspective, 940
- audio captioning
 - retrieval-augmented, 2542
- audio codec, *see* neural audio codec
- audio deepfake detection
 - definition, 929
- audio deepfakes, 939
- audio detection, 929–937
 - arms race, 936
 - features, 935
 - fundamental limit, 935
 - generalisation, 936
 - problem formulation, 929
 - statistical methods, 934–936
 - threat models, 930
- audio diffusion
 - classifier-free guidance, 956
 - composite loss, 853
- audio editing, 902–904
 - instruction-based, 903
- audio embeddings, 2470
- audio evaluation metrics, 915–923
 - CLAP score, 919
 - disagreement, 923
 - MOS, 920
 - summary, 922
- audio fingerprinting, 2438–2441
 - collision probability, 2441
 - definition, 2439
 - exercises, 2446
 - matching algorithm, 2440
- audio forensics
 - history, 937
- audio generation, 16, 765–959
 - applications, 765
 - commercial systems, 910–915
 - comparison with images, 766

- compress-diffuse-decode, 833
- connections to other chapters, 940–944
- cross-modal, 897
- data challenges, 855
- diffusion architectures, 841–847
- diffusion process, 848–852
- dimensionality, 765
- diversity, 926
- ethics, 937–940
- evaluation, 915–923
- exercises, 949–959
- history, 766
- latency, 799, 945
- latent diffusion, 832–841
- legal status, 938
- motivation, 765
- musicality, 923–927
- open problems, 945–949
- representation bottleneck, 766
- retrieval-conditioned, 2542
- timeline, 768
- training, 853–862
- two-stage pipeline, 799
- unified view, 765
- video parallel, 943
- audio inpainting, 902–904
 - boundary artefacts, 903
 - definition, 902
 - posterior sampling, 904
- audio provenance, 939–940
 - open problem, 948
- audio representation
 - compression hierarchy, 778
 - design choice, 779
 - dimensions, 779
 - taxonomy, 778
- audio signal processing, 769–779
- audio watermarking, 929–937
- audio-language retrieval, 2447–2457
 - examples, 2454
- audio-native RAG, 2541
- audio-text alignment, 803
- audio-text pairs
 - scarcity, 855
- audio-text retrieval
 - benchmarks, 2601
- audio-visual generation, 744
- AudioCaps, 807
- AudioLCM, 875–880, 953
 - speedup, 876
 - text-to-audio, 877
- AudioLDM, 767, 835
 - inference, 837
 - training objective, 836
- AudioLDM2, 837–839
 - conditioning expressivity, 839
 - unified framework, 839
- AudioMAE, 837
 - definition, 838
- AudioSeal, 932–934
 - detection error, 933
- auditing
 - federated generative models, 4651–4652
- augmentative and alternative communication, 4690, 4758–4769, 4917
 - definition, 4760
- augmented reality
 - CVD correction, 4779–4789
- AUPRC, 3810
- AURA platform, 3899–3903
 - architecture, 3899
 - evaluation metrics
 - exercise, 3909
 - figure, 3901
 - patient cohort, 3899
 - pipeline, 3900
- AURA system, 3913

- AUROC, 3810
 - detection metric, 2843
- Aurora
 - Earth system foundation model, 4377
 - extreme event evaluation, 4401
 - Perceiver architecture, 4378
 - PM2.5 air quality, 4379
- autism spectrum conditions, 4690
- autism spectrum disorder (ASD), 4789
- auto-differentiation
 - pipeline stages, 5176
- auto-tuning
 - parallelism, 5143
- AutoAgents, 1340
- autocorrelation, 2696
- autoencoder, 385–386
 - definition, 386
 - exercise, 399
 - linear, 386–387
 - definition, 387
 - PCA equivalence, 387
 - exercise, 399
 - sparse, 1310
- autoformalization, 2959–2960, 3039, 3064–3080, 3085, 3117–3118
 - ambiguity exercise, 3077
 - challenges, 3065–3066
 - definition, 2959
 - difficulty levels, 3065
 - difficulty prediction, 3078
 - early approaches, 3066
 - early work, 2960
 - error analysis, 3077
 - examples, 3072–3074
 - future directions, 3078
 - GPT-4 example, 3067
 - hard example, 3073
 - key limitations, 3074
 - limitations, 3074–3075
 - LLM-based, 3067
 - medium example, 3073
 - Mizar, 3003
 - outlook, 3078–3080
 - pipeline, 2960, 3076
 - pipeline definition, 3076
 - pre-LLM, 3039
 - simple example, 3072
 - success rates, 3074–3075
 - synthetic data, 3071–3072
 - two-way bridge, 3069–3070
 - verification, 2960
- AutoGen, 1325–1327
 - coding example, 1326
 - group chat, 1326
 - nested conversations, 1326
- autograd
 - computation graph, 4989
- autograd engine
 - distributed, 5100
 - execution order, 5100
- automated circuit discovery, 3975
- automated design of agentic systems, 2181–2190
- automated discovery, 2755
- automated induction, 3005
- automated reasoning, 15
 - aesop, 2990
 - history, 2937–2942
- automated scientific discovery, 2277
 - challenge exercise, 2297
 - sufficiency conjecture, 2278
 - testbeds, 2278
- automated segmentation
 - quantitative impact, 3489
- automated theorem provers, 3016
- automated theorem proving, 3015–3031
 - limitations, 3028–3031
 - symbolic, 2954
 - symbolic methods, 2955–2956

- Automath, [2944](#)
- automatic differentiation, [95](#)
 - distributed, [5097–5162](#)
 - in PIDMs, [3278](#)
 - PDE residuals, [3205](#)
 - single device, [5098](#)
- automatic indexing, [2626](#)
- automatic speech recognition,
[2434–2435](#)
 - contextual captions, [4701](#)
- autonomous AI
 - device-level memory, [4630](#)
- autonomous evolution
 - limits, [2256–2267](#)
- autonomous mathematical research,
[3173–3178](#)
 - future, [3180](#)
- autonomous research, [2215–2223](#)
 - loop, [2215–2217](#)
- autonomous research agent, [2215](#)
- autoregressive
 - propagation pathway, [3894](#)
- autoregressive chunk generation,
[698–706](#)
 - definition, [698](#)
 - error accumulation, [699](#)
 - exercise, [705](#)
- autoregressive chunking
 - audio, [884–888](#)
 - parameter selection, [886](#)
- autoregressive diffusion, [3281](#)
- autoregressive error accumulation, [4257](#)
- autoregressive factorization
 - tabular, [2805](#)
- autoregressive flow, [456–458](#)
 - Jacobian, [457](#)
- autoregressive forecasting
 - 12-hour steps, [4313](#)
- autoregressive generation
 - as recurrence, [1029](#)
 - chunk overlap, [760](#)
 - computational bottleneck, [814](#)
 - cost analysis, [815](#)
- autoregressive model, [4](#), [7](#), [57](#), [427–448](#)
 - attention, [103](#)
 - definition, [428](#)
 - EEG, [3888](#)
 - EEG exercise, [3898](#)
 - factorisation, [428](#)
 - modern developments, [444–445](#)
- autoregressive modelling, [421](#)
- autoregressive models
 - attribution problem, [3993](#)
 - audio, [767](#)
 - audio generation, [943](#)
 - causal masking, [4105–4113](#)
 - mathematical reasoning, [3047](#)
 - sequential dependency, [5081](#)
 - time series, [2801](#)
 - vs. discrete diffusion, [4538–4549](#)
- autoregressive music generation,
[809–816](#)
- autoregressive prior, [415](#)
- autoregressive rollout, [4189](#)
 - error accumulation, [3324](#)
 - training, [4269](#)
- autoregressive tactic generation, [3128](#)
- autoregressive-diffusion hybrid,
[726–733](#)
 - architecture, [731](#)
 - comparison, [730](#)
 - convergence, [730](#)
 - exercise, [731](#)
 - frame-level, [726](#)
 - definition, [727](#)
 - information bottleneck, [732](#)
 - latency, [727](#)
 - systems, [728](#)
 - temporal AR, spatial diffusion, [728](#)
 - definition, [728](#)

- tokenisation exercise, [733](#)
- auxiliary classifier, [2732](#)
 - loss, [2734](#)
- auxiliary construction
 - geometry, [3121](#)
- auxiliary loss
 - load balancing, [5142](#)
- average accuracy, [1178](#)
- average precision, [2359](#), [3505](#)
- aviation
 - weather, [4182](#)
- Avida, [2241](#)
 - metabolic rewards, [2241](#)
- AWQ, [1537](#)
- AWS
 - EFA, [4945](#)
- AWS EFA, [4968](#)
- Axelrod, Robert, [2281](#)
- axial self-attention
 - MetNet, [4240](#)
- Axiom Math, [3150–3151](#)
- axiom of choice, [2974](#)
- AXLE (Axiom Lean Engine), [3150](#)
- axonal routing, [1854](#)
- B200
 - Blackwell architecture, [4935](#)
 - specification, [4935](#)
 - vs H100 comparison, [4939](#)
- Baars, Bernard, [4148](#)
- back-translation
 - autoformalization, [3071](#)
 - mathematics, [3071](#)
 - quality exercise, [3077](#)
- backjumping, [3018](#)
- backpropagation, [95](#)
 - activation storage, [5070](#)
 - early meteorology, [4232](#)
 - Rumelhart 1986, [4233](#)
 - through time, [100](#)
- backpropagation network
 - minimum temperature, [4235](#)
- backup replicas, [5150](#)
- backward pass
 - gradient independence across layers, [4989](#)
 - overlap with communication, [5025](#)
- backward stochastic differential
 - equations, [3439](#)
- backward transfer, [1178](#)
- bag of visual words, [2400–2406](#)
 - definition, [2400](#)
 - kernel interpretation, [2401](#)
 - vocabulary size, [2405](#)
- bag-of-words
 - limitations, [2364](#)
- Baker, [16](#)
- BalancedBaggingClassifier, [3807](#)
- Baldwin effect, [2034–2036](#)
 - landscape smoothing, [2035](#)
 - meta-learning connection, [2035](#)
- Baldwin, James Mark, [2034](#)
- ball tree, [2377](#)
- Banach contraction mapping theorem, [1032](#)
- Banach fixed point theorem, [5235](#)
- Banach fixed-point theorem, [459](#)
- Banach space, [5214](#)
 - operator learning, [3211](#)
- banana distribution, [239](#)
- bandit-based resource routing, [2198](#)
- bandwidth
 - centralisation bottleneck, [4449](#)
 - communication, [4951](#)
 - communication bottleneck, [4930](#)
 - distributed vs federated, [5119](#)
 - effective vs raw, [4943](#)
 - latency trade-off, [5040](#)
- bandwidth bottleneck
 - exercise, [1113](#)

- bandwidth hierarchy, [4927](#)
- bandwidth optimality
 - ring all-reduce, [4955](#)
- bandwidth ratio
 - intra vs inter node, [4928](#)
- bandwidth-bound
 - all-reduce, [4954](#)
 - crossover, [4956](#)
- bandwidth-optimal collective, [5033](#)
 - definition, [5033](#), [5035](#)
- Bansal, Kshitij, [3036](#)
- Banzhaf, Wolfgang, [2282](#)
- Barenblatt solution, [3356](#)
- Barlow, H. B., [3969](#)
- barotropic vorticity equation, [4210](#)
- Barricelli, Nils Aall, [2279](#)
- barrier synchronisation
 - hiding, [4987](#)
- BART, [2521](#)
- base LLM
 - task vector transfer, [4080](#)
- base models
 - introspection, [4143](#)
- base rate fallacy, [1607](#)
- baselines
 - inconsistent, [2835](#)
- basis (topology), [5216](#)
- batch conditioning, [2783](#)
- batch normalisation, [96](#), [2715](#)
 - deep training, [4242](#)
 - information loss, [355](#)
 - loss landscape smoothing, [1489](#)
- batch size
 - gradient inversion, [4592](#)
 - gradient variance, [5110](#)
 - learning rate scaling, [5153](#)
- batch size scaling, [5026](#)
- Bath (England)
 - IMO 2024, [3112](#)
- bathymetry, [3320](#)
- Bayes by Backprop, [257](#)
- Bayes theorem, [38](#)
- Bayes' rule
 - conditional score, [2776](#)
- Bayes' theorem
 - GMM, [150](#)
- Bayesian continual learning, [1179–1182](#)
 - definition, [1179](#)
 - intractability, [1180](#)
 - optimality, [1180](#)
- Bayesian Gaussian mixture model, [192](#)
 - specification, [192](#)
- Bayesian GMM, [158](#)
- Bayesian inference, [46](#)
 - for inverse problems, [329](#)
 - inverse problems, [3399](#)
- Bayesian inverse problem, [3299](#)
- Bayesian inversion, [3399](#)
- Bayesian learning rule, [251](#)
- Bayesian linear regression, [187](#)
 - exercise, [201](#)
- Bayesian logistic regression, [223](#)
 - SVGD, [248](#)
- Bayesian Model Averaging, [4228](#)
 - BMA, [4228](#)
- Bayesian network, [126](#)
 - construction, [142](#)
 - definition, [126](#)
 - medical diagnosis, [127](#)
 - meteorology, [4234](#)
 - probabilistic weather forecasting, [4235](#)
- Bayesian networks
 - synthetic data, [2711](#)
- Bayesian neural operators, [3229](#)
- Bayesian neural networks
 - VI, [257](#)
- Bayesian neural operators, [3411](#)
- Bayesian optimisation, [2226–2229](#)
- Bayesian source separation, [890](#)

- Bayesian structure learning, 2873
- BBVI
 - logistic regression, 265
- BCI
 - generative language models, 4868
- BCI Competition IV
 - motor imagery, 3824
- Beacon selection protocol, 1338
- Beale–Kato–Majda criterion, 3429
- beam blockage, 4363
- beam search
 - tactic generation, 3128
- beamforming
 - federated generative synthesis, 4624
- beat synchronisation, 839
- beat-synchronous mixup, 956
- Bedrock Claude, 4728
- behaviour descriptor
 - design, 2087
- behaviour space
 - discretisation, 2084
- behavioural distance, 2082
- BEIR, 2596–2597
 - domain transfer, 2607
 - evaluation protocol, 2596
- BEIR benchmark, 2371
- Belew, Richard, 2282
- belief propagation, 134, 136, 251–255
 - algorithm, 252
 - Bethe approximation, 252
 - exact on trees, 254, 267
 - exercise, 143
 - fixed points, 253
 - loopy, 136
- belief state, 1280
- Benamou–Brenier formula, 3872
- benchmark
 - cross-disability, 4910
 - fine-grained discrimination, 2606
 - generative epilepsy models, 3952
 - theorem proving, 3036
- benchmark contamination, 2605
 - detection, 2607
 - example, 2606
- benchmark datasets, 22, 2854–2855
- benchmark overfitting
 - accessibility, 4893
- benchmark saturation, 2605
- benchmark treadmill, 2606
- benchmarking, 2854–2855
 - accessibility AI, 4885–4897
 - multi-agent, 1414–1418
- benchmarks
 - design exercise, 3425
 - design principles, 3423, 3737
 - disability sector, 4907
 - inconsistency, 2907
 - InverseBench, 3406
 - limitations, 2837
 - metric monoculture, 3737
 - multimodal prompting, 3736
 - PDE, 3416–3427
 - PDEBench, 3416
 - PINNacle, 3418
 - single-prompt limitation, 3737
 - static evaluation, 3737
- Bengio, Y., 3962
- Berger, Hans, 3775
- Bernoulli distribution, 30
 - exponential family form, 205
 - MLE, 56
- BERT, 2366, 2629
 - accessibility, 4701
 - for ranking, 2589
 - for reranking, 2488
 - in DPR, 2366
 - parameter count, 4926
- BERTScore, 4145
 - clinical text, 3924
- best-first search

- complexity, 3094
- in proof automation, 2990
- theorem proving, 3041–3042, 3083
- best-of-S, 3496
- Beta distribution, 34
- beta waves, 3793
- β – λ correspondence, 1242
- beta-TCVAE
 - EEG, 3943
- beta-VAE, 397, 4515–4516
 - EEG, 3824
 - exercise, 400
 - federated scheduling, 4515–4516
- β -VAE, 256, 404
 - segmentation, 3562
- Bethe entropy, 253
- Bethe free energy, 251–255
 - definition, 253
 - stationary points, 253
- betweenness centrality
 - epilepsy, 3869
- Bezout’s identity, 2976
- BF16
 - exponent range, 5108
 - gradient communication, 5037
 - practical advantage, 5075
 - vs FP16, 5075
- BFloat16
 - vs FP16, 1556
- bfloat16, 5075, 5108
- BGE embedding, 2387
- bi-encoder, 2365, 2629
 - architecture search, 2611
- bi-level optimisation, 3708
 - exercise, 3714
- Bi-LSTM
 - WGAN-GP discriminator, 3820
- bias, 19, 44
 - accessibility AI, 4889
 - circuit problem, 4111
 - key insight, 4111
- circuit surgery
 - exercises, 4114
- diffusion models, 4111–4113
- disability, 4699, 4841–4853
- feature identification, 4171
- features
 - identification, 4111
- frequency, 4372
- mechanistic intervention, 4174
- nationality, 4138
- steering
 - experimental results, 4112
- steering vector, 4112
- synthetic data, 2856–2858
- bias amplification, 2910
 - federated, 4650–4651
 - mode collapse, 2910
 - synthetic data, 2910–2912
- bias audit
 - disability, 4915
- bias auditing, 4137–4138
- bias benchmark
 - disability, 4920
- bias correction
 - MOS, 4225
 - systematic, 4223
- bias detection
 - vision models, 4093
- bias field, 3524
- bias-variance tradeoff
 - meteorology, 4236
- BIC, 155
 - exercise, 161
- biconjugate theorem, 5258
- bidirectional attention, 116
- bidirectional LSTM
 - EEG encoding, 3855
- bidirectional vulnerability, 1695
 - theorem, 1695

- bifurcation
 - vascular, 3513
- BigBird, 1087
 - connectivity, 1087
 - exercise, 1110
- BikeButler, 4719
- bilateral filter, 330
- bilevel optimisation, 1875, 1921
 - convergence, 1887
 - DARTS, 1922
 - linear example, 1879
 - NAS, 1886–1891
- bilevel optimization
 - single-level reduction, 1994
- bilinear game, 282, 309
 - general, 324
 - Jacobian, 315
 - SimGDA analysis, 312
- bilinear interpolation
 - differentiable, 4240
- binary cross-entropy, 275
- binary independence model, 2351–2352
- binary mask
 - federated learning, 4599
- binary tree broadcast, 4954
- binary trees
 - inductive definition, 2986
- binomial distribution, 31
- biofluid simulation, 3731
- biological learning, 1175
- biological neural networks
 - as retrieval systems, 2623
- biomarker
 - identifiable EEG, 3826
- BiomedCLIP, 3520, 3670–3671
 - quantitative improvements, 3671
 - zero-shot radiology, 3670
- biometric data
 - edge processing, 4814
- bipartite matching
 - cell tracking, 3654
 - exercise, 3664
- Birkhoff polytope, 1500
 - exercise, 382
- bisection bandwidth, 4943, 5170
 - definition, 4943
 - exercise, 4950
 - interpretation, 4944
- bistability
 - epilepsy, 3892
- bit depth, 770
- bit vector theory, 3023
- BKKKZ26 result, 3134–3135
- black box, 3962
 - neural network interpretability, 4234
- black box problem, 3962
 - definition, 3962
- black hole imaging, 3406
- black-box access, 1583
- black-box attacks
 - adversary definition, 1660
 - atypicality weighting, 1664
 - challenge problem, 1669
 - conditioning mismatch, 1668
 - conditioning optimisation, 1665
 - diffusion models, 1659–1670
 - exercises, 1668
 - generation-based signals, 1660
 - large diffusion models, 1663
 - large models, 1666
 - limitations, 1666
 - multi-scale querying, 1665
 - perceptual metrics, 1668
 - practical considerations, 1666
 - query efficiency, 1661
 - query-efficient, 1664
 - score-based, 1662
 - scoring API, 1660
 - threat model, 1660

- black-box detection, 930
- black-box variational inference, 219–224
 - algorithm, 222
 - logistic regression, 223
- blending
 - raised cosine, 2792
- blending function, 2792
- BLEU, 22
 - accessibility evaluation, 4887
- BLEU score
 - clinical text, 3924
- blind and low vision, 4706–4716
- blind navigation
 - hallucination, 4919
 - offline, 4818
- BLIP, 2412–2414
 - caption bootstrapping, 2413
 - captioning as denoising, 2413
- BLIP model, 4710
- BLIP-2, 2412–2414
 - efficiency, 2414
 - Q-Former, 2413
- bloat problem, 2045
- blockchain
 - provenance tracking, 5130
- blocking
 - ensemble spread, 4229
- blocking communication
 - MPI, 4962
- blocking synchronization
 - CG inner loop, 4986
- blood cell
 - synthetic generation, 3587–3589
- blood–brain barrier, 3483
- blow-up
 - 3D Navier–Stokes, 3473
- BLT-DP-FTRL, 4626
- Blue Gene
 - 3D torus, 4945
- blur
 - nowcasting, 4369
- blurring
 - vs uncertainty, 4301
- blurring problem, 4194–4202, 4297–4307
 - figure, 4432
 - theoretical explanation, 4298
- BM25, 2351–2354, 2364, 2627
 - continued relevance, 2362
 - exercise, 2631
 - limiting cases, 2353
 - on BEIR, 2597
 - retrieval algorithm, 2361
 - scoring function, 2353
- BM3D, 330
- BMA
 - comparison, 4241
- Boltzmann policy, 3057
- Bolzano–Weierstrass theorem, 2992–2995
- Bolzano–Weierstrass theorem, 5201
- Bonn EEG
 - limitations, 3921
- Bonn EEG dataset, 3921
- book-keeping technique, 1677
- Boole, George, 3016
- Boolean algebra, 3016
- Boolean circuit, 1036
- Boolean Pythagorean triples, 2955
- Boolean reflection, 3000
- Boolean variables, 3017
- bootstrap, 2212
- bootstrap confidence intervals, 2212
- Borg scheduler, 2114
- Borgeaud, Sebastian, 2630
- Borges principle, 2335
- Borges, Jorge Luis, 2334–2335
- Born–Oppenheimer approximation, 3364
- Born-Again Networks, 1534
- Borst and Theunissen, 1858

- bottleneck, 386
 - distributed training, 4929
- boundary, 5197
- boundary ambiguity, 3490
- boundary conditions
 - atmospheric equations, 4208
 - Dirichlet, 3296
 - Neumann, 3296
- boundary dissolution, 2187
- boundary hallucination, 3721
- boundary layer
 - NWP parameterisation, 4236
- boundary precision
 - adversarial training, 3574
- boundary uncertainty, 4290
- boundary-aware segmentation, 3604
- bounded model checking, 3023
- bounded sequence, 2993
- bounding box annotations, 3520
- bounding box prompts, 3519
- bowling outcome prediction (example), 4121–4122
- Box–Muller transform, 53
- Boyer–Moore theorem prover, 3004
- BPTT (backpropagation through time), 4129
- Bradley–Terry model, 1130–1133
 - definition, 1130
 - exercise, 1161
- Bradley–Terry model, 4135
- Braille
 - history, 4693
 - Louis Braille, 4693
- brain
 - cranial volume, 1854
 - information throughput, 1859
 - metabolic cost, 1853
 - physical constraints, 1853
- brain digital twin
 - definition, 3931
- brain dynamics
 - advanced OT, 3869–3879
- brain fingerprint
 - EEG, 3827
- brain MRI
 - privacy-preserving synthesis, 3590
 - scanner harmonisation, 3583
 - synthesis, 3589–3590
- brain network dynamics, 3862–3868
- brain segmentation, 3511–3512
- brain state
 - topological, 3865
- brain structure
 - transformers, 3889–3896
- brain tumour
 - automated diagnosis, 3907
 - classification, 3905
 - SegAN, 3575
 - segmentation, 3482
 - surgical planning example, 3483
- brain tumour MRI
 - federated generation, 4612
 - five-hospital federation, 4612
- brain tumour segmentation, 3511
 - Attention U-Net, 3541
 - BraTS, 3552
 - example, 3516
 - federated, 3720
 - PHiSeg, 3565
- brain-computer interface, 4868
 - LLM-augmented, 4868
 - retrieval, 2481
- branch-and-bound, 3023
- branched optimal transport, 3870–3871
 - branching exponent
 - exercise, 3878
- branching
 - DPLL, 3018
- branching factor
 - Go vs chess, 3114

- theorem proving, [3116](#)
- BraTS
 - reproducibility model, [3927](#)
- BraTS challenge, [3483](#), [3511](#), [3552–3571](#)
 - example, [3516](#)
 - inter-rater variability, [3484](#)
- BraTS dataset, [3920](#)
- BraTS2021
 - Hilbert-VLM results, [3905](#)
- breast tumour segmentation
 - complete pipeline, [3705](#)
- bred vector
 - definition, [4288](#)
 - vs singular vectors, [4288](#)
- Bregman divergence, [205](#), [5271](#)
- Bregman projection, [373](#), [5273](#)
- Bremermann, Hans-Joachim, [2280](#)
- Brenier map
 - real-time, [3937](#)
- Brenowitz et al. 2025, [4302](#)
- Bricken, T., [3969](#)
- Brier Score
 - heatwave, [4316](#)
- brightness constancy equation, [528](#)
- Brin, Sergey, [2628](#)
- broadcast
 - alpha-beta cost, [4954](#)
 - parameter synchronisation, [4988](#)
- bromide
 - epilepsy treatment, [3775](#)
- Brooks, Harold E., [4234](#)
- Broyden's method, [1054](#)
- BSDE
 - deep methods, [3439](#)
- BTCV dataset, [3512](#)
- bubble fraction
 - computation, [5059](#)
- Buchberger's algorithm, [3032](#), [3033](#)
- bucketing, [656](#)
- bulk image accessibility, [4877](#)
- bundle classification
 - white matter, [3891](#)
- Burgers equation, [3311–3319](#), [3416](#)
 - exact solution, [3314](#)
 - FNO implementation, [3319](#)
 - latent flow matching, [3290](#)
 - neural operators, [3316](#)
 - optimal control, [3317](#)
 - PDEformer graph, [3259](#)
 - PINN, [3204](#)
 - PINNsFormer, [3262](#)
 - Rankine–Hugoniot, [3312](#)
 - shock formation, [3312](#)
 - small viscosity challenge, [3316](#)
 - viscous, [3311](#)
- Burgers' equation
 - Cole–Hopf transformation, [3454](#)
 - FNO implementation, [3461](#)
 - PINN loss, [3451](#)
- Burges, Christopher, [2628](#)
- Bush, Vannevar, [2349](#), [2626](#)
- butterfly all-reduce
 - cost, [4954](#)
 - hypercube, [4946](#)
- butterfly attention, [5157](#)
- butterfly effect, [4183](#), [4185](#)
- Buzzard, Kevin, [2950](#), [3033](#)
- ByT5, [3083](#)
- Byzantine fault tolerance, [1383](#)
 - exercise, [1454](#)
- C-CLIP, [1202](#), [1225](#)
- C2PA, [19](#)
 - audio, [939](#)
 - manifest, [939](#)
- CA-CG, [5003](#)
 - algorithm, [5004](#)
 - communication savings, [5004](#)
 - numerical stability, [5004](#)

-
- CAB, *see* Cross-Attention Alignment Block
 CaDiCaL, 3018
 Cahn–Hilliard equation, 3357–3364
 conservation properties, 3358
 formulation, 3358
 image inpainting, 3363
 PINN challenges, 3360
 Cajal, Santiago Ramón y, 2004
 Cajal, Santiago Ramon y, 1849
 CakeML, 3004
 Calculus of Constructions, 2999
 Calculus of Inductive Constructions, 2944, 2948–2949, 2984
 calculus of variations, 168
 mean-field, 185
 calculus ratiocinator, 2938, 3016
 calibration
 base models, 4143
 marginal vs. joint, 4329
 nowcasting, 4371
 segmentation, 3496
 synthetic PET, 3860
 uncertainty, 3413
 Callimachus, 2338, 2625
 CALM, 2196–2203, 2286
 algorithm, 2197
 comparison with standard RL, 2198
 connection to ADAS, 2188
 Cambrian explosion
 computational analogue, 2165
 Cambridge University Press
 environmental ML, 4236
 camera conditioning
 exercise, 759
 camera control, 675–683
 conditioned diffusion, 677
 definition, 677
 Plücker coordinates, 676
 ray diagram, 678
 representations, 675
 types, 680
 CameraCtrl, 677
 Cameron–Martin theorem, 3380
 cancer immunotherapy
 cell tracking, 3662
 candidate generation, 2575
 Cantor, Georg, 3142
 cap set problem, 2285, 3175
 capability
 interpretability tradeoff, 4160
 capability ceiling, 1305–1308
 definition, 1305
 capability ledger, 1338
 capacity allocation, 2746
 capacity factor, 625
 capacity gap, 1859
 capacity-plasticity tradeoff, 1891
 caption augmentation, 856
 captioning
 importance for video, 657
 carbon footprint
 federated generative AI, 4653
 large-scale training, 5161
 cardiac imaging
 4D MRI, 3585
 cardiac motion, 3513
 cardiac MRI
 4D segmentation, 3508
 segmentation, 3512
 cardiac segmentation, 3477, 3512–3513
 wall boundaries, 3549
 care execution gap
 epilepsy, 3913
 care pathway
 reimagined, 3911
 caregiver support
 AI, 4799
 Carlini extraction attack, 1710
 CASC, 3020

- cascaded diffusion
 - audio, 845
 - definition, 845, 4322
 - error propagation, 846
 - exercise, 864
 - nowcasting, 4369
- cascaded generation
 - flow matching, 867
- cascaded models
 - weather forecasting, 4274
- CasCast, 4322–4369
 - comparison to single-stage, 4369
- cat-and-mouse dynamic, 936
- catamenial epilepsy
 - bias, 3942
- catastrophic forgetting, 656, 1135, 1171, 1183–1186, 1872, 4072
 - architecture change, 1998
 - Bayesian interpretation, 1180
 - causes, 1183
 - collective bound, 1436
 - definition, 1183
 - EEG, 3838, 3932–3933
 - evolutionary view, 2249
 - exercise, 4075
 - FengWu, 4277
 - fine-tuning, 3888
 - in coevolution, 2071–2073, 2162
 - in NAS, 1881
 - measurement, 1184
 - mitigation, 3095
 - population-based protection, 2250
 - retrieval bound, 2551
 - retrieval-based mitigation, 2550
 - scale, 4074
 - theorem proving, 3087
 - video models, 749
 - weather rollout, 4278
- categorical columns, 2693
 - diffusion, 2760
- categorical distribution
 - precipitation, 4241
- categorical posterior, 2764
 - exercise, 2773
- categorical weighting, 2771
- category forgetting, 2762
- Cauchy sequence, 5199
- Cauchy–Schwarz inequality, 42
- Cauchy–Schwarz inequality, 5213
- causal 3D VAE, 562–574
 - architecture, 571
 - ELBO, 566
 - mixed image-video training, 574
 - total loss, 569
 - training algorithm, 573
 - training loss, 569
 - training objective, 566
 - training procedure, 573
 - training schedule, 574
- causal convolution, 564–566
 - definition, 564
 - implementation, 564
 - receptive field, 565, 757
- causal discovery, 2873
 - EEG, 3936
- causal explanation, 3963
- causal generative model
 - open problem
 - epilepsy, 3936
- causal generative models
 - epilepsy, 3935–3936
- causal graph
 - preservation, 2862
- causal hierarchy
 - Pearl, 3935
- causal inference
 - neuroimaging, 3842
- causal interventions
 - diffusion models, 4100–4102
 - cascade effect, 4100

- mechanistic interpretability, 3974
- causal language modelling
 - loss, 2807
- causal masking
 - transformer, 3887
- causal misleads
 - hallucination, 4831
- causal retrieval, 2619–2620
 - definition, 2619
 - research directions, 2619
- causal structure, 2862–2863
 - Bayesian network, 4235
- causal tracing, *see* activation patching, 4059
- causal training
 - implementation, 3464
 - temporal weights, 3457
 - variational formulation, 3459
- causal training (PINNs), 3208
- CAVI, 190–203
 - algorithm, 190
 - Bayesian Gaussian mixture, 263
 - Bayesian GMM, 192
 - Bayesian GMM algorithm, 194
 - Bayesian GMM exercise, 201
 - Bayesian linear regression, 263
 - cluster assignment update, 192
 - comparison with EM, 195
 - convergence, 191
 - convergence diagnostics, 197
 - exercise, 201
 - exponential family, 206
 - implementation, 268
 - initialisation, 197
 - local optima, 191
 - mean update, 193
 - mixing weight update, 193
 - monotonicity, 263
 - monotonicity exercise, 201
 - natural gradient connection, 197
- Cayley digraph, 3165
 - in Claude’s solution, 3167
- CCPA, 4443–4447
 - federated learning, 4647–4649
 - key provisions, 4446
- CDCL, 2955, 3018
- CE marking
 - EU MDR, 3723
 - medical AI, 3722
- CelebA
 - HOLD++ defense, 1682
 - likelihood ratio, 1624
- CelebA-HQ, 22
- cell division, 3510, 3653
- cell nuclei
 - segmentation, 3482
- cell segmentation, 3509–3510, 3642–3650
 - benchmarks, 3643
 - challenges, 3510, 3643
 - dense clusters, 3510
 - gradient flow representation, 3645
 - SEG metric, 3643
- cell tracking, 3510, 3650–3666
 - association methods, 3653–3655
 - bipartite matching, 3752
 - error cascading, 3653
 - evaluation metric design, 3665
 - example, 3517
 - formal definition, 3650
 - foundation model, 3764
 - key idea, 3643
 - paradigms, 3652
 - TRA metric, 3644
- Cell Tracking Challenge, 3518, 3643–3652
 - combined score, 3644
 - history, 3651
 - metrics evolution, 3651
- cell-based search, 1893
- CellFinder, 3649

- Cellpose, 3645–3646
 - exercise, 3663
- Cellpose-SAM, 3648
- CellSAM, 3649–3650
- cellular pathology
 - network synchrony, 3894–3895
- Census Bureau
 - synthetic data, 2690
 - Synthetic Longitudinal Business Database, 2711
- census data, 2689
 - exercise, 2890
 - results, 2890
 - synthetic, 2889–2891
- Central Limit Theorem, 41
- centralisation
 - failure modes, 4448
- centrality
 - spatio-temporal shift, 3869
- Centroidal Voronoi Tessellation, 2092
- cerebral palsy, 4690
- cerebrovascular segmentation, 3511, 3513
- certified neural surrogate, 3438
- CFL condition, *see*
 - Courant–Friedrichs–Lewy condition
- chain (graphical model), 129
- chain of thought
 - length scaling, 3136
- chain rule
 - probability, 428–429
- chain rule of probability, 37, 57
- Chain-of-Thought
 - exercise, 1068
- chain-of-thought, 1038–1044, 3971
 - architecture design, 1952
 - computational class, 1041
 - computational power, 3048
 - computational power exercise, 3063
 - definition, 1039
 - error amplification, 4070
 - exercise, 3062
 - factual circuits, 4069
 - faithfulness, 3971
 - for mathematics, 2957
 - impossibility theorem, 1040
 - mathematical reasoning, 3136
 - multiplication example, 3049
 - prompting, 1039
 - proof decomposition, 3049
 - sufficiency theorem, 1041
 - theoretical foundations, 1040
 - vs mechanistic process, 4143
 - vs. Tree of Thoughts, 1044
 - why it works, 3048
 - zero-shot, 1040
- chain-of-thought prompting, 3048–3050, 3146
- challenge problems
 - PDE generative models, 3469–3474
- Chalmers University, 3002
- change of variables, 42, 450–452, 5226
 - multivariate, 44
 - normalizing flows, 236
 - push-forward, 58
- channel capacity
 - biological neurons, 1857–1860
 - definition, 1857
- channel gating, 1956
- channel scoring
 - dynamic selection, 3882
- channel selection
 - adaptive, 3880
 - set transformer, 3880
- Channel-Aware Set Transformer, 3880–3884
 - analysis exercise, 3896
 - architecture figure, 3883
- chaos

- in RNNs, 1035
- chaos theory, 4183
 - Lorenz equations, 4185
- chaotic systems
 - conditional mean smoothness, 4298
- Chapter 32, *see* PDE chapter
- Chapter 35, *see* interpretability chapter
- Chapter 37
 - federated learning reference, 5121
- chapter connections
 - accessibility, 4703
 - weather, 4198
- chapter roadmap, 2691
 - accessibility, 4704
 - Chapter 37, 4464
 - figure, 4464
 - interpretability, 3977
 - weather, 4198
- chapter summary
 - weather prediction, 4422
- characteristic kernel, 2840
- characteristica universalis, 3016
- Charlton, James
 - Nothing About Us Without Us, 4701
- Charney, Jule, 4184, 4210
- ChatGPT, 17
- CHB-MIT
 - Diff-EEG results, 3834
 - limitations, 3919
 - patient-specific generation, 3818
 - performance table, 3835
 - preictal augmentation example, 3836
 - RCVAE, 3826
 - unsupervised anomaly detection, 3839
- CHB-MIT dataset, 3781, 3800, 3818, 3880–3882, 3919
 - annotation cost, 3887
 - demographic composition, 3941
 - description, 3800
- Chebyshev’s inequality, 41
- checkerboard artifacts, 98
- checkpoint
 - definition, 5147
 - frequency, 5147
 - frequency trade-off, 5147
 - parallel write, 5148
- checkpoint-free recovery, 5163
- checkpointing
 - distributed training, 5145
 - strategy, 5147
- chemical accuracy, 3368
- chemical potential, 3358
- Chen, Shyi-Ming, 4234
- Chen, Ting
 - simple diffusion, 633
- Chen–Chen theorem, 3212
- Cheng and Zhu 2026, 4344–4345
- chest X-ray
 - Grad-CAM, 3985
 - lung fields, 3510
 - privacy attack example, 4597
 - segmentation, 3506
- chest X-ray synthesis
 - federated, 4611
- CheXpert
 - MGDS attack, 4596
- chi-squared distribution, 43
- chi-squared test, 2850–2851
- childhood absence epilepsy, 3777
- Children’s Hospital Boston EEG dataset, 3880
- China
 - data localisation, 4449
- Chinchilla
 - exercise, 4933
 - scaling laws, 4926
- Chinchilla scaling law, 5028

- Chinese restaurant process, 158
- Chinese Room
 - introspective version, 4147
- Cholesky decomposition, 40
 - reparameterization, 228
- Chollet, François, 3034
- Chow, Tommy W.S., 4233
- chroma similarity, 938
- chromagram, 813
- Chronos, 2801
- chunked cross-attention, 2525
- Church, Alonzo, 3016
- CIC, *see* Calculus of Inductive
 - Constructions, *see* Calculus of Inductive Constructions, 3025
- CIDeR
 - accessibility evaluation, 4887
- CIFAR-10, 22
 - differential privacy, 1672
 - gradient inversion, 4592
 - loss histogram, 1617
 - membership inference, 1615
 - SecMI vulnerability, 1630
- CIMD, 3607–3610
 - Ambiguity Controlling Network, 3607
 - Ambiguity Modeling Network, 3607
 - ELBO, 3608
 - exercise, 3617
 - LIDC-IDRI, 3609
- cingulum bundle, 3891
- circuit
 - comparison metric, 4155
 - definition, 3972, 4031
 - formal definition, 4031
 - formal verification, 4159
 - linguistic task, 4173
 - minimality, 4037
- circuit complexity, 1035–1038
 - classes, 1036
 - mathematical reasoning, 3047
- circuit discovery
 - computational complexity, 3976
- circuit fidelity, 3980
- circuit finding
 - generalisation across models, 4037
 - IOI, 4028–4038
 - limitations, 4035
- circuit obfuscation, 4141
- circuit universality, 3973
- circuits, 3973
 - alignment audit, 4137
 - cross-architectural metrics, 4155–4156
 - deceptive alignment, 4135
 - definition, 3972
 - neural network, 3971
- citation verification, 2569
- city-scale accessibility auditing, 4722
 - example, 4723
- CLAP, 802–809, 2447–2457, *see*
 - contrastive language-audio pretraining, 2541
 - architecture, 2447–2449
 - audio encoder, 805, 2448
 - contrastive loss, 803
 - definition, 803
 - evaluation example, 808
 - exercise, 830
 - exercises, 2455
 - mutual information bound, 804
 - score, 805
 - temperature parameter, 805
 - text encoder, 805, 2448
 - training data, 807, 2448
- CLAP score, 919–920
 - definition, 919
 - exercise, 928
 - likelihood interpretation, 920

- class activation mapping, 3985–3986
- class activation maps, 3520
- class imbalance, 2688, 2887, 3803–3813
 - abdominal segmentation, 3512
 - augmentation comparison table, 3808
 - ensemble methods, 3807
 - exercise, 3789
 - extreme, 3781
 - federated augmentation, 4457
 - federated GAN, 4492
 - generative augmentation, 3807
 - imbalance ratio, 3804
 - ML failure modes, 3805
 - naive classifier bias, 3805
 - patient heterogeneity, 3804
 - practical considerations, 3811
 - segmentation, 3486
 - seizure prediction, 3882
 - seizures, 3781
 - summary, 3812
 - tabular GAN, 2721
 - temporal proximity bias, 3804
 - traditional solutions, 3805–3807
 - limitations, 3807
- class weighting, 3805, 3806
- class-balanced buffer, 1193
- class-conditional generation, 2734
- class-free guidance
 - debiasing, 3942
- class-incremental learning, 1177
- classical assistive technology, 4698
- classical segmentation methods, 3487
- classifier guidance, 2776
 - PDE applications, 3275
 - physics constraints, 3719
- classifier predicates, 4726
- classifier-free guidance, 2776, 5091
 - audio, 819–821, 956
 - audio training, 860–862
 - Bayesian interpretation, 664
 - compositional, 663
 - compositional schedule, 664
 - computational cost, 664
 - conditional dropout, 666
 - definition, 3834
 - dual conditioning, 870
 - dynamic schedule, 667
 - EEG, 3834
 - EEG conditioning, 3852
 - geometric illustration, 668
 - limitations, 2775
 - medical segmentation, 3598
 - multi-condition, 662
 - definition, 662
 - multi-condition Bayesian, 665
 - multi-condition posterior, 665
 - negative prompts, 668
 - posterior sharpening, 821
 - scale schedule, 820
 - scale selection, 861
 - score interpretation, 662
 - score sharpening, 861
 - single-condition review, 662
 - tempered posterior, 664
 - time-dependent, 667
 - video, 538, 661–669
 - weight selection, 666
- Claude
 - introspection, 4150–4151
 - introspective markers, 4150
- Claude Opus 4.6, 3166
- Claude’s Cycles, 3165–3168
- clausal form, 2955
- clause learning, 3018
- clauses, 3017
- clausification, 3020
- Clay Mathematics Institute, 3427
- Clay Millennium Prize, 3473
- Clay Millennium Prize Problem, 3198

- Clearview AI, [1571](#)
- client drift, [4451](#), [4573](#)
 - convergence analysis, [4573](#)
 - definition, [4473](#), [4573](#)
 - exercises, [4578](#)
 - SCAFFOLD correction, [4472](#)
- climate change
 - MOS validity, [4226](#)
- climate downscaling
 - resolution, [4324](#)
- climate modelling, [3440–3442](#)
 - century-scale, [3442](#)
 - diffusion models, [3282](#)
 - machine learning, [4281](#)
- climate science
 - federated generative models
 - research challenge, [4683](#)
- climatological mean
 - regression to, [4194](#)
- climatological unbiasedness
 - MOS, [4225](#)
- Climavision
 - AI weather, [4389](#)
- clinical accuracy
 - LLM evaluation, [3924](#)
- clinical decision support
 - uncertainty, [3769](#)
- clinical notes
 - federated discrete diffusion, [4546–4547](#)
- clinical pipeline
 - segmentation, [3480](#)
- clinical recommendation
 - integrated, [3916](#)
- clinical relevance
 - epilepsy AI, [3930](#)
 - segmentation metrics, [3547](#)
- clinical semiology
 - seizure, [3891](#)
- clinical target, [3509–3515](#)
 - generative interventions, [3675](#)
 - overview table, [3675](#)
- clinical testing
 - LLM, [4763](#)
- clinical text
 - epilepsy, [3934](#)
- clinical translation, [3715–3727](#)
 - design principle, [3939](#)
 - epilepsy, [3939–3943](#)
 - trust, [3715](#)
- clinical triage
 - synthetic EEG, [3949](#)
- clinical trial simulation, [2887](#)
- clinical trials
 - privacy-preserving data sharing, [4618](#)
- clinical utility
 - synthetic PET, [3859](#)
- clinical workflow
 - automated, [3704](#)
 - segmentation role, [3477](#)
- clinician in the loop
 - regulatory advantage, [3940](#)
- CLIP, [2390](#), [2408–2411](#), [2630](#), [3668–3680](#)
 - accessibility, [4704](#)
 - accessibility applications, [4877](#)
 - architecture, [2408](#)
 - bias
 - detection method, [4093](#)
 - bias features, [4093](#)
 - findings, [4093](#)
 - comparison with BoVW, [2417](#)
 - correlation maps, [3695](#)
 - frozen encoder, [4996](#)
 - gender bias
 - example, [4093](#)
 - limitations in medical imaging, [3670](#)
 - loss function, [2408](#)
 - mutual information, [2409](#)

- retrieval example, 2410
- temperature, 2370, 2410
- temporal consistency, 738
- text encoder, 615
 - frozen, 5083
- text-guided segmentation, 3520
- training, 2408
- training algorithm, 2409
- zero-shot classification, 3669
- zero-shot transfer, 2409
- CLIP encoder
 - video diffusion, 685
- Clip-Higher, 1149
 - definition, 1149
 - exercise, 1163
- CLIP4Clip, 2419–2421
 - architecture, 2420
 - MSR-VTT results, 2420
- ClipBERT, 2419–2421
 - sparse sampling, 2420
- clipping
 - PPO, 3054
- CLIPScore, 22
- clique, 131
- closed set, 5196
 - characterisation, 5197
- closed-box detection, 930
- closed-loop DBS
 - OT-guided, 3874
- closed-loop neuromodulation, 3873
 - seizure prediction, 3930
- closed-loop neurostimulation, 3916
- closure, 5197
 - via limit points, 5198
- Clotho, 807
- cloud computing
 - accessibility, 4809–4820
- cloud latency
 - decomposition, 4811
- cloud microphysics, 4209
- cloud pricing
 - spot vs on-demand, 5149
- cloud properties
 - generative estimation, 4381
- cloud vs edge
 - comparison figure, 4818
- Cloud4D
 - cloud property estimation, 4381
- cloze completion
 - MDM vs ARM, 4110
- CLS theory
 - generalisation, 1218
- CLS token
 - information flow figure, 4090
 - layer evolution, 4088
 - spatial specificity, 4088
- CLS token dynamics, 4088–4095
- Clune, Jeff, 2083, 2284
- cluster management, 2114
- clustering
 - evaluation, 2856
- CMA
 - AI weather ecosystem, 4388
- CMA-ES, 2041–2044
 - definition, 2042
 - natural gradient connection, 2043
- CNF, 3017
- CNN
 - accessibility, 4701
 - lightweight edge model, 4821
- CNN features
 - for retrieval, 2406–2407
 - vs. SIFT, 2407
- CNN insight, 2730
- co-design
 - accessibility, 4701
 - accessibility AI, 4894
- co-evolution, 1373
 - algorithms and language models, 2196

- algorithms and models, 2188
- conjecturer and prover, 3107
- co-evolutionary ecosystem of AI, 2255
- co-player inference, 1372
- coarsening, 3359
- CoCoGen, 3282–3295
 - forward and inverse problems, 3283
 - incompressible flow, 3284
 - method, 3283
 - sampling algorithm, 3283
- CoDA-NO, 3255–3256, 3329, 3375
 - transfer learning, 3256
- code as genome, 2182
- code embedding
 - for novelty detection, 2147
- code generation
 - architecture design, 1951
 - success rates, 2935
 - testability, 2934
 - vs. theorem proving, 2962–2963
- code release
 - absence, 3923
- codebook
 - EEG, 3833
 - exponential moving average
 - update, 798
 - factorised, 794
 - utilisation, 577
 - video, 575
- codebook collapse, 415, 576, 798
- codebook flattening, 810–812
 - delayed pattern, 810
 - exercise, 831
 - interleaved pattern, 810
 - parallel pattern, 811
 - sequence length, 811
- codebook patterns
 - complexity analysis, 951
- codec artifacts, 935
- codec bottleneck, 947
- CodeEvolve, 2121–2123
 - elitist migration, 2122
 - meta-prompting, 2121
 - transparent evaluation, 2123
- CODES
 - epidemic model, 1454
- coding projects
 - PDE generative models, 3460–3469
- codomain attention, 3255, 3375
 - definition, 3255
- coevolution, 2019–2067, 2301
 - advantages, 2113
 - alignment, 2296
 - and quality-diversity, 2094
 - biological, 2154
 - convergence theorem, 2065
 - cycling, 2289
 - difficulty escalation, 2220
 - disengagement, 2069–2071
 - environment-agent, 2242
 - exercise, 2289
 - GAN training, 2251
 - hall of fame, 2076
 - host–parasite, 2281
 - in autonomous research, 2219–2220
 - multi-block, 2111–2113
 - pathologies, 2067–2079, 2252
 - summary, 2078
 - pathology diagnosis, 2078
 - Red Queen, 2030
 - remedies, 2075–2079
 - search and solution, 2113
- coevolutionary algorithms, 2057–2067
- coevolutionary arms race, 2154
- coevolutionary cycling, 2161
- coevolutionary forgetting, 2071–2073, 2162
 - definition, 2072
 - example, 2072
 - measurement, 2073

- coevolutionary game, 2065
- coevolutionary Nash equilibrium, 2065
- coevolutionary paradigm, 2169
- coevolutionary pathologies, 2161–2163
- coevolutionary system
 - competitive, 2059
 - cooperative, 2062
- Cofino, Antonio S., 4235
- cognitive copilot (GenAI), 4791
- cognitive disabilities, 4690
- cognitive load
 - evaluation, 4886, 4887
- cognitive memory hierarchy, 1098
- CogVideoX, 4115
 - first-frame conditioning, 582
 - joint attention, 604
 - system design, 688
 - text encoding, 686
 - VAE, 571
- Cohen’s kappa, 3550
- coherence propagation
 - theorem, 4333
- coherent sampling, 3559
- COLA condition, 772
- ColBERT, 2368–2373, 2630
 - exercise, 2636
 - expressiveness, 2369
- ColBERTv2, 2373
- cold front
 - CorrDiff example, 4355
- cold start
 - expert iteration, 3099
- cold-start problem, 2445, 2514, 2582–2583
 - music recommendation, 2582
 - segmentation, 3485
- Cole–Hopf transformation, 3313–3314, 3454
 - derivation, 3318
 - statement, 3313
- collaboration
 - in mathematics, 2941
- collaborative filtering, 2445, 2576–2578
 - federated, 4510
- collaborative generation
 - federated, 4457
- collaborative systems, 1308
- collaborative training, 5124–5134
 - crossover threshold, 5131
 - hybrid future, 5133
 - insight, 5133
 - Petals, 5124
 - vs distributed, 5131
- CollaFuse, 5129
 - privacy, 5129
- collective communication
 - algorithm comparison exercise, 4959
 - algorithm selection, 4943
 - all-reduce, 5031–5041
 - alpha-beta costs, 4953
 - cost table, 4955
 - taxonomy, 5031
- collective intelligence, 1270–1277, 3553
 - artificial, 1276
 - biological, 1270
 - definition, 1272
 - federated generative models, 4457
 - human organisations, 1274
 - superadditivity conditions, 1273
- collective knowledge capacity, 1428
 - bandwidth bound, 1430
 - bound, 1429
 - exercise, 1456
- collective operations
 - cost table, 4967
 - figure, 4966
 - latency vs. bandwidth, 4991
 - MPI, 4964
- Collectively Intelligent Medical

- Diffusion, *see* CIMD
- collider, 129
- collocation points
 - sequential, 3260
- colonoscopy
 - polyp segmentation, 3514
 - rare pathology augmentation, 3611
- color refinement
 - high-fidelity, 4781
- color shifting
 - global, 4772
- color translation
 - text legibility, 4775
- color vision deficiency, 4769–4779
 - biology, 4770
 - formal definition, 4770
 - GAN-based correction, 4773
 - subway map, 4771
 - subway map recoloring, 4777
 - transformers, 4779–4789
 - types, 4770
- ColorBlindGAN, 4773–4913
 - architecture, 4776
- colorectal cancer screening, 3514
- colour contrast
 - accessibility, 4692
- colourisation
 - exercises, 340
- column dependencies, 2745
- column dependency
 - problem, 2746
- column ordering
 - invariance, 2752
- column pair trends metric, 2849
- column shapes metric, 2849
- column shuffling, 2807
- column type ambiguity, 2695
- column types, 2693
 - formal definition, 2693
 - in practice, 2693
 - visualization, 2694
- column-parallel linear, 5043–5050
 - definition, 5043
- column-specific activation, 2753
- column-wise metrics, 2837–2838
- combinatorial explosion, 3016
 - in theorem proving, 3028
- combinatorial hashing, 2439
- combinatorial optimization, 2865
- combined loss
 - segmentation, 3486
- commercial music generation, 910–915
 - comparison, 914
 - evolution, 915
 - Google, 911
 - Suno, 912
 - Udio, 913
- commercial weather forecasting
 - AI, 4389
- commitment loss, 411, 576
- Common Crawl, 22
 - training data, 4926
- common spatial pattern filters, 3886
- common spatial patterns (CSP), 3797
- CommonCrawl, 4444
- communication
 - agent, 1308–1320
 - channel capacity, 1309
 - classifier-free guidance, 5091
 - complexity, 1319
 - compression, 1460
 - diffusion training, 5089–5097
 - dominant bottleneck, 4931
 - efficiency, 1310
 - optimal protocol, 1440
- communication avoiding, 4999–5007
 - definition, 5001
 - diffusion models, 5014
 - diffusion multi-step batching, 5014
 - generative models, 5007–5020

- grocery analogy, 5000
- machine learning, 5005
- communication complexity
 - federated generative models
 - exercise, 4465
 - lower bounds, 5165
- communication cost, 2884
 - exercise, 4465
 - federated diffusion, 4459
 - federated methods comparison
 - exercise, 4680
 - FedLLM, 4460
 - hybrid parallelism, 5139
 - topology dependence, 4943
- communication efficiency, 4475–4483
 - definition, 4475
 - federated diffusion, 4660
- communication hiding, 4981–4999
 - 13B LLM example, 4993
 - analogies, 4982
 - analogy, 4981
 - analogy limits, 4982
 - bus-reading analogy, 4982
 - combined with avoiding, 5018
 - DDP backward pass, 4988
 - deep pipelining, 4987
 - definition, 4983, 5001
 - diffusion training, 5091
 - does not reduce volume, 4984
 - download-while-edit analogy, 4982
 - formal framework, 4983
 - framework comparison, 4997
 - FSDP, 5180
 - generative model training, 4990
 - latency bound, 4988
 - overlap efficiency, 5171
 - pipelined tail exercise, 4990
 - pipelining, 4984
 - shared resource limitation, 4989
 - speedup example, 4983
 - speedup exercise, 4989
 - time model, 4983
 - versus avoiding, 5001
- communication locality
 - principle, 4928
- communication lower bound
 - n-parameter models, 5166
- communication model
 - alpha-beta, 4951–4959
- communication optimality
 - auto-differentiation, 5162
- communication optimisation
 - taxonomy, 5016
- communication overhead, 1342
- communication protocol
 - definition, 1303
 - information-theoretic, 1302
- communication time
 - exercise, 4933
 - lower bound, 4930
 - prediction, 4951
- communication volume
 - definition, 4944
- communication wall, 4925
 - definition, 4930
- communication-accuracy tradeoff, 4480
- communication-avoiding
 - attention, 5156
 - attention design, 5178
 - diffusion training, 5096
 - local SGD, 5172
- communication-avoiding algorithms
 - GPT example, 5013
 - linear algebra, 5002
 - matrix multiply, 5003
 - transformer, 5013
- communication-avoiding training, 5091
- communication-computation overlap,
 - 4958
 - exercise, 4959, 5031

- communication-hiding conjugate
 - gradient, 4986
 - algorithm, 4986
 - numerical stability, 4987
- communicator
 - definition, 4961
 - MPI, 4960
- community-in-the-loop, 4910
- commutative ring, 2989
- commutativity
 - of addition, 2967
 - of conjunction, 2969
 - of disjunction, 2969
- compact set, 5202
- compactness
 - continuous image, 5206
 - equivalence with sequential compactness, 5203
- compactness arguments, 2936
- comparison
 - tabular methods, 2772
- comparison table
 - MLWP systems, 4279
- CompCert, 3000
- competence
 - model self-assessment, 4145
- competition, 1370–1375
- competition mathematics, 2935, 3060
 - formalization, 3073
- competitive coevolution, 2058–2061
 - definition, 2059
- competitive data silo, 4448
- competitive data silos
 - federated resolution, 4457
- competitive fitness sharing, 2060, 2077
- competitive systems, 1308
- compilation
 - Lean 4, 2997
- complementarity problem
 - EEG–fMRI, 3849
- complementary learning systems,
 - 1175–1177
 - definition, 1176
- complementary slackness, 5266
- complete agent system, 1426
 - capstone, 1462
 - exercise, 1456
 - generality, 1427
- complete intelligent system, 1437
- complete pipeline
 - breast tumour example, 3705
 - medical segmentation, 3703–3713
- complete space, 5200
- complete-data log-likelihood, 150
- completeness
 - interpretability, 3997
 - of metric spaces, 5200
 - of proof search, 3016
 - of $\mathbb{R}^n$, 5200
- completeness axiom, 3987, 5192
- complex geometry
 - PDE solver, 3387
- complexification, 2050
- complexity
 - increasing via self-play, 3109
 - open-ended increase, 2026
- complexity barrier, 2243–2246
 - theorem, 2245
- composite loss
 - definition, 853
 - usage patterns, 854
- composition
 - K-composition, 4022
- composition interface
 - generative models, 3938
- composition theorem, 1600
- compositional architecture
 - epilepsy, 3938
- compositional generation
 - formal statement, 754

- multiphysics, 3372
 - open problem, 754
- Compositional Pattern Producing Network, 2052
- compositional retrieval, 2618–2619
 - definition, 2619
- compositional safety, 1978
- compositional understanding
 - medical, 3712
- compound word
 - pop music, 907
- comprehensibility horizon, 2263–2267
 - definition, 2263
 - exercise, 2266
 - formal definition, 2263
- compressed sensing
 - EEG, 3821–3822
- compression
 - quality tradeoff, 535
- compression ratio
 - examples, 535
 - exercise, 588
 - video autoencoder, 559
- computability
 - limits on evolution, 2264
- computation vs proof, 3032
- computational chemistry
 - history, 2893
- computational complexity, 1035–1038
 - attention, 3237
 - chain-of-thought, 3049
 - limits on evolution, 2264
- computational cost
 - concrete examples, 549
 - cross-modality comparison, 554
 - evaluation, 2863
 - PDE solvers, 3199–3203
 - quality tradeoff, 552
 - scaling with resolution, 552
 - video diffusion, 548–556
- Wasserstein distance, 3868
- computational economics
 - of evolution, 2198
- computational electromagnetics, 3345–3352
- computational fluid dynamics, 15, 2210, 3325–3352
- computational graph
 - distributed, 5097
 - formal definition, 5099
 - PDE representation, 3258
 - score network, 1936
- computational irreducibility, 2196
- computational pathology
 - segmentation, 3480, 3515
- computational requirements
 - mathematical AI, 3141
- computational resources
 - AlphaProof, 3113
- computational subgraph, 4031
- computational verification, 2209–2211
- compute budget
 - formula, 4926
- compute limitations
 - 1990s, 4238
- compute wall, 4925
 - definition, 4930
- compute–memory crossover, 1079
- compute–quality trade-off, 2827
- compute-bound
 - definition, 4936
- compute-communication overlap, 5171
- compute-to-communication ratio, 4936
- computed tomography
 - segmentation, 3481
 - volumetric, 3506
- computer algebra systems, 3031–3033
- computer vision, 10
 - inverse problems, 327
- CON-PREDIFF, 4544–4545

- federated algorithm, 4545
- staged aggregation
 - exercise, 4549
- concatenation
 - naive, 2789
- concatenation conditioning, 614, 617, 727
 - definition, 617
 - flexibility, 617
- concatenative synthesis, 767
- concept discovery
 - impossibility in ATP, 3029
- concept prompts, 3629–3630
- condition manifold, 2776
- conditional correlation, 2700
- conditional diffusion, 3280–3295
 - concatenation, 3598
 - cross-attention, 3598
 - definition, 4310
 - demographic conditioning, 3942
 - segmentation, 3597
 - strategies, 3280
 - weather forecasting, 4310
- conditional distribution, 37
 - as manifold slice, 2775
 - weather forecasting, 4197
- conditional dropout
 - joint vs independent, 667
 - schedule, 666
 - typical rates, 667
 - video, 666
- conditional expectation
 - smoothness, 4298
- conditional fidelity, 2838–2840
- conditional flow matching, 865
 - definition, 4340
 - linear interpolant, 4340
 - weather, 4416
- conditional GAN, 284, 344–346
 - debiasing, 3942
 - definition, 3580
 - exercise, 287, 359
 - generator optimality, 345
 - medical imaging, 3580–3596
 - MNIST example, 345
 - mode collapse, 361
 - MRI-to-PET, 3857
 - objective, 344, 3580
 - optimal discriminator, 344
 - seizure EEG, 3816–3818
- conditional generation, 103, 406
 - comparison, 2787
 - LLM-based, 2814
 - medical images, 3597
 - privacy, 2925
 - retraining cost, 2775
 - scalability, 2708
 - without retraining, 2776
 - zero-shot, 2775–2787
- conditional generative modelling
 - weather, 4197
- conditional generator, 2715
- conditional independence
 - Bayesian network, 4235
- conditional loss, 2725
- conditional mean
 - blurring problem, 3227
 - exercises, 3232
 - limitations, 3265
 - weather forecasting, 4195
- conditional memory, 1092–1096
- conditional normalizing flow, 3855
- conditional probability, 36
- conditional regularity, 3429
- conditional score, 2776
 - manifold approximation, 2781
- conditional tabular GAN, 2713–2729
- conditional VAE, 406
 - architecture, 3824
 - EEG, 3824–3825

- for SAM taming, 3698–3701
 - taming SAM, 3699
- conditional variational autoencoder
 - segmentation, 3495
- conditional vector
 - construction, 2722
- conditioning
 - audio generation, 816–823
 - batch-level, 2783
 - cross-attention vs. concatenation, 3754
 - diffusion model, 4311
 - dual-stream EEG, 3838
 - entropy reduction, 617–620
 - theorem, 618
 - exercise, 619
 - Gaussian entropy bound, 618
 - hard, 2783
 - independence assumption, 822
 - modalities for audio, 816
 - multi-modal, 821–823
 - pathways, 613
 - practical dimensions, 616
 - soft, 2783
 - subject-adaptive diffusion, 3834
 - taxonomy, 613
 - video diffusion, 613–620
 - video generation, 537
- Condorcet jury theorem, 1367
 - correlation threshold, 1459
 - weighted, 1452
- cone effect, 2464
- cone photoreceptors, 4770
- cone sensitivity
 - personalised model, 4782
- confabulation, 3963
 - base models, 4143
- confidence masking, 1817
 - practical deployment, 1818
- CoNFILD, 3337, 3392
 - channel flow, 3338
- conflict-driven clause learning, 3018
- conformal prediction, 2879–2881, 3414
 - distribution shift, 2880
 - exercise, 2880
 - FCD detection, 3860
 - pipeline, 2881
 - prediction set, 2880
- confusion lines
 - geometry, 4778
- confusion map
 - computation, 4788
- confusion matrix, 3810
- confusion-prone region
 - definition, 4781
- confusion-prone regions, 4780
- conjecture
 - generative architecture
 - epilepsy, 3937–3938
 - nature of, 3178
 - scale dominance, 4407
- conjecture filtering, 3108
- conjecture generation, 3107, 3178–3181
 - self-play, 3179
- conjecturer-prover, 3107
- conjugacy, 180
 - variational inference, 182
- conjugate function
 - of norm, 5277
- conjugate gradient
 - communication avoiding, 5003
 - communication bottleneck, 4986
- conjugate prior, 46, 206–209
 - definition, 206
 - natural parameter update, 264
- conjunction, 2987
- conjunctive normal form, 3017
- Conmy, A., 3975
- connected set, 5210
 - path-connected implies, 5210

- connectionism, 2833, 3969
 - revival, 4233
- connectives
 - logical, 3017
- connectome
 - breakdown, 3912
 - epilepsy, 3863
- connectome perturbation
 - pathology, 3895
- connector hub, 3912
- consciousness
 - AI models, 4149–4150
 - frontier models, 4149
 - IIT
 - frontier models, 4148
- consciousness and interpretability, 3963
- consensus, 1363–1370
 - agent opinions, 1366
 - graph Laplacian, 1364, 1452
- consensus label
 - limitations, 3553
- consensus optimisation, 322
 - convergence, 322
 - definition, 322
 - exercises, 325
- consent
 - voice cloning, 939
- conservation laws, 3311–3319
 - atmosphere, 4207
 - enforcement, 3437
 - failure, 722
 - guidance, 720
 - learned weather models, 4403
 - residuals, 3421
 - structure, 3317
- conservation of mass
 - atmosphere, 4204
- conserved quantity, 316
 - exercises, 325
- consistency, 44
 - automated segmentation, 3488
 - of arithmetic, 2939
- consistency distillation, 646–652
 - audio, 875, 953
 - definition, 646
 - exercise, 760, 896
 - medical segmentation, 3729
 - tabular, 2877
- consistency function, 2876
 - audio, 875
- consistency models, 2876–2877
 - audio, 873–880
 - consistency function, 2876
 - distillation, 2878
 - exercise, 652
 - medical imaging, 3762
 - multi-step, 647, 2877
 - tabular data, 2876, 2926
 - tabular exercise, 2895
 - video, 646, 760
- ConsistencyTTA, 877–880
- constant overlap-add, 772
- constant-power property, 2793
- constellation map, 2439
- constitutional AI, 4144
 - for evolutionary systems, 2260
 - introspection, 4144
- constitutional fitness
 - exercise, 2266
 - optimality, 2261
- constrained decoding
 - clinical NLP, 3902
- constraint discovery
 - automatic, 2926
- constraint language
 - definition, 2825
 - tabular, 2824–2826
- constraint loss, 2703
- constraint satisfaction, 3024
 - evaluation, 2705

- geometry, 2829
- constraints
 - implicit, 2693
 - satisfaction, 2703–2704
- constructive existence proofs, 3132
- consumer AI tools
 - BLV, 4712
 - cognitive disabilities, 4802
- containment
 - control mechanisms, 2325
 - evolutionary systems, 2260–2261
- containment protocol
 - definition, 2260
- content restructuring
 - neurodivergence, 4701
- content transformation
 - exercise, 4697
- content-addressable memory
 - definition, 2345
- content-based filtering, 2445
- content-based image retrieval, 2397–2406
- content-based recommendation, 2578–2579
- context engineering, 1347–1354
 - definition, 1348
 - entropy management, 1348
- context length
 - audio models, 882
- context management
 - comparison, 2549
 - strategies, 2675
- context parallelism, 5056–5060, 5138
- context relevance, 2533
- context window, 1081–1085, 2548–2554
 - and agent horizon, 1280
 - as retrieval problem, 2548
- context-memory conflict, 4068
 - exercise, 4075
- contextual bandit, 2150
- contextual knowledge, 4068
- contiguity, 2334
- continental United States
 - precipitation forecasting, 4240
- contingency similarity, 2838
- continual agent, 1285–1289
 - definition, 1285
- continual compositionality, 1203–1204
 - definition, 1203
- continual fine-tuning, 1185, 1209–1210
 - pipeline, 1209
- continual learning, 1171–1263, 2072
 - architecture, 3086
 - architecture adaptation, 1959–1960, 1970
 - capacity scaling, 1960
 - connection to forgetting, 4073
 - connections, 1240–1246
 - cross-agent, 1460
 - curriculum connection, 2231
 - definition, 1171
 - EEG diffusion, 3838
 - epilepsy, 3932
 - evolutionary perspective, 2249–2251
 - exercises, 1250–1263
 - federated, 4662
 - open problem, 4662
 - for LLMs, 1204–1211
 - history, 1247–1249
 - in agents, 1284–1289
 - mathematical foundations, 1177–1182
 - method comparison, 1960
 - motivation, 1172–1177
 - multi-agent, 1441
 - open problem
 - EEG, 3933
 - open problems, 1244
 - philosophy, 1173

- problem formulation, 1177
- regularization methods, 1186–1193
 - comparison, 1191
- retrieval connection, 2612
- retrieval for, 2550–2554
- sleep connection, 1874
- structured data, 2899–2904
- taxonomy, 1185–1186
- unified framework, 1229–1239
- video diffusion, 749
- weather, 4404
- continual LoRA, 1201
- continual pre-training, 1205
 - definition, 1205
 - medical domain, 1205
- continual pretraining, 1185
- continuity
 - epsilon-delta, 5205
 - hierarchy, 5209
 - sequential characterisation, 5205
 - uniform, 5206
- continuity equation, 635
 - atmospheric, 4204
- continuous columns, 2693
 - non-Gaussian, 2718
- continuous ensemble
 - definition, 4324
- continuous ensemble forecasting, 4324
- continuous normalising flows, 3286
 - PDEs, 3285
- continuous normalizing flow, 460–463
 - PDEs, 3384–3396
- Continuous Ranked Probability Score,
see CRPS
- continuous ranked probability score
 - CRPS, 4229
- continuous relaxation, 1900–1904
 - approximation error, 1901
 - key insight, 1902
 - visualisation, 2008
- continuous tokens, 732
- continuous vs discrete search, 3139
- continuous wavelet transform
 - EEG, 3832
 - time-frequency, 3832
- continuous wavelet transform (CWT)
 - definition, 3798
- continuous-time diffusion, 496
- continuous-time dynamics, 316
- continuous-time Markov chain, 2771
- continuum memory system, 1217
- contour-based segmentation, 3701
- ContourMS, 3701–3702
 - architecture, 3701
 - exercise, 3714
 - inference, 3702
 - training, 3701
- contraction mapping, 1031, 5234
- contraction mapping principle, *see*
Banach fixed point theorem
- contrast
 - in retrieval, 2334
- contrast variation
 - abdominal CT, 3512
- contrastive decoding, 4074, 4835
 - exercise, 4840
 - MARINE, 4836
- contrastive language-audio pretraining,
2469
- Contrastive Language-Image
Pre-training, *see* CLIP
- contrastive learning
 - audio-text, 802–809
 - exercise, 2632
 - for embeddings, 2392
 - for retrieval, 2369–2370
 - in DPR, 2366
 - InfoNCE, 3669
 - information bottleneck, 2562
 - optimal temperature, 2369

- premise retrieval, 3083
- unsupervised, 2391
- vision-language, 3668
 - pretraining, 3679
- vs generative, 2392
- contrastive loss
 - premise retrieval, 3094
- control signals
 - taxonomy, 537
- control theory
 - in evolutionary computation, 2176
- control variate, 221
 - optimal coefficient, 265
- controllability, 523
- controlled natural language, 3066
- ControlNet
 - video, 537
 - video extension, 679
- convection parameterisation, 4191, 4209
- convection parameterization
 - machine learning, 4406
- convective initiation, 4362
 - nonlinear, 4230
- convective-scale prediction, 4403
- convergence
 - architecture diffusion, 1950
 - bilevel NAS, 1887
 - diffusion PDE solvers, 3290
 - EM algorithm, 153
 - function-space diffusion, 3384
 - latent vs. spectrogram diffusion, 860
 - of sequences, 5199
 - of theorem provers, 3009
 - open problems, 3295
 - strong implies weak, 5218
 - weak, 5217
 - weak but not strong, 5218
- convergent evolution, 2159
- convergent subsequence, 2993
- Conversable Agent, 1325
- convex combination, 5243
- convex conjugate, *see* Fenchel conjugate
- convex function, 5246
 - examples, 5251
 - first order condition, 5249
 - operations preserving, 5251
 - second order condition, 5250
 - strictly convex, 5246
- convex hull, 5244
- convex set, 5243
 - operations preserving, 5244
- ConvLSTM, 4238
 - comparison, 4241
 - exercise, 4243
 - formal definition, 4239
 - forward pass exercise, 4413
- convolution
 - 2D, 97
- convolution theorem, 3218
- convolutional neural network, 96
 - analogy, 2747
 - features for retrieval, 2406
 - tabular data, 2729
 - vs. Transformer
 - weather, 4376
- convolutional neural networks
 - limitations for PDEs, 3234
- ConvSwiGLU, 1061
- cooperation, 1370–1375
 - spontaneous, 1371
- cooperative coevolution, 2061–2064, 2282
 - convergence, 2062
 - definition, 2062
- cooperative game theory, 1361
 - Shapley values, 3981
- cooperative games
 - fair division, 3982
- coordinate ascent

- mean-field, [184](#)
- coordinate ascent variational inference,
 see CAVI
- coordination, [1363–1370](#)
- copilot
 - theorem proving, [3084](#)
- COPPA
 - federated learning compliance, [4629](#)
- copula, [543](#)
- copyright
 - audio generation, [937–938](#)
- copyright detection, [1787](#)
- Coq, [2934](#), [2944](#), [2999–3000](#), [3080](#)
 - key features, [3000](#)
 - notable formalizations, [3000](#)
- Coq (proof assistant), [3025](#)
- CoqGym, [3013](#)
- CoqHammer, [3027](#)
- Coquand, Thierry, [2944](#), [3000](#)
- Core War, [2155](#)
 - as AI testbed, [2155](#)
 - bombing strategy, [2159](#)
 - definition, [2155](#)
 - history, [2156](#)
 - multi-threaded strategy, [2159](#)
 - replication strategy, [2159](#)
 - scanning strategy, [2159](#)
 - strategy taxonomy, [2156](#)
 - virtual machine, [2155](#)
- COREVQA benchmark, [4889](#)
- Coriolis force, [3320](#), [4203](#)
- Cornelis et al. (2019), [4987](#)
- coronary artery segmentation, [3513](#)
- CorpusLM, [2515–2516](#)
 - question answering, [2516](#)
 - unified framework, [2516](#)
- CorrDiff, [15](#), [3282](#), [4353](#)
 - architecture, [4353](#)
 - computational efficiency, [4355](#)
 - exercise, [4360](#)
 - inference exercise, [4416](#)
 - pipeline figure, [4435](#)
 - research challenge, [4420](#)
 - training, [4354](#)
- correctness
 - subjectivity in NLP, [2933](#)
- correlation, [35](#)
 - challenge, [2700–2702](#)
 - preservation, [2701](#)
- correlation difference, [2838](#)
- correlation map, [3695](#)
- correlation matrix distance, [2839](#)
- correlation preservation, [2743](#)
- correlation structure
 - formal definition, [2701](#)
- cortical development, [1846–1849](#)
- cortical stimulation
 - Fritsch and Hitzig, [3775](#)
- corticospinal tract, [3890](#)
- cosine noise schedule
 - vulnerability, [1631](#)
- cosine similarity
 - LSH, [2376](#)
- cost
 - edge AI devices, [4827](#)
- cost-loss model, [4371](#)
- cost-sensitive learning, [3806](#)
- CoT faithfulness, [3971](#)
- counterexample, [3024](#)
 - marginals vs. joints, [4331](#)
- counterfactual methodology, [4123](#)
- counterfactual probing
 - video transformer, [4123](#)
- counterfactual query
 - epilepsy, [3935](#)
- counterfactual reasoning
 - epilepsy, [3934](#)
 - neuroimaging, [3842–3844](#)
- counting
 - failure, [722](#)

- coupled cluster theory, 3370
- coupled PDEs, 3371–3377
- coupling, 80, 369
 - joint distribution, 3865
- coupling flow, 452–455
 - affine coupling, 452
 - Jacobian, 453
- coupling polytope, 369
 - exercise, 382
- Courant–Friedrichs–Lewy condition, 4190
- covariance, 35
- covariance matrix, 36
- Coverttype dataset, 2854
- CPPN, 2052, 2283
- CPRA, 4446
- CPU
 - latency optimisation, 4934
- CPU offloading, 5068
- CPU pipeline
 - communication hiding analogy, 4983
- Cranfield experiments, 2349
- Cray Slingshot, 4946
- creativity
 - and hallucination, 3171
 - in mathematics, 2933
 - vs optimisation, 2222
- credit assignment, 1375–1380
 - counterfactual, 1376
 - in LLM training, 2133
 - temporal, 1296
- credit assignment problem, 2133
- Credit dataset, 2854
- critic, 294
 - density ratio, 306
- critic iterations, 300
- critical periods, 1847
- Critical Success Index, 4365
 - CSI
 - nowcasting, 4239
- cross-attention, 103
 - alignment interpretation, 817
 - AR-diffusion, 727
 - attribution problem, 3995
 - audio conditioning, 817
 - bidirectional, 3852
 - chunk conditioning, 698
 - conditional expectation
 - interpretation, 817
 - conditioning, 613
 - definition, 615
 - definition for audio, 817
 - dictionary interpretation, 615
 - diffusion
 - EEG explanation, 3942
 - exercise, 619, 831
 - integral kernel, 3361
 - lesion segmentation, 3905
 - multi-scale fusion, 3895
 - text conditioning, 614–620
- Cross-Attention Alignment Block, 3851–3852
 - definition, 3852
 - exercise, 3860
- cross-attention maps, 3520
- cross-border federated learning, 4656–4657
- cross-chapter connections
 - medical, 4618
- cross-datacenter training, 5159
- cross-device federated learning, 4462–4463
 - definition, 4462
 - design exercise, 4466
- cross-domain transfer, 2275–2276
 - mathematics, 3039
- cross-encoder, 2373, 2488–2489
 - for ranking, 2589
 - formal definition, 2591

- vs bi-encoder, 2595
- cross-entropy, 69
 - conditional vector, 2725
- cross-entropy loss
 - class imbalance, 3499
 - segmentation, 3485
 - tabular, 2754
 - vs. Dice loss, 3548
- cross-hospital federated synthesis, 4610–4612
- cross-library transfer, 3079
- cross-modal alignment, 3995–3996, 4076–4085
 - attribution failure, 3995
- cross-modal attention
 - audio-visual, 744
- cross-modal circuits, 4157
- cross-modal EEG translation, 3825
- cross-modal generation, 897–904
 - accessibility, 4696
 - audio-to-audio, 902
 - federated
 - open problem, 4663
 - lyrics, 900
 - video-to-music, 897
- cross-modal interpretability benchmark, 4174
- cross-modal recommendation, 2578
- cross-modal reconstruction
 - information-theoretic bound, 3856
- cross-modal retrieval, 2467–2474, 2630
 - asymmetry, 2468
 - exercise, 2633
 - exercises, 2455
 - formal definition, 2467
 - hard negatives, 2468
 - implementation exercise, 2485
 - information bounds, 2560
 - information loss theorem, 2560
 - information-theoretic limits, 2473
- cross-modal synthesis, 3849–3856
 - EEG-fMRI evaluation, 3948
 - EEG-MRI, 3934
 - epilepsy challenge, 3954
 - exercise, 3790
- cross-modal task representations, 4078–4084
- cross-modal translation, 4876–4877
 - epilepsy, 3783
- cross-modality alignment, 3712
- cross-modality segmentation, 3583
- cross-patient evaluation, 3921
- cross-patient generalisation, 3920
- cross-patient generalization, 3932–3933
- cross-silo federated learning, 4462–4463
 - definition, 4462
 - design exercise, 4466
- cross-temporal retrieval, 2482
 - architecture, 2482
- cross-validation
 - stratified, 3811
- crosscoder
 - definition, 4092
 - feature taxonomy, 4092
 - taxonomy exercise, 4094
- crosscoders, 4092–4095
- crossover
 - LLM-guided, 2103–2105
 - semantic, 2104
 - system-aware merging, 2141
- crossover point
 - alpha-beta, 4956
 - exercise, 4959
- CRPS, 4292
 - as training loss, 4305
 - computation, 4294
 - decomposition, 4293
 - definition, 4292, 4314, 4330
 - EMOS fitting, 4229
 - ensemble form, 4330

- exercise, [4232](#)
- exercises, [4337](#)
- Gaussian forecast exercise, [4415](#)
- GenCast, [4314](#)
- joint blindness, [4330](#)
- nowcasting, [4365](#)
- proper scoring rule, [4229](#), [4330](#)
- reliability–resolution–uncertainty
 - decomposition, [4293](#)
 - weather evaluation, [4397](#)
- CSI, *see* Critical Success Index
- exercise, [4372](#)
- extreme thresholds, [4366](#)
- precipitation verification, [4323](#)
- CT
 - low-dose denoising, [4611](#)
- CT synthesis
 - federated, [4610](#)
- CTAB-GAN+, [2729–2743](#)
 - architecture, [2731–2734](#)
 - auxiliary classifier, [2734–2736](#)
 - decoder, [2732](#)
 - differential privacy, [2736–2740](#)
 - discriminator, [2732](#)
 - encoder, [2732](#)
 - experimental results, [2740–2743](#)
 - generator, [2732](#)
 - information loss, [2734–2736](#)
 - innovations, [2730–2731](#)
 - training procedure, [2732](#)
 - when to use, [2742](#)
- CTGAN, [2713–2729](#)
 - architecture, [2714–2716](#)
 - combined loss, [2726](#)
 - comparison with CTAB-GAN+, [2742](#)
 - comparison with GOGGLE, [2757–2758](#)
 - comparison with LLMs, [2804](#)
 - conditional vector exercise, [2916](#)
 - contributions, [2729](#)
 - convergence, [2726](#)
 - discriminator, [2715](#)
 - generator, [2715](#)
 - limitations, [2728–2729](#)
 - loss function, [2725–2728](#)
 - membership inference, [2899](#)
 - scalability, [2728](#)
- cubed sphere grid, [4214](#)
- Cubical Agda, [3002](#)
- cubical type theory, [3002](#)
- CUDA cores
 - H100, [4934](#)
- CUDA streams
 - NCCL, [4972](#)
- cultural representation
 - robustness, [4136](#)
- culturally authentic generation, [4899](#)
- cumulative distribution function, [30](#)
- curb cut effect, [4691](#)
- curb ramp detection, [4719](#)
- Curie–Weiss theory, [183](#)
- curiosity reward, [2217](#)
- curiosity-driven evolution, [2270–2272](#)
- curiosity-driven exploration
 - theorem proving, [3105](#)
- curriculum learning, [2229–2231](#), [2963](#)
 - AlphaProof, [3119](#)
 - autoformalization, [3079](#)
 - definition, [2229](#)
 - mathematical reasoning, [3107](#)
 - strategy, [3087](#)
 - theorem proving, [3087](#)
 - video, [655](#)
- Curry–Howard correspondence, [2965](#), [2984](#)
- Curry–Howard correspondence, [2206](#), [2946–2953](#)
 - definition, [2946](#)
 - diagram, [2948](#)

- curse of dimensionality, [5](#), [54](#), [157](#), [162](#), [2375](#)
 - evaluation, [2864](#)
 - in evolutionary exploration, [2236](#)
 - MAP-Elites, [2092](#)
 - MOS predictors, [4230](#)
 - PDEs, [3199](#), [3438–3440](#)
 - quantum mechanics, [3365](#)
- Cuthill–McKee algorithm, [2744](#)
- CutMix, *see* regularization, CutMix
- CVC-ClinicDB dataset, [3514](#)
- CVC5, [3022–3023](#)
- CVD correction
 - comparison table, [4786](#)
- CVD simulation
 - differentiable, [4774](#)
- CVT-MAP-Elites, [2092](#)
- Cybenko’s theorem, [3213](#)
- cybernetics, [2026](#)
- cybersecurity
 - coevolution, [2165](#)
- cycle consistency, [3582](#)
 - definition, [3583](#)
 - medical images, [3751](#)
- cycle consistency loss
 - definition, [3583](#)
- cycle-consistency loss
 - analysis, [4778](#)
- CycleGAN, [12](#), [348–350](#), [3582–3585](#)
 - architecture diagram, [3584](#)
 - brain MRI harmonisation, [3583](#)
 - cycle consistency loss, [3583](#)
 - exercise, [359](#), [3595](#)
 - framework, [3582](#)
 - identity loss, [350](#)
 - inverse mapping theorem, [349](#)
 - limitations, [3585](#)
 - medical domain adaptation, [3751](#)
 - mode collapse, [360](#)
 - objective, [349](#)
 - stain normalisation, [3515](#)
 - steganographic encoding, [3585](#)
- cyclic dynamics, [2067–2069](#)
- cyclic encoding, [2795–2796](#)
 - definition, [2795](#)
 - exercise, [2923](#)
 - properties, [2795](#)
- cyclic groups
 - prime order, [3073](#)
- cyclone
 - asymmetric flow, [4240](#)
- cylindrical Wiener process, [3378](#)
 - definition, [3378](#)
- CytoTracker, [3655–3658](#)
 - CytoEmbedding, [3655](#)
 - diffusion motion predictor, [3655](#)
 - mitosis-heavy tracking, [3656](#)
 - transformer associator, [3656](#)
- d-separation, [128](#)
 - definition, [130](#)
 - exercise, [141](#)
 - soundness, [130](#)
- DAC
 - codebook learning, [794](#)
 - configuration, [794](#)
- DAC (Descript Audio Codec), [793](#)
- DAG workflow, [1335](#)
- DAG-GNN
 - epilepsy, [3936](#)
- DALL-E
 - training data, [4443](#)
 - visual aids, [4792](#)
- Daltonization, [4769–4779](#)
 - classical, [4772](#)
 - simulation module, [4774](#)
 - subway map example, [4777](#)
 - text preservation, [4775](#)
- Daly formula
 - optimal checkpoint interval, [5148](#)

-
- DAPO, 1149–1153
 - ablation, 1153
 - algorithm, 1152
 - Clip-Higher, 1149
 - Dynamic Sampling, 1150
 - Overlong Reward Shaping, 1151
 - Token-Level Loss, 1151
 - Darcy flow, 3296–3303, 3416
 - 2D benchmark, 3302
 - benchmark, 3297
 - DeepONet, 3217
 - diffusion model, 3464
 - DiffusionPDE, 3276, 3300
 - equation, 3297
 - FNO, 3222
 - FNO implementation, 3302
 - Galerkin Transformer, 3241
 - inverse problem, 3382
 - Darcy’s law, 3297
 - DARL, 3613
 - DARTS, 1887, 1920–1926, 2254
 - algorithm, 1921
 - approximation error bound, 1922
 - architecture parameters, 1900
 - bilevel optimisation, 1921, 1922
 - collapse, 1902
 - continuous relaxation, 1900, 1920
 - continuous relaxation illustration, 2010
 - convergence rate, 1923
 - discretisation gap, 1924
 - failure modes, 1924–1926
 - first-order approximation, 1922
 - hypergradient, 1922
 - mixed operation, 1920
 - performance collapse, 1924
 - relaxation artefacts, 1924
 - second-order approximation, 1922
 - skip connection dominance, 1924
 - variants comparison, 1925
 - DARTS+, 1925
 - Darwin, Charles, 2019, 2279
 - data assimilation, 4185–4202, 4215–4221
 - 4D-Var, 4374
 - AI bypass, 4373–4383
 - computational cost, 4217
 - cost function, 4411
 - diffusion models, 3280, 3331, 4323
 - flow matching, 4344–4345
 - generative models, 4339–4349
 - gradient exercise, 4348
 - latent diffusion, 4343–4344
 - ML-based, 4380
 - NWP bottleneck, 4374
 - with diffusion, 3281
 - data at the edge, 4447
 - data attribution, 1787
 - data augmentation
 - adaptive attacks, 1685
 - beat-synchronous, 840, 956
 - common techniques, 1685
 - comparison with DP, 1685
 - cross-silo federated, 4456
 - diffusion-based, 3605
 - FCD neuroimaging, 3844
 - federated, 3720
 - generative, 3492
 - epilepsy, 3783
 - medical segmentation, 3535
 - membership advantage, 1684
 - paired generation, 3703
 - pix2pix, 3581
 - privacy, 1684–1686
 - rare pathologies, 4445
 - seizure detection, 3945
 - structured data, 2690
 - synthetic EEG, 3817
 - text-guided, 3610
 - theorem proving, 3092
 - data breaches

- historical examples, 4448
- data centralisation
 - why it fails, 4448
- data centre
 - architecture, 4927
 - hierarchy, 4927
 - TikZ figure, 4928
- data centres
 - multi-region training, 5119
- data constraints, 2703–2704
- data curation
 - exercise, 660
 - pipeline, 657
 - video, 656
- data deduplication, 1708
- data exfiltration
 - agentic systems, 1822
- data extraction
 - Lean, 3081
- data flywheel, 2901, 2959
- data heterogeneity
 - mode collapse, 4453
- data hunger, 4444
- data leakage
 - pre-training, 2813
 - temporal, 3922
 - theorem proving, 3081
- data loader
 - distributed, 5026
- data localisation laws, 4448
- data locality
 - privacy fallacy, 4590
- data manifold, 2775, 4444
- data mixing, 1197
 - for CPT, 1205
- data paradox, 4444
- data parallel training
 - MPI, 4968
- data parallelism, 5020–5031
 - backward pass, 5022
 - definition, 4929, 5021
 - diffusion models, 5084
 - effective batch size, 5173
 - formal definition, 5021
 - forward pass, 5021
 - gradient equivalence, 5030
 - history, 4931, 5020
 - LLM, 5028
 - parameter update, 5022
 - protocol, 5021
 - scaling property, 5029
 - throughput, 5029
 - token budget, 5028
- data preprocessing, 9
 - TabDDPM, 2771
- data privacy
 - regulation, 4443–4467
 - regulatory landscape, 4445
- data processing inequality, 78, 1590
 - membership leakage, 1590
 - retrieval, 2555
- data reweighting
 - disability bias, 4849
- data scarcity, 2687
 - 1990s meteorology, 4238
 - epilepsy, 3781
 - histopathology, 3895
 - in formal mathematics, 2963
 - medical imaging, 3490
- Data Science Bowl 2018, 3644
- data sharing
 - synthetic alternatives, 2687
- data silo, 4447
 - definition, 4447
 - formal definition, 4448
- data sovereignty, 4448
 - as constraint, 4449
 - biometric data, 4814
 - edge AI, 4815
- data stealing

- federated diffusion, [4593](#)
- data types
 - landscape, [2691–2698](#)
- data warehouses, [2690](#)
- data-driven methods
 - vs. physics-driven, [3446](#)
- data-parallel training, [5115](#)
- dataset bias
 - medical segmentation, [3726](#)
- datasets
 - accessibility
 - multimodal, [4879](#)
 - epilepsy catalog, [3919](#)
- datasphere
 - 200 zettabytes, [4447](#)
- Davis, Martin, [3018](#)
- DCG (discounted cumulative gain), [2360](#)
- DCGAN, [280](#), [2732](#)
 - brain MRI, [3589](#)
 - brain tumour generation, [3589](#)
 - EEG synthesis, [3818](#)
 - influence on CTAB-GAN+, [2731](#)
- DCT
 - gradient compression, [4627](#)
- DDIM, [2768](#)
 - acceleration, [3837](#)
 - definition, [4313](#)
 - EEG, [3836](#)
 - epilepsy application, [3931](#)
 - for video, [645](#)
 - step reduction, [646](#)
 - versus flow matching, [646](#)
 - video formulation, [645](#)
 - weather sampling, [4313](#)
- DDIM inversion
 - content preservation, [894](#)
 - video, [717](#)
- DDP
 - efficiency, [4941](#)
- gradient bucketing exercise, [5031](#)
- PyTorch, [4974](#)
- DDPM, [479](#), [491](#)
 - data parallelism, [5028](#)
 - definition, [4308](#)
 - EEG, [3945](#)
 - EEG anomaly detection, [3839](#)
 - forward process, [3596](#)
 - impact, [2775](#)
 - noise schedule, [3596](#)
 - cosine, [3597](#)
 - exercise, [3616](#)
 - linear, [3597](#)
 - reverse process, [3596](#)
 - sampling algorithm, [493](#)
 - segmentation, [3596–3600](#)
 - training algorithm, [493](#)
 - video extension, [544](#)
 - video vs. image training, [545](#)
- DDPO, [1155](#)
 - variance, [1155](#)
 - variance exercise, [1164](#)
- de Bruijn, Nicolaas, [2944](#)
- De Jong, Kenneth, [2062](#), [2282](#)
- De Morgan, Augustus, [3016](#)
- de Moura, Leonardo, [2945](#)
- De Novo
 - AI device, [3939](#)
- De Novo pathway, [3939](#)
- de-identification
 - EEG, [3826](#)
- dead codes, [798](#)
- dead ends
 - proof search, [3088](#)
- dead features, [4053–4055](#), [4089](#)
 - complementarity, [4089](#)
 - exercises, [4057](#)
 - layer dependence, [4089](#)
- deadlock
 - MPI, [4964](#)

- Deaf community, [4690](#)
- Deb, Kalyanmoy, [2283](#)
- debate
 - scalable oversight, [4147](#)
- debiasing
 - deployment-level, [4851](#)
 - disability, [4849](#)
 - disability representation, [4902](#)
 - generative models
 - EEG, [3942](#)
- decentralised diffusion
 - exercise, [5134](#)
- decentralised diffusion models, [5128](#)
 - training, [5128](#)
- decentralised training, [5015](#), [5121](#)
 - overview, [5124–5134](#)
- decentralized diffusion models, [5092](#)
- decentralized training
 - bandwidth constraints, [5097](#)
- deception
 - internal state, [4139–4140](#)
- deceptive alignment, [4134–4135](#)
 - definition, [4134](#)
- deceptive fitness landscapes, [2081](#)
 - novelty search, [2082](#)
- decidable propositions, [2990](#)
- decide tactic, [2990–2991](#)
- decision algorithm
 - epilepsy, [3949](#)
- decision problem, [3016](#)
- decision procedure vs proof procedure, [3032](#)
- decision theory
 - forecast value, [4371](#)
- decision tree
 - metric selection, [3925](#)
- declarative memory, [1175](#)
- decoder, [386](#), [388](#)
- deconvolution, *see* transposed convolution
- decorrelation time, [2799](#)
- decorrelation timescale, [530](#)
- deductive database (DD+AR), [3122](#)
- Dedukti, [3009](#)
- deep brain stimulation, [3873–3874](#)
 - closed-loop, [3916](#), [3936–3937](#)
 - limitations, [3874](#)
- deep equilibrium model, [1054](#)
 - backpropagation, [1054](#)
 - example, [1056](#)
 - exercise, [1069](#)
- deep image embedding, [2406](#)
- deep image retrieval, [2406–2417](#)
- Deep Leakage from Gradients, [4518–4519](#)
- deep learning, [91](#)
 - accessibility, [4693](#)
 - epilepsy, [3776](#)
 - medical segmentation, [3487](#), [3531–3548](#)
 - opacity, [3962](#)
- deep learning revolution
 - meteorology, [4241](#)
- deepfake, [17](#)
 - election interference, [18](#)
 - financial fraud, [17](#)
 - non-consensual imagery, [18](#)
 - sign language avatars, [4742](#)
- deepfake detection
 - audio, [929–937](#)
- deepfakes
 - federated, [4655–4656](#)
 - medical imaging, [3593](#)
- DeepMath, [3034–3035](#)
 - paradigm, [3038](#)
- DeepMind
 - AI for math ecosystem, [3131–3146](#)
 - unified view, [3139](#)
 - AI for Math Initiative, [3132](#)
 - Aletheia, [3133](#)

- AlphaProof, [3112–3131](#)
- DeepNorm, *see* normalization,
 - DeepNorm
 - scaling derivation, [1552](#)
- DeepONet, [3214–3218](#)
 - architecture, [3214](#)
 - connection to FNO, [3232](#)
 - mathematical formulation, [3214](#)
 - physics-informed, [3216](#)
 - Poisson equation, [3462](#)
 - stacked, [3218](#)
 - training, [3215](#)
 - universal approximation, [3458](#)
 - variants, [3217](#)
- DeepSeek, [114](#)
 - comparison with OpenAI, [3149](#)
 - mathematical AI approach, [3136](#)
 - MLA, [115](#)
- DeepSeek-Math, [3057–3058](#)
 - architecture, [3058](#)
 - GRPO hyperparameters, [3058](#)
- DeepSeek-Prover, [2961](#), [3136](#)
- DeepSeek-R1, [1049–1051](#), [1148](#), [3052–3053](#)
 - emergent behaviours, [3052](#)
 - GRPO, [1047](#)
 - reward design, [1050](#)
 - training pipeline, [1049](#)
 - training recipe, [3052](#)
- DeepSpeed
 - 3D parallelism, [5136](#)
- DeepSpeed ZeRO, [4927](#)
- Deerwester, Scott, [2627](#)
- default mode network, [4150](#)
 - analogy, [4150](#)
- default settings
 - accessibility, [4703](#)
- DEffusion, [4318–4329](#)
 - CRPS improvement, [4321](#)
- definitional equality, [2981](#)
- defuzzification, [4235](#)
- DeGroot learning
 - exercise, [1455](#)
- Dehaene, Stanislas, [4148](#)
- delay pattern, [951](#)
- DeLTa, [2815–2834](#)
 - architecture, [2820–2834](#)
 - generation algorithm, [2822](#)
 - healthcare example, [2823](#)
 - motivation, [2816](#)
 - pipeline, [2821–2823](#)
- delta waves, [3793](#)
- dementia
 - generative AI, [4799](#)
 - prevalence, [4690](#)
- Demmel lower bound, [5002–5003](#)
 - machine learning, [5003](#)
 - statement, [5002](#)
- demographic bias
 - EEG datasets, [3941–3942](#)
- demographic inference
 - disability cues, [4844](#)
- demographic parity, [2777](#), [2856](#)
 - generative, [2785–2786](#)
 - projection, [2785](#)
- demographic parity gap, [2856](#)
- DEN (Dynamically Expandable Networks), [1960](#)
- Dennett, Daniel, [2020](#)
- denoiser
 - spatiotemporal, [545](#)
 - spatiotemporal requirement, [545](#)
- denoiser mean, [490](#)
- denoising, [330–331](#)
 - as feature learning, [2610](#)
 - classical methods, [330](#)
 - iterative, [2759](#)
 - mixed data, [2763–2768](#)
 - problem formulation, [330](#)
- denoising autoencoder, [330](#), [386](#)

- exercise, 399
- denoising diffusion implicit models, 3837
- denoising diffusion probabilistic model, 491, 2758
- denoising diffusion probabilistic models
 - forward process, 3745
- denoising error spectrum, 1649
- denoising errors
 - spectral analysis, 1649
- denoising loss
 - estimation variance, 1612
 - membership signal, 1611, 1617
- denoising objective, 489
- denoising score matching, 87, 5089
 - discretised fields, 3459
- denoising step
 - computation-communication overlap, 4995
- dense environments
 - hallucination, 4830
- dense model
 - memory implications, 4932
- Dense Passage Retrieval, *see* DPR
- Dense Passage Retriever, 2520
- dense retrieval, 2364–2374
 - exercises, 2374
 - formal definition, 2365
 - paradigm shift, 2365
 - revolution, 2364
 - theorem proving, 3083
 - vs. sparse
 - exercise, 2635
- DenseNet, *see* skip connections, DenseNet
 - dense block, 1498
 - growth rate, 1498
 - parameter count, 1552
- density estimation
 - function space, 3383
- density functional theory, 3370
- density ratio estimation, 276
 - exercise, 287
- dependency graph, 2745
 - comparison, 2919
- dependent function type, *see* Pi-type
- dependent type, 2947, 2972
- dependent types, 2949, 2984–2999
 - definition, 2984–2985
- depth
 - trainability 1990s, 4237
- DEQ, *see* deep equilibrium model
- derivative-free optimisation, 2056
- dermatology segmentation
 - fairness, 3725
- dermatoscopy
 - segmentation, 3506
- DermoSegDiff, 3604–3605
 - boundary attention, 3604
 - boundary loss, 3604
 - boundary width, 3605
- Descript
 - motor accessibility, 4872
- design of experiments
 - history, 2235
- detailed balance
 - GFlowNet, 2867–2873
- detailed balance condition, 2868
- detection
 - synthetic data, 2841
- detection metrics, 2843–2844
- detection sensitivity
 - generative scaling, 3912
- deterministic chaos, 4183
- deterministic MLWP
 - comparison with generative, 4305
 - limitations, 4195
- Deterministic Nonperiodic Flow, 4185
- deterministic sampling, 645

- deterministic transitions, 3097
- deutanopia, 4770
 - cycle consistency, 4913
 - transit map, 4777
- developmental interpretability, 4141
- developmental sample
 - MOS, 4223
- device heterogeneity, 4579–4589, 5119
 - landscape, 4579
 - spectrum, 4580
- device-level data silo, 4448
- Devlin, Jacob, 2629
- Dewey Decimal Classification, 2626
- Dewey, Melvil, 2626
- DGMR, 4364
 - architecture, 4364
 - evaluation, 4365
 - formal definition, 4364
- DGX
 - interconnect topology, 4989
- DGX H100
 - node specification, 4927
- DGX SuperPOD, 4949
- Dhariwal, Prafulla, 3035
- DHH
 - ethical concerns, 4741–4742
- diabetic retinopathy, 4690
 - federated GAN, 4490
- diagnosis
 - segmentation role, 3479
- dialect diversity
 - NLP, 4445
- dialogue management
 - ECA, 4769
- diameter
 - communication cost, 4943
 - network, 4943
- diastole, 3513
- Dice coefficient, 3486, 3505
 - BraTS, 3905
- clinical relevance, 3547
- definition, 3544
- inter-observer variability, 3483
- MRI evaluation, 3923
- properties, 3498, 3740
- relationship with IoU, 3544
- Dice loss, 3486
 - class imbalance, 3499, 3537
 - gradient, 3546, 3740
 - optimisation properties, 3546
 - V-Net, 3537
 - vs. cross-entropy, 3548
- Dice score
 - FCD, 3842
- DICOM
 - RT structure set, 3484
- Diff-EEG, 3832–3835
 - augmentation example, 3836
 - pipeline figure, 3835
- DiffAug, 3610–3612
 - colonoscopy application, 3611
 - latent-space validation, 3611
 - pipeline, 3610
- DiffDA, 4323
 - verification, 4324
- diffeomorphism
 - contour deformation, 3714
- difference map
 - pseudo-healthy
 - FCD, 3956
- difference of Gaussians, 2398
- difference rewards, 1377
- differentiable architecture search, 1920–1926
- differentiable dynamical core, 4280
- Differentiable Neural Computer, 1090
- differentiable physics, 3308
 - segmentation, 3732
- differentiable quality-diversity, 2088–2094, 2284

- Differentiable Search Index, 2508–2511, 2630
 - exercise, 2635
- differentiable solvers, 3309
- differential circuit audit, 4138
- differential privacy, 19, 1595–1607, 2736–2740, 2845–2846, 4521–4522
 - (ϵ, δ) -DP, 1596
 - absence in CTGAN, 2728
 - advanced composition, 1600
 - budget computation, 2919
 - comparison with VAEFL, 4521
 - composition, 1673
 - computational cost, 1677
 - conditional generation, 2925
 - definition, 1596, 2706, 2736, 4495
 - diffusion models, 1670–1678
 - dimensionality curse, 1677
 - discrete diffusion, 4544
- DP-SGD
 - ghost clipping algorithm, 1809
 - LLM scale, 1808
- DP-SGD guarantee, 2737
- EEG, 3830
- EEG generation, 3940
- effect on utility, 2742
- evolutionary tuning, 2253
- exercise, 1701
- federated, 2884
- federated GAN, 4495–4505
- federated learning, 4451, 4599
- fine-tuning advantage, 1812
- fine-tuning vs pre-training, 1811
- formal definition, 2736
- Gaussian mechanism, 1598
- generative models, 3765
- ghost clipping, 1808
- GPT-2 example, 1811
- history, 1703
- interpreting ϵ , 1597
- language model training, 1808
- large language models, 1807–1814
- limitations, 4661
- LLaMA fine-tuning, 1812
- LoRA, 1812
- lower bound, 1675
- membership advantage bound, 1597
- membership inference, 4599
- membership inference audit, 3951
- minimax bound, 1689
- multi-agent composition, 1826
- NAS, 2000
- optimal model size, 1814
- output perturbation, 1818
- parameter settings, 1813
- Pareto optimisation, 2253
- per-sequence cost, 1810
- personalized federated, 4645
- practical considerations, 1814
- privacy accounting
 - next-token prediction, 1810
- privacy-utility tradeoff, 1600
- propagation through generator, 4498
- quality-privacy tradeoff, 1674–1677
- robustness guarantee, 1688
- scaling laws, 1813
- score matching, 1670
 - definition, 1671
 - guarantee, 1671
- segmentation quality, 3757
- sequence truncation, 1810
- synthetic data guarantee, 4456
- theory-practice gap, 1691
- total privacy cost, 1810
- tradeoff, 2705–2708
- unlearning, 1805
- utility bounds, 2906

- difficulty
 - human vs AI, 3126
- difficulty estimation, 3108
- diffusion
 - comparison with flow matching, 865
 - denoising steps
 - pipeline, 5054
 - discrete forward process, 2917
 - interpolation, 547
 - source separation, 891
 - temporal transformer, 3323
- diffusion architecture
 - audio, 841–847
 - Moûsai, 846
 - U-Net vs. DiT, 842
- diffusion features, 2414–2416
 - definition, 2415
 - intuition, 2414
- diffusion language model
 - exercise, 1070
- diffusion language models
 - attack surface, 1695
 - medical applications, 1696
 - membership inference, 1692–1703
 - music connection, 944
 - privacy implications, 1697
 - privacy risk, 1695
- diffusion loss
 - conditional, 3851
- diffusion model, 7, 475
 - cell tracking, 3655
 - EEG-to-fMRI, 3851
 - energy prediction, 500
 - epilepsy applications, 3786
 - ϵ -prediction, 498
 - gradient inversion, 4592
 - lesion synthesis, 3786
 - Markovian, 477
 - memorisation, 4590
 - federated, 4593
 - motion prediction, 3656
 - MRI pipeline, 3914
 - MRI super-resolution, 3786
 - noise schedule, 4415
 - parameterisation, 497
 - score prediction, 500
 - toy implementation, 5178
 - vs. FGN, 4332
 - weather downscaling, 4353
 - x_0 -prediction, 497
- diffusion models
 - 1D PDE fields, 3463
 - 24B decentralized training, 5094
 - accessibility, 4703, 4883
 - activation memory, 4932
 - advantages for PDEs, 3266
 - alignment, 1153–1158
 - architecture bottleneck, 1935
 - architecture comparison, 1942
 - architecture configuration, 1936
 - architecture generation, 1949
 - architecture search, 1895, 1935–1943
 - attribution problem, 3993
 - bias
 - mechanistic interpretability, 4111–4113
 - bias mitigation, 4174
 - boundary hallucination, 3721
 - causal intervention
 - exercises, 4103
 - coarse-to-fine segmentation, 3599
 - collaborative training, 5128
 - communication hiding, 4995–4999
 - communication patterns, 5089–5097
 - comparative analysis, 3292
 - comparison with GANs, 358, 3590
 - computational cost, 3599
 - concept emergence, 4103
 - concept schedule, 4096

- conditioning, 3597–3600
- continual learning, 1219–1224
- CVD refinement, 4781
- data parallelism, 5028
- denoising trajectory, 4096
- design space, 1935
- distributed backward, 5104
- distributed training, 4932, 4977, 5081–5089, 5142
- EEG, 3831–3840
 - Diff-EEG, 3832
- EEG augmentation, 3945
- elliptic inverse problems, 3299
- embarrassing parallelism, 5088
- ensemble challenge, 3617
- evolutionary architecture search, 2252
- feature injection, 4100–4102
- features for retrieval, 2610
- federated, 4459, 4526–4538, 4574
 - billion-parameter, 4659
 - challenges, 4526
 - open problem, 4659
- federated compatibility, 4459
- for prompt generation, 3697
- full-sequence, 2801
- function space, 3377–3384, 3455
- function spaces, 3267–3295
- generation steering, 4102
- heat equation analogy, 3267
- historical development, 3616
- hybrid discrete-continuous, 4544–4545
- hybrid forward process, 2896
- image description, 4703
- inference cost, 3616
- interpretability, 4095–4102
- inverse problems, 3401
- layout preservation
 - exercises, 4104
 - layout steering, 4102
- log-probability, 1154
- mask sampling, 3606
- MDP exercise, 1164
- MDP formulation, 1154
- medical imaging, 3745, 5122
- medical imaging example, 5122
- medical segmentation, 3596–3618
- memorisation, 1575
- motivation for PDEs, 3265
- neuroimaging, 3840–3847
- PDE control, 3317
- PDE fields, 3459
- PDEs, 3264–3295
- physics-informed, 3467
- physics-informed distillation, 3354
- pipeline parallelism, 5054
- privacy, 1608–1616
- probabilistic framework, 4308
- real-time inference, 3762
- SAE, 4098–4102
- segmentation methods, 3600–3606
- seizure forecasting, 3836
- self-supervised segmentation, 3612–3613
- semantic layout stage, 4097
- sleep-inspired NAS, 2000
- speed-quality conjecture, 3729
- speed-quality tradeoff, 3728
- style stage, 4097
- style steering, 4102
- tabular data, 2758–2775, 2896–2904
- temporal analysis, 4168
- temporal concept schedule, 4097
 - definition, 4097
- texture clamping, 4102
- texture stage, 4097
- time-aware manipulation, 4101–4102
 - example, 4101

- timestep, 4096–4102
- training cost, 1936
- training loss, 4996
- training steps, 5082
- training vs LLM training, 5081
- turbulence, 3284
- uncertainty quantification, 3412
- universality, 2775
- unsupervised segmentation, 3612–3613
- VI perspective, 256
- VLB derivation, 267
- VP process on L^2 , 3272
- vs. other families, 2709
- weather, 4198
- weather application, 4197
- weather ecosystem, 4318–4329
- weather forecasting, 4305, 4307–4329
- well-posedness, 3273
- zero-shot segmentation, 3612
- diffusion parameterisation
 - audio, 850
 - exercise, 863
 - gradient analysis, 852
 - implicit weighting, 851
 - three types, 850
- Diffusion Posterior Sampling, 3402
 - exercise, 3407
- diffusion posterior sampling, 4323, 4358
 - function space, 3381
- diffusion process
 - audio latent, 848–852
- diffusion steps
 - number of, 2771
- diffusion tensor imaging
 - definition, 3890
 - epilepsy, 3889–3896
- diffusion timestep
 - pipeline heterogeneity, 4996
- diffusion trajectory, 2780
- Diffusion Transformer, 590–598, 4367
 - block architecture, 593–598
 - block definition, 594
 - exercise, 598
 - history, 597
 - nowcasting, 4367
 - patchification, 590
 - practical scaling, 597
 - scaling laws, 596
 - strengths, 1941
 - token count, 591
 - vs U-Net, 596–598
 - weather, 4341
- diffusion transformer
 - tensor parallelism, 5048
- Diffusion Transformer (DiT), 1941, 4115
- diffusion transformers
 - bias circuits, 4112
- diffusion weather models
 - comparison, 4324
- Diffusion-Augmented Representation Learning, *see* DARL
- Diffusion-DPO, 1156–1157
 - exercise, 1165
 - loss, 1157
- diffusion-reaction equation, 3416
- diffusion-reaction equations, 3303–3311
 - definition, 3305
- DiffusionNAG, 1937–1939
 - algorithm, 1939
 - architecture representation, 1937
 - conditional generation, 1938
 - denoising network, 1938
 - graph diffusion process, 1937
 - pipeline, 2013
 - quality-diversity tradeoff, 1938
 - results, 1939
- DiffusionPDE, 3274–3295, 3300, 3330
 - architecture, 3275

- cylinder flow, 3330
- implementation exercise, 3293
- observation guidance, 3330
- PDE constraint guidance, 3275
- posterior consistency, 3277
- sparse observation guidance, 3275
- DiffusionPipe, 4996, 5054–5060, 5085–5086
 - bubble analysis, 5088
 - speedup analysis, 5175
 - topology-aware, 5167
- DiffWave, 767
- DIFT, 2414–2416, 2610
 - cross-domain retrieval, 2416
 - definition, 2415
 - timestep selection, 2415
- digamma function, 194
- digital accessibility, 4690–4697
- digital audio, 769
 - definition, 770
- digital brain twin, 3931–3932
- digital divide
 - accessibility, 4826
 - accessibility technology, 4860
 - disability, 4689
 - WCAG compliance, 4692
- digital organisms, 2241
- Digital Red Queen, 2154–2170
 - algorithm, 2156
 - archive regret, 2163
 - arms race, 2159
 - avoiding pathologies, 2161
 - convergence, 2164
 - convergent evolution, 2159
 - cross-domain results, 2166
 - dual-use, 2167
 - emergent complexity, 2165
 - full algorithm, 2157
 - generalisation, 2165
 - LLM mutation, 2158
 - mathematical analysis, 2163
 - monotonicity, 2163
 - responsible use, 2167
- digital twins, 3433
- dilated causal convolution, 442
 - definition, 442
- DiLoCo, 5126–5127
 - connection to FedAvg, 5127
 - exercise, 5133
 - formal definition, 5126
- dimensionality
 - explosion, 521
 - video, 521
- dimensionality reduction, 386
- direct effects, 4013
 - definition, 4013
 - exercises, 4019
- direct ensemble forecasting, 4320
- direct numerical simulation
 - grid requirements, 3456
 - impossibility, 4209
- direct preference optimization, 1138–1144
- direct reward fine-tuning, 1156
- directed acyclic graph, 126
 - Bayesian network, 4235
 - GFlowNet, 2866
- directed matrix-tree theorem, 3170
- Dirichlet distribution, 193
- Dirichlet process, 158
 - mixture model, 158
- Dirichlet smoothing, 2354
- disabilities
 - AI-assisted, 4689
- disability
 - aging population, 4689
 - co-design, 4920
 - global prevalence, 4689
 - individual variation, 4703
 - spectrum, 4690

taxonomy, 4690, 4694
 disability justice
 AI, 4852
 disability-led design, 4894
 disaster preparedness
 weather, 4183
 discrete diffusion, 2771
 backward generation
 exercise, 4548
 federated, 4538
 federated advantages, 4542–4549
 federated training, 4543–4544
 forward process, 4541
 inference speed, 4546
 masked diffusion loss
 exercise, 4548
 masking schedule
 exercise, 4678
 open questions, 4547
 privacy
 gradient leakage, 4542
 symbolic music, 909, 944
 vs. autoregressive
 comparison, 4546
 discrete features
 projection, 2786
 discrete latent space, 408
 discrete masked diffusion
 language model
 definition, 4540
 discrete metric, 5195
 discretisation
 NWP, 4213
 discretisation consistency, 3272
 discretisation error, 316, 2782
 discretisation invariance
 diffusion models, 3470
 discretisation-agnostic generation,
 3431–3433
 discretised logistic mixture, 441

discrimination information, *see* KL
 divergence
 discriminative model, 4
 discriminative models
 accessibility harm, 4699
 accessibility limitations, 4697–4706
 exercise, 4704
 limitations for segmentation, 3490
 structural insufficiency, 4699
 vs generative interpretability, 3966
 vs. generative
 epilepsy, 3783
 discriminative segmentation
 comparison with generative, 3493
 definition, 3491
 discriminative vs. generative
 accessibility comparison, 4701
 discriminator, 274
 optimal, 275
 exercise, 285
 prediction gap, 4597
 discriminator gap
 membership inference, 4597
 disengagement, 2069–2071
 definition, 2069
 fitness gap condition, 2071
 simulation exercise, 2071
 disentangled representation, 405
 disentanglement, 353
 latent space
 EEG, 3943
 disjoint support, 82
 disjunction, 2987
 Dissertatio de Arte Combinatoria, 3016
 distance function, *see* metric space
 DistBelief, 5020
 Google, 4931
 DistilBERT, 2396
 distillation
 audio, 873–880

- exercise, 3396
- PDE solvers, 3387–3391
- physics-guided, 3289
- video, 649
- video diffusion, 750
- distributed AI
 - bottleneck analysis, 4931
 - career relevance, 4927
 - practitioner knowledge, 4927
- distributed auto-diff
 - postal system analogy, 5104
- distributed checkpointing, 5148
- distributed computational graph
 - definition, 5099
- distributed computing
 - motivation, 4925
- distributed data parallel
 - implementation, 5178
- distributed generative AI, 4925–4942
- distributed gradient averaging, 5171
- distributed memory
 - programming model, 4959–4971
- distributed optimisation
 - definition, 4929
- distributed problem solving, 1374
- distributed representations, 3969
 - biology, 3969
- distributed training, 4925
 - 7B recipe, 5112
 - all-reduce, 5031–5041
 - autodiff, 5097
 - chapter summary, 5181
 - communication bottleneck, 5040
 - communication hiding, 4981–4999
 - comparison table, 5118
 - complete recipe, 5112
 - data parallelism, 5020–5031
 - definition, 4929, 5117
 - diffusion models, 5081–5089
 - fault tolerance, 5144–5155
 - formal definition, 4929, 5116
 - history, 4931
 - hybrid strategies, 5135–5144
 - key idea, 5122
 - low-communication, 5126
 - memory optimization, 5060–5069
 - milestones, 4931
 - necessity, 4926
 - open problems, 5155–5168
 - hybrid parallelism search, 5155
 - paradigm, 5115
 - topology, 4942–4951
 - vs federated learning, 5115–5124
 - when to use, 5118
- DistributedDataParallel, 5023
 - algorithm, 4975
 - communication hiding, 4988
 - gradient bucketing, 4991–4999
- DistributedSampler, 5026
- distribution preservation, 2792
- distribution comparison, 2850–2852
- distribution over segmentations, 3554
- distribution perspective
 - segmentation, 3491
- distribution shift, 1184, 2507
 - e-commerce, 2507
 - expert iteration, 3099
 - nowcasting, 4368
 - segmentation, 3490
 - time series, 2802
- distribution skew, 4569
 - definition, 4569
- distribution-aware fairness, 3724
- distribution-aware sampling, 1205
- distributional mismatch
 - visual tokens, 4077
- DiT, *see* Diffusion Transformer
 - 6.5B distributed training example, 5087
 - audio diffusion, 842

- Diffusion Transformer, 5083–5084
- diffusion transformer, 5143
- parallelism design, 5089
- pipeline parallelism, 4996
- Stable Audio, 843
- tensor parallelism, 5048, 5084
- DiTTO, 3323
- diurnal temperature range
 - NWP bias, 4227
- divergence, 67
 - role in generation, 59
- divergence-free
 - generation, 3284
- diversity
 - conjecture generation, 3108
 - in generation, 2865
 - synthetic data, 2705
 - video generation, 524
- diversity maintenance, 2029
- diversity metrics, 926–927
 - quality trade-off, 926
- divide and conquer agent, 2185
- divisibility, 2996
 - in Lean, 2976
- division of labour, 1275
- DLG attack, 4518–4519, 4591
 - batch size exercise, 4598
 - DLG seminal, 4591
- DLSS, 12
- DNS, *see* direct numerical simulation
- Doc-to-LoRA, 1202–1203
 - amortisation, 1202
 - definition, 1202
- DocID
 - numeric, 2509
 - semantic hierarchical, 2509
 - semantic structure, 2510
 - string, 2509
- document chunking, 2389
- document understanding
 - accessibility, 4701
- domain adaptation
 - adversarial, 3578
 - clinical significance, 3579
 - CycleGAN, 3582
 - formal definition, 3758
 - GAN-based vs. diffusion-based, 3758
 - limitations, 3579
 - unpaired translation, 3751
- domain constraints
 - learning, 2824–2826
 - tabular data, 2820
- domain decomposition
 - inverse problems, 3405
- domain expert
 - graph specification, 2747
- domain gap
 - MRI, 4611
- domain generalisation
 - medical imaging, 3726
 - theoretical bounds, 3727
- domain knowledge
 - encoding, 2754
- domain shift
 - EEG, 3802
 - federated learning, 3719
 - mathematical formalization, 3758
 - medical imaging, 3490, 3726
 - microscopy, 3510
 - scanner-specific, 3726
- domain-adaptive pretraining, 1185
- domain-incremental learning, 1177
- domain-specific evaluation, 2864
- domain-specific reasoning, 3022
- domestication
 - as evolutionary precedent, 2265
- dominance conjecture
 - weather modeling, 4407
- dominated convergence theorem, 5228

- dosimetry
 - segmentation, [3479](#)
- Doswell, Charles A., [4234](#)
- double binary tree, [4972](#), [5036](#)
 - algorithm, [5036](#)
- double effectiveness hypothesis, [3448](#)
- double-well potential, [3357](#)
- doubly stochastic matrices, [1552](#)
- doubly stochastic matrix, [375](#), [1500](#)
 - exercise, [382](#)
- Downpour SGD, [5020](#)
- downscaling
 - classical methods, [4351](#)
 - generative, [4349](#)
 - generative approach, [4352](#)
 - generative vs. classical, [4359](#)
 - method comparison, [4359](#)
 - problem formulation, [4350](#)
 - statistical, [4223](#)
 - MOS, [4225](#)
 - statistical definition, [4350](#)
- downstream evaluation, [2855–2856](#)
- downstream task
 - performance gap, [2924](#)
- DP score matching, [1671](#)
- DP-FedGAN, [4495–4505](#)
- DP-FL, [4456](#)
- DP-FTRL
 - Gboard, [4626](#)
- DP-SGD, [1598](#), [2736–2740](#), [4496](#)
 - algorithm, [1599](#)
 - CIFAR-10 experiments, [1672](#)
 - diffusion training algorithm, [1672](#)
 - EEG generation, [3941](#)
 - exercise, [2919](#), [4504](#)
 - federated learning, [4456](#), [4602](#)
 - Gboard, [4626](#)
 - noise calibration
 - exercise, [4675](#)
 - noise computation example, [1599](#)
 - scalability, [1677–1678](#)
 - score matching, [1670](#)
- DPLL algorithm, [3018](#)
- DPLL($\mathcal{T}$), [3023](#)
- DPO, [1138–1144](#), [1762–1771](#)
 - equivalence to RLHF, [1140](#)
 - exercise, [1162](#)
 - gradient, [1140](#)
 - gradient analysis, [1162](#)
 - implicit reward, [1141](#)
 - limitations, [1144](#)
 - loss function, [1139](#)
 - partition function cancellation, [1139](#)
 - reparameterization, [1138](#)
 - RLHF equivalence, [1163](#)
 - temperature β , [1143](#)
 - vs. PPO, [1143](#)
- DPR, [2366–2373](#), [2629](#)
 - training algorithm, [2367](#)
- DQD, [2088](#)
 - definition, [2088](#)
- DRaFT, [1156](#)
 - exercise, [1165](#)
 - objective, [1156](#)
 - truncated backprop, [1156](#)
- Draft-Sketch-Prove, [3068–3069](#)
 - exercise, [3077](#)
- drafting-execution pipeline, [1341](#)
- dragonfly topology, [4945–4946](#)
- Dravet syndrome, [3777](#)
- dream indexing, [2483](#)
- DreamBooth
 - video, [692](#)
 - video loss, [692](#)
 - with LoRA, [692](#)
- dreaming
 - as data augmentation, [1968](#)
- drift
 - diffusion, [2776](#)
- drizzle everywhere artefact, [4195](#)

- DROP benchmark, 2186
- dropout, *see* regularization, dropout, 2715
- ensemble interpretation, 1554
- in TabDDPM, 2764
- privacy, 1683
- dropout rate, 1517
- DropPath, *see* regularization, DropPath
- DRQ, 2286
- drug design
- case study, 2891
 - generative models, 2891–2896
 - pipeline, 2891
- drug discovery, 14
- federated, 4666
 - federated generative models, 4666
- drug-resistant epilepsy, 3777, 3885
- consequences, 3777
 - definition, 3899
 - ILAE criteria, 3899
 - ILAE definition, 3777
 - prevalence, 3773
 - surgical referral
 - exercise, 3909
- DrVD-Bench, 3686
- VLM reasoning failures, 3687
- DSI, 2508–2511, 2630
- constrained decoding, 2511
 - DocID design, 2509–2510
 - Natural Questions, 2510
 - training, 2510
- DTW, *see* dynamic time warping
- dual encoder, 2408, 4511
- dual function framework
- neurodivergent AI, 4791
- dual text encoder, 615–620
- definition, 615
 - rationale, 616
- dual text encoding
- ablation, 689
 - definition, 685
- dual-encoder, 2365
- efficiency, 2373
- dual-loop architecture, 2190
- dual-use
- federated generative AI, 4655–4656
 - synthetic EEG, 3933
- dual-use technology
- voice synthesis, 939
- duality
- duality gap, 5262
 - Lagrange dual function, 5262
 - strong duality, 5262
 - weak duality, 5262
- Dubinski et al. attack, 1663
- DuoLearn, 1815
- algorithm, 1815
 - GPT-2 experiment, 1816
 - theoretical guarantee, 1816
- duoT5, 2490
- Duplicate Token Heads, 4031
- walkthrough, 4034
- dynamic agent generation, 1340
- dynamic architecture function, 1955
- dynamic channel selection, 3882
- analysis, 3896
- dynamic conditional coding, 3600
- definition, 3601
- dynamic difficulty adjustment (DDA), 4793
- dynamic loss scaling, 5107
- PyTorch, 5108
- dynamic network expansion, 1869
- dynamic networks, 1955–1960
- input-dependent, 1955
 - Pareto dominance, 1956
- dynamic programming
- architecture search, 1927
- Dynamic Sampling, 1150
- definition, 1151

- exercise, 1164
- dynamic time warping, 2427–2429
 - complexity, 2428
 - definition, 2428
 - exercises, 2437
 - subsequence variant, 2429
- dynamic workload
 - topology, 5157
- dynamical core
 - differentiable, 4280
- dynamical downscaling, 4351
- dynamical isometry, 1476
- dynamical systems
 - and RNNs, 1035
- dynamically expandable networks, 1960
- DynamiCrafter, 674
- dynamics modelling, 535–540
 - spectrum, 536
- dysarthria, 4690, 4747
 - voice cloning, 4746
- dysarthric speech
 - enhancement, 4701
 - training data, 4703
- dyscalculia, 4789
- Dyscover, 4765, 4796
 - limitations, 4767
- dyslexia, 4690, 4789
 - AI intervention, 4765
 - handwriting analysis, 4796
- E (theorem prover), 2955
- E prover, 3020–3021
- e-commerce
 - federated personalisation, 4629
- E/I imbalance
 - network dynamics, 3894
- E4M3, 5108
- E5 embedding, 2386
- E5-Mistral, 2392
- E5M2, 5108
- early exit
 - definition, 4822
 - networks, 1955
 - neural networks, 4822
 - policy, 1956
- early ML methods
 - comparison, 4241
- early stopping
 - privacy, 1615
- earth mover distance, 80
- Earth Mover’s distance, 290
 - intuition, 291
- earth mover’s distance, *see* Wasserstein distance
- Earth system foundation models, 4373–4383
- Earth-2
 - NVIDIA AXA, 4389
- Earth-specific positional bias, 4256
 - definition, 4256
- Earth-Specific Transformer, 4254–4264
- Ebbinghaus forgetting curve, 1174–1175
 - definition, 1174
- ECC errors
 - predictive failure detection, 5152
- ECHO, 3321, 3393
- echocardiography
 - segmentation, 3512
- Eckart–Young theorem, 387
- ECMWF, 4181
 - AIFS, 4304, 4387
 - Ensemble Prediction System, 4307
 - ensemble system, 4291
- ECMWF ENS, 4196–4197
 - comparison, 4314
 - emulation, 4318
- ECMWF HRES
 - GraphCast comparison, 4270
- ECMWF IFS, 4218

- ecological validity, 4905
- economic impact
 - weather, 4182
- economic value
 - weather forecasting, 4391
 - exercise, 4393
- edge AI
 - accessibility, 4809–4820
 - definition, 4813
 - formal definition, 4813
 - affordable, 4901
 - federated generative models, 4621–4635
 - latency energy tradeoff, 4916
 - real-time generation, 4667
- Edge Attribution Patching, 3975
 - complexity, 3976
 - description, 3976
- edge computing, 4447
 - federated generative models
 - hardware-aware, 4667
- edge deployment, 3717
 - CVD, 4783
 - EEG seizure prediction, 3953
- edge inference latency
 - definition, 4812
- edge-cloud hybrid, 4820–4829
 - policy, 4908
- edit fidelity-creativity trade-off, 713
- education
 - federated curriculum generation, 4629
- educational data, 2688
- EEG
 - annotation, 3795
 - artefacts, 3781, 3797–3798
 - band power, 3794
 - benchmark datasets, 3800–3803
 - reproducibility, 3802
 - table, 3801
 - biometric identification, 3826
 - cross-channel dependencies, 3879
 - cross-dataset generalisation, 3802
 - data challenges, 3780
 - demographic underrepresentation, 3941
 - diffusion exercises, 3847
 - diffusion models, 3831–3840
 - dimensionality, 3797
 - federated datasets, 3802
 - forgery, 3933
 - frequency bands, 3793
 - table, 3793
 - fundamentals, 3790–3803
 - high dimensionality, 3796
 - history, 3803
 - inter-rater variability, 3795
 - invention, 3775
 - matrix representation, 3791
 - micro-states, 3795
 - missing documentation, 3900
 - montage, 3791
 - multimodal fusion, 3934
 - noise, 3797
 - non-stationarity, 3796
 - normalisation, 3811
 - phase-sensitive representations, 3799
 - recording principle, 3791
 - reference electrode, 3791
 - report generation
 - seq2seq, 3902
 - sampling rate, 3793
 - seizure states, 3794
 - spectral evolution, 3794
 - temporal resolution, 3849
 - TUH corpus, 3902
- EEG channel count
 - effect on synthesis quality, 3854
- EEG reconstruction

- compressed sensing, 3821
- EEG synthesis, 3815–3823
- EEG-DIF, 3836–3837
 - exercises, 3847
 - inpainting interpretation, 3837
- EEG-fMRI
 - cross-modal generator, 3954
 - nonlinear mapping, 3852
 - simultaneous recording, 3850
 - synthesis metrics, 3948
- EEG-to-fMRI synthesis, 3849–3856
 - evaluation, 3852
 - NeuroBOLT, 3851
- EF-Diffusion, 3851–3852
 - benchmark results, 3854
- effective batch size
 - definition, 5026
- effective dimensionality
 - latent video, 587
- effective parameters, 1865
- effective temporal dimension
 - exercise, 634
 - practical estimation, 632
- efficiency
 - in architecture adaptation, 1885
 - video generation, 524
- efficient NAS, 1926–1935
- EfficientNet, 1893
- EHR
 - generation pipeline, 2886
 - synthetic, 2885–2887
 - unstructured notes, 3900
- Eiffel Tower example, 4063
- ejection fraction, 3479, 3508
- El Niño–Southern Oscillation, 4182
- elaboration
 - Lean, 3081
 - Lean 4, 2992
- elastic depth, 1945
- elastic training, 5144–5155
 - definition, 5148
 - formal definition, 5149
- elastic transformer, 1945
- Elastic Weight Consolidation, 1187–1188
 - definition, 1187
 - example, 1188
 - Laplace approximation, 1187
- elastic weight consolidation, 3087
 - analogy, 2072
 - diffusion model adaptation, 3838
 - EEG, 3838
 - generative models, 3933
 - NAS perspective, 1886
- ELBO, 7, 175–183, *see* evidence lower bound, 4506
 - as training objective, 180
 - decomposition, 2765
 - decomposition exercise, 3829
 - decompositions, 177
 - definition for segmentation, 3562
 - diffusion model, 480
 - diffusion models, 4311
 - dominance over raw loss, 1621
 - EM algorithm, 154
 - exercise, 161, 200, 2773
 - exponential family, 181
 - gap, 567
 - Gaussian conjugate model, 178
 - Gaussian example, 178
 - Jensen derivation, 262
 - Jensen’s inequality derivation, 175
 - KL decomposition, 176, 262
 - membership inference, 1619–1621
 - membership inference algorithm, 1621
 - membership score, 1620
 - model comparison, 199
 - monitoring convergence, 194
 - numerical example, 180
 - Poisson model, 200

- score composition, 5093
- segmentation, 3562
- tabular VAE, 2922
- tightness, 180, 262
- VAE-EEG derivation, 3947
- video, 566
- video diffusion, 548
- with normalizing flows, 237
- Eldredge, Niles, 2033
- electrical impedance tomography, 3406
- electricity demand forecasting, 4233
- electrocardiogram
 - synthetic, 2697
- electrocorticography, 3780
- electrode selection
 - transformer, 3946
- electrodes
 - scalp, 3791
- electroencephalography (EEG), 3790–3803
 - definition, 3791
- electromagnetic inverse problem, 3851
- electron microscopy
 - segmentation, 3509
- electronic health record
 - seizure summary, 3914
- electronic health records, 2885–2887
 - LLM parsing, 3899–3903
- ElevenLabs, 4755
- Elhage, N., 3969
- elimination ordering, 136
- elliptic PDE, 3295–3303
 - inverse problems, 3299
 - parametric, 3296
- ELM, 2285
- EM algorithm, 138
 - as special case of VI, 195
 - BMA, 4229
 - comparison with CAVI, 195
 - exercise, 143
- GMM, 149
 - algorithm, 152
- hand computation, 160
- initialisation, 161
- M-step, 151
- monotonicity, 160
- EMA codebook update, 798
- embedding
 - as lossy channel, 2555–2565
 - capacity example, 2557
 - growing corpus conjecture, 2564
 - information loss, 2558–2565
 - irrelevant information, 2559
 - memorization, 2613
 - memory-bound, 4937
- embedding bottleneck theorem, 2473, 2556–2678
 - exercise, 2487
 - practical implications, 2474
- embedding capacity
 - challenge, 2638
- embedding channel capacity, 2473
- embedding dimension
 - exercise, 2634
- embedding distance
 - music, 938
- embedding models, 2385–2396
 - exercises, 2396
- embedding quality profile, 2600
- embedding space
 - geometry
 - exercise, 2636
 - universal lingua franca, 2482
- embedding spaces
 - multimodal, 2457–2467
- embedding training pipeline, 2391–2392
 - contrastive pre-training, 2391
 - distillation, 2391
 - fine-tuning, 2391
 - pre-training, 2391

- embedding-mediated generation, [2483](#)
- embedding-mediated translation, [2483](#)
- embodied conversational agent
 - definition, [4762](#)
- embodied conversational agents,
 - [4758–4769](#)
 - post-stroke, [4761](#)
 - rehabilitation pipeline, [4765](#)
- embodiment
 - evolutionary systems, [2276–2277](#)
 - LLM-guided evolution, [2277](#)
- embryo development
 - μ SAM tracking, [3648](#)
- emergence, [1273](#)
 - prediction, [4161](#)
- emergent capabilities
 - interpretability, [4161–4162](#)
- emergent cognition
 - video models, [4116](#)
- emergent reasoning
 - video models, [4115–4124](#)
- emerging synthesis, [3910](#)
- EMOS, [4228](#)
- emotion
 - model self-report, [4150](#)
- emotional retrieval, [2481](#)
- Emu Video, [730](#)
- ENAS, [1906](#), [1916–1919](#), [2254](#)
 - amortised training, [1917](#)
 - approximation bound, [1918](#)
 - convergence analysis, [1918](#)
 - parameter sharing insight, [1916](#)
 - ranking correlation, [1919](#)
 - ranking preservation, [1919](#)
 - results comparison, [1920](#)
 - supernet, [1916](#)
 - supernet illustration, [2009](#)
 - training algorithm, [1917](#)
 - training procedure, [1917](#)
- EnCodec, [791–793](#)
 - architecture, [792](#)
 - balancer intuition, [792](#)
 - bitrates, [793](#)
 - loss balancer, [792](#), [857](#)
- encoder, [386](#)
- encoder-decoder architecture
 - U-Net, [3532](#)
- encoder-processor-decoder
 - GraphCast, [4267](#)
- encoding
 - biological, [2343](#)
- end-to-end pipeline, [3703](#)
- end-to-end training
 - data assimilation, [4421](#)
 - weather, [4406](#)
 - challenges, [4374](#)
- end-to-end weather forecasting,
 - [4373–4383](#)
 - Aardvark, [4375](#)
- end-to-end weather prediction
 - definition, [4376](#)
- endoscopy
 - challenges, [3514](#)
 - segmentation, [3514](#)
 - synthetic data generation, [3582](#)
 - video segmentation, [3507](#)
- energy
 - weather dependence, [4182](#)
- energy conservation
 - primitive equations, [4207](#)
- energy consumption
 - edge AI, [4812](#)
 - federated training, [4653](#)
- energy efficiency
 - distributed training, [5161](#)
 - weather forecasting, [4279](#)
 - weather models, [4355](#)
- energy functional
 - active contours, [3526](#)
- energy norms, [3420](#)

- energy score
 - definition, 4335
- energy spectrum, 3421
 - damping by MSE training, 4300
- energy systems, 2801
- energy-based model, 133, 500, 501
- energy-based models
 - multiphysics, 3372
- energy-latency tradeoff, 4813
- Engel classification
 - surgical outcome, 3895
- Engram, 1092–1096
 - complexity, 1093
 - connection to sparse memory, 1209
 - hash collisions, 1111
 - scaling law, 1094
 - vs. MoE, 1114
- engram, 2342–2343
 - connection to retrieval, 2550
 - search for, 2343
- enhancing tumour, 3552
- ENIAC, 4184
 - weather prediction, 4210
- ensemble
 - bias-variance decomposition, 1453
 - variance inflation, 4359
- ensemble calibration
 - exercise, 4201
- ensemble distillation, 4584
- ensemble emulation, 4194–4202
- ensemble envelope
 - definition, 4318
 - formal definition, 4318
 - nonparametric coverage, 4319
- ensemble forecasting, 4219
 - diffusion, 4307–4329
 - operational NWP, 4307
- ensemble Kalman filter, 4216
- ensemble methods, 3229, 3412
 - limitations, 2879
 - synthetic data, 2879
- Ensemble Model Output Statistics
 - EMOS, 4229
- ensemble NWP
 - computational cost, 4296
 - SAPEMOS predictor, 4228
- ensemble post-processing, 4228
- ensemble prediction, 4185–4202
 - cost, 4193
 - generative models, 4193
- ensemble prediction system, 4285–4297
 - conceptual framework, 4290
 - definition, 4290
 - operational, 4287
 - trajectory divergence, 4431
- ensemble prediction systems, 4196–4197
- ensemble size
 - exercise, 4202
 - optimal, 4287
- ensemble spread
 - diffusion sampling, 4313
 - MOS predictor, 4228
 - underdispersion, 4228
- ensemble weather forecasting
 - MLWP, 4276
- ENSO, *see* El Niño–Southern Oscillation
- enterprise data, 2690
- entropic regularization, 371–373
 - definition, 372
 - exercise, 382
 - KL formulation, 373
 - limiting behaviour, 372
 - uniqueness, 372
- entropy, 67
 - ambiguous segmentation, 3500
 - conditional, 618
 - differential, 68
 - Shannon, 68
 - source coding, 68
- entropy collapse, 1150

- exercise, [1163](#)
- entropy decomposition
 - segmentation, [3492](#)
- entry threshold
 - MOS, [4224](#)
- Entscheidungsproblem, [3016](#)
- environmental complexity, [2243](#), [2272](#)
 - biological vs. artificial, [2243](#)
- environmental sound retrieval,
 - [2451–2453](#)
- Envision glasses, [4826](#)
- epigraph, [5254](#)
 - characterisation of convexity, [5255](#)
- EPILEPSIAE database, [3816](#)
- EPILEPSIAE dataset, [3800](#), [3816](#), [3920](#)
 - description, [3801](#)
- epilepsy, [3773](#)
 - benchmarks, [3918–3929](#)
 - care bottleneck
 - logistics, [3900](#)
 - care gaps, [3899](#)
 - care logistics, [3913](#)
 - classification, [3776](#)
 - clinical management
 - LLMs, [3899–3903](#), [3910](#)
 - clinical overview, [3774](#)
 - clinical translation, [3939](#)
 - cross-modal synthesis, [3849–3856](#)
 - data challenges, [3780](#)
 - datasets, [3918–3929](#)
 - definition, [3774](#)
 - diagnosis
 - VLM, [3903–3908](#)
 - economic burden, [3773](#)
 - GAN-based data augmentation,
 - [3815–3823](#)
 - generative AI synthesis, [3910–3918](#)
 - global prevalence, [3773](#)
 - history, [3775](#)
 - ILAE definition, [3774](#)
 - metabolic imaging, [3857–3860](#)
 - network disorder, [3778](#), [3863](#)
 - network geometry, [3912](#)
 - open problems, [3929–3939](#)
 - optimal transport, [3862–3868](#)
 - registry
 - autonomous, [3899](#)
 - seizure propagation, [3850](#)
 - transformer architectures,
 - [3879–3889](#)
 - types, [3777](#)
 - VAE, [3823–3829](#)
- epilepsy chapter
 - exercises, [3789](#)
 - roadmap, [3787](#)
- epilepsy pipeline
 - integrated, [3914](#)
- epilepsy surgery, [3773](#)
 - history, [3775](#)
 - pathway, [3778](#)
 - resection, [3780](#)
- epilepsy syndrome, [3777](#)
- epileptic connectome, [3863](#)
- epileptogenic zone, [3774](#)
 - probabilistic model, [3789](#)
- epipolar constraint, [682](#)
- episode length
 - theorem proving, [3116](#)
- episodic memory, [2341–2342](#)
 - biological basis, [2621](#)
 - definition, [2342](#)
 - in agents, [1281](#)
- episodic retrieval
 - embodied agents, [2621](#)
- epistasis, [2023](#)
- epistemic uncertainty, [2878](#), [3410](#)
 - inverse EEG, [3855](#)
 - segmentation, [3490](#)
- ϵ -parameterisation
 - audio, [850](#)

-
- Epstein, Edward
 - stochastic-dynamic prediction, [4285](#)
 - EPYC
 - server CPU, [4934](#)
 - equal-area grids, [4266](#)
 - equalised odds, [2856](#)
 - equalised odds gap, [2857](#)
 - equation of state
 - atmosphere, [4206](#)
 - equilibrium, [1035](#)
 - equitable threat score, [4195](#)
 - ERA-Interim
 - reanalysis, [4238](#)
 - training data, [4242](#)
 - ERA5, [4188](#), [4217](#), [4218](#)
 - benchmark evaluation, [4240](#)
 - smoothing, [4396](#)
 - smoothing bias, [4251](#), [4261](#)
 - training data, [4249](#), [4354](#)
 - ERA5 reanalysis
 - GraphCast, [4269](#)
 - erasure coding
 - checkpointing, [5151](#)
 - Erdos problems, [3161](#)–[3165](#)
 - Christmas 2025 breakthrough, [3161](#)
 - Problem 728, [3162](#)–[3163](#)
 - Erdos problems
 - AI solutions, [3135](#)
 - Erdos, Paul, [3161](#)
 - Erdos, Paul, [3039](#), [3135](#)
 - error accumulation
 - autoregressive chunking, [699](#)
 - error categorisation
 - theorem proving, [3091](#)
 - error estimation
 - a posteriori, [3438](#)
 - error feedback
 - theorem proving, [3090](#)
 - ESC-Net, [3695](#)–[3697](#)
 - architecture, [3695](#)
 - polyp detection example, [3696](#)
 - escalator problem, [4832](#)–[4898](#)
 - benchmark, [4839](#)
 - ESRGAN
 - brain MRI, [3589](#)
 - super-resolution, [3589](#)
 - essential tremor, [4690](#)
 - etcd
 - distributed coordination, [5149](#)
 - ethics
 - closed-loop stimulation, [3918](#)
 - federated generative models, [4647](#)–[4658](#)
 - generative AI
 - epilepsy, [3939](#)–[3943](#)
 - EU AI Act, [3964](#), [3968](#), [4443](#)–[4447](#), [4647](#)–[4649](#)
 - key provisions, [4446](#)
 - risk categories, [4446](#)
 - EU Medical Device Regulation, [3723](#)
 - EU of the Deaf
 - position on AI avatars, [4741](#)
 - EU-US Data Privacy Framework, [4449](#)
 - Euclid's theorem, [2976](#)
 - Euclidean geometry
 - AI, [3121](#)
 - Euclidean metric, [5194](#)
 - Euler method
 - discretisation error, [644](#)
 - Euler–Lagrange equation
 - active contours, [3526](#)
 - European Centre for Medium-Range Weather Forecasts, [4181](#)
 - European heatwave
 - GraphCast forecast, [4271](#)
 - European Union of the Deaf, [4741](#)
 - evaluation
 - benchmarking, [2854](#)–[2855](#)
 - challenges, [2834](#)
 - crisis, [2835](#)–[2837](#)

- cross-system, 3013
- data contamination, 3812
- evolution, 2711
- five axes, 2706
- open challenges, 2864–2865
- scalability, 2863
- standardized protocols, 2907
- synthetic data, 2704–2705
- tabular data, 2834–2847
- evaluation awareness, 4134
- evaluation bottleneck, 2129
- evaluation contamination, 2507
- evaluation design, 4918
- Evaluation Fidelity, 3965
- evaluation frameworks, 2849–2865
- evaluation gap
 - patient-specific vs cross-patient, 3921
- evaluation infrastructure
 - weather ML, 4243
- evaluation metrics, 19
 - ambiguous segmentation, 3748
 - clinical interpretation, 3810
 - epilepsy, 3923
 - event-based, 3811
 - exercise, 2921
 - human evaluation, 741
 - imbalanced EEG, 3809–3813
 - information retrieval, 2359–2364
 - limitations of L^2 , 3421
 - medical segmentation, 3544–3547
 - output structure dependence, 3505
 - PDE, 3416–3427
 - seizure prediction, 3811
 - summary, 741
 - user-centred, 4703
 - VBench, 737
 - video, 733–741
 - window-based, 3811
- evaluation paradigm
 - accessibility AI, 4909
 - evaluation problem, 2933
 - evaluation protocol
 - exercise, 3927
 - synthetic EEG, 3952
 - evaluation protocols
 - epilepsy, 3921
 - evaluation report, 2861–2862
 - evaluative paradigm shift, 4909
 - evaluator psychology, 4133
 - evasion attack
 - audio detection, 935
 - event-based training, 4368
 - evidence lower bound, *see* ELBO, 389–391
 - CIMD, 3608
 - definition, 389
 - EEG, 3823
 - exercise, 400, 401
 - segmentation, 3747
 - TabDDPM, 2765–2767
- evolution of cooperation, 2281
- evolution strategies, 2040–2044, 2280
 - definition, 2041
 - exercise, 2288
 - policy search, 2054
 - reinforcement learning, 2284
- Evolution Through Large Models, 2285
- evolutionary AI, 2019–2301
 - as unifying framework, 2255
 - comparative analysis, 2294
 - concept map, 2330
 - connections to other topics, 2300
 - consolidation, 2249–2256
 - exercises, 2287–2298
 - for retrieval, 2611
 - motivation, 2024–2027
 - open problems, 2267–2278
 - mind map, 2328
 - summary, 2298–2301

- timeline, [2279–2287](#)
- evolutionary algorithms
 - classical, [2037–2057](#)
 - early NAS, [1880](#)
- evolutionary alignment, [2272–2274](#)
- evolutionary arms race, [2159](#), [2220](#)
- evolutionary biology
 - foundations, [2027–2036](#)
- evolutionary branching
 - and progressive networks, [2250](#)
- evolutionary computation, [2019](#)
 - acceleration of progress, [2286](#)
 - computational limits, [2274–2275](#)
 - grand narrative, [2300](#)
 - historical patterns, [2287](#)
 - history, [2026](#), [2279–2287](#)
 - classical period, [2280](#)
 - foundations, [2279](#)
 - LLM era, [2285](#)
 - quality-diversity era, [2283](#)
 - timeline figure, [2329](#)
 - LLM-guided, [2095–2108](#)
 - new paradigm, [2131](#)
 - progressive liberation, [2329](#)
 - static schedules, [2171](#)
 - system comparison, [2179](#)
 - theoretical advances 2025, [2286](#)
- evolutionary curriculum design, [2229–2231](#)
 - algorithm, [2230](#)
- evolutionary design
 - philosophy, [2019–2027](#)
- evolutionary discovery
 - soundness, [2207](#)
 - verification, [2203–2215](#)
 - with formal verification, [2208](#)
- evolutionary exploration, [2224–2237](#)
 - synthesis, [2235](#)
- evolutionary programming, [2280](#)
 - LLM-guided, [3132](#)
- evolutionary sandboxing, [2260–2261](#)
- evolutionary search
 - for conjectures, [3179](#)
 - structured problems, [2248](#)
- evolutionary system
 - definition, [2020](#)
- evolutionary-generative hybrid, [1953–1955](#)
 - algorithm, [1953](#)
 - efficiency, [1954](#)
 - framework, [1953](#)
- evolvability
 - evolution of, [2243](#)
- EVOLVE-BLOCK markers, [2111](#)
- EvoX, [2286](#)
- EvoX framework, [2190–2203](#)
 - algorithm, [2191](#)
- EWC
 - diffusion models, [1220](#)
- excitation-inhibition ratio, [3892](#)
- execution trajectory, [2135](#)
- executive dysfunction
 - task decomposition, [4802](#)
- exercise
 - multi-step retrieval, [2537](#)
 - multimodal RAG, [2543](#)
 - RAG variants, [2531](#)
 - RETRO analysis, [2531](#)
 - speech RAG, [2544](#)
 - theorem proving with retrieval, [2536](#)
- exercises
 - alignment and interpretability, [4141](#)
 - benchmarks and reproducibility, [3927–3929](#)
 - class imbalance
 - focal loss, [3814](#)
 - imbalance ratio, [3813](#)
 - SMOTE limitations, [3813](#)
 - cross-modal information loss, [2565](#)

- cross-modal verification, [2574](#)
- distributed training, [5169–5182](#)
- EEG
 - frequency bands, [3813](#)
 - manifold estimation, [3814](#)
 - STFT design, [3815](#)
- epilepsy
 - generative AI, [3943–3957](#)
- evaluation metrics
 - AUROC vs. AUPRC, [3814](#)
- information theory of retrieval,
 - [2564](#)
- introspection, [4151](#)
- mathematical AI chapter, [3186–3188](#)
- medical image segmentation,
 - [3739–3762](#)
- memory-augmented retrieval, [2554](#)
- optimal transport
 - epilepsy, [3877](#)
- PDE generative models, [3450–3460](#)
- RAG error propagation, [2565](#)
- retrieval for verification, [2573](#)
- self-verification, [2573](#)
- transformer architectures for
 - epilepsy, [3896–3898](#)
- weather prediction chapter,
 - [4410–4441](#)
- existential
 - proof in Lean, [2969](#)
- existential quantification
 - as Sigma-type, [2985](#)
- existential risk, [2263–2267](#)
 - measured perspective, [2265](#)
- ExIt, *see* expert iteration
- expectation, [35](#)
- Expected Calibration Error, [4151](#)
- expected calibration error
 - segmentation, [3769](#)
- expected improvement, [2227](#)
- expected model change, [2225](#)
- experience replay, [1175](#), [1193–1194](#)
 - as retrieval, [2612](#)
 - coevolutionary analogue, [2072](#)
 - definition, [1193](#)
 - EEG, [3839](#)
 - retrieval formulation, [2612](#)
- expert
 - in MoE, [2816](#)
- expert collapse, [2818](#)
- expert dispatch, [4977](#)
- expert iteration, [2956](#), [3042–3043](#),
 - [3097–3099](#)
 - algorithm, [3098](#)
 - convergence analysis, [3110](#)
 - definition, [3042](#)
 - exploration phase, [3098](#)
 - training phase, [3098](#)
 - why it works, [3099](#)
- expert knowledge
 - constraints, [2824](#)
- expert parallelism, [5138](#)
 - formal definition, [5138](#)
- expert systems, [3962](#)
- explainability, [3961–3981](#)
 - classical methods, [3981–3990](#)
 - definition, [3965](#)
 - diffusion models
 - EEG, [3942–3943](#)
 - formal definition, [3964](#)
 - recurrent models, [4130](#)
- explaining away, [129](#)
- explicit density model, [6](#)
- explicit generative model, [52](#)
- exploration-exploitation trade-off, [2055](#)
 - unified view, [2231–2235](#)
- exponential distribution, [34](#)
 - sampling, [53](#)
- exponential family, [203–209](#)
 - definition, [181](#), [203](#)
 - ELBO, [181](#)

- log-partition function, 263
- exponential moving average, 659
- exposure bias, 443
- expression complexity
 - reward shaping, 3105
- expressiveness
 - architecture-dependent, 1984
- extended-range forecasting, 4274–4284
- external memory, 1090
- extra-gradient method, 317–319
 - algorithm, 317
 - computational cost, 319
 - convergence, 318
 - definition, 317
 - exercises, 324
 - idea, 317
- extraction attack
 - algorithm, 1711
- extreme events
 - AI model underprediction, 4261
 - GenCast, 4315
 - nowcasting warning, 4371
 - underprediction by deterministic MLWP, 4300
 - underrepresentation, 4234
- extreme precipitation
 - NowcastNet, 4366
 - tail statistics, 4356
 - Zhengzhou 2021, 4367
- extreme value theorem, 2936, 5206
- extreme weather
 - machine learning, 4394–4410
 - MOS failure, 4230
- extreme weather events, 4193–4202
 - tail distributions, 4194
- eye tracking
 - generative text prediction, 4865
- eye-gaze communication system
 - generative, 4865
- f-divergence
 - exercise, 286
 - variational representation, 277
- f -divergence, 76
- f-GAN, 277
 - objective, 278
- F1 score, 3809
 - definition, 3810
- face generation
 - federated DP-GAN, 4500
- facial action units, 745
- facial animation, 744
- FACS, *see* facial action units
- fact-checking
 - algorithm, 2569
 - retrieval-based, 2568–2574
- factor graph, 134
 - definition, 134
- factored attention, 551, 598–607
 - alternating blocks, 602
 - alternating transformer, 602
 - approximation bound, 601
 - complexity analysis, 756
 - computational cost, 601
 - definition, 600
 - exercise, 607
 - information propagation, 604
 - quality, 602
 - savings, 551
 - spatial-then-temporal, 600
 - speedup, 601
- factored dynamics, 536
- Factorised FNO, 3223
- factual association tuple, 4060
- factual circuits, 4069
- factual editing, 4058–4067
- factual knowledge
 - localisation, 4059
- factuality score, 4837
 - exercise, 4841

- FAD, *see* Frechet Audio Distance
- fair synthetic data, [2911](#)
- FairDARTS, [1926](#)
- FairNAS, [1906](#)
- fairness
 - auditing, [2690](#)
 - federated generation, [4650–4651](#)
 - generative models, [2785–2786](#)
 - synthetic data, [2856–2858](#)
 - synthetic data generation, [2911](#)
- fairness constraint
 - definition, [2785](#)
- fairness constraints, [2777](#)
- Faiss, [2379](#), [2382](#), [2384](#)
- faithfulness, [2532–2538](#)
 - attention weights, [3988](#)
 - bound, [2532](#)
 - classical XAI failure, [3999](#)
 - definition, [3997](#)
 - formal definition, [2532](#)
 - RAGAS metric, [2533](#)
 - vs. correctness, [2532](#)
- FALC (Facile a lire et a comprendre), [4793](#)
- false alarm
 - home monitoring, [3923](#)
- False Alarm Rate
- FAR
 - nowcasting, [4239](#)
- false alarms per hour, [3811](#)
- false positive rate
 - seizure prediction, [3882](#)
- false positive rate (FPR)
 - definition, [3810](#)
- false positive rate per hour, [3923](#)
- Fano's inequality, [1593](#), [1864](#)
 - membership inference, [1593](#)
- Fast Fourier Transform
 - in FNO, [3219](#)
- fast loop
 - propagation prediction, [3916](#)
- fast sampling
 - diffusion, [4313](#)
 - PDE solvers, [3387–3391](#)
 - PDEs, [3285](#)
- fat-tree
 - Clos network, [4945](#)
 - full bisection bandwidth, [4945](#)
 - scaling exercise, [4950](#)
- fat-tree network
 - bisection bandwidth, [5170](#)
- fat-tree topology, [4945](#)
- FateZero, [716](#)
- fault diagnosis
 - distributed training, [5152](#)
- fault tolerance, [1380–1390](#), [5144–5155](#)
 - Byzantine, [1383](#)
 - checkpoint frequency, [5177](#)
 - definition, [5145](#)
 - formal definition, [5145](#)
 - four-tier, [1336](#)
 - self-healing, [5163](#)
- FCD, *see* focal cortical dysplasia
- FCD detection
 - generative AI challenge, [3955](#)
- FCD localisation
 - data augmentation, [3845](#)
 - evaluation, [3842](#)
 - exercises, [3848](#)
 - performance table, [3843](#)
 - pseudo-healthy, [3841–3842](#)
- FCN, *see* fully convolutional network
- FDA
 - 510(k) pathway, [3722](#)
 - AI regulation, [3489](#)
 - AI regulatory pathway, [3939–3940](#)
 - De Novo pathway, [3722](#)
 - medical AI, [3722](#)
- FDG-PET
 - cerebral metabolism, [3857](#)

- epilepsy, 3780, 3857
- feature caching, 647–652
 - exercise, 651
 - speedup bound, 648
 - strategies, 649
- feature compression
 - split inference, 4823
- feature dictionary, 3972
- feature direction, 4040
- feature discovery, 4046–4058
- feature engineering
 - limitations, 3530
 - medical images, 3528
- feature fixing, 2777
- Feature Frequency Parser, 3601, 3754
- feature importance
 - preservation, 2856
- feature interactions
 - local, 2730
- feature map, 1514
- feature maps
 - diffusion U-Net, 3612
- feature matching loss, 785, 3574
- feature provenance
 - federated models, 4464
- feature skew, 4570
 - definition, 4570
- feature-value pairs, 2805
- FedAdam, 4471
- FedAvg, 4444–4475, 5117
 - algorithm, 4449
 - convergence, 4468
 - non-IID, 4574
 - convergence analysis
 - exercise, 4674
 - convergence rate, 4470
 - definition, 4468
 - exercise, 5133
 - LLM incompatibility, 4551
 - segmentation, 3757
- FedCF, 4510
 - movie example, 4514–4515
- FedDDPM, 4533–4535, 4574
 - algorithm, 4575
 - convergence and privacy
 - exercise, 4537
 - coverage improvement, 4535
 - exercises, 4579
 - FedDDPM+ algorithm, 4534
 - synthetic augmentation, 4576
 - research challenge, 4682
 - synthetic auxiliary dataset, 4533
- FedDDPM+, 4576
- FedDF, 4584
 - ablation, 4589
 - method, 4585
- FedDiffusion, 4459
- FedDifRC, 4528–4532
 - algorithm, 4530
 - convergence, 4532
 - hyperparameter sensitivity
 - exercise, 4537
 - noise-driven consistency, 4529
 - text-driven contrasting, 4529
- FedDT, 4478, 4582–4583
 - results, 4583
- FedDUL, 4599, 4604
 - forget direction, 4605
 - global filter, 4605
 - local update, 4605
- federated continual learning, 1224–1229
 - definition, 1226
- federated data augmentation, 4456
- federated diffusion model
 - art studio example, 4535
 - data extraction example, 4597
 - definition, 4527
 - text-driven contrasting, 4528–4532
 - trajectory mismatch, 4528
 - exercise, 4537

- federated discrete diffusion
 - definition, [4543](#)
- federated edge generation, [4626](#)
 - definition, [4626](#)
- federated fairness
 - definition, [4650](#)
- federated financial synthesis
 - five-bank example, [4633](#)
- federated fine-tuning
 - foundation models, [4451](#)
- federated flow matching, [4639–4643](#)
 - definition, [4640](#)
- federated GAN, [4483–4495](#)
- federated generative agents, [4630–4632](#)
- federated generative model
 - definition, [4458](#)
 - formal, [4458](#)
- federated generative models, [4443–4467](#)
 - chapter summary, [4684](#)
 - connections to prior chapters, [4463](#)
 - continual learning, [4662–4663](#)
 - core principles, [4685](#)
 - cross-modal, [4663–4664](#)
 - edge computing, [4667–4668](#)
 - emergence as field, [4451](#)
 - exercises, [4673–4685](#)
 - mathematical, [4673](#)
 - formal definition, [4458](#)
 - healthcare, [4609–4621](#)
 - long-term vision, [4685](#)
 - motivation, [4455](#)
 - open problems, [4659–4673](#)
 - privacy beyond DP, [4661–4662](#)
 - quality verification, [4665–4666](#)
 - research challenges, [4668, 4680](#)
 - scaling, [4659–4661](#)
 - scientific discovery, [4666–4667](#)
 - taxonomy, [4460](#)
 - taxonomy figure, [4462](#)
 - taxonomy table, [4461](#)
 - taxonomy tree figure, [4461](#)
- trilemma, [4668](#)
 - conjecture, [4668](#)
 - unique challenges, [4453](#)
 - unlearning, [4664–4665](#)
 - why, [4455](#)
- federated learning, [2882–2884, 4443–4467](#)
 - accessibility, [4703](#)
 - architecture, [2883, 4449](#)
 - architecture figure, [4452](#)
 - as regulatory compliance, [4447](#)
 - asynchronous diffusion, [5130](#)
 - brain tumour segmentation, [3720](#)
 - broken promise, [4590](#)
 - challenges, [2883](#)
 - clinical NLP, [4556](#)
 - clinical notes
 - example, [4546](#)
 - collaborative benefits, [4457](#)
 - communication cost, [3718](#)
 - comparison table, [5118](#)
 - convergence, [3737](#)
 - convergence thesis, [4672](#)
 - core insight, [4449](#)
 - definition, [3717, 5117](#)
 - diffusion models, [4526–4538](#)
 - domain shift, [3719](#)
 - EEG, [3802](#)
 - exercise, [2884](#)
 - centralisation trade-off, [4454](#)
 - formal definition, [5116](#)
 - fundamentals, [4467–4475](#)
 - GAN, [3593](#)
 - generative augmentation, [3720](#)
 - generative model, [2882](#)
 - gradient bias, [5123](#)
 - heterogeneous devices, [4579–4589](#)
 - heterogeneous latent spaces,
 - [4508–4510](#)

- history, 4450–4451
 - FedAvg, 4450
- key idea, 5122
- key insight, 4450
- large language models, 4549–4558
- LLM optimization, 4558–4568
- LLM scale challenge, 4550
- medical imaging, 5122
- medical segmentation, 3717–3721
- non-IID, 4568–4579
- paradigm, 5115
- personalisation, 4704
- personalised
 - diffusion, 4532–4533
- privacy defenses, 4599–4609
- privacy promise, 4590
- privacy vulnerabilities, 4589–4599
- regulatory objective, 4455
- segmentation, 3757
- structure, 5116
- synthetic auxiliary data, 4533–4535
- system architecture, 4451
- text generation, 4538–4549
- unlocking inaccessible data, 4457
- variational autoencoder, 4506–4517
- vs distributed training, 5115–5124
- when to use, 5118
- federated LLM
 - optimization comparison, 4566
- federated LoRA, 4460
- federated medical image synthesis, 4612
 - definition, 4612
- federated personalisation
 - LLM pipeline, 4627
- federated privacy
 - taxonomy, 4607–4608
- federated prompt tuning, 4555
 - definition, 4555
 - exercise, 4558
- federated synthetic data, 2882–2884
- federated tabular synthesis
 - definition, 4492
- federated transfer learning, 4451
- federated unlearning, 4599–4609
 - definition, 4603
 - difficulty, 4604
 - efficiency, 4606–4607
 - efficiency proposition, 4606
 - proof sketch, 4606
 - verification exercise, 4609
- federated vs distributed
 - decision, 5119
- FedFlow, 4460
- FedGAN, 4458–4459
 - definition, 4485
 - exercise, 4494
 - MNIST
 - implementation challenge, 4681
 - mode collapse
 - exercise, 4675
- FedKSeed, 4479, 4561
 - definition, 4562
 - exercise, 4567
- FedLLM, 4451–4467
- FedLoRA
 - training pipeline
 - research challenge, 4683
- FedMD, 4584
 - algorithm, 4584
 - method, 4584
- FedMeZO, 4559
 - definition, 4560
 - memory, 4561
 - memory analysis, 4561
- FedP2EFT, 4565
 - definition, 4566
 - exercise, 4568
- FedProx, 4471
 - cardiac CT, 4615
- FedRDMA, 4479, 4565

- FedVAE, [4459](#), [4507–4516](#)
 - definition, [4508](#)
 - dual encoder, [4511](#)
 - ELBO decomposition exercise, [4465](#)
 - gating network, [4511–4512](#)
- feedback loop
 - adaptive learning, [4794](#)
- feedback mechanisms
 - accessibility AI, [4703](#)
- feedback quality
 - APOLLO, [3091](#)
- feedforward network
 - interpretability, [4016](#)
- feedforward neural network, [91](#)
- Feit–Thompson theorem, [3025](#)
- Feit–Thompson theorem, [2944](#)
- Fenchel conjugate, [277](#), [5257](#)
 - definition, [277](#)
- Fenchel inequality, [5258](#)
- Fenchel–Young equality, [5258](#)
- Feng26 paper, [3174](#)
- Fenghe
 - meteorological LLM, [4388](#)
- FengWu, [4274–4284](#)
 - architecture, [4277](#)
 - energy consumption, [4279](#)
 - multi-modal architecture, [4277](#)
 - operational model, [4388](#)
- Fenton reaction, [3369](#)
- Fermat’s Last Theorem, [3039](#)
 - formalisation, [2950](#)
 - Wiles’ gap, [2941](#)
- FermiNet, [3366](#)
- fermions
 - antisymmetric wavefunctions, [3455](#)
- FERPA, [2688](#)
- few-shot fine-tuning
 - privacy, [1793](#)
- few-shot generation
 - tabular, [2910](#)
- few-shot learning
 - brain twin, [3932](#)
 - example, [2812](#)
 - PDE operators, [3256](#)
 - statistical inference, [2813](#)
 - tabular, [2812](#)
- few-shot personalisation
 - epilepsy, [3931](#)
- few-shot segmentation
 - GenSeg, [3710](#)
 - medical imaging, [3493](#)
- Feynman–Kac formula, [3439](#)
- FFA-LoRA, [4554](#)
 - exercise, [4557](#)
- FFJORD, [462](#)
 - algorithm, [463](#)
- FFM-GOT, [4641–4642](#)
- FFM-LOT, [4640–4641](#)
- FFN memory bank, [1101](#)
 - efficiency, [1102](#)
- FFN parallelism, [5045](#)
- FGN training loss, [4334](#)
- FID, [20](#), [62](#), [283](#), *see* Fréchet Inception
 - Distance
 - blood cell images, [3588](#)
 - cross-modal EEG, [3948](#)
 - differential privacy, [1674](#)
 - DP-SGD
 - EEG, [3941](#)
 - exercise, [286](#)
 - federated generative model quality, [4458](#)
 - local SGD degradation, [5092](#)
 - privacy cost, [1634](#)
 - privacy tradeoff, [1632](#)
- fidelity, [2704](#)
 - weight sharing, [1905](#)
- Fields Medal
 - IMO alumni, [3112](#)
- FiLM conditioning, [818–819](#)

- affine interpretation, 819
- comparison with cross-attention, 819
- definition, 818
- diffusion step, 4312
- exercise, 831
- filter
 - in Lean, 2977
- filter paradigm, 3038
- Fin type, 2985
- finance
 - synthetic data, 2887–2889
- financial data
 - synthetic, 2688, 2800
- financial time-series
 - federated synthesis, 4633–4634
- fine-tuning, 1123–1167
 - for tabular generation, 2807
 - theorem proving, 3091
- finite differences, 4213
 - Hessian approximation, 1922
 - NWP, 4213
- finite element method
 - cost comparison, 3458
 - Galerkin, 3238
- finite group, 2996
- FIRST, 2495–2496
 - key insight, 2495
- First Proof benchmark, 3157–3161
 - problem creators, 3157
 - results, 3158
 - solution analysis, 3159
- first wave
 - ML meteorology, 4233
- first-frame conditioning, 581–582
 - consistency, 582
 - definition, 581
- first-order logic
 - automated proving, 3020–3021
 - decision problem, 3016
 - expressiveness, 3020
 - in proof assistants, 3003
- Fisher divergence, 85, 506
 - definition, 86
- Fisher information
 - as Riemannian metric, 249
 - forgetting bound, 1182
 - member separation, 1639
 - membership inference, 1639
 - per-sample, 1639
- Fisher information matrix, 216, 249
 - definition, 1187
- Fisher vector, 2403–2406
 - definition, 2403
 - fine-grained recognition, 2404
- Fisher’s fundamental theorem, 2029
- Fisher, R. A., 2028
- Fisher, Ronald, 2279
- Fisher–KPP equation, 3305
- FISTA, 5276
- fitness
 - definition, 2028
 - relational, 2058
 - relative, 2059
- fitness landscape, 2022–2024
 - definition, 2022
 - discovery barrier, 2222
 - exercise, 2292
 - rugged
 - lower bound, 2274
- fitness sharing
 - competitive, 2077
 - exercise, 2288
- Fitzpatrick scale, 3725
- fixed architecture, 1890
- fixed environment
 - complexity bound, 2245
- fixed output vocabulary
 - limitation, 4699
- fixed point, 1031–1035

- Banach theorem, 1032
- definition, 1031
- example, 1033
- fixed-depth computation, 3047
- fixed-point iteration
 - exercise, 1066
- FLAIR
 - synthetic generation, 3844
- FLAME face model, 4736–4738
- Flash Attention, 109
 - exercise, 121
- FlashAttention, 553
 - AlphaEvolve optimisation, 2114
 - roofline position, 4938
 - SRAM tiling, 4935
- flat minima, 2034
- flat organisation, 1275
- flattened pattern, 951
- Fleiss' kappa, 3550
- Flesch-Kincaid
 - readability, 4914
- Flexiffusion, 1940–1941
 - acceleration bound, 1941
 - architecture policy, 1940
 - timestep-dependent architecture, 1940
 - zero-cost proxy, 1940
- FlexLoRA, 4554, 4585–4588
 - aggregation, 4586
 - algorithm, 4587
 - communication, 4587
 - definition, 4586
 - design exercise, 4589
 - drift reduction, 4588
- Flickr30k benchmark, 2417
- floating point
 - FP16 training, 5060
 - FP16 vs BF16, 5073
- floating point formats, 1527
- floating-point
 - non-determinism, 5023
- floating-point verification, 3004
- FLOP
 - 7B model calculation, 4940
 - silent corruption, 5152
 - training estimate, 4926
- flow conservation, 2867
- flow function, 2867
- flow matching, 463–465, 3384
 - advantages, 4341
 - audio, 864–873
 - audio generation, 941
 - audio latents, 952
 - audio objective, 867
 - basics, 3286
 - conditional, 464
 - connection to diffusion, 465
 - continuity equation, 4348
 - definition, 635
 - DiT, 4341–4343
 - ensemble, 4324
 - equivalence to diffusion, 865
 - Euler sampling, 642
 - exercise, 644, 896, 3395, 4416
 - federated, 4636–4643
 - Gaussian-preserving interpolation, 867
 - higher-order solvers, 642
 - latent geometry, 866
 - latent space, 3289
 - normalizing flow connection, 750
 - objective, 463, 3286, 3385
 - optimal transport interpretation, 747
 - OT path, 464, 636
 - path design, 757
 - PDE solvers, 3384–3396
 - PDEs, 3285–3295
 - quantum systems, 3369
 - relationship to diffusion, 635, 865

- sample efficiency, 870
- sampling algorithm, 642
- theory, 4340–4341
- time convention, 637
- timestep weighting, 641
- training algorithm, 640, 641
- truncation, 3288
- versus noise prediction, 867
- video, 635–644
- video objective, 636
 - definition, 636
- vs. diffusion, 4347
- vs. score matching, 3287
- weather, 4339–4349
- weather forecasting, 4324
- flow matching condition, 2867
- flow reconstruction, 3284–3295
 - architecture, 3285
 - problem setup, 3284
 - results, 3285
- FlowDA, 4344–4345
 - definition, 4344
 - observation conditioning, 4345
 - vs. Appa, 4345
- fluid dynamics
 - guidance, 720
- fluid–structure interaction, 3371
- fluid-structure interaction, 3328
- fluorescence microscopy, 3509
- Flyspeck project, 2950, 3001, 3025
- flywheel effect
 - AI for mathematics, 3141
- FMA, 22
- fMRI
 - spatial resolution, 3849
- FNO, *see* Fourier Neural Operator
- focal cortical dysplasia, 3776, 3892
 - anomaly detection, 3859
 - dataset, 3920
 - detection, 3912
- diffusion models, 3840–3847
 - MRI, 3781
 - MRI-negative, 3857
 - MRI-negative case, 3782
 - pseudo-healthy synthesis, 3955
 - Type IIb, 3778
- focal epilepsy
 - dataset overrepresentation, 3942
- focal loss, 3806
 - derivation, 3814
- focal seizure, 3776
- Fogel, Lawrence, 2280
- Fokker–Planck equation, 3267
- Fokker-Planck equation
 - brain, 3872
- FOL, *see* first-order logic
- folic acid
 - deficiency documentation, 3900
- Folk theorem
 - exercise, 1452
- folk theorem, 1357
- forecast operator, 4188–4192
 - definition, 4188
 - probabilistic, 4189
- forecast skill, 4190–4191
- forecast uncertainty
 - sources, 4290
- foreign keys
 - synthetic data, 2698
- forgetting
 - architecture change, 1883
 - continual learning, 4073
 - cross-modal, 1228
 - diffusion models, 1219
 - human, 1174–1175
 - mechanistic basis, 4072
 - selective, 1884
 - two components, 1884
- forgetting bound
 - HSD, 1233

- forgetting budget
 - EEG, 3932
- forgetting matrix, 1184
- forgetting measure, 1178
- forgetting-adaptation tradeoff, 1881–1891
- fork (graphical model), 129
- form labels
 - accessibility, 4692
- formal verification
 - level 5, 2213
- Formal Abstracts, 3032
- formal language
 - for mathematics, 2942
- formal libraries, 3011
- formal mathematics, 3064
 - benchmarks, 3061
- formal proof, 2206
 - definition, 2943
- formal proof generation
 - dedicated models, 3068
- formal proofs
 - from natural language, 3064–3080
 - informalization, 3070
 - requirement, 3033
- formal verification, 2205–2209, 2270, 2942–2953
 - history, 2944–2946
 - mathematical, 2566
 - neural circuits, 4158–4159
 - theorem proving, 3080
 - timeline, 2946
- formalisation
 - scale, 3025
- formalization
 - as training signal, 2940
 - completeness in principle, 2949
 - current state, 2950
 - definition, 2942–2944
 - error sources, 3118
 - informal to formal, 3069
 - scope, 2949
 - worked example, 2992–2997
- formalization bottleneck, 3141
- formalization gap, 2997
- formalization of mathematics, 2942–2953
- FormulaNet, 3044
- fornix
 - seizure propagation, 3891
- forward diffusion
 - marginal distribution, 4309
 - MRI noise injection, 3841
- forward diffusion process
 - audio, 848
- forward KL, 170
- forward model, 327
 - examples, 328
- forward process, 476, 479
 - closed form, 2760
 - exercise, 2773
 - Gaussian, 2760
 - information retention, 1608
 - video, 540
 - theorem, 540
- forward SDE, 2770
 - definition, 4339
- forward selection, 4223
- forward transfer, 1178
- foundation model
 - weather
 - timeline, 4241
- foundation models
 - annotation, 3521
 - architecture, 2909
 - architecture search, 1934
 - as segmentation backbone, 3711
 - Earth system, 4377
 - EU AI Act, 4447
 - federated, 4672

- for tabular data, 2804
- integration with generative models, 3694–3715
- medical SAM, 3627–3641
- medical segmentation, 3493
 - landscape, 3641
- MedSAM, 3631–3632
- microscopy, 3764
- no universal dominance, 3729
- PDEs, 3253, 3434–3436, 3472
- SAM, 3618–3627
- SAM timeline, 3627
- scaling laws, 3730
- SegVol, 3639–3640
- self-optimization, 1967
- sleep, 1873
- tabular data, 2908–2912
- time series, 2801
- universal medical, 3729
- wave equations, 3351
- foundational differences, 3009
- foundational embeddings
 - accessibility, 2476
 - applications, 2474–2487
 - as computational primitive, 2475
 - content moderation, 2477
 - creative applications, 2475
 - e-commerce, 2478
 - medical applications, 2479
 - open problems, 2484
 - scientific discovery, 2478
 - visionary applications, 2480
- foundations
 - of theorem provers, 3011
- foundations of mathematics, 2945
- Four Color Theorem
 - formalization, 3000
- four-colour theorem, 2944, 3029
 - Kempe’s error, 2941
- FourCastNet, 4195, 4244–4254
 - architecture, 4248
 - Fourier layer complexity, 4413
 - performance benchmarks, 4251
 - training objective, 4249
 - version 2, 4251
 - wind forecast example, 4250
- Fourier Neural Operator, 15, 3218–3223, 3417
 - 1D Burgers, 3461
 - approximation theory, 3298
 - benchmarks, 3222
 - Burgers equation, 3316
 - Burgers implementation, 3319
 - complexity, 3221, 4413
 - computational complexity, 3222
 - connection to DeepONet, 3232
 - Darcy flow, 3297
 - Fourier layer, 3219
 - Fourier-space kernel, 3218
 - full architecture, 3220
 - heat equation, 3308
 - implementation, 3302
 - kernel properties, 3452
 - key insight, 3218
 - layer definition, 3219, 4246
 - lifting layer, 3220
 - mode selection, 3231
 - PDEBench benchmark, 3468
 - periodic domains, 3224
 - projection layer, 3220
 - resolution invariance, 3298
 - super-resolution, 3221, 3456
 - uncertainty quantification, 3415
 - universal approximation, 3224
 - variants, 3223
- Fourier-type attention, 3239
- fourth-order PDEs
 - PINN difficulties, 3360
- FP16
 - gradient overflow, 5038

- overflow, [5106](#)
- underflow problem, [5106](#)
- FP16 all-reduce
 - bandwidth analysis, [5175](#)
- FP32
 - gradient accumulation, [5037](#)
- FP4
 - inference, [4935](#)
- FP4 training, [5158](#)
- FP8
 - H100 training, [4935](#)
 - training, [5108](#)
- FP8 training, [5075–5076](#), [5158](#)
- Fréchet Audio Distance
 - Wasserstein interpretation, [941](#)
 - Wasserstein proof, [953](#)
- Fréchet distance, [20](#)
- Fréchet Inception Distance, *see* FID, [735](#)
 - as W_2 , [306](#)
- Fréchet Video Distance, [734–737](#)
 - computation pipeline, [735](#)
 - definition, [734](#)
 - limitations, [736](#)
 - practical considerations, [736](#)
 - proof exercise, [758](#)
 - Wasserstein interpretation, [735](#)
- Fréchet-Hoeffding theorem, [543](#)
- fractional anisotropy, [3894](#)
- Fractions Skill Score, [4365](#)
- fractions skill score, [4195](#)
- frame interpolation, [533](#)
- frame sampling
 - MLLMs, [4832](#)
- frame-level annotation gap, [3514](#)
- fraud detection, [2688](#)
 - class imbalance, [2722](#)
 - federated, [4462](#)
 - pipeline, [2888](#)
 - results, [2888](#)
 - synthetic data, [2887–2889](#)
- Frechet Audio Distance, [915–917](#)
 - definition, [915](#)
 - exercise, [927](#)
 - sample size, [917](#)
 - Wasserstein interpretation, [916](#)
- Fredholm integral operator, [3236](#)
- free bits, [397](#)
- free boundary problems
 - Stefan condition, [3454](#)
- free energy
 - Ginzburg–Landau, [3357](#)
 - variational, [181](#)
- FreeInit, [709](#)
- FreeNoise, [700](#)
- Frege, Gottlob, [2942](#), [3016](#)
- frequency bands
 - EEG, [3885](#)
- frequency chirp
 - seizure, [3794](#)
- frequency-domain attacks, [1649–1659](#)
 - challenge problem, [1659](#)
 - combined approach, [1656](#)
 - data augmentation robustness, [1657](#)
 - exercises, [1658](#)
 - fingerprint, [1653](#)
 - high-frequency, [1650](#)
 - Liu et al., [1652](#)
 - orthogonality, [1657](#)
 - threshold selection, [1652](#)
- From Zero to Turbulence, [3337](#)
- frontal analysis
 - statistical, [4222](#)
- frontal lobe epilepsy, [3777](#)
- frontal lobe FCD, [3860](#)
- frontier map
 - open problems, [4408](#)
- Frontier supercomputer, [4946](#)
- Frontier-CS benchmark, [2171](#)
- FrontierMath, [3060–3061](#), [3148](#)
- Froude number, [3320](#)

- frozen encoders
 - pipeline bubble filling, 4996
- Fréchet Inception Distance, 283
- Fréchet Inception Distance
 - EEG, 3818
- FSDP, 5065–5069
 - all-gather
 - shard-gather cycle, 4996
 - all-gather prefetch, 4988, 4996–4999
 - flat parameters, 5065
 - formal definition, 5066
 - hybrid sharding exercise, 5069
 - prefetch analysis exercise, 4999
 - prefetching analysis, 5180
 - sharding lifecycle, 5066
 - sharding strategies, 5066
- FSS, *see* fractions skill score, *see*
 - Fractions Skill Score
 - scaling properties, 4373
- full configuration interaction, 3366
- full waveform inversion, 3404
- fully automatic segmentation, 3518
- fully connected network
 - limitations for tabular, 2746
 - spatial information, 4237
- fully convolutional network (FCN), 3532
 - key insight, 3532
- Fully Sharded Data Parallel, 5065–5069
 - communication hiding, 4996–4999
- function autoencoder, 3382
- function spaces
 - diffusion models, 3267
- function type, 2947
- function-preserving expansion, 1869
- function-preserving transformations, 1957
 - definition, 1957
 - Net2Net, 1957
 - zero-forgetting, 1958
- function-space diffusion, 3377–3384
- forward SDE, 3379
 - implementation, 3396
- functional verification, 2213
- functional dependencies, 2700, 2825
 - constraints, 2703
- functional equations, 2959, 3034
- functional generative network, 4329–4337
 - 32-dimensional noise, 4332
 - coherence mechanism, 4333
 - definition, 4331
 - formal definition, 4332
 - latent noise, 4332
 - modulation, 4332
 - spatial scales, 4338
 - training, 4333–4334
- fundamental limits
 - on evolution, 2264
- fundamental theorem of algebra, 2945
- fundamental theorem of arithmetic, 2935
- fundamental theorem of natural selection, 2028–2030
- FunDiff, 3382
- FunDPS, 3381
- fundus photography
 - segmentation, 3481
- FunSearch, 15, 2285, 3175
 - exercise, 2291
- Furhat
 - robotic platform, 4762
- Fusion-in-Decoder, 2526–2527
 - attention complexity, 2527
 - computational cost, 2527
- FuXi, 4274–4284
 - cascaded architecture, 4274
 - end-to-end DA, 4380
 - forecast example, 4282
 - model handoffs, 4276
 - training, 4276

- FuXi-ENS, [4276](#)
- fuzzy logic
 - temperature forecasting, [4234](#)
- fuzzy sets
 - temperature, [4234](#)
- fuzzy time series
 - meteorology, [4234](#)
- fuzzy transition matrix, [4235](#)
- FVD, *see* Fréchet Video Distance
 - Wasserstein distance, [747](#)
- FVSBN, [433](#)
 - definition, [433](#)
- Gödel
 - implications for formalization, [2950–2951](#)
- Gödel’s incompleteness theorems, [2264](#)
- Gödel, Kurt, [2939](#)
- GABA-ergic deficiency, [3892](#)
- Gabor filters, [3529](#)
- Gabor limit, [774](#)
- GAE, *see* Generalised Advantage Estimation
- Galerkin attention, [3239](#), [3301](#)
 - projection property, [3239](#)
- Galerkin method
 - classical, [3238](#)
 - neural attention, [3238](#)
- Galerkin projection, [3238](#)
- Galerkin Transformer, [3238](#)
 - architecture, [3240](#)
- Galerkin-type attention, [3239](#)
- Gallina, [3000](#)
- game Jacobian, [315](#)
 - decomposition, [316](#)
 - definition, [315](#)
- game theory, [310](#), [1354–1363](#)
 - and coevolution, [2065–2067](#)
- game-playing
 - coevolution, [2165](#)
- GameGen-X, [723](#)
- GamePad, [3035–3036](#)
 - paradigm, [3038](#)
- gamma waves, [3793](#)
- GAN, [4](#), [8](#), *see* generative adversarial networks
 - adversarial training, [3816](#)
 - architecture search, [1895](#)
 - as coevolution, [2251](#)
 - audio generation, [942](#)
 - coevolutionary exercise, [2256](#)
 - conditional, [344–346](#)
 - CycleGAN, [348](#)
 - discriminator averaging problem, [4459](#)
 - EEG synthesis, [3785](#)
 - EEG variants comparison, [3823](#)
 - epilepsy applications, [3785](#)
 - federated, [4458–4459](#), [4483–4495](#)
 - fundamental tension, [3816](#)
 - GigaGAN, [358](#)
 - gradient leakage exercise, [4455](#)
 - latent space interpretability, [3994](#)
 - latent space structure, [3994](#)
 - limitations, [2758](#)
 - medical imaging, [16](#)
 - membership inference, [4501](#), [4596](#)
 - memorisation, [4590](#)
 - mode collapse, [3820](#)
 - MRI-to-PET exercise, [3861](#)
 - MRI-to-PET translation, [3857](#)
 - pathologies, [2252](#)
 - pix2pix, [346](#)
 - precipitation downscaling, [4355](#)
 - progressive growing, [351–352](#)
 - SPADE, [355](#)
 - StyleGAN, [352–355](#)
 - training dynamics
 - EEG, [3943](#)
 - video autoencoder, [748](#)

- vision applications, 343–359
- GAN training
 - as alternating GDA, 314
 - optimisation strategies, 325
 - practical methods, 320
- GAN vs diffusion, 3590–3592
 - comparison table, 3676
 - exercise, 3595
 - quantitative comparison, 3592
- GANs
 - vs. other families, 2709
 - weather, 4198
- GAOT, 3329
 - geometry-conditioned attention, 3329
- gate mechanism
 - ConvLSTM, 4239
- gated activation, 441
- gated attention, 111
 - expressiveness, 1551
- Gated Linear Unit, 113
- gated linear unit
 - gradient flow, 1551
- gated sparse autoencoder, 4050–4051
 - definition, 4050
- gating, 1481–1488
 - gated attention, 1485
 - expressiveness, 1486
 - general mechanism, 1482
 - GLU, 1484
 - definition, 1484
 - GRU, 1483
 - highway networks, 1483
- gating network, 2816
 - personalised, 4511–4512
- Gaudi 3
 - Intel, 4939
- GauGAN, 355–357
- Gauss–Newton, 2779
- Gaussian diffusion, 2758
 - for continuous columns, 2759–2760
- Gaussian distribution, 33
 - exponential family form, 204
 - MLE, 56
 - multivariate, 38
- Gaussian filter, 330
- Gaussian grid, 4214
- Gaussian mechanism, 1598
- Gaussian mixture model, 145
 - agent interpretation, 1461
 - covariance types, 148
 - definition, 146
 - Fisher vector, 2403
 - generative, 156
 - sampling, 161
 - singularity, 149, 161
- Gaussian noise
 - DP-SGD, 2737
- Gaussian process, 2226
- Gaussian random field
 - score function, 3453
- Gaussian RBF kernel, 2840
- gaze input
 - text generation, 4694
- GB200 NVL72, 4935
- Gboard
 - federated language model, 4626–4627
 - federated learning, 4450
- GCD
 - in Lean, 2976
- GDAS, 1926
- GDPR, 2885, 3717, 3968, 4443–4447
 - EEG, 3940
 - enforcement, 4446
 - federated generative AI, 4647–4649
 - key provisions, 4445
 - machine learning compliance
 - exercise, 4455
 - right to be forgotten, 1797, 1831

- synthetic data, [2911](#)
- Gecko, [2393](#)
- GED
 - exercise, [3928](#)
 - graph edit distance, [3924](#)
- GEFS, [4196–4197](#)
- GEM, [1196](#)
- GeM pooling, [2406](#)
 - generalisation, [2407](#)
- Gemini
 - fault tolerance, [5145](#)
- Gemini (language model)
 - AlphaEvolve, [3132](#)
 - fine-tuning for formalization, [3117](#)
 - formalization, [3113](#)
 - foundation for math AI, [3140](#)
- Gemini Embedding, [2390–2391](#), [2461–2467](#)
 - Matryoshka representations, [2461](#)
 - task-type conditioning, [2462](#)
- Gemini Flash, [2109](#)
- Gemini Pro, [2109](#)
- Gemini Ultra
 - training compute, [4925](#)
- GEMM
 - throughput, [4934](#)
- GenAI Explainability Triangle, [3965](#)
- GenASL, [4728–4729](#)
- GenCast, [3](#), [13](#), [3282](#), [3440](#), [4193](#), [4307–4329](#)
 - architecture, [4311](#)
 - diffusion exercise, [4415](#)
 - ERA5 training, [4308](#)
 - formal definition, [4310](#)
 - verification, [4314](#)
- GenCFD, [3338](#)
- GenEEG, [3838–3839](#)
 - dual conditioning, [3838](#)
- GenEPS, [4321–4644](#)
 - verification, [4322](#)
- General Neural Operator Transformer,
 - see* GNOT
- generalisation
 - climate zones, [4357](#)
- generalisation depth, [724](#)
- generalisation gap, [1591](#), [3726](#)
 - membership advantage, [1614](#)
 - quantification, [3726](#)
- Generalised Advantage Estimation, [3102](#)
- generalised energy distance, [3496](#)
- generalised mean pooling, *see* GeM
 - pooling
- generalised seizure, [3776](#)
- generalist strategy
 - robustness, [2161](#)
- generality, [4064](#)
- generalization
 - generative models, [1982](#)
- generalization bounds
 - continual learning, [1181–1182](#)
 - VC dimension, [1182](#)
- generalized belief propagation, [254](#)
- Generalized Energy Distance, [3566](#)
 - base distance, [3567](#)
 - computation exercise, [3570](#)
 - definition, [3566](#)
 - proper scoring rule, [3566](#)
 - sample size, [3567](#)
- generalized energy distance, [3747](#)
 - properties, [3748](#)
- generate-then-retrieve, [2529](#)
- generation
 - complementarity with verification, [3038](#)
- generation proximity
 - membership test, [1661](#)
- generation proximity statistic, [1660](#)
- generation window, [697](#)
- generation-retrieval duality, [2622](#)

- generation-retrieval spectrum, 2622
 - examples, 2622
- generation-verification asymmetry, 2204
- generative accessibility model
 - definition, 4700
 - formal definition, 4700
 - training exercise, 4706
- generative adversarial network, 3, 273
 - algorithm, 279
 - Daltonization, 4773
 - data requirements, 4443
 - discriminator features, 2609
 - EEG synthesis, 3815–3823
 - evaluation, 283
 - objective, 274
 - timeline, 284
 - training, 279
 - training dynamics, 286
 - training instability, 282
 - vs VAE, 287
- generative adversarial networks, 3572
 - and coevolution, 2061
 - conditional objective, 3580
 - disengagement, 2070
 - fundamentals, 3572
 - game-theoretic view, 2066
 - medical imaging timeline, 3592
 - medical segmentation, 3572–3580
 - Red Queen dynamics, 2031
 - training dynamics, 3744
 - turbulence, 3336
- generative AI
 - accessibility, 4706–4716
 - multimodal, 4875
 - disability solutions taxonomy, 4694
 - individuality, 4703
 - interface adaptation, 4866
 - scale requirements, 4925
 - structural change for accessibility, 4695
 - synthesis vs. classification, 4693
- generative amplification, 3736
- generative architecture DAG, 1896
- generative audio
 - legal considerations, 937–940
- generative augmentation, 3807–3813, 4457
 - augmentation ratio, 3808
 - comparison table, 3808
 - figure, 3809
 - validation strategy, 3812
 - vs. SMOTE, 3808
- generative Daltonization, 4775
 - formal definition, 4775
- generative data engine, 3711
- generative design
 - prosthetics, 4869
 - limitations, 4870
- generative downscaling
 - exercise, 4416
- generative flow network, 2865–2876
- generative latent neural PDE solver, 3386
- Generative Mechanism Transparency, 3965
- Generative Medical Segmentation, *see* GMS framework
- generative medical segmentation, 3706
 - overview, 3494
 - three pillars, 3492
- generative model, 3
 - definition, 4
 - evaluation, 62
 - unlearning difficulty, 4606
- generative model training
 - communication hiding, 4990–4999
- generative model zoo, 4458
- generative modeling, 4, 51
 - problem definition, 59
- generative models

- accessibility advantages, 4697–4706
- and formal libraries, 3013
- architecture search, 1894
- as design tools, 2892
- as NAS tools, 1949–1955
- challenges, 2698–2713
- comparison, 2709
 - epilepsy, 3786
- connection to evolution, 2105
- data need vs. discriminative models, 4445
- data requirements, 4444
- distributed training, 5020–5041
- distribution necessity, 4932
- epilepsy, 3773
 - future directions, 3929–3939
 - history, 3776
 - motivation, 3782
- evolutionary perspective, 2251–2253
- explainability challenges, 3965
- explainability gap, 3990–4002
- federated training challenges, 4453
- future directions, 3427–3442
- harder to interpret, 3966
- interpretability, 3961–3981
- mathematics, 2933–2964
- medical imaging, 3477–3500
- medical segmentation history, 3487
- medical segmentation revolution, 3489
- motivation for PDE solving, 3230
- necessity for accessibility, 4696
- PDEs, 3191–3450
- privacy attacks, 4589–4599
- relationship to deterministic models, 4303
- scientific discovery history, 2893
- structured data, 2687–2691
- taxonomy
 - epilepsy, 3785
 - federated, 4458
 - taxonomy exercise, 3523
 - weather, 4181–4202
 - weather motivation, 4192–4202
- generative PDE solvers, 3264–3295
 - motivation, 3230
- generative prior, 330, 337–338
 - definition, 337
 - evolution, 337
 - exercises, 340
 - GAN projection, 338
- generative recommendation, 2581–2583
 - formal definition, 2582
- generative replay, 2899–2904
 - dreaming analogy, 1968
 - epilepsy, 3933
 - error accumulation, 2900
 - tabular streams, 2900
- generative reranking, 2487–2499
- generative retrieval, 2333–2520, 2630
 - computational limits
 - challenge, 2639
 - corpus updates, 2516
 - definition, 2509
 - exercise, 2635
 - explainability, 2517
 - hallucinated IDs, 2516
 - limitations, 2516–2520
 - memory capacity, 2517
 - open problems, 2517
 - recommendation, 2513
 - scalability, 2516
 - scaling conjecture, 2518
 - theory, 2517–2518
- generative segmentation
 - comparison with discriminative, 3493
 - definition, 3491
- generative superensemble, 4643–4644

- generative technologies
 - convergence, 3914
- generative theorem proving, 3036–3037
 - definition, 3036
- generative training
 - for embeddings, 2392
- generative uncertainty, 2878
- generative weather model, 4297–4307
 - definition, 4301
 - formal definition, 4301
- generator, 274
 - data vault metaphor, 4598
- generator averaging
 - failure modes, 4453
- generator paradigm, 3038
- genetic algorithm, 2280
 - exercise, 2287
- genetic algorithms, 2037–2040
 - definition, 2037
- genetic programming, 2045–2047, 2281
 - bloat, 2045
 - crossover, 2045
 - definition, 2045
 - human-competitive, 2046
 - mutation, 2045
- Genie, 723
- genre transfer, 895
- GenSeg, 3707–3711
 - convergence
 - practical implications, 3710
 - convergence theorem, 3708
 - exercise, 3714
 - implementation, 3708
 - low-data regime, 3710
 - Nature Communications, 3707
- GenSelfDiff-HIS, 3613
- geo-distributed training, 5119–5120, 5132
 - latency, 5123
- Geo-FNO, 3223
- geo-visual agents, 4717
- geodesic
 - connectivity, 3914
 - probability space, 3936
 - Wasserstein space, 3872
- geometric distribution, 31
- geopolitics
 - weather AI, 4283
- geospatial accessibility agent
 - definition, 4718
 - design exercise, 4724
- geospatial AI
 - accessibility, 4716–4725
- GEPA, 2133–2144, 2286
 - algorithm, 2138
 - benchmark results, 2142
 - coach LLM, 2135
 - convergence rate, 2141
 - crossover operator, 2141
 - execution trajectory, 2135
 - full algorithm, 2138
 - guided mutation, 2137
 - Pareto selection, 2139
 - pipeline, 2139
 - prompt genome, 2134
 - prompts as genomes, 2134
- GFlowNet, 2865–2876
 - amortized MCMC, 2871
 - applications, 2874–2876
 - backward policy, 2869
 - Bayesian structure learning, 2873, 2874
 - binary string example, 2866
 - biological sequence design, 2875
 - combinatorial optimization, 2873
 - comparison with RL and MCMC, 2873
 - connection to MCMC, 2871
 - DAG example, 2868
 - DAG formulation, 2866–2873

- detailed balance, [2868](#)
- detailed balance exercise, [2876](#)
- diversity, [2875](#)
- drug design exercise, [2895](#)
- exercise, [2875](#)
- flow function, [2867](#)
- flow matching, [2867–2873](#)
- flow matching condition, [2867](#)
- flow matching exercise, [2918](#)
- flow matching loss, [2868](#)
- forward policy, [2869](#)
- initial state, [2866](#)
- molecule generation, [2873, 2874](#)
- proportional sampling, [2866](#)
- relational data, [2924](#)
- sampling distribution, [2867](#)
- state space, [2866](#)
- structured data, [2872–2873](#)
- sub-trajectory balance, [2870](#)
- TB loss, [2869](#)
- terminal state, [2866](#)
- training algorithm, [2874](#)
- trajectory balance, [2869–2871](#)
- trajectory balance condition, [2869](#)
- trajectory visualization, [2870](#)
- variational interpretation, [2872](#)
- vs. reinforcement learning, [2875](#)
- ghost clipping, [1677](#)
- ghost generator
 - FedGAN, [4459](#)
- Ghysels, Pieter, [4986](#)
- Ghysels-Vanroose algorithm, *see*
 - communication-hiding
 - conjugate gradient
- Gibbs distribution, [133](#)
- Gibbs sampling, [131](#)
- Gibbs' inequality, [70](#)
- GIDM
 - medical images, [4593](#)
 - optimisation objective, [4593](#)
 - timestep selection, [4593](#)
- GIDM attack, [4592–4593](#)
- GigaGAN, [357](#)
- gigapixel image, [3515](#)
- Glahn, Harry, [4221–4231](#)
 - 1972 paper, [4222](#)
 - biography, [4231](#)
- gland segmentation, [3515](#)
- Glawion et al. (2025), [4357](#)
- Gleason grading, [3515](#)
- glioblastoma
 - segmentation, [3483](#)
- glioma
 - segmentation, [3511](#)
 - subregion segmentation, [3516](#)
- GLM, [114, 116](#)
- global coherence, [701](#)
- global context
 - tractography, [3891](#)
- global dot products
 - synchronisation overhead, [4986](#)
- global downscaling, [4355](#)
- global encoder, [4511](#)
- global filtering
 - cone projection, [4605](#)
 - federated unlearning, [4604](#)
- Global Forecast System
 - GFS
 - MOS, [4231](#)
- global gradient norm, [5109](#)
- global health
 - imaging access, [3484](#)
- global precipitation downscaling, [4357](#)
- global synchronisation
 - hiding latency, [4987](#)
- global utility, [2840–2841](#)
- global variable, [214](#)
- global workspace, [4148](#)
- Global Workspace Theory, [4148–4149](#)
- Gloo

- algorithm, 4976
- CPU backend, 4976
- specifications, 4977
- GloVe, 2366
- Glow, 455–456
 - 1x1 convolution, 455
 - ActNorm, 456
 - architecture, 456
 - LU decomposition, 456
- GLU, *see* gating, GLU
- GMS framework, 3613–3615, 3706–3707
 - architecture, 3613, 3706
 - generative interpretation, 3706
 - key insight, 3706
 - knowledge transfer, 3615
 - parameter efficiency, 3615
 - performance, 3707
- GNN
 - equivariance, 2752
 - message passing, 2749–2752
 - message passing exercise, 2920
- GNoME, 15
- GNOT, 3242–3375
 - computational complexity, 3244
 - heterogeneous attention, 3242
 - Inductor2D, 3347
 - Maxwell’s equations, 3347
 - multi-scale architecture, 3243
 - multiple PDE settings, 3243
 - problem setup, 3242
- Go (game)
 - AlphaGo, 3114
- goal
 - in proof assistant, 2987
- goal count
 - reward shaping, 3105
- goal extraction, 3094
- goal misgeneralization, 4134
- goal reduction reward, 3105
- Goblin Tools, 4802
- Gödel incompleteness theorems, 3079
- Goedel’s incompleteness theorem, 2974
- GOGGLE, 2745–2758
 - architecture, 2747
 - comparison with CTGAN, 2757–2758
 - decoder, 2752–2753
 - encoder, 2752–2753
 - expert constraints, 2754–2755
 - graph discovery, 2747–2749
 - loss function, 2753
 - sampling, 2755
 - training algorithm, 2755
- Goldberg, David, 2281
- Goldilocks zone, 2242
- Goldowsky-Dill, N., 3974
- Gong et al. (2024), 4369
- Good, I. J., 2258
- Good, I. J.
 - intelligence explosion, 1989
- Goodhart’s Law
 - constitutional fitness, 2261
 - in evolution, 2257
- Goodhart’s law, 1132
- Google
 - MetNet, 4240
- Google DeepMind
 - AI for science, 4272
 - mathematical AI approach, 3135
 - WeatherNext 2, 4334–4336
- Google Image Search, 2404
- Google Street View
 - accessibility, 4717
- Google TPU
 - training, 4937
- Gordon, Mike, 3003
- gossip protocols, 5015–5016
 - definition, 5015
 - topology, 5015
- gossip SGD, 5121

- definition, [5121](#)
- mixing speed, [5123](#)
- Gould, Stephen Jay, [2033](#)
- governance
 - federated generative AI, [4647–4658](#)
- government data
 - synthetic, [2889–2891](#)
- Goyal et al. 2017, [5026](#)
- GPAI models
 - regulatory requirements, [4447](#)
- GPipe, [5051–5060](#)
 - bubble overhead, [5173](#)
 - memory cost, [5051](#)
 - vs SWARM, [5126](#)
- GPT, [7](#)
 - EEG pretraining, [3887](#)
 - federated fine-tuning, [4460](#)
 - text simplification, [4792](#)
- GPT-2
 - data extraction, [1574](#)
 - data parallelism, [5028](#)
 - mechanistic analysis, [3977](#)
 - parameter count, [4926](#)
 - SAE features, [4055](#)
- GPT-2 Small
 - IOI circuit, [4028–4038](#)
- GPT-3
 - 175 billion parameters, [4926](#)
 - data parallelism, [5028](#)
 - mathematical limitations, [3047](#)
 - Megatron-LM training, [5142](#)
 - pipeline configuration, [5141](#)
 - training FLOPs, [4926](#)
 - training infrastructure, [5135](#)
- GPT-4, [17](#)
 - 1.8 trillion parameters, [4926](#)
 - accessibility, [4704](#)
 - autoformalization, [3067](#)
 - lesson plan generation, [4766](#)
 - lesson plans, [4796](#)
 - mathematical ability, [3146](#)
 - memory requirements, [4930](#)
 - training compute, [4925](#)
 - training data, [4443](#)
- GPT-4V
 - CVD clinical reasoning, [4782](#)
- GPT-5.2 Pro
 - graviton discovery, [3170](#)
- GPT-5.4 Pro, [3052](#), [3148](#)
- GPT-f, [2956](#), [3004](#), [3036–3037](#)
 - paradigm, [3038](#)
- GPT-NAS, [1946–1947](#)
 - algorithm, [1947](#)
 - architecture tokenisation, [1946](#)
 - effectiveness, [1947](#)
 - fine-tuning, [1947](#)
 - pipeline, [2013](#)
 - token sequence, [1946](#)
- GPT-Neo
 - membership inference, [1720](#)
- GPTQ, [1537](#)
 - error analysis, [1556](#)
- GPU
 - absence in 1990s, [4232](#)
 - architecture, [4933](#), [4942](#)
 - comparison table, [4936](#)
 - deep learning revolution, [4242](#)
 - memory hierarchy, [4934](#)
 - per-node count, [4927](#)
 - throughput optimisation, [4934](#)
- TPU
 - ASIC mixed cluster, [5164](#)
 - vs CPU, [4934](#)
- GPU access gap, [3716](#)
 - health equity, [3717](#)
- GPU cluster
 - anatomy, [4927](#)
- GPU communication libraries,
 - [4971–4980](#)
- GPU health checks

- pre-training, 5152
- GPU health monitoring, 5152
- GPU memory
 - capacity limits, 5042
 - HBM, 5135
- GPU-Direct RDMA, 4971
- GPU-to-GPU communication, 4971–4980
- Gröbner basis, 2990
- Grad-CAM, 3966, 3985–3986
 - axiomatic analysis exercise, 3990
 - comparison figure, 3986
 - derivation, 3985
 - medical imaging example, 3985
- gradient, 5220
- gradient accumulation, 659, 1530, 5101
 - definition, 5102
 - diffusion, 5085
 - diffusion training, 5090
 - elastic training, 5153
 - equivalence, 1531, 5153
 - exercise, 5105
 - FID trade-off, 5096
 - variance reduction, 5090
- gradient balancing, 857
- gradient bucket
 - definition, 4991
- gradient bucketing, 4991–4999, 5023, 5140
 - bucket size exercise, 4997
 - DDP, 5025
 - definition, 4991
 - optimal bucket size, 4992
 - proposition, 4993
- gradient checkpointing, 553, 5070–5081
 - algorithm, 5071
- gradient clipping, 1530
 - bias, 1677
 - distributed
 - exercise, 5114
 - distributed settings, 5109
 - DP-SGD, 2737
 - spike rollback, 5152
- gradient compression, 3718, 4475–4483, 5009–5010
 - all-reduce, 5041
 - convergence guarantees, 5166
 - motivation, 4958
- gradient decomposition
 - optimality, 1642
 - statistic, 1641
 - timestep, 1641
- gradient descent, 94
 - deceptive alignment, 4134
 - definition, 94
 - limitations, 2024
- gradient descent ascent
 - alternating, 313
 - algorithm, 313
 - bilinear analysis, 314
 - convergence, 324
 - definition, 313
 - comparison, 314
 - simultaneous, 311
 - algorithm, 312
 - convergence criterion, 315
 - definition, 311
 - divergence, 312
 - eigenvalue analysis, 323
- gradient direction
 - alignment, 1639
- gradient divergence
 - non-iid, 4450
- Gradient Episodic Memory, 1190–1191
 - definition, 1191
 - QP formulation, 1191
- gradient estimator
 - choosing, 229
- gradient flow
 - Cellpose, 3645

- distributed, 5176
- SVGD, 247
- gradient hooks
 - DDP, 5023
 - exercise, 5105
 - PyTorch, 5100
 - PyTorch DDP, 4989
- gradient inversion
 - analytical invertibility, 4590
 - attack, 4590, 4591
 - definition, 4591
 - diffusion models, 4592–4593
 - exercise, 4455
 - medical imaging example, 4597
 - PRISM intractability, 4601
- gradient inversion attack
 - definition, 4518
 - exercise, 4676
- gradient inversion attacks, 4453
- gradient leakage
 - defense, 4517–4526
- gradient noise
 - distributed training, 5110
- gradient norm
 - membership statistic, 1638
- gradient overlap
 - diffusion training, 5091
- gradient penalty, 296, 2725
 - computation example, 2727
 - definition, 297
 - derivation, 304
 - interpolation, 297
 - not needed, 2769
 - one-sided vs. two-sided, 297
 - WGAN-GP, 3819
- gradient quantization, 4476, 5009
 - 1-bit, 5010
 - ternary, 5010
- gradient reversal layer, 3578
 - implementation, 3579
- gradient sparsification, 4466, 4476, 5009
- gradient staleness
 - pipeline, 5059
- gradient synchronisation, 5022
- gradient underflow, 5106
- gradient variance
 - 1/B scaling, 5111
- gradient vector flow, 3527
- gradient-based attacks, 1636–1649
 - forensic fingerprints, 1640
 - membership signals, 1638
- gradient-boosted trees
 - for detection, 2843
- gradients
 - partitioning, 5062
- GradInversion attack, 4592
- Gram matrix, 334
 - exercises, 339
- grammatical evolution, 2282
- grand challenge
 - PDE solving, 3449
- grandmother cell hypothesis, 3969
- graph attention
 - HAMLET, 3250
- graph attention network
 - brain regions, 3894
 - connectome, 3894
- graph automorphism
 - in NAS, 1899
- graph convolutional network
 - attention, 3869
 - EEG, 3885–3887
 - spectral, 3886
- graph diffusion
 - architecture generation, 1937
- graph edit distance, 3924
- graph factorisation
 - Bayesian network, 4235
- graph generation
 - molecules, 2891

- graph hubness
 - propagation, 3869
- graph Laplacian, 1364
- graph neural network
 - cell tracking, 3655
 - message passing, 4268
 - tabular data, 2920
 - tabular generation, 2745–2758
 - weather, 4311
- graph neural networks
 - PDE solving, 3248
 - weather forecasting, 4265–4273
- graph partitioning
 - communication volume, 5100
- graph stabilisation, 2755
- graph structure
 - PDE domains, 3249
- graph transformer
 - PDE solving, 3248
 - spread prediction, 3914
- graph-cut algorithms, 3521
- graph-structured data, 2696
- GraphCast, 13, 3440, 4181, 4265–4273
 - architecture, 4267
 - blurring at long lead times, 4299
 - Google DeepMind, 4241
 - message passing exercise, 4414
 - open source, 4272
 - training, 4269
 - verification targets, 4270
- graphs
 - generative models, 2696
- graviton amplitude, 3148, 3168–3173
- gravity wave drag, 4210
- gravity waves, 3320
- gray zone
 - definition, 4361
 - nowcasting, 4361
- Gray–Scott system, 3305
- gray-box access, 1583
- greatest lower bound, *see* infimum
- Green’s function, 3234
- grid input
 - flattening, 4237
- grid types
 - NWP, 4214
- grid-cell averaging
 - NWP bias, 4226
- Griewank–Walther algorithm, 5072
- Griffin-Lim algorithm, 773
- Gröbner basis, 3032
- gross tumour volume, 3482
- Grothendieck, Alexander, 3142
- ground clutter, 4363
- ground truth
 - limitations, 3549–3557
- ground-glass opacity, 3549
- ground-state energy, 3364
- ground-state sampling, 3369
- grounding, 2531–2538
 - in evolutionary search, 2117
- group normalisation, *see* normalization,
 - group normalisation
- Group Relative Policy Optimization, 3053–3058
- group relative policy optimization, 1145–1149
- group theory
 - formalization, 2995–2997
 - in Lean, 2979
- group-relative advantage, 3055
- GroupConceptualCaptions dataset, 4879
- grouped query attention
 - parallelism constraint, 5049
 - tensor parallelism, 5047
- GroupNorm
 - special cases, 1551
- GroupWikiArt dataset, 4879
- grow-prune architecture, 1971

- growing capacity, [1232](#)
- GRPO, [1047–1149](#), [2133](#), [3053–3058](#)
 - algorithm, [1147](#), [3054](#), [3103](#)
 - binary rewards, [1147](#)
 - comparison with PPO, [3149](#)
 - convergence, [3057](#)
 - definition, [1047](#), [3054](#), [3149](#)
 - example, [1147](#)
 - exercise, [1068](#), [3062](#)
 - group-relative advantage, [3103](#)
 - intuition, [3055](#)
 - mathematical reasoning, [3102–3104](#)
 - memory comparison, [1163](#)
 - objective, [1145](#)
 - practical implementation, [3057](#)
 - theoretical properties, [3056](#)
 - training dynamics, [3058](#)
 - unbiasedness, [1048](#), [3056](#)
 - variance, [1163](#)
 - variance analysis, [3110](#)
 - vs. PPO, [1048](#), [3055](#)
 - vs. PPO exercise, [3064](#)
- GRPO (Group Relative Policy Optimization), [3136](#)
- GRU, [1029–1031](#), *see* gating, GRU
 - convolutional TrajGRU, [4240](#)
 - definition, [1029](#)
 - exercise, [1066](#)
 - gradient, [1030](#)
 - sequence modelling, [3885](#)
 - vs. LSTM, [1030](#)
- GSM8K, [3060](#)
- GTE embedding, [2387](#)
- guidance strength, [2781](#)
- guidance term, [2776](#)
- guided diffusion, [3401](#)
- guided evolution, [2265](#)
- Gumbel-Softmax, [229](#)
 - architecture search, [1926](#)
 - density, [265](#)
 - in NAS, [1902](#)
 - properties, [1902](#)
- Gumbel-softmax, [2715](#)
- GWT, *see* Global Workspace Theory
- H-MEM, [1100](#)
 - complexity, [1101](#)
 - exercise, [1112](#)
 - retrieval, [1100](#)
- H-SAM, [3633–3636](#)
 - architecture diagram, [3635](#)
 - exercise, [3641](#)
 - results, [3635](#)
- H100
 - FP8 support, [5075](#)
 - hardware failures, [5151](#)
 - NVLink bandwidth, [4978](#)
 - ridge point, [4936](#)
 - ridge point calculation, [4937](#)
 - specification, [4935](#)
 - Transformer Engine, [4935](#)
 - vs TPU, [4939](#)
- H100 GPU
 - communication hiding, [4993](#)
- H200
 - HBM3e, [4934](#)
- Hadamard conditions, [328](#), [3398](#)
- haematological analysis
 - GAN, [3587](#)
- haematoxylin and eosin staining, [3481](#), [3515](#)
- haemodynamic response function, [3850](#)
 - inversion, [3851](#)
- Haken, Wolfgang, [3029](#)
- Hales, Thomas, [2950](#)
- Hall of Fame, [2282](#)
- hall of fame, [2076](#)
- Hall, Timothy, [4234](#)

-
- hallucination, 4067–4076
 - abstention, 4074
 - accessibility AI, 4829–4841
 - accessibility summary, 4839
 - AI
 - BLV users, 4713
 - as discovery, 3148
 - as mathematical discovery, 3171
 - avoiding in geometry, 3122
 - circuit exercise, 4076
 - circuit perspective, 4070
 - clinical AI, 4807
 - coping strategies
 - blind users, 4860
 - detection exercise, 4915
 - downscaling, 4360
 - elimination via verification, 3085
 - evaluation, 4838
 - failure patterns, 4830
 - feature combinations, 2814
 - formal definition, 4070
 - in evolutionary search, 2204
 - in language models, 2520
 - in synthetic data, 2507
 - mathematical, 2958
 - mechanistic definition, 4070
 - minority culture, 4138
 - mitigation, 4073, 4835
 - multimodal, 3689–3691, 4903
 - clinical danger, 3689
 - detection exercise, 3693
 - evaluation, 3691
 - mitigation, 3690
 - open problem, 3690
 - sources, 3689
 - out-of-distribution, 4071
 - problem, 2531
 - reduction via RAG, 3683
 - retrieval head failure, 4071
 - risk decomposition, 4835
 - safety filter, 4919
 - safety standard, 4910
 - spurious co-occurrence, 4071
 - stakes, 4829
 - tabular data, 2814
 - visual entailment, 4834
 - vs deception, 4140
 - Hamilton–Jacobi–Bellman equation, 3438
 - Hamiltonian cycle, 3165
 - Hamiltonian game, 316
 - HAMLET, 3248–3375
 - fusion, 3250
 - geometry encoder, 3249
 - graph structure, 3249
 - graph Transformer layers, 3250
 - modular encoder, 3249
 - multi-relational graph, 3249
 - PDE info encoder, 3249
 - hammering, 3025–3028
 - Hammersley–Clifford theorem, 132
 - handwriting analysis
 - ML, 4765
 - handwriting recognition
 - dyslexia, 4765
 - Hann window
 - COLA, 773
 - hapax legomena, 4763
 - definition, 4764
 - Haralick texture features, 3529
 - hard problem of consciousness, 4149
 - hard conditioning, 2783
 - hard negative mining, 2393–2468
 - BM25 negatives, 2393
 - cross-encoder negatives, 2393
 - gradient analysis, 2394
 - in DPR, 2367
 - self-negatives, 2393
 - hard negatives
 - synthetic, 2506

- hardware
 - AI training, [4933](#)
 - neural PDE accelerators, [3434](#)
 - parallelism strategy, [4940](#)
- hardware cost
 - accessibility barrier, [4826](#)
- hardware design
 - evolutionary optimisation, [2115](#)
- hardware failures
 - large-scale training, [5145](#)
- hardware heterogeneity, [4580](#)
- hardware scaling
 - challenge, [4942](#)
- hardware stratification, [3716](#)
- hardware utilisation
 - efficiency, [4927](#)
 - MFU, [4930](#)
- hardware verification, [3004](#)
 - ACL2, [3004](#)
- hardware-aware NAS, [1855](#), [1909](#),
[1928–1930](#)
 - differentiable, [1929](#)
 - latency lookup tables, [1929](#)
 - objective, [1928](#)
 - results comparison, [1930](#)
- harmful behavior
 - concealment, [4138](#)
- Harmonic AI, [3152](#)
- harmonic functions, [3296](#)
- harmonisation
 - cross-scanner, [3583](#)
- Harpoon, [2775–2787](#)
 - algorithm, [2779–2787](#)
 - condition types, [2783–2787](#)
 - convergence, [2782](#)
 - discrete features, [2786](#)
 - fairness, [2785–2786](#)
 - full algorithm, [2781](#)
 - guidance strength, [2786](#)
 - hyperparameters, [2781](#)
 - imputation, [2783–2787](#)
 - manifold geometry, [2776–2787](#)
 - practical considerations, [2786](#)
 - privacy analysis, [2925](#)
 - projection operator, [2777–2787](#)
 - scalability, [2787](#)
 - theoretical justification, [2781](#)
 - theory, [2781–2787](#)
- Harrison, John, [3003](#)
- Hartree–Fock theory, [3370](#)
- Hassabis, [16](#)
- Hausdorff distance
 - 95th percentile, [3545](#)
 - definition, [3545](#)
 - loss, [3486](#)
- HBM
 - bandwidth, [4934](#)
 - High Bandwidth Memory, [4934](#)
 - memory errors, [5151](#)
- HBM2
 - specification, [4934](#)
- HBM3
 - specification, [4934](#)
- HCCL
 - Intel interconnect, [4939](#)
- He initialisation
 - derivation, [1551](#)
- head composition, [4006–4009](#)
- head importance score, [1944](#)
- head pruning
 - performance bound, [1945](#)
- health equity
 - computational, [3716](#)
- healthcare
 - expert constraints, [2754](#)
 - synthetic data, [2885–2887](#)
 - time series, [2800](#)
- healthcare access
 - metabolic imaging, [3857](#)
- healthcare data

- access restrictions, 2687
- generation, 2823
- hearing aids
 - history, 4693
- hearing loss
 - prevalence, 4690
- heat equation, 3303–3311
 - 2D PINN implementation, 3460
 - as generative model, 3311
 - backward, 3399
 - DeepONet, 3231
 - definition, 3303
 - diffusion model analogy, 3267
 - discontinuous IC, 3307
 - neural operators, 3308
 - parameter estimation, 3408
 - PINN loss, 3451
 - PINNs, 3307
 - properties, 3304
 - smooth initial condition, 3307
- heatwave
 - spatial coherence, 4336
- heatwaves
 - machine learning prediction, 4401
- Hebb, Donald, 1849, 2344
- Hebbian learning, 1176, 2344–2346
- Hebbian learning rule, 1850
 - pruning dynamics, 1850
- Heine–Borel theorem, 2995
- Heine–Borel theorem, 5202
- Heine–Cantor theorem, 2958, 5207
- Heisenberg uncertainty
 - spectrogram, 826
- Heisenberg uncertainty principle
 - time-frequency, 3799
- helicity, 3168
- helium atom
 - variational calculation, 3371
- Hellmann–Feynman theorem, 3369
- Helmholtz equation, 3341–3352
- 1D, 3345
- applications, 3344
- definition, 3341
- diffusion models, 3343
- high-frequency regime, 3342
- seismic imaging, 3343
- spectrum preservation, 3344
- hemispheric asymmetry
 - FDG-PET, 3858
- heredity, 2019–2022
- heritability
 - definition, 2028
- Hessian matrix, 5221
- heterogeneity
 - hospital example, 4577
 - taxonomy, 4569
 - unified view, 4588
- heterogeneous attention, 3242
- heterogeneous clusters, 5164
- heterogeneous features, 2689
 - challenge, 2698
- HGEFS, *see* NOAA Project EAGLE
 - ensemble weighting, 4392
- hidden Markov model, 142
 - definition, 2430
 - keyword spotting, 2429
- hierarchical adaptive optimization, 2170–2180
- hierarchical All-Reduce
 - exercise, 5134
- hierarchical codebook, 413
- hierarchical decoding, 3633–3636
- hierarchical decomposition, 1459
- hierarchical generation
 - audio, 882–888
 - compression, 883
 - definition, 882
 - training, 883
- hierarchical keyframe generation, 700–706

- complexity, 703
- definition, 701
- exercise, 706
- hierarchical latent variables
 - segmentation, 3563
- hierarchical matrices, 3252
- hierarchical memory, 1098–1102
 - H-MEM, 1100
- Hierarchical Memory Transformer, 1099
- Hierarchical Navigable Small World, *see* HNSW
- hierarchical organisation, 1275
- hierarchical planning, 1290–1294
 - search reduction, 1291
- Hierarchical Reasoning Model, 1059
- hierarchical reinforcement learning, 1059
- hierarchical search, 1893
- hierarchical sparse-dense framework, 1233–1239
- hierarchical spatial structure
 - atmosphere, 4237
- hierarchical VAE, 475
- HiFi-GAN, 767, 4751
 - architecture, 4752
 - configuration, 782
 - discriminator, 783–787
 - discriminator analysis, 801
 - feature matching, 785
 - GAN connection, 942
 - GAN theory connection, 787
 - generator, 781–787
 - definition, 781
 - MPD vs. MSD, 784
 - multi-period discriminator, 783, 4758
 - multi-scale discriminator, 784
 - receptive field, 782
 - training algorithm, 786
 - training losses, 784–787
- high-cardinality features, 2693
- high-dimensional PDEs, 3438–3440
- high-dimensional tables, 2708, 2908
- high-frequency components
 - membership signals, 1650
 - signal amplification, 1650
- high-frequency oscillations
 - EEG, 3884
- high-frequency oscillations (HFOs), 3794
- high-stakes decisions
 - accessibility, 4908
- higher-order dependencies, 2864
- higher-order interactions, 2700
- higher-order logic, 2945, 3000
- highway networks, *see* gating, highway networks
- Hilbert (Apple), 3153
- Hilbert curve
 - 3D MRI, 3904–3907
 - locality bound
 - exercise, 3908
 - locality preservation, 3904
- Hilbert space, 5214
 - diffusion models, 3270
 - diffusion on, 3378
 - operator learning, 3211
- Hilbert’s problems, 2939
- Hilbert’s programme, 2938
- Hilbert, David, 2938, 3016
- Hilbert-VLM, 3904–3907
 - example, 3907
 - figure, 3906
- Hillis, W. Daniel, 2058
- Hillis, W. Daniel, 2281
- hindsight experience replay, 3106–3107
 - example, 3106
 - implementation, 3111
- Hinton, 16
- Hinton, G., 3962

-
- HIPAA, [2885](#), [3717](#), [4443–4447](#)
 - de-identification, [4446](#)
 - EEG, [3940](#)
 - key provisions, [4446](#)
 - synthetic data, [2911](#)
 - synthetic data compliance, [4616](#)
 - hippocampal atrophy, [3912](#)
 - hippocampal replay, [1872](#), [2612](#)
 - hippocampal sclerosis, [3776](#)
 - hippocampal seizure
 - scalp visibility, [3855](#)
 - hippocampal volume, [3479](#)
 - hippocampus, [1176](#), [2340–2341](#)
 - encoding, [2340](#)
 - ROI weighting, [3858](#)
 - seizure onset, [3891](#)
 - Hippocrates
 - epilepsy, [3775](#)
 - histology
 - segmentation, [3515](#)
 - histopathology
 - segmentation, [3481](#), [3506](#)
 - HKO-7 dataset, [4239](#), [4370](#)
 - HMCA
 - Hilbert-Mamba Cross-Attention, [3905](#)
 - HMM
 - accessibility, [4701](#)
 - HMT, *see* Hierarchical Memory Transformer
 - compression analysis, [1112](#)
 - efficiency, [1100](#)
 - HNSW, [2379](#)
 - complexity, [2381](#)
 - construction, [2380](#)
 - exercises, [2384](#)
 - search algorithm, [2379](#)
 - HNSWlib, [2384](#)
 - HOL, *see* higher-order logic
 - HOL Light, [3003–3004](#), [3025](#)
 - kernel size, [3004](#)
 - HOL4, [3003–3004](#)
 - HOLD++, [1679–1686](#)
 - AUC reduction, [1682](#)
 - CelebA experiments, [1682](#)
 - convergence analysis, [1680](#)
 - convergence theorem, [1681](#)
 - dynamics, [1679](#)
 - exercise, [1702](#)
 - method, [1679](#)
 - momentum analogy, [1680](#)
 - one-dimensional example, [1680](#)
 - HOList, [3036](#)
 - paradigm, [3038](#)
 - Holland, John, [2280](#)
 - Homer, Nils, [1577](#)
 - homogeneous data types, [2688](#)
 - homomorphic encryption, [4522](#)
 - comparison with VAEFL, [4521](#)
 - federated, [4602](#)
 - federated learning, [4599](#)
 - homotopy type theory, [3002](#)
 - Hong Kong
 - load forecasting, [4234](#)
 - Hong Kong Observatory
 - radar archive, [4239](#)
 - Hong Kong University of Science and Technology, [4238](#)
 - hop size, [771](#)
 - Hopfield, [16](#)
 - Horizon AI
 - hurricane forecasting, [4389](#)
 - horizontal federated learning, [4451](#)
 - Horsley, Victor, [3775](#)
 - hospital consortium
 - federated learning, [4462](#)
 - hospital data, [2746](#)
 - hospital federation, [4577](#)
 - host-parasite dynamics, [2058–2061](#)
 - hosted evaluation server, [3923](#)

- HoTT library, [3003](#)
- Hounsfield units, [3481](#)
- How2Sign dataset, [4728](#)
- HRM, *see* Hierarchical Reasoning Model
- HRRR
 - High-Resolution Rapid Refresh, [4241](#)
- Hsieh, William W., [4236](#)
 - timing, [4236](#)
- HT-Fed-GAN, [4491](#)
 - exercise, [4494](#)
- Hu et al. 2023, [4345–4346](#)
- Huang, Daniel, [3035](#)
- hub node
 - stimulation target, [3916](#)
- hub nodes
 - epilepsy, [3864](#)
- Hubel, D. H., [3969](#)
- Hubel–Wiesel experiments, [3969](#)
- HuBERT, [2432–2433](#)
- hubness distribution, [3869](#)
- Huet, Gérard, [3000](#)
- Huet, Gérard, [2944](#)
- Human Connectome Project, [3891](#)
- human evaluation
 - video, [741](#)
- human intuition
 - role of, [3448](#)
- human motion generation, [742–743](#)
- human preferences
 - inconsistency, [2262](#)
- human ratings
 - introspection, [4144](#)
- human values, [4133](#)
 - diversity, [4135](#)
- human-consensus bound, [3649](#)
- human-in-the-loop, [2754–2755](#)
 - segmentation, [3489](#)
- human-in-the-loop segmentation, [3518](#)
- Hungarian algorithm, [3505](#), [3654](#)
 - cell tracking, [3752](#)
 - exercise, [3664](#)
- hurricane
 - Beryl, [4399](#)
 - Helene, [4399](#)
 - Lee, [4399](#)
 - Milton, [4399](#)
- Hurricane Beryl
 - GenCast ensemble, [4315](#)
- Hurricane Lee
 - forecast example, [4260](#)
 - track prediction, [4260](#)
- Hurricane Milton
 - AI forecast, [4389](#)
- Hutchinson’s trace estimator, [462](#), [518](#), [5223](#)
- Hwang, Jeng-Ren, [4234](#)
- hybrid accessibility architecture
 - definition, [4821](#)
- hybrid architecture, [1106](#)
 - accessibility, [4820–4829](#)
 - CVD real-time, [4783](#)
 - edge deployment pipeline, [4785](#)
 - paradigm, [4821](#)
- hybrid architectures, [1962](#)
 - diffusion models, [1942](#)
 - medical segmentation, [3543](#)
- hybrid CNN-Transformer
 - seizure prediction, [3885–3887](#)
- hybrid forecasting
 - AI and NWP, [4327](#)
- hybrid forward process, [2896](#)
- hybrid generative models
 - epilepsy, [3823–3829](#)
- hybrid methods
 - challenge, [2773](#)
 - latent and function space, [3395](#)
- hybrid models
 - autoregressive-diffusion, [726–733](#)
- hybrid NWP-diffusion, [4644–4645](#)

- hybrid NWP-ML
 - architectures, [4304](#)
- hybrid parallelism, [4932](#), [5135–5144](#)
 - axis commutativity, [5136](#)
 - communication complexity, [5139](#)
 - configuration proposal, [5179](#)
 - configuration search, [5155](#)
 - decision algorithm, [5140](#)
 - decision framework, [5140](#)
 - definition, [5135](#)
 - diffusion models
 - definition, [5086](#)
 - formal definition, [5136](#)
- hybrid pipeline
 - latency analysis, [4788](#)
- hybrid retrieval, [2372–2373](#)
- hybrid weather model, [4280](#), [4409](#)
- HyDE, [2502–2505](#)
 - algorithm, [2502](#)
 - approximation quality, [2504](#)
 - cross-lingual, [2505](#)
 - definition, [2502](#)
 - example, [2503](#)
 - failure modes, [2504](#)
 - query-document gap, [2503](#)
 - vs fine-tuned retrieval, [2504](#)
- hydrostatic approximation, [4208](#)
- HyFLM, [4532–4533](#)
 - personalised objective, [4533](#)
- hygiene
 - macro, [2992](#)
- hypercube topology, [4946](#)
- hypergradient, [1922](#)
- HyperNEAT, [2051](#), [2283](#)
- hypernetwork, [1202](#)
- hyperparameter reporting, [3923](#)
- hyperparameter search
 - 1990s cost, [4238](#)
 - parallelism configuration, [5140](#)
- hypothesis generation
 - LLM, [3134](#)
- hypothesis testing, [1578](#)
- hypothetical document embeddings, [2502–2505](#)
- I-projection, [73](#), [171](#)
- I2VGen-XL, [673](#)
- IAF, [239](#)
- IAF (Inverse Autoregressive Flow), [457](#)
- ICI
 - Inter-Chip Interconnect, [4939](#)
- ICLR 2026
 - memorization attack, [4593](#)
- ICON, [3256–3260](#)
 - approximation guarantee, [3258](#)
 - architecture, [3257](#)
 - convergence rate, [3258](#)
 - coupled multiphysics, [3260](#)
 - formal definition, [3257](#)
- ICON model, [4218](#)
- icosahedral grid, [4214](#), [4215](#)
 - 0.25 degree, [4312](#)
- icosahedral mesh, [4265–4273](#)
- icosahedral tessellation, [4266](#)
- ictal discharge propagation, [3890](#)
- ictal EEG, [3794](#)
- ictal propagation velocity, [3869](#)
- ictal SPECT, [3857](#)
- IDC
 - datasphere projection, [4447](#)
- idea generator, [3090](#)
- ideal continual learner, [1237](#)
- ideal gas law, [4206](#)
- identity
 - vocal, [4745](#)
- identity consistency, [707](#)
- identity fraud
 - voice, [939](#)
- identity preservation
 - talking head, [745](#)

- IDGenRec, 2514–2515, 2581–2583
 - movie recommendation, 2515
 - textual identifier, 2514
- iDLG attack, 4592
- iEEG, 3854
- iEEG.org, 3781
- IFS, *see* Integrated Forecasting System
 - comparison to Pangu-Weather, 4259
- IIT, *see* Integrated Information Theory
- ill-posed problem, 328
 - exercises, 341
 - in vision, 329
- ill-posedness, 3398
- illumination, 2084
- image captioning
 - accessibility, 4694
 - VLM, 4701
- image collection description, 4711
- image conditioning
 - video, 537
- image inpainting
 - diffusion exercises, 3847
 - EEG forecasting, 3836
 - phase-field methods, 3363
- image labelling
 - ImageNet, 4698
- image matching
 - example, 2399
- image retrieval, 2397–2406
 - definition, 2397
- image synthesis, 10
 - medical, 3580–3596
- image–text matching, 3679
 - score, 3679
- image-level labels, 3520
- image-text retrieval
 - benchmarks, 2417, 2601
- image-to-image translation, 12, 336–337
 - examples, 336
 - exercises, 340
 - paired, 346–348, 3580
 - paired vs. unpaired, 336
 - problem formulation, 336
 - unpaired, 348–350
- image-to-text retrieval, 2467
- image-to-video, 669–675
 - autoregressive extension, 675
 - concatenation, 617, 671
 - concatenation conditioning, 670
 - conditioning mechanisms, 670
 - cross-attention conditioning, 671
 - definition, 670
 - hybrid conditioning, 672
 - information flow, 672
 - information propagation, 672
 - logarithmic depth, 673
 - noise-free conditioning, 671
 - pipeline, 674
 - problem formulation, 670
 - systems, 673
 - training objective, 675
- ImageBind, 2450–2451, 2458–2467
 - cross-modal retrieval, 2451
 - emergent alignment, 2450, 2459
 - emergent retrieval, 2460
 - exercises, 2455
 - image-anchored alignment, 2458
 - scaling laws, 2460
 - star topology, 2459
- ImageGPT, 444
- ImageNet, 22
- images
 - regular grid structure, 2688
- ImageSet2Text, 4711–4878
- imaging modalities
 - overview, 3516
- imbalance ratio
 - definition, 3804
- immune system
 - as multi-agent system, 1271

IMO

AI performance, 2961

IMO 2024, 3112

Problem 1, 3125

Problem 2, 3125

Problem 3, 3125

Problem 4, 3125

Problem 5, 3126

Problem 6, 3126

problem difficulty, 3113

problem-by-problem analysis,
3124–3126

story, 3112–3113

IMO geometry problems

benchmark, 3123

IMO Grand Challenge, 3033–3034

implantable device

generative EEG, 3953

implantable devices

generative models, 3930

implicit density model, 8

implicit function theorem, 1054, 2782

implicit generative model, 52, 2088

implicit knowledge

mathematics, 3065

implicit model, 139, 273

implicit motion blindness, 4898

importance ratio, 3054

importance sampling, 55

pseudo-likelihood, 1747

importance score, 1885

importance weighting, 241–243

self-normalized, 243

importance weights, 55

importance-weighted autoencoder

exercise, 401

importance-weighted ELBO, 202

imputation

conditional, 2783

in-batch negatives, 2366

in-context learning

as agent adaptation, 1286–1287

Bayesian, 1446

Bayesian interpretation, 1286

emergence, 4026

mechanistic basis, 4020–4028

memory analogy, 4150

operator learning, 3257

tabular data, 2812

in-context operator learning, 3256–3260

Inception Score, 21, 283

exercise, 287

InceptionV3

EEG transfer learning, 3818

inclusive classroom, 4880–4882

example, 4882

incommensurability, 4135

incompleteness, 2974

implications, 2950–2951

incompleteness theorems, 2939

incremental compilation, 2997

independence, 38

continuous and categorical noise,
2765

independence from ZFC, 2951

independence map, 140

independent component analysis (ICA)

EEG, 3797

independent frame diffusion, 542

independent frame generation

failure, 542

independent replication

performance gap, 3922

India

AI weather

farmers, 4390

indirect effects, 4013

definition, 4013

exercises, 4019

measurement, 4014

- indirect encoding, [2051](#)
- Indirect Object Identification, [4028–4038](#)
 - task definition, [4029](#)
- indirect object identification, [3973](#)
- inducing points
 - set transformer, [3891](#)
- induction
 - in Lean, [2966](#)
- induction head, [4008–4009](#)
 - associative recall, [4126](#)
 - attention pattern characterisation, [4023](#)
 - definition, [4022](#)
 - information flow example, [4010](#)
 - pattern matching, [4166](#)
 - random tokens, [4024](#)
 - traced example, [4024](#)
- induction heads, [4020–4028](#)
 - pattern matching, [4022](#)
- inductive bias
 - complementarity, [2903](#)
 - periodic, [793](#)
 - structured, [2747](#)
 - structured data, [2689](#)
- inductive types, [2948](#), [2985–2987](#)
- inference
 - energy cost, [4812](#)
 - graphical model, [134](#)
 - memory capacity constraint, [4939](#)
 - NP-hard, [166](#)
- inference cost
 - GenCast TPU, [4314](#)
- inference network, [230](#)
- inference–learning duality, [170](#)
- inference-time scaling, [3120](#)
- infilling, [2799](#)
- infimal convolution, [293](#)
 - properties, [305](#)
- infimum, [5191](#)
- InfiniBand, [4966](#), [5115](#)
- bandwidth, [4949](#), [5137](#)
- cluster topology, [4928](#)
- crossover point, [4957](#)
- diagram, [4968](#)
- fat-tree, [4945](#)
- HDR
 - communication hiding, [4993](#)
 - latency and bandwidth, [4951](#)
 - NDR 400Gbps, [4928](#)
 - specifications, [4967](#)
- infinite cosine schedule, [1207](#)
 - plasticity, [1207](#)
- infinite descent, [2936](#)
- infinite-dimensional diffusion, [3377–3384](#)
- infinitely many primes, [2976](#)
- influence functions
 - diffusion models, [1644](#)
 - exercises, [1648](#)
 - membership inference, [1644](#)
 - self-influence, [1645](#)
- InfoNCE
 - exercise, [2632](#)
- InfoNCE loss, [2366](#), [2369](#), [3669](#)
 - audio-text, [803](#)
 - in CLIP, [2408](#)
 - temperature analysis, [3692](#)
- informal mathematics, [3064](#)
- informal proof
 - alignment with formal, [3092](#)
- informal proofs
 - as scaffolding, [3068](#)
- informal reasoning
 - ATP limitations, [3029](#)
 - limitations, [2958](#)
- informal to formal translation, [3085](#)
- informal vs. formal proofs, [2936](#)
- informal-to-formal translation, [3040](#)
- informal-to-Lean translation, [3067](#)
- informalization, [3069–3070](#)

- applications, 3070
- example, 3070
- formal to informal, 3070
- information bottleneck, 405, 1860
 - audio codec, 802
 - controversy, 1863
 - convexity, 1993
 - deep learning, 1862
 - Gaussian case, 1866
 - practical trade-off, 2562
 - pruning connection, 1863
 - retrieval, 2561
 - theorem, 1860
- information bound
 - continual learning, 1231
- information curve, 1862
- information flow
 - graph-controlled, 2750
 - patch to CLS, 4089
 - retrieval pipeline, 2677
- information gain
 - in scientific discovery, 2218
 - prompt enhancement, 685
- information geometry, 205
- information loss, 2735
 - as regulariser, 2735
- information movement
 - OV circuit, 4005
- information retention, 1608
 - definition, 1609
 - exercises, 1616
 - monotonic decay, 1609
 - non-Gaussian inputs, 1610
- information retrieval, 2333
 - foundations, 2348–2364
 - history, 2348–2349, 2524
- information routing, 4012
- information theory
 - cross-modal reconstruction, 3856
 - in evolutionary search, 2103
 - membership inference, 1587
 - neural systems, 1857–1868
 - retrieval analysis, 2554–2565
- information-preserving transformation, 4696
- information-theoretic bound
 - exercise, 3862
- information-theoretic lower bound
 - distributed training, 5165
- information-theoretic metrics, 2852–2854
- inherently interpretable architectures, 4156–4157
- initial condition perturbations
 - NWP vs. diffusion, 4313
- initial condition uncertainty, 4290
- initial conditions
 - NWP, 4215
- initial noise
 - membership inference, 1653
- initialisation
 - sensitivity, 3531
- initialization, 1472–1481
 - Fixup, 1478
 - Glorot, *see* initialization, Xavier
 - He, 1475
 - definition, 1476
 - Kaiming, *see* initialization, He
 - LSUV, 1477
 - muP, 1478
 - definition, 1479
 - hyperparameter transfer, 1479
 - orthogonal, 1476
 - definition, 1476
 - variance propagation, 1473
 - Xavier, 1474
 - definition, 1474
 - variance preservation, 1474
- inner product space, 5213
- inpainting, 11, 333–334

- classical methods, 333
- definition, 11
- degrees of freedom, 333
- diffusion PDE, 333
- exercises, 339
- generative, 333
- patch-based, 333
- problem formulation, 333
- InPars, 2500–2501
 - domain adaptation, 2500
 - quality bound, 2501
 - query generation, 2500
- inprocessing, 3018
- input definition
 - generative models, 3991
- INRIA, 3000
- inspiration porn
 - AI generation, 4843
- instance normalisation, *see*
 - normalization, instance
 - normalisation
- instance segmentation, 3503–3504
 - cell counting, 3504
 - definition, 3478, 3504
 - vs. semantic segmentation, 3643
- instantaneous change of variables, 461
- institutional data silo, 4448
- InstructGPT, 1136–1137
 - results, 1137
 - reward model, 1132
- instruction pipelining
 - latency hiding, 4983
- instruction tuning
 - federated, 4556
- instruction-conditioned embeddings, 2386–2387
- Instructor embedding, 2386
- instrument rendering, 947–948
- INT8
 - accessibility calibration, 4824
- INT8 quantization
 - hybrid systems, 4824
- integer encoding
 - pitfalls, 2699
- integral operators
 - attention approximation, 3452
 - attention as, 3236
 - attention connection, 3234
- Integrated Forecasting System, 4189
- Integrated Gradients, 3987–3988
 - axioms, 3987
 - baseline sensitivity exercise, 3990
 - path integral, 3987
- integrated gradients, 3966
- Integrated Information Theory, 4148
- integration by parts, 5224
- integration challenge, 3918
- integrity constraints
 - learning, 2926
- Intel
 - verification, 3005
- Intel Loihi, 4816
- intellectual disability, 4690
- intelligence explosion, 2258
- intensity prediction
 - hurricane, 4399
- inter-annotator variability, 3760
- inter-column dependencies, 2700–2702, 2728
- inter-node communication, 4928
- inter-observer variability, 3484, 3549–3557
 - averaging problem, 3498
 - clinical practice, 3549
 - distribution perspective, 3491
- inter-rater agreement
 - Cohen’s kappa, 3550
 - Fleiss’ kappa, 3550
- inter-rater variability
 - lung nodule example, 3515

- lung nodules, 3510
- inter-regional coupling
 - seizure propagation, 3893
- interaction terms
 - exponential growth, 2702
- interactive autoformalization, 3079
- interactive generation, 2829
- interactive segmentation, 3519
 - definition, 3519
- interactive theorem proving
 - automation, 3025–3028
 - history, 2999–3015
- interchange intervention, *see* activation
 - patching
- interference score, 4073
- interictal EEG, 3794
- interictal epileptiform discharge
 - automated detection, 3825
 - cross-modal detection, 3825
 - synthesis, 3824
- interictal epileptiform discharges (IEDs), 3795
- interictal PET, 3857
- interior, 5197
- interleaved 1F1B, 5052–5060
- interleaved AR-diffusion, 731
- interleaved retrieval, 2537
 - advantage, 2537
- intermediate reasoning tokens, 3048
- intermediate value theorem, 2937, 5211
- internal chain-of-thought, 3051
- internal covariate shift, 1489
- internal-output consistency, 4139
- international collaboration
 - federated learning, 4656–4657
- International League Against Epilepsy, 3774
- International Mathematical Olympiad, 2961, 3033–3034, 3112–3131
- International Medical Device
 - Regulators Forum, 3940
- Internet of Things, 4447–4448
- interoperability
 - theorem provers, 3008
- interpolation, 398
- interpolation scheme, 643
- interpretability, 3961–3981
 - alignment, 4133
 - bias detection, 4171
 - boundary hallucination, 3722
 - chapter summary, 4176
 - concept schedule, 4097
 - conjecture, 4162
 - connection to federated learning, 4463
 - definition, 3964
 - diffusion
 - steering, 4102
 - diffusion models, 3721–4102
 - diffusion temporal schedule, 4168
 - emergence, 4161
 - ensemble explanations, 3722
 - evaluation metrics, 4155
 - exercises, 4164–4176
 - formal definition, 3964
 - gaming of, 4140
 - mathematical AI, 3142
 - motivation, 3962
 - multimodal benchmark, 4174
 - multimodal generation, 4157
 - non-reproducibility, 3721
 - NWP vs neural network, 4234
 - of evolved systems, 2264
 - open problems, 4152–4164, 4177
 - oversight protocol, 4175
 - performance tradeoff, 3732–3734
 - plurality conjecture, 4163
 - research challenges, 4172–4176
 - research frontier, 4162
 - scaling, 4153–4154

- tabular generation, 2745
- temporal structure, 4096
- training-time integration, 4156
- with high performance, 3732
- interpretability desiderata, 3997–4000
 - definition, 3999
 - exercise, 4001
 - impossibility theorem, 3999
- interpretability wall, 4154
- interpretability-capability tradeoff, 4160–4161
- interpretability-first alignment, 4137
- Intersection over Union
 - definition, 3544
 - relationship with Dice, 3544
- intersection over union, 3486, 3740
- intersectional bias
 - definition, 4845
 - detection exercise, 4853
 - detection impossibility, 4847
 - disability, 4844
 - formal definition, 4845
- intra-node bandwidth, 4935
- intra-node communication, 4928
- intracranial EEG
 - data volume, 3780
 - synthesis, 3818, 3854
- intractability
 - posterior, 166
- intractable integral, 60
- intransitive dynamics
 - persistence, 2068
- intransitive superiority, 2068
- intransitivity, 2067–2069
 - definition, 2068
- intraoperative guidance, 3728
 - latency requirements, 3728
- intraoperative segmentation, 3762
- intrinsic dimensionality
 - video, 532
- intrinsic motivation, 2270–2272
 - scientific, 2217
- intrinsic physical states, 3245
- introspection
 - AI models, 4142–4152
 - base models, 4143
 - definition, 4142
 - model self-knowledge, 4171
 - philosophy of mind, 4142
- introspection gap, 3963
 - language models, 3971
- introspective awareness
 - definition, 4144–4145
- Intrusion dataset, 2854
- inverse CDF, 52
 - theorem, 52
- inverse design
 - diversity–accuracy trade-off, 3350
- inverse document frequency, 2350
- inverse problem
 - downscaling as, 4352
 - generative models and, 4352
- inverse problems, 327–338, 3397–3427
 - diffusion model, 3464
 - diffusion priors, 3401
 - function-space diffusion, 3381
 - general framework, 327
 - mathematical formulation, 3397
 - PDE-constrained, 3397
 - permeability recovery, 3299
- inverse STFT, 772
- InverseBench, 3406
- inversive geometry, 3123
- inverted file index, *see* IVF
- inverted index, 2360–2361
 - definition, 2361
 - with SPLADE, 2371
- IOI circuit, 4028–4038
 - ablation exercise, 4036
 - Duplicate Token Heads, 4031

- generalisation, 4037
- head roles, 4030
- implications, 4035
- limitations, 4035
- logit difference example, 4034
- Name Mover Heads, 4030
- previous token heads, 4031
- S-Inhibition Heads, 4030
- template, 4029
- walkthrough, 4032
- IoT
 - federated generation, 4621–4635
 - synthetic data, 2801
 - VAEFL deployment, 4523–4524
- IoT anomaly detection
 - federated, 4622
- IoT devices
 - 40 billion by 2030, 4447
- Iotti et al. (2025), 4355
- IoU, *see* Intersection over Union
- iResNet, 459
- iron–sulfur clusters, 3369
- irrationality
 - of $\sqrt{2}$, 2935
 - of $\text{sqrt}(2)$, 2981
 - rapidly converging series, 3134
- irregular meshes
 - GNOT, 3242
 - Transformer handling, 3235
- irregular sampling, 2696
- irritative zone, 3775
- Irving, Geoffrey, 3034
- Isabelle, 2934, 3080
- Isabelle/HOL, 2945, 3000–3002, 3025
 - key features, 3001
 - notable formalizations, 3001
 - proof style, 3001
- Isar proof language, 3001
- isCompact_Icc, 2995
- ISIC archive
 - MGDS attack, 4596
- Ising model, 132, 185
 - exercise, 143
- island model
 - exercise, 2294
 - migration strategies, 2118
 - spectral analysis, 2130
 - topology, 2127–2128
- Isola et al. 2017, 3857
- isolated point, 5196
- Itakura–Saito distance, 5273
- iterative projection, 2779
- iterative reasoning operator, 1063
- iterative refinement, 1051–1056, 3519
 - flow matching, 3288
 - theorem proving, 3090
- IVF, 2381–2382
 - multi-probe, 2382
 - residual encoding, 2381
 - with PQ, 2382
- IWAE, 241–243
 - bias, 242
 - gradient, 243
 - objective, 241
 - signal-to-noise ratio, 242
 - SNR degradation, 243
 - tightness, 241
 - tightness hierarchy, 241
 - tightness proof, 266
- Jaccard index, 3486, *see* Intersection over Union
- Jacobian, 44
 - normalizing flow, 3855
- Jacobian covariance, 1931
- Jacobian matrix, 5220
- jailbreaking, 4139
- Jain and Wallace 2019, 3988
- Jamba, 1106
- JAX

- pmap, [4927](#)
- Jelinek-Mercer smoothing, [2354](#)
- Jensen's gap, [175](#), [3354](#), [3390](#)
 - elimination, [3355](#)
- Jensen's inequality, [42](#), [175](#), [5247](#), [5276](#)
- Jensen–Shannon divergence, [75](#), [2741](#), [2852](#)
 - conditional, [345](#)
- Jensen-Shannon divergence
 - GAN, [276](#)
 - exercise, [285](#)
 - saturation, [287](#)
- JFT-3B, [4444](#)
- jina-embeddings, [2389](#)
- JMA GSM, [4218](#)
- Joachims, Thorsten, [2628](#)
- job management
 - automated restart, [5151](#)
- Johnson–Lindenstrauss lemma, [4042](#)
 - exercises, [4046](#)
 - statement, [4043](#)
- Johnson-Lindenstrauss lemma, [3971](#)
- joint attention, [604–607](#)
 - conditioning strength, [605](#)
 - definition, [604](#)
 - exercise, [607](#)
 - vs cross-attention, [605](#)
- joint coherence
 - WeatherNext 2, [4336](#)
- joint distribution, [36](#)
 - challenge, [4329–4330](#)
 - evaluation, [2835](#)
 - weather forecasting, [4329–4337](#)
- joint fidelity, [2838–2840](#)
- joint image-mask synthesis, [3844–3846](#)
 - exercises, [3849](#)
- joint label fusion, [3530](#)
- joint probabilistic forecasting
 - from marginals, [4337](#)
- Jost's law, [1174](#)
- Journal of Applied Meteorology, [4231](#)
- JRA-55, [4217](#), [4218](#)
- Jukebox, [417](#), [809](#)
 - lyrics conditioning, [901](#)
- Jumper, [16](#)
- juvenile myoclonic epilepsy, [3777](#)
- k-anonymity
 - synthetic medical images, [4615](#)
- K-composition, [4006](#), [4022](#)
 - definition, [4008](#)
 - Frobenius norm, [4028](#)
 - mechanism, [4023](#)
- k-means, [154](#)
 - as GMM limit, [154](#), [160](#)
 - initialisation, [153](#)
- k-means++, [153](#)
- k-space
 - under-sampling, [3586](#)
- Kaggle seizure prediction challenge, [3801](#)
- Kaliszyk, Cezary, [2956](#)
- Kalman filter
 - ensemble, [4216](#)
- Kalnay, Eugenia, [4287](#)
- Kantorovich duality, [371](#)
 - exercise, [383](#)
 - theorem, [371](#)
- Kantorovich formulation, [369–371](#)
- Kantorovich relaxation, [3864](#)
- Kantorovich, Leonid, [290](#), [3865](#)
- Kantorovich–Rubinstein duality, [81](#), [291](#), [371](#)
 - proof, [291](#)
- Kantorovich–Rubinstein theorem, [2726](#)
- kappa
 - Cohen's, [3550](#)
 - Fleiss', [3550](#)
 - pitfalls, [3550](#)
- Karpukhin, Vladimir, [2629](#)

- Karush-Kuhn-Tucker conditions, *see*
KKT conditions
- Kass, Michael, 3525
- Kauffman, Stuart, 2022
- Kaufmann, Matt, 3004
- KB-VQA, 2539
example, 2540
- KD-tree, 2377
- Kendall tau
propagation ranking, 3894
- Kendall's tau, 2856
- Kendall's τ , 1919
- Kepler conjecture, 2950
formalization, 3001
- kernel
size comparison, 3011
- kernel integral operator, 3236
- kernel methods
atmospheric science, 4236
- kernel verification
pipeline, 3094
- kernelized Stein discrepancy, 245
definition, 245
- KernelSHAP, 3983
- key-value memory, 4060
MLP, 4017
MLP layers, 4150
- keyframe spacing
optimal, 702
- keypoint estimation
sign language, 4729
- keystroke efficiency, 4760
- keyword spotting, 2429
- KFAC, 251
- Khan Academy
AI tutoring, 4805
- Khanmigo, 4805
- Khattab, Omar, 2630
- Kikuchi approximation, 254
- KILT, 2598–2599
provenance, 2599
- Kim et al. 2026, 4302
- Kimina-Prover, 3153
- Kimura, Motoo, 2033
- Kinetics, 22
- Kissat, 3018
- kissing numbers, 3132
- KKT conditions, 5265
example, 5278
- KL annealing, 397
- KL divergence, 21, 69
anomaly score, 3846
as Bregman divergence, 5273, 5279
bound, 2782
definition, 70, 170
EEG VAE, 3947
evaluation, 2852–2854
exercise, 200, 400
forward, 170
forward vs reverse, 262
forward vs. reverse, 73
Gaussian, 72, 261, 393–394, 489
closed form, 393
Gaussian example, 173
in distillation, 1533
in GOGGLE, 2754
in variational inference, 170
joint vs. product, 543
multivariate Gaussian, 72
non-negativity, 70
not a metric, 71
per-step, 2767
projection, 372
properties, 174
regularization, 3055
reverse, 170
sensitivity analysis, 1166
unbiased estimator, 1146
VAE regularisation, 3824
- Knothe–Rosenblatt rearrangement, 451

- knowability
 - limits on evolution, [2264](#)
- knowledge
 - parametric vs. non-parametric, [2520](#), [2523](#)
- knowledge bottleneck, [3029](#)
- knowledge capacity, [1231](#)
- knowledge conflicts, [4067–4076](#)
- knowledge distillation, [1196](#), [1532–1535](#), [3389](#)
 - as consolidation, [1966](#)
 - black-box, [1535](#)
 - definition, [1532](#), [4816](#)
 - diffusion models, [3762](#)
 - edge deployment, [4816](#)
 - federated, [4582](#)
 - federated learning, [4454](#), [4478](#)
 - for embeddings, [2391](#)
 - for ranking, [2591](#)
 - in continual learning, [1190](#)
 - loss function, [1532](#)
 - medical segmentation, [3717](#)
 - mutual learning, [1534](#)
 - PDE solvers, [3389](#)
 - self-distillation, [1534](#)
 - temperature, [1556](#)
 - temperature scaling, [1533](#)
- knowledge editing
 - evaluation, [4064](#)
 - history, [4066](#)
 - ROME, [4167](#)
 - side effects, [4066](#)
- knowledge evolution, [4071](#)
- knowledge graph
 - retrieval, [2620](#)
- knowledge importance score, [1885](#)
- knowledge index, [2521](#)
- knowledge storage, [4058–4067](#)
- knowledge-intensive tasks
 - MDM vs ARM, [4110–4113](#)
- Knuth, Donald, [3165](#)
- Kohn, Walter, [3370](#)
- Kolmogorov
 - 5/3 law, [3334](#)
- Kolmogorov barrier, [3225](#)
- Kolmogorov complexity, [2238](#)
 - in open-ended evolution, [2238](#)
- Kolmogorov flow
 - generative surrogates, [3466](#)
 - truncated sampling, [3389](#)
- Kolmogorov microscale, [3453](#)
- Kolmogorov n -width, [3225–3226](#)
 - definition, [3225](#)
 - examples, [3225](#)
- Kolmogorov spectrum, [3334](#), [4352](#)
 - computation, [3340](#)
- Kolmogorov superposition theorem, [3213](#)
- Kolmogorov–Smirnov statistic, [2741](#), [2837](#)
- Kolmogorov–Smirnov test, [2850](#)
 - exercise, [2921](#)
- Kong et al. attack, [1664](#)
- Kotelnikov, Akim, [2758](#)
- Koza, John, [2045](#), [2281](#)
 - human-competitive results, [2046](#)
- Kraskov estimator, [2853](#)
- Krylov basis
 - s-step computation, [4987](#)
- Krylov subspace methods
 - communication hiding, [4987](#)
- KS statistic, *see* Kolmogorov–Smirnov statistic
- Kullback–Leibler divergence, *see* KL divergence
- KV cache, [107](#)
 - exercise, [121](#)
 - memory cost, [1078](#)
 - memory estimate, [550](#)
 - memory formula, [4932](#)

- sizing, 1109
- tensor parallelism, 5047
- video, 550
- KV-cache
 - video diffusion, 748
- Kvasir-SEG dataset, 3514
- L-C-WGAN-GP, 3821–3822
- L1 cache
 - GPU, 4934
- L1 loss
 - MRI-to-PET, 3858
- L1 penalty
 - motivation, 4048
 - SAE, 4048
 - shrinkage bias, 4049
- label noise
 - audio captions, 856
- label skew, 4569
 - definition, 4570
 - simulation, 4578
- label smoothing
 - calibration, 1555
- label switching, 192
- label-free tracking, 3662–3663
- label-only access, 1584
- labelling bottleneck, 3490
- lagged inputs
 - MLP limitation, 4237
- Lagrange multiplier, 5261
- Lagrange’s theorem, 2995–2997
 - in Lean, 2979
- Lagrangian, 5261
- Lagrangian relaxation
 - for agent alignment, 1297
- LAION-5B, 4444
- LAION-Audio, 2449–2450
- lake
 - cache, 2998
- Lamarckian inheritance, 2034
 - in NAS, 2254
- LambdaMART, 2587–2589, 2628
 - algorithm, 2593
 - commercial deployment, 2589
- LambdaRank, 2587
 - gradient derivation, 2595
- LAMOL, 1195
 - definition, 1195
 - quality degradation, 1195
- LAMP, 710
- land-surface scheme
 - NWP error, 4227
- Landauer’s principle, 2259, 4812
 - accessibility, 4812
 - retrieval, 2623
- landmark retrieval, 2407
- Lang et al. 2026, 4304
- Langevin dynamics, 505
 - higher-order, 1679
- language games, 2335
- language model, 430
- language models
 - for retrieval, 2354–2356
- Language Server Protocol, 3008
- Language-Oriented Audio, 838
- Laplace approximation, 1187, 3412
- Laplace equation, 3295–3303
 - definition, 3296
- large language model
 - architecture, 114
 - cloud accessibility, 4821
 - epilepsy, 3786
 - triage pipeline, 3914
- large language models
 - accessibility, 4689
 - adaptive architecture, 1889
 - architecture design, 1943
 - architecture limitations, 1889
 - architecture search, 1895, 1943–1949
 - as architecture designers, 1951

- as mutation operators, 2095–2108
- as rankers, 2592–2593
- autoformalization, 3067
- baseline performance, 3047
- care execution gap, 3913
- compression, 1944
- distributed training, 5028
- epilepsy, 3899–3903
- federated, 4549–4558
- for recommendation, 2579–2585
- hallucination risk, 1952
- listwise ranking, 2593
- mathematics, 3046–3064
- memorisation, 1575
- PDE solving, 3259
- pruning results, 1946
- scaling, 4925
- self-adapting architecture, 2001
- tabular data, 2804–2815
- text simplification, 4704
- variational inference, 257
- vs trained rankers, 2592
- large-batch training
 - instability, 5027
- large-scale training, 2708, 4925
- laser interstitial thermal therapy, 3780
- Lashley, Karl, 2343
- last mile problem
 - scientific discovery, 2222
- last-10-meters problem, 4717
 - definition, 4717
- Late Chunking, 2389
- late interaction, 2368, 2630
 - efficiency, 2373
- latency
 - accessibility, 4809–4820
 - audio generation, 874
 - communication, 4951
 - decomposition, 875
 - seizure prediction, 3880
 - tolerance budget, 4811
- latency hiding, *see* communication hiding
- latency lookup tables, 1929
- latency-accuracy tradeoff
 - hybrid, 4825
- latency-bound
 - all-reduce, 4954
 - crossover, 4956
- latency-bound vs bandwidth-bound, 4951
- latent compression
 - predictability, 4349
- latent diffusion, 421
 - audio, 832–841
 - AudioLDM, 835
 - AudioLDM2, 837
 - compression ratio, 3600
 - compression-quality tradeoff, 835
 - computational advantage, 834
 - computational savings, 3599
 - differential privacy, 1676
 - exercise, 862, 3617
 - fMRI synthesis, 3851
 - medical imaging, 3599–3600
 - memory efficiency, 5167
 - MusicLDM, 839
 - PDE solvers, 3289
 - PDEs, 3391–3396
 - pipeline, 3392
 - pipeline definition, 833
 - SEEDS, 4319
 - systems lesson, 5083
 - temporal structure, 584
 - three-stage pipeline, 833
 - training algorithm, 858
 - weather, 4339–4349
- latent diffusion models, 5082
 - federated personalisation, 4459
- Latent Dirichlet Allocation, 212–213

- stochastic inference, 217
 - variational EM, 212
- latent Dirichlet allocation, 2357–2359
 - definition, 2357
 - posterior intractability, 2358
- latent initialisation, 698
- latent interpolation, 582–590
 - linear, 582
 - norm dip, 583
 - smoothness, 583
 - spherical, 583
- latent optimisation
 - data assimilation, 4346
 - definition, 4346
 - objective, 4346
- latent reasoning hypothesis, 1065
 - exercise, 1070
- Latent Semantic Indexing, 2627
- latent semantic indexing, 2356–2357, 2366
- latent space
 - alignment, 4508–4510
 - as meeting ground, 4513
 - continuous vs. discrete, 578
 - desiderata, 588
 - dimensionality, 3396
 - exercise, 399
 - fragmentation
 - FedVAE, 4459
 - implications for diffusion, 586
 - interpretability, 3966
 - EEG, 3942
 - misalignment, 4509
 - normalisation, 588
 - PDE methods, 3391–3396
 - PDE solver, 3386
 - quality, 587
 - smoothness, 3823
 - structure, 398–399
 - temporal structure, 579–590
- latent space illumination, 2089
- latent space opacity, 3994–3995
 - consequence, 3995
- latent trajectory, 585–590
 - definition, 586
 - diameter, 581
- latent variable, 5
 - GMM, 146
 - segmentation style, 3495
- latent variable model, 59, 387–389
 - definition, 388
- latent video diffusion, 556–590
- latent-space diffusion, 534
- lateral adapter, 1198
- lateral connections, 1959
- latitude-longitude grid
 - pole singularity, 4266
- Law of Large Numbers, 40
- layer collapse, 1966
- layer importance, 1945
- layer normalisation, 96
 - fold-LN approximation, 4010
- layer partitioning
 - edge-cloud, 4822
- layer-based search, 1893
- layer-wise relevance propagation (LRP), 4129
 - Mamba, 4129
- LayerNAS, 1926–1928
 - algorithm, 1928
 - complexity, 1927
 - decomposition, 1927
 - dynamic programming, 1927, 1996
- LayerNorm, 3981
- lazy evaluation
 - proofs, 2998
- LDA, *see* latent Dirichlet allocation, 2366
- Le, Quoc, 1911
- lead-time embedding
 - MetNet, 4240

- leaderboard
 - epilepsy, 3923
- leaky ReLU, 2715
- Lean, 2934, 2945
 - lemma retrieval, 2566
 - proof assistant, 3080
- Lean (proof assistant), 3025
 - AlphaProof, 3112
- Lean 4, 2206–2209, 2945, 2964–2983
 - advanced topics, 2984–2999
 - advantages, 3005–3008
 - AI-friendliness, 3008
 - as programming language, 3007
 - axioms, 2974
 - common errors, 2981
 - compilation, 2982
 - compiler, 3007
 - debugging, 2981
 - examples, 2976–2983
 - first theorem, 2965
 - formalization example, 3117
 - lemma search, 2980
 - logic, 2968
 - performance, 2982, 2997–2998
 - tactics, 2969–2971
 - type hierarchy, 2966
 - typeclasses, 2973–2974
 - Vector type, 2984
- Lean Copilot, 3084–3086
 - architecture, 3084
 - latency analysis, 3095
 - modes of operation, 3085
 - user experience, 3086
- Lean REPL, 3008
- Lean tactics, 3116
- Lean type checker
 - verification, 3116
- LeanAgent, 3086–3088
 - results, 3088
- LeanCopilot, 2207
- LeanDojo, 2957, 3013, 3081–3084
 - benchmark, 3081
 - benchmark splits, 3082
 - data extraction, 3081
- LeanProgress, 3088–3089
 - architecture, 3089
 - results, 3089
- Leanstral, 3152–3153
- learned embeddings
 - cell tracking, 3654
- learned parameterisations, 4280
- learned sparse retrieval, 2370
- learning
 - graphical model, 137
- learning rate, 94
 - and forgetting, 1206–1208
 - audio diffusion, 858
 - elastic scaling, 5153
 - forgetting bound, 1206
 - infinite cosine, 1207
 - rewarming, 1205
 - scaling, 5026
 - video training, 659
- learning rate scaling
 - batch size, 5111
- learning rate schedule, 1522–1526
 - comparison, 1555
 - cosine annealing, 1523
 - cyclical, 1524
 - one-cycle, 1525
 - range test, 1524
 - warmup, 1522
 - gradient noise, 1523
 - WSD, 1525
- learning rate warmup, 5027
 - necessity, 5027
- learning to rank, 2362, 2585–2595, 2628
 - classification, 2586
 - listwise, 2587–2589
 - pairwise, 2586–2587

- pairwise complexity, [2595](#)
- pairwise-nDCG connection, [2594](#)
- pointwise, [2585–2586](#)
 - limitations, [2586](#)
- regression, [2585](#)
- Learning without Forgetting, [1190](#)
 - connection to alignment, [1190](#)
 - definition, [1190](#)
- least upper bound, *see* supremum
- LeCun, Y., [3962](#)
- Lee Sedol, [3114](#)
- left cosets, [2996](#)
- legal
 - data movement constraints, [4449](#)
- Legendre transform, *see* Fenchel
 - conjugate
- Lehman, Joel, [2081](#), [2283](#)
- Leibniz integral rule, [5229](#)
- Leibniz, Gottfried, [2938](#)
- Leibniz, Gottfried Wilhelm, [3016](#)
- Leith, Cecil
 - Monte Carlo forecasts, [4286](#)
- Leitner system, [1174](#)
- length generalisation, [611](#)
- length scalability, [2789](#)
- Lennox-Gastaut syndrome, [3777](#)
- Leray, Jean, [3429](#)
- lesion detection
 - small lesions, [3549](#)
- lesion localization
 - diffusion models, [3840–3847](#)
- lesional zone, [3775](#)
- Leung, C.T., [4233](#)
- Leutbecher, Martin, [4290](#)
- level set, [2777](#)
- level set methods
 - connection to snakes, [3742](#)
 - segmentation, [3487](#)
- level sets, [3527–3528](#)
 - evolution equation, [3527](#)
- level shifts, [2790](#)
- level- K reasoning
 - exercise, [1447](#)
- Level- K reasoning, [1292–1294](#)
 - convergence, [1293](#)
 - definition, [1293](#)
- level-set methods, [3521](#)
- LEVI, [2123–2126](#)
 - adaptive allocation, [2125](#), [2130](#)
 - stratification principle, [2124](#)
 - tier classification, [2125](#)
- Lewis, Patrick, [2629](#)
- lexical diversity
 - metrics, [4763](#)
- Li et al. 2024, [3855](#)
- Li, Tian
 - federated learning theory, [4450](#)
- library incompatibility, [3009](#)
- Library of Alexandria, [2338](#), [2625](#)
- Library of Babel, [2334–2335](#)
- library search
 - importance, [2997](#)
- LibriSpeech, [22](#)
- LibriTTS, [4754](#)
- LIDC-IDRI, [3551](#)
 - blinded read, [3551](#)
 - unblinded read, [3551](#)
- LIDC-IDRI dataset, [3510](#), [3551–3571](#), [3609](#)
 - example, [3515](#)
 - probabilistic segmentation, [3563](#)
- lifelong learning
 - biological, [1853](#)
 - theorem proving, [3086](#)
- lifelong learning system, [1230](#)
- Lifshitz–Slyozov coarsening, [3359](#)
- likelihood ratio test
 - algorithm, [1623](#)
 - computational cost, [1624](#)
 - diffusion models, [1622–1624](#)

- optimality, 1622
- reference model, 1623
- LIME, 3964, 3966, 3983–3984
 - comparison figure, 3986
 - instability, 3984
 - locality, 3984
 - stochastic output exercise, 4001
 - surrogate model, 3984
- limit point, 5196
- linarith tactic, 2995
- lineage tracking, 3659–3660
- lineage tree
 - definition, 3653
- linear algebra
 - in Lean, 2979
- linear arithmetic
 - automation, 2988
 - theory, 3023
- linear map
 - in Lean, 2979
- linear model
 - limitations
 - meteorology, 4229
- linear noise schedule
 - vulnerability, 1631
- linear probe, 4136
 - deception, 4140
- linear program
 - dual, 5278
- linear representation hypothesis, 4040
- linear scaling rule, 5026–5153
 - derivation, 5026
 - exercise, 5030
 - learning rate, 5173
- linguistic minorities
 - language models, 4445
- linguistic variables
 - fuzzy, 4235
- lip reading
 - disambiguation, 4917
- lip synchronisation, 745
- lip synthesis, 4735–4745
 - audio encoding, 4738
 - dental realism, 4737
 - ethics, 4741–4742
 - evaluation, 4740
 - FLAME, 4736
 - VisualSpeaker, 4736
- Lip Vertex Error
 - limitations, 4740
- lip-reading
 - viseme basis, 4735
- lip-reading loss, 4739–4740
- Lipschitz constant, 1032
- Lipschitz constraint
 - comparison of methods, 299
 - enforcement, 295
 - gradient penalty, 3829
- Lipschitz continuity, 5208
 - composition, 304
 - gradient characterisation, 5208
 - latent space, 580
 - latent temporal, 580
 - neural networks, 580
- Lipschitz function, 82
- Lipschitz operator, 3224
- LiRA, 1623, 1721–1727
 - algorithm, 1722
 - computational cost, 1725–1727
 - cost estimation, 1726
 - definition, 1721
 - Gaussian assumption, 1723–1724
 - Gaussian form, 1721
 - number of reference models, 1724
 - optimality, 1723
 - reference models, 1724–1725
- ListMLE, 2588
- ListNet, 2588
 - implementation, 2595
- lists

- inductive definition, 2986
- listwise reranking
 - definition, 2491
 - permutation generation, 2491
- literals, 3017
- literature mining
 - for mathematics, 3178
- Liu et al. attack, 1652
- Liu, Tennison, 2775, 2834
- LIVECell dataset, 3644
- liver segmentation, 3480
- LLaDA, 4539–4549
- LLaMA
 - 70B memory budget, 5077
 - context extension, 1084
 - federated fine-tuning, 4460
 - fine-tuning for tables, 2807
 - memory calculation, 5067
 - training time estimate, 4940
- Llama 3
 - 405B parameters, 5151
 - context parallelism, 5138
 - failure analysis, 5151
 - fault tolerance, 5145
- LLaMA-3
 - 15 trillion tokens, 4926
- LLaMA-70B
 - all-reduce, 5039
 - memory analysis, 5048
 - tensor parallelism, 5048
- LLaVA
 - membership inference, 1776
- LLaVA-Med, 3672–3673
 - concept alignment, 3672
 - instruction tuning, 3672
 - limitations, 3673
- LLM
 - 13B parameter
 - communication hiding, 4993
 - as universal representation, 2246
 - caregiver coaching, 4801
 - clinical NLP pipeline, 3900
 - context injection, 3916
 - disability representation, 4915
 - federated, 4460
 - fixed architecture limitations, 1889
 - for mathematics, 2957–2959
 - hallucination
 - clinical risk, 3902–3903
 - patient safety, 3902
 - mathematical reasoning, 2954
 - open-ended output, 4701
 - prior over programs, 2194
 - prompt engineering
 - clinical, 3900
 - task planning
 - robotics, 4867
 - training memory, 5077
- LLM agent, 1277–1284
 - agent loop, 1279
 - and continual learning, 1283
 - code-writing example, 1284
 - definition, 1278
 - formal definition, 1278
 - memory, 1281
- LLM forgetting, 1210–1211
 - case study, 1210
- LLM hallucination
 - in evolutionary search, 2107
- LLM synthetic data
 - on-device fine-tuning, 4627–4628
- LLM training
 - gradient synchronization, 4977
- LLM-as-a-judge
 - accessibility evaluation, 4878, 4890
- LLM-based generation
 - exercise, 2921
- LLM-evolutionary loop, 2099–2101
 - advantages, 2101
- LLM-guided evolution, 2301

- algorithm, 2100
- challenge exercise, 2296
- challenges, 2128–2131
- design trade-offs, 2126
- exercise, 2290
- information theory, 2297
- information-theoretic limits, 2275
- open-source, 2118–2131
- paradigm shift, 2097
- scaling, 2132
- system comparison, 2126
- LLM-guided mutation, 2174
- LLM-inspired PDE methods, 3259
- LLM2Vec, 2392
- LLMs
 - accessibility connection, 4703
 - distributed training requirements, 4932
 - theorem proving, 3080–3095
- LO-SDA, 4346–4347
- LOA, *see* Language-Oriented Audio
- load balancing, 1380–1390
 - balls into bins, 1382
 - exercise, 1455
 - hybrid parallelism, 5141
 - MoE, 2818
 - NAS, 1948
- load balancing loss, 622
 - exercise, 627
- load forecasting
 - neural network, 4233
- loan approval
 - bias, 2856
- local polytope, 253
- local binary patterns, 3529
- local encoder
 - personalised, 4511
- local feature descriptor, 2397
- local features, 2397–2400
 - comparison, 2399
- local optima, 2025
- local SGD, 5007–5009
 - communication savings, 5017
 - convergence, 5008, 5172
 - diffusion, 5095
 - definition, 5007
 - diffusion training, 5091–5092
 - diffusion training example, 5016
 - practical considerations, 5009
- local surrogates, 3983–3984
- local variable, 214
- Local-SGD, 5120
 - convergence, 5120
- locality, 4064
- locality bias, 2730
- locality-sensitive hashing, 2375
 - collision probability, 2376
 - definition, 2376
 - exercises, 2384
- localized watermarking, 932–934
 - definition, 932
- locally stationary process, 3796
- location-guided diffusion, 3841–3842
- location-invariant convolution
 - limitation, 4240
- location-variant connection
 - spatiotemporal, 4239
- locked algorithm
 - FDA, 3940
- locked-in syndrome
 - AAC, 4865
- log-cluster metric, 2841
- log-likelihood, 20
 - evaluation, 62
- log-normal distribution, 43
- log-partition function, 203
 - properties, 204
- log-space
 - flow matching, 2868
- log-sum-exp

- conjugate, 5277
- Logemann, George, 3018
- Logic Theorist, 2939
- logical connectives
 - as inductive types, 2987
 - in Lean, 2968
 - type constructors, 2947
- logical constraints, 2703, 2820
- logical kernel, 3033
- logistic regression
 - precipitation, 4225
 - transparency, 3970
- logistics co-pilot, 3914
- logit, 205
- logit attribution
 - decomposition theorem, 4009
- logit lens, 4014
 - definition, 4014
 - exercises, 4019
 - stage detection
 - exercise, 4085
 - three stages, 4083
 - VLM, 4082–4085
 - VLM failure, 4077
- long context
 - memory cost, 5071
 - parallelism, 5056
- long video generation, 697–706
 - autoregressive, 698
 - consistency problem, 752
 - fundamental tension, 697
 - hierarchical, 700
 - hybrid, 704
 - hybrid algorithm, 705
 - noise schedule, 704
 - open problem, 752
- long-context embeddings, 2389
- long-context inference
 - memory, 4932
- long-context training
 - communication analysis, 5176
- long-form generation
 - audio, 881–888
 - audio-video analogy, 947
 - challenge, 881
 - computational cost, 882
- long-form music generation, 946–947
 - formal statement, 946
- long-horizon accuracy, 722
- long-range dependencies, 2788
 - PDEs, 3234
- long-term depression, 1849
- long-term potentiation, 1849
- LongAgent, 1327
 - conflict resolution, 1327, 1450
- Longformer, 1087
- longitudinal evaluation, 4905
- longitudinal monitoring
 - segmentation, 3480
- lookup-free quantisation, 577–579
 - capacity, 578
 - definition, 577
- Loos, Sarah, 3036
- LoRA, 1126–1129, 4552
 - adapter interference, 696
 - approximation bound, 690
 - approximation quality, 690
 - composed rank, 693
 - composition, 693, 1222
 - continual learning, 1201
 - definition, 689, 1127, 4552
 - diffusion models, 1222
 - exercise, 1161
 - federated, 4553, 4585
 - federated fine-tuning, 4553
 - federated LLM, 4460
 - fine-tuning algorithm, 695
 - initialisation, 690
 - layer selection, 695
 - learning rate, 694

- LLaMA example, 1128
- medical segmentation, 3716
- membership inference, 1791
- membership mitigation, 1791
- membership mitigation proof, 1791
- merging, 1129
- motion adaptation, 692
- on-device fine-tuning, 4628
- parameter count, 690, 1127
- QLoRA, 1127
- quantised, 694
- rank analysis, 696, 1165
- rank selection, 694, 761, 4557
- storage for CL, 1201
- temporal vs spatial, 696
- variants, 4554
- video diffusion, 689–696
- video DiT, 691
- vision-language CL, 1225
- weight update, 689
- Lorenz attractor, 4187
- Lorenz, Edward, 4183–4188, 4192
- loss balancing, 792, 857–862
 - gradient equalisation, 857
 - gradient-based, 857
- loss distribution
 - separation, 1618
 - variance, 1618
- loss function, 94
 - normalised MSE, 4269
- loss scaling, 1529, 1556, 5074, 5105
 - distributed training, 5038
 - dynamic, 5074
 - exercise, 5114
 - gradient communication, 5037
 - mechanism, 5107
- loss spike
 - rollback, 5152
- loss weighting
 - diffusion parameterisation, 851
- loss-based attacks, 1616–1625
 - algorithm, 1617
 - basic, 1617
 - comparison, 1624
 - membership score, 1617
 - practical considerations, 1619
 - statistical separation, 1618
- loss-based selection, 1193
- lost-in-the-middle effect, 4106
 - exercises, 4113
- Lottery Ticket Hypothesis, 1540
 - sparsity, 1557
- Loveland, Donald, 3018
- low learning rate
 - fine-tuning, 3888
- low-dimensional manifold hypothesis, 4444
- low-rank adaptation, 689, *see* LoRA
- low-rank approximation
 - gradients, 4478
- Lowe, David, 2397
- Lowry, Dale, 4222
 - 1972 paper, 4231
- ℓ^p metric, 5195
- LPIPS, 561
- LSH, *see* locality-sensitive hashing
 - AND-OR composition, 2376
- LSI, *see* latent semantic indexing, 2366, 2627
- LSTM, 100
 - bidirectional, 3885
 - convolutional generalisation, 4239
 - exercise, 1066
 - flattening problem, 4239
 - gradient highway, 1034
 - TimeGAN generator, 3821
 - vs. Transformer, 100
- LSTM-GRU hybrid
 - seizure prediction, 3884–3885
- Ltac, 3000

- Luhn, Hans Peter, 2626
- lung nodule
 annotation variability, 3515, 3551
 detection, 3479
 segmentation, 3482
 segmentation example, 3482
- lung nodule segmentation, 3609
- lung segmentation, 3510–3511
- Luo et al. 2025, 4078
- Lustre
 filesystem, 5147
- Lyapunov exponent, 4187–4202
 definition, 4187
 exercise, 4191
- lyrics alignment, 901
- lyrics-conditioned generation, 900–902
- M-projection, 73, 171
- machine learning
 accessibility generations, 4701
 early meteorology, 4232–4243
 environmental science textbook, 4236
 Navier–Stokes regularity, 3430
- machine learning utility, 2705, 2841–2843
- machine learning weather prediction, 4189–4202
- machine unlearning, 1796–4455
 arms race, 1803
 completeness, 1797
 completeness exercise, 1806
 definition, 1797
 differential privacy, 1805
 efficiency, 1804
 exact vs approximate, 1797
 exercise, 1806
 failure against strong attacks, 1802
 false negatives, 1800
 federated
 large scale, 4664
 open problem, 4664
 generative models, 4603–4609
 gradient ascent, 1801
 history, 1807
 legal considerations, 1805
 MIA duality, 1798
 MIA duality proof, 1798
 MIA duality theorem, 1798
 MIA verification, 1798
 multi-attack verification, 1799
 right to be forgotten, 1797
 sequential requests, 1804
 strong attacks, 1802
 TOFU benchmark, 1800
 verification algorithm, 1799
 verification gap, 1802
 verification standard, 1802
- machine-assisted mathematics, 3031
- Macready, William, 2282
- macro-architecture, 1893
- macros
 Lean 4, 2992
- MADA
 majority vote, 4601
- MADA aggregation, 4599
 description, 4600
 exercise, 4608
- MADE, 436–437
 definition, 436
 order agnosticism, 437
- MAF, 239
- MAF (Masked Autoregressive Flow), 457
- MAF–IAF duality, 457
- magnetic resonance imaging
 segmentation, 3481
 volumetric, 3506
- magnetoencephalography, 3776
 spike mapping, 3780

- magnitude spectrogram, 773
- MAGVIT-v2, 578, 729
- Mahalanobis distance, 39
- majority vote
 - limitations, 3554
 - silencing minority opinions, 3556
- majority voting, 1367
- majority-vote sampling, 3059
- MALLM, 1350–1354
 - configuration, 1450
 - configuration space, 1350
 - navigation, 1351
- Mamba, 1104, 1962, 4124–4133
 - architecture, 4125
 - associative recall mechanism, 4127
 - associative treecall performance, 4127
 - hierarchical retrieval limitation, 4128
 - interpretability, 4170
 - selective gates emulate attention, 4128
 - selective state space, 4125–4126
- Mamba SSM
 - medical imaging, 3904–3907
 - selective state space, 3904
 - selectivity
 - exercise, 3909
- MambaLRP, 4129
 - comparison with attention, 4130
 - exercise, 4132
 - implementation, 4130
 - relevance propagation, 4129
- MAML
 - EEG personalisation, 3932
- mammography
 - segmentation, 3477
- mandatory watermarking, 939
- Manhattan metric, *see* taxicab metric
- manifold
 - condition set, 2776–2787
- manifold guidance, 2775–2787
 - algorithm, 2779
- manifold hypothesis, 56, 532
 - and divergences, 83
 - video, 532
- manifold learning
 - exercise, 400
- manifold projection, 2777
- Manshausen 2025, 4343–4344
- manual segmentation
 - history, 3487
- many-electron problem, 3364–3371
- many-objective optimization, 1987
- MAP (mean average precision), 2359
- MAP estimation, 46, 329
 - exercises, 338
 - Gaussian denoising, 329
- MAP inference, 135
- MAP-Elites, 2079–2094, 2109, 2284
 - algorithm, 2083–2088
 - applied to proofs, 3109
 - archive, 2084
 - coverage rate, 2091
 - curse of dimensionality, 2092
 - exercise, 2290
 - implementation exercise, 2087
 - in AlphaEvolve, 2109
 - in DRQ, 2157
 - monotone improvement, 2086
 - properties, 2085
 - robotics example, 2086
 - scaling, 2090
 - selection pressure, 2092
 - within-generation, 2158
- Maple, 3031
- Marchuk
 - DiT backbone, 4341
 - efficiency, 4342
 - model, 4341–4343

- predictability exercise, [4349](#)
 - S2S results, [4342](#)
 - VAE compression, [4341](#)
- Mardani et al. (2025), [4353](#)
- marginal CRPS
 - vs. joint coherence, [4329](#)
- marginal distribution, [36](#)
 - diffusion, [485](#)
- marginal distributions
 - evaluation, [2834](#)
 - insufficiency of, [2835](#)
 - vs. joint, [2701](#)
- marginal fidelity, [2702](#), [2837–2838](#)
- marginal inference, [135](#)
- marginal likelihood, [166](#)
 - intractability, [388](#)
- marginal polytope, [252](#)
- marginal training
 - joint calibration, [4337](#)
- marginal trap, [2702](#)
- marginalisation, [37](#)
 - over retrieved documents, [2522](#)
- MARINE
 - algorithm, [4838](#)
 - contrastive decoding, [4836](#)
 - factuality reranking, [4837](#)
 - hallucination mitigation, [4835](#)
- marine layer
 - coastal jets, [4230](#)
- market mechanisms, [1276](#)
- Markov blanket, [131](#)
 - exercise, [142](#)
- Markov chain, [477](#)
 - global balance, [2869](#)
- Markov chain Monte Carlo, [137](#)
- Markov Decision Process
 - theorem proving, [2961](#)
- Markov decision process
 - denoising, [1154](#)
 - theorem proving, [3096](#)
- Markov property
 - pairwise, [132](#)
- Markov random field, [131](#)
 - definition, [131](#)
 - exercise, [143](#)
- Markov's inequality, [41](#)
- Marks, S., [3976](#)
- Marmousi model, [3405](#)
- Maron, M. E., [2626](#)
- Marques-Silva, Joao, [3018](#)
- Marr, D., [3972](#)
- Martin-Löf type theory, [3002](#)
- MaSh, [3025](#)
- mask averaging
 - failure, [3554](#)
- mask-to-image synthesis, [3581](#)
- masked autoencoder
 - audio, [837](#)
- masked convolution, [439](#)
 - definition, [439](#)
- masked diffusion models
 - attention, [4104–4113](#)
 - attention floating, [4107](#)
 - exercises, [4113](#)
 - bidirectional attention, [4105–4113](#)
 - bipartite mechanism, [4107–4113](#)
 - deep layers, [4109](#)
 - knowledge-intensive tasks, [4110–4113](#)
 - shallow layers, [4107](#)
- masked language modelling
 - vision-language, [3679](#)
 - vs. discrete diffusion, [4540](#)
- masked language models
 - pseudo-likelihood, [1750](#)
- masked self-attention, [103](#)
- mass conservation, [3320](#)
 - primitive equations, [4207](#)
- mass editing, [4065](#)
- materials science, [15](#)

- case study, [2892](#)
- crystal generation, [2892](#)
- federated, [4666](#)
- generative models, [2891–2896](#)
- MATH benchmark, [3060](#)
- Math Inc., [3153](#)
- Mathematica, [3031](#)
- mathematical AI
 - competition, [3150–3156](#)
 - four eras, [2940](#)
 - landscape, [2938](#)
 - taxonomy, [2953–2964](#)
 - timeline, [3184](#)
- mathematical benchmarks, [3059–3061](#), [3156](#)
 - exercise, [3063](#)
- mathematical capabilities
 - frontier models, [3052](#)
- mathematical competitions, [3033](#)
- Mathematical Components, [3000](#)
- mathematical conventions, [2936](#)
- mathematical creativity
 - AI limitations, [3141](#)
- mathematical discovery
 - AlphaEvolve, [3131](#)
 - via self-play, [3108](#)
- mathematical errors
 - historical, [2941](#)
 - LLMs, [2958](#), [3047](#)
- mathematical formalisation
 - at scale, [3133](#)
- mathematical formalization, [3064–3080](#)
- mathematical knowledge
 - breadth, [3039](#)
- mathematical language
 - ambiguity, [3065](#)
- mathematical language models, [3059](#)
 - landscape, [3059](#)
- mathematical libraries
 - dependence, [3065](#)
- mathematical olympiad
 - AI performance, [2958](#)
- mathematical proofs
 - decomposition, [3049](#)
- mathematical reasoning, [2933–2942](#)
 - chain-of-thought, [3048–3050](#)
 - LLMs, [3046–3064](#)
 - OpenAI, [3146](#)
 - pipeline, [2937](#)
 - pre-LLM gaps, [3039–3040](#)
 - pre-LLM landscape, [3031–3046](#)
 - reasoning models, [3050–3064](#)
- mathematical understanding, [2933](#)
 - lack of, [3028](#)
- mathematics
 - appeal for generative models, [2933–2942](#)
 - for AI, [2933–2942](#)
 - social aspects, [2941](#)
 - vs. code generation, [2935](#)
- Mathlib, [2945](#), [2951–2953](#), [2980](#), [3025](#), [3081](#)
 - as training data, [2952](#)
 - Bornology.IsBounded, [2994](#)
 - build time, [2998](#)
 - comparison with other libraries, [3011](#)
 - coverage, [2953](#)
 - Filter.Tendsto, [2994](#)
 - matrix definitions, [2985](#)
 - pre-training, [3118](#)
 - scale, [2951](#)
 - significance, [3007](#)
 - SimpleGraph, [2987](#)
 - Subgroup.card_sub-group_dvd_card, [2997](#)
 - training data, [3117](#)
- matrices
 - dependently typed, [2985](#)
- matrix determinant lemma, [238](#)

- matrix factorization, 2576–2578
 - as two-tower model, 2584
- matrix multiplication
 - 48-multiplication algorithm, 2114
 - AlphaEvolve, 2113, 3132
- matrix multiply
 - GPU efficiency, 4934
- Matryoshka embedding, 2462
 - rate-distortion, 2474
 - retrieval algorithm, 2465
 - speedup, 2466
- Matryoshka embeddings, 2631
- Matryoshka Representation Learning, 2388–2389
 - exercises, 2396
- Matryoshka representation learning, 2461
- maximal update parameterisation, *see* initialization, muP
- maximum a posteriori, 329
- maximum activation examples, 3972
- maximum inner product search, 2382, 2576
- maximum likelihood
 - GMM, 149
- maximum likelihood estimation, 45
 - for generation, 55
 - Gaussian, 45
- Maximum Mean Discrepancy
 - PRISM, 4602
- maximum mean discrepancy, 2840
 - EEG, 3821
 - exercise, 2921
- maximum temperature
 - MOS, 4224
 - MOS example, 4226
- MaxRenyi-K%, 1773–1780
 - algorithm, 1776
 - definition, 1774
 - LLaVA experiment, 1776
 - optimality, 1775
 - optimality proof, 1775
- MaxSim, 2368
- Maxwell’s equations, 3345–3352
 - differential form, 3346
 - Green’s function, 3352
 - inverse design, 3347
 - physics-guided diffusion, 3349
 - TE mode, 3350
- maze navigation
 - video models, 4116
- MC-Dropout, 257
- McClelland, J. L., 3962
- McCulloch, W. S., 3969
- McMahan, H. Brendan, 4444
 - FedAvg, 4450
- MCMC, 167
 - connection to GFlowNets, 2871
 - underdamped Langevin, 1680
- MCTS
 - AlphaZero, 3115
 - backpropagation, 3042
 - computational cost, 3102
 - evaluation, 3042
 - expansion, 3042
 - selection, 3042
 - theorem proving, 3042
- MCUNet, 4824
- MD Anderson
 - 4D cardiac CT, 4614–4616
- MD-GAN, 4459
- MDP
 - theorem proving formulation, 3110
- MEAD dataset, 4740
- mean ablation, 3974–3981
 - definition, 3975
 - vs zero ablation, 3980
- mean ablation patching, 4119
- mean absolute percentage error
 - MAPE, 4234

- mean average precision, 2605
- mean curvature, 3527
- mean curvature flow, 3358
- mean diffusivity, 3894
- mean field
 - as inner approximation, 252
 - optimal factor, 262
 - vs structured VI, 267
- mean field games
 - boundary conditions
 - exercise, 3879
 - brain dynamics, 3872
- Mean Opinion Score, 920–922
 - definition, 920
 - example comparison, 922
 - statistical testing, 921
- mean opinion score, 4754, 4913
- mean parameter, 205
- mean parameters, 250
- mean pooling
 - information loss, 2419
- mean reciprocal rank, 2604
- mean regression, 3265
- mean squared error
 - variance suppression, 3228
 - weather, 4194
- mean state
 - fine-scale, 4353
- mean time between failures, 5145
 - checkpointing, 5177
- mean-field
 - exponential family updates, 206
- mean-field approximation, 183–190
 - Bayesian linear regression, 187
 - bivariate Gaussian, 186
 - definition, 183
 - exercise, 200
 - Ising model, 185
 - optimal factors, 184
 - segmentation, 3494
- mean-squared error
 - MOS screening, 4223
- meaning
 - contextual, 2335
- mechanical reasoning, 2937–2942, 3015–3031
 - history, 3016–3017
- mechanism design, 1359
- mechanistic interpretability, 3971–3981, 4117, 4133
 - alignment program, 4137
 - circuit discovery, 4173
 - completeness, 3980
 - definition, 3971
 - formal definition, 3972
 - future directions, 4152–4164
 - gap figure, 3998
 - paradigm shift, 3996
 - programme, 4035
 - sparse autoencoders, 4046–4058
 - superposition, 4038–4046
 - taxonomy, 3977
 - transformer circuits, 4002–4012
 - vs feature attribution, 3996
- mechanistic oracle, 4145–4146
- Med-SA, *see* Medical SAM Adapter
- median heuristic, 248
- medical data
 - correlations, 2700
 - evaluation example, 2835
- medical image segmentation, 3477–3500
 - ambiguity, 3549–3557
 - annotation bottleneck, 3483
 - automation, 3488
 - bridge metaphor, 3480
 - chapter roadmap, 3497
 - chapter structure, 3497
 - classical methods, 3523–3531
 - limitations, 3530
 - clinical applications, 3479

- clinical importance, 3477
- clinical translation, 3715–3727
- computational cost, 3716
- connection to federated learning, 4463
- deep learning, 3531–3548
- definition, 3478
- example tasks, 3482
- fairness, 3724
- formal definition, 3478
- foundation model synergy, 3694–3715
- generative approach, 3489
- historical development, 3487
- history, 3487
- interpretability, 3721
- motivation, 3477
- open problems, 3727–3739
- probabilistic, 3557–3571
- taxonomy, 3501–3523
- vision-language models, 3666–3680
- medical images
 - data extraction
 - federated, 4597–4598
- medical imaging
 - data volume, 3484
 - distributed vs federated, 5122
 - federated synthesis, 4609–4621
 - modalities, 3481
- medical model of disability, 4691
- medical records
 - serialization, 2806
- medical retrieval
 - case-based, 2479
- Medical SAM 2, 3716
- Medical SAM Adapter, 3636–3637
 - placement, 3636
 - vs full fine-tuning, 3637
- medical time series, 2697
- MedicalNarratives, 3687
- mouse traces, 3688
- mediocre stable state, 2073–2075
 - definition, 2073
 - detection difficulty, 2075
 - escape time, 2074
- mediocre stable states, 2162
- medium loop
 - connectivity monitor, 3916
- MedSAM, 3493, 3519, 3631–3632
 - cross-modality, 3632
 - dataset, 3631
 - training, 3631
- MedSegDiff, 3600
 - dynamic conditional coding, 3754
 - FF-Parser, 3601
 - optic cup segmentation, 3601
- MedSegDiff-V2, 3603–3604
 - anchor condition, 3603
 - computational complexity, 3604
 - transformer backbone, 3603
- Medusa decoding, 1545
- MedVQA
 - benchmark limitations, 3908
- MedVQA-TREE, 3907–3908
 - accuracy, 3908
 - retrieval, 3908
- Megatron-LM, 4927, 5043–5050, 5135–5144
 - communication cost, 5047
 - distributed auto-diff, 5103
 - GPT-3 recipe, 5142
 - tensor parallel overlap, 4997
 - tensor parallelism, 4931
- Megill, Norman, 3004
- mel filterbank, 776
 - information loss, 777
 - properties, 800
- mel scale, 775–802
 - asymptotics, 775
 - definition, 775

- origin, 775
- mel spectrogram
 - frame count, 949
 - frequency resolution, 828
 - generation, 823
- mel-frequency spectrogram
 - EEG, 3799
- mel-spectrogram, 775–802, 4751
 - definition, 776
 - voice cloning, 4751
- melanoma detection
 - protanopia, 4782
- MELD project, 3859
- melodic contour, 938
- melody conditioning
 - flow matching, 869
- MelodyFlow, 869–873
- membership advantage, 1585, 2844
 - AUC bound, 1586
 - differential privacy bound, 1597
 - mutual information bound, 1589
- membership classifier, 4596
- membership inference, 1571–1842, 2705, 2844–2846
 - adaptive attacks, 1686, 1821
 - advantage metric, 1585
 - alignment data, 1762–1771
 - architecture choice, 1972
 - architecture-dependent bound, 1972
 - arms race, 1686–1692
 - as coevolution, 2253
 - attack modality comparison, 1759
 - attack surface, 1576
 - attention-based attacks, 1754–1762
 - AttenMIA, 1756
 - AUC-ROC, 1585
 - base rate fallacy, 1607
 - benchmark results, 1633
 - benchmarks, 1697–1700
 - standardisation, 1698
 - biometric data, 1573
 - black-box attacks, 1659–1670
 - chapter roadmap, 1576
 - chapter synthesis, 1833
 - chest X-ray example, 4598
 - comparison table, 1625
 - computational cost, 1624, 1699
 - connection to federated learning, 4449
 - continuous, 1830
 - copyright, 1573
 - data processing inequality, 1590
 - defense design exercise, 1702
 - defenses
 - composition, 1819
 - confidence masking, 1817
 - data augmentation, 1684
 - differential privacy, 1670–1678
 - DuoLearn, 1815
 - evaluation, 1820
 - output perturbation, 1817
 - regularization, 1679–1686
 - token-level, 1814–1821
 - definition, 1579
 - differential privacy, 1595
 - diffusion language models, 1692–1703
 - diffusion models, 1575, 1608–1616
 - summary, 1700
 - discriminative vs. generative, 1575
 - dual role, 1831
 - EEG, 3941
 - ELBO-based, 1619
 - ethical considerations, 1668
 - evaluation, 1697
 - pitfalls, 1699
 - evaluation methodology, 1604
 - evaluation metrics, 1604
 - evaluation pitfalls, 1606

- evaluation protocol, 1625
- evolution from pretraining to alignment, 1769
- exercises, 1835–1842
 - diffusion loss, 1836
 - DP composition, 1838
 - DP scaling law, 1842
 - DP-SGD, 1835
 - fine-tuning, 1841
 - Fisher information, 1837
 - frequency domain, 1837
 - HOLD++, 1838
 - LiRA, 1839
 - Min-K%, 1839
 - Neyman-Pearson, 1835
 - perplexity threshold, 1838
 - SaMIA, 1840
 - SecMI, 1836
 - shadow models, 1836
 - SPV-MIA, 1840
 - token consistency, 1841
 - unlearning verification, 1841
- false sense of security, 1574
- federated GAN, 4596
- federated gradient attacks, 4464
- federated learning, 4590
- financial data, 1573
- fine-tuning, 1789–1796
 - 10-shot example, 1793
 - amplification, 1789
 - amplification proof, 1789
 - amplification theorem, 1789
 - epoch effect, 1796
 - exercise, 1806
 - few-shot, 1793
 - few-shot exercise, 1807
 - few-shot vulnerability, 1793
 - Fisher information, 1793
 - LoRA, 1791
 - LoRA comparison, 1792
 - magnifying glass, 1790
 - medical example, 1791
 - PEFT warning, 1792
 - practical amplification, 1790
 - practical focus, 1795
 - pre-training comparison, 1794
 - privacy-utility trade-off, 1796
 - QLoRA, 1792
- formal definition, 1579
- frequency-domain attacks, 1649–1659
- fundamental limits, 1593
- fundamental question, 1572
- fundamental tension, 1701
- future directions, 1692
- future outlook, 1703
- generalisation gap, 1591
- generative model landscape, 1575
- generative models, 3765
- healthcare, 1572
- historical context, 1636, 1669
- history, 1577
- hypothesis test, 1580
- hypothesis testing, 1578
- impossibility results, 1594
- indirect, 1823
 - via prompt injection, 1823
- influence generalization, 1830
- information hierarchy, 1760
- information theory, 1587
- information-theoretic summary, 1595
- interpreting results, 1606
- label-only attacks, 1754–1762
 - PETAL, 1754
- large language models, 1575, 1704–1713
- likelihood ratio, 1622
- likelihood ratio attack, 1721–1727
- loss distributions, 1716–1718

- loss separation, 1716
- loss-based attacks, 1616–1625, 1713–1720
- machine unlearning, 1796
- membership oracle, 1580
- memorisation differences, 1576
- method comparison, 1833
- Min-K% Prob, 1727–1738
- minimax formulation, 1688–1692
- mitigation exercise, 4599
- motivation, 1572
- mutual information bounds, 1588
- Neyman–Pearson framework, 1581
- open problems, 1690, 1828–1834
 - benchmarks, 1832
 - beyond binary, 1830
 - fundamental tradeoffs, 1829
 - influence estimation, 1831
 - inherent privacy, 1691
 - multimodal, 1832
 - Pareto frontier, 1830
- optimal threshold, 1719
- overfitting, 1591
- overfitting as root cause, 1592
- per-token analysis, 1714–1716
- perfect generalization, 1589
- perplexity algorithm, 1719
- perplexity attacks, 1713–1720
- perplexity score, 1714
- practical approximation, 1582
- pre-training
 - dilution, 1794
- preference data, 1762–1771
- privacy challenge, 1571
- reference-based attacks, 1720–1727
- regulatory implications, 1831
- retrieval systems, 2612
- retrieval-specific, 2613
- SAMA, 1693
- sampling-based attacks, 1746–1753
- self-influence, 1645
- self-prompted verification, 1738–1746
- shadow models, 1601
- statistical leakage, 4590
- step-wise error, 1626–1636
- synthetic EEG, 3951
- tabular data, 2898–2904
- threat models, 1582
- threshold selection, 1718–1720
- token-level, 1781–1788
 - aggregation, 1784
 - algorithm, 1784
 - algorithms, 1784
 - applications, 1787
 - consistency, 1783
 - consistency proof, 1783
 - copyright detection, 1787
 - data attribution, 1787
 - definition, 1781
 - email example, 1782
 - exercise, 1806
 - granularity, 1787
 - interpretability, 1784
 - mixed sequences, 1781
 - motivation, 1781
 - open problems, 1788
 - part-of-speech, 1785
 - score definition, 1782
 - statistical framework, 1782
 - visualisation, 1786
 - Wikipedia example, 1786
- token-level signal, 1715
- TPR at low FPR, 1606
- unified comparison, 1665
- unified framework, 1771
- unlearning verification, 1798, 4605
 - limits, 1802
- vision-language models, 1576, 1771–1780

- architecture, 1772
- attack surfaces, 1772
- cross-modal amplification, 1777
- cross-modal signals, 1777
- definition, 1772
- exercise, 1805
- guilt by association, 1778
- MaxRenyi, 1773
- medical data, 1779
- modality signals, 1778
- optimal strategy, 1780
- projection layer, 1777
- visual vs text tokens, 1778
- white-box attacks, 1636–1649
- white-box vs. black-box, 1591
- Yeom’s bound, 1592
- zero-knowledge attack, 1762
- membership inference attack
 - exercise, 4505
 - GAN, 4501
- membership inference attacks
 - connection to federated learning, 4464
- membership signals
 - generation-based, 1660
 - gradient-based, 1638
 - score-based, 1662
- Memex, 2349, 2519, 2626
- MEMIT, 4065
 - extension exercise, 4067
 - scaling, 4065
- memorisation, 1591
 - diffusion models, 1610
 - generative models, 4454, 4590
 - information-theoretic bound, 1613
 - LLM-based methods, 2859
 - mathematical characterisation, 1612
 - model capacity, 1614
 - rare examples in federated setting, 4454
 - score function, 1611
 - vs generalisation, 1866
- memorisation index, 1612
 - definition, 1613
 - exercises, 1616
- memorisation rate
 - small datasets, 4596
 - training data, 4593
- memorisation score, 4594
- memorization
 - approximate, 1709
 - architecture dependence, 1972
 - data deduplication, 1708–1709
 - duplication effect, 1708
 - eidetic, 1709
 - empirical rates, 1706
 - extraction attacks, 1710–1713
 - formal definitions, 1709–1710
 - large language models, 1704–1713
 - phase transition, 1706
 - scaling law, 1705
 - scaling laws, 1704–1708
 - tabular GANs, 2898
 - verbatim, 1709
- Memorization-Focused Guidance
 - MFG, 4594
- memorization-guided attack
 - definition, 4594
- memorization-guided data stealing, 4593
 - pipeline figure, 4595
- memory
 - agent duality, 1457
 - agent hierarchy, 1447
 - as reconstruction, 2333
 - as retrieval problem, 2337
 - model training breakdown, 5060
 - neuroscience, 2340–2348
 - reconstructive nature, 2342
 - retrieval duality, 2613

- vector embeddings, [2551–2554](#)
- video diffusion, [550](#)
- memory bandwidth, [1079](#)
 - bottleneck, [1079](#)
 - limiting factor, [4937](#)
- memory budget
 - video diffusion, [553](#)
- memory complexity, [1078](#)
- memory compression, [1099](#)
- memory consolidation, [1176](#), [1872](#), [2343–2344](#)
 - sleep, [1964](#)
- memory efficiency
 - ZeRO, [4932](#)
- memory footprint
 - exercise, [4932](#)
 - formula, [4930](#)
- memory hierarchy
 - as retrieval hierarchy, [2614](#)
- memory limitation
 - fuzzy time series, [4235](#)
- memory management, [1075–1114](#)
- memory optimisation
 - comparison table, [5077](#)
 - techniques table, [5078](#)
- memory per GPU
 - ZeRO stages, [5174](#)
- memory relevance
 - birthday effect, [2547](#)
- memory relevance score, [2547](#)
- memory strength, [1174](#)
- memory system, [1106](#)
- memory wall, [553](#), [4925](#)
 - definition, [4929](#)
 - model parallelism, [5042](#)
- memory–CL duality, [1241](#)
- memory–refinement decomposition, [1186](#)
- memory-architecture duality, [1969](#)
- memory-augmented architectures,

- [2549–2554](#)
- memory-augmented network,
 - [1090–1098](#)
- memory-augmented retrieval,
 - [2545–2554](#)
 - algorithm, [2552](#)
 - architecture, [2676](#)
 - formal definition, [2546](#)
- memory-bound
 - definition, [4936](#)
- memory-computation trade-off, [1107](#)
- memory-compute tradeoff, [553](#)
- memory-retrieval duality, [2614](#)
- memory-retrieval equivalence,
 - [2546–2554](#)
 - diagram, [2676](#)
- Meno dialogue, [2333](#)
- mental models
 - AI
 - BLV users, [4713](#)
 - study design exercise, [4716](#)
- mental rotation
 - video models, [4116](#)
- MePo filter, [3025](#)
- MERRA-2, [4217](#), [4218](#)
- mesh refinement levels, [4266](#)
- mesh-free methods
 - generative models, [3470](#)
- meshes
 - unstructured, [3235](#)
- mesial temporal lobe epilepsy, [3891](#)
- mesosphere, [4208](#)
- message passing, [134](#), [197](#), [2749–2752](#)
 - GraphCast, [4414](#)
 - layer, [2749](#)
 - multi-layer, [2750](#)
 - worked example, [2751](#)
- message passing interface, [4959–4971](#)
- message passing neural network
 - weather, [4268](#)

- Met Office
 - ensemble, [4196](#)
- Meta
 - GDPR fine, [4446](#)
- Meta Agent Search, [2183](#)
- Meta AI
 - infrastructure, [5151](#)
- meta-evolution, [2190–2203](#)
 - cross-domain transfer, [2202](#)
 - exercises, [2202](#)
 - future directions, [2200](#)
 - novel operator discovery, [2193](#)
 - objective discovery, [2271](#)
 - open-endedness conjecture, [2195](#)
 - optimizer convergence, [2192](#)
 - safety, [2202](#)
 - taxonomy, [2200](#)
 - theoretical analysis, [2193](#)
 - unification, [2200](#)
- meta-evolutionary framework, [2191](#)
- meta-learned personalised CVD
 - correction, [4783](#)
- meta-learning, [1203](#)
 - adaptation exercise, [4789](#)
 - architecture search, [1979](#)
 - evolutionary, [2025](#)
 - personalised CVD correction, [4782](#)
- meta-prompting, [2121–2122](#)
 - definition, [2121](#)
- metabolic bound, [1853](#)
- metabolic efficiency, [1856](#)
- metacognition, [4145](#)
- metamaterials, [3350](#)
 - generative models, [3350](#)
- Metamath, [3004](#), [3036](#)
 - and AI, [3004](#)
 - neural contributions, [3037](#)
- metaprogramming
 - custom tactics, [2991](#)
 - Lean 4, [2992](#)
 - levels, [2992](#)
- metasurface
 - diffusion-based design, [3347](#)
 - example, [3348](#)
 - inverse design, [3348](#)
- method comparison, [2858–2861](#)
- metis, [3027](#)
- MetNet, [4240–4241](#)
 - 12-hour limit, [4241](#)
 - exercise, [4243](#)
- MetNet-2, [4240–4241](#)
 - comparison, [4241](#)
 - output representation, [4240](#)
 - vs HRRR comparison, [4241](#)
- metric selection
 - decision framework, [3925](#)
- metric shopping, [2835](#)
- metric space, [5194](#)
- metric topology, [5217](#)
- metric-loss alignment, [3547](#)
- metrics zoo, [3923](#)
- MeZO
 - algorithm, [4559](#)
- MFCC, [4751](#)
 - computation, [4751](#)
- MFG
 - TFG exercise, [4598](#)
- MFG system
 - neural populations, [3872](#)
- MFU
 - analysis challenge, [4942](#)
 - GPT-3, [5142](#)
 - realistic estimate, [4941](#)
- MGDS
 - chest X-ray example, [4597](#)
 - defense gap, [4602](#)
- MGDS attack
 - invisibility, [4595](#)
- MGSM benchmark, [2186](#)
- micro-architecture, [1893](#)

- micro-batch
 - minimum count, [5141](#)
 - pipeline fill, [5140](#)
- micro-batches
 - trade-off, [5102](#)
- micro-to-macro
 - epilepsy, [3894–3895](#)
- microcontroller
 - AI inference, [4815](#)
- microscopy
 - cell segmentation, [3642–3650](#)
 - cell tracking, [3650–3666](#)
 - cross-modality tracking, [3764](#)
 - segmentation, [3509–3510](#)
 - synthetic data generation, [3660](#)
- Microsoft
 - and Lean, [3008](#)
- Microsoft Research, [2945](#)
 - Aurora, [4377](#)
- middle padding, [2808](#)
 - advantages, [2808](#)
- MIDI, [904–906](#)
 - event sequence, [905](#)
- Midjourney, [10](#)
- migration
 - elitist, [2122](#)
- migration policy
 - exercise, [2294](#)
- migration topology, [2127](#)
- Mikolov, Tomas, [2628](#)
- Millennium Prize Problems, [3427](#)
- millennium problems, [3731](#)
- Min-K% Prob, [1727–1738](#)
 - algorithm, [1729](#)
 - consistency, [1731–1733](#)
 - consistency theorem, [1731](#)
 - definition, [1728](#)
 - finite sample bound, [1732](#)
 - key idea, [1728–1729](#)
 - optimal k , [1731](#)
 - order statistics, [1729–1731](#)
 - practical implementation, [1733–1738](#)
 - Pythia evaluation, [1733](#)
 - separation, [1730](#)
- min-max game, [3572](#)
 - GAN, [3572](#)
- Minerva, [2957](#), [3059](#)
 - training data, [3059](#)
- mini-batch
 - sharding, [5021](#)
- MiniF2F, [3061](#)
- miniF2F benchmark, [3137](#)
- minimality
 - interpretability, [3997](#)
- minimax
 - and coevolution, [2061](#)
 - equilibrium theorem, [1688](#)
 - exercise, [1702](#)
 - membership inference, [1688](#)
 - numerical examples, [1689](#)
- minimax game, [274](#)
- minimax inequality, [5268](#), [5279](#)
- minimax optimality, [3383](#)
- minimax optimization, [309–326](#)
 - method comparison, [323](#)
- minimax theorem, [311](#), [5264](#)
 - Von Neumann, [5269](#)
 - von Neumann, [311](#)
- minimising communication, [5005](#)
- minimum bandwidth problem, [2744](#)
- minimum description length, [1864](#)
 - polynomial example, [1868](#)
- minimum temperature
 - ANN forecasting, [4235](#)
 - boundary layer, [4236](#)
 - MOS, [4224](#)
- minimum vertex cover, [3980](#)
- minimum-norm solution, [2779](#)
- MiniSat, [3018](#)

-
- Minkowski sum, 5245
 - minority class manifold, 3814
 - MIONet, 3217
 - MIPS, 2521
 - MIR, *see* music information retrieval
 - Miras, 1214–1215
 - equivalence theorem, 1214
 - unification, 1214
 - Mirzakhani, Maryam, 3112
 - misattribution of affect, 3963
 - misinformation, 17, 18
 - audio, 939
 - missing code
 - reproducibility, 3922
 - missing completely at random, 2783
 - missing data imputation, 2783–2787
 - missing sequence imputation, 3517
 - missing values, 2693
 - Mistral, 1086
 - Mistral AI, 3152
 - mitosis
 - definition, 3653
 - detection, 3659–3660
 - mitosis detection, 3510
 - mitotic index, 3504
 - mixed image-video training, 656
 - ratio, 656
 - mixed operation, 1920
 - mixed precision
 - all-reduce exercise, 5041
 - bandwidth savings, 5175
 - numerical exercise, 5080
 - mixed precision training, 1526–1532, 5070–5081, 5105
 - algorithm, 1528
 - all-reduce, 5037
 - BF16, 5073
 - communication savings, 5076
 - figure, 5184
 - formal definition, 5073, 5109
 - FP16, 5073
 - loss scaling, 1529
 - memory savings, 1529
 - training flow, 5076
 - mixed-type columns, 2693
 - mixed-type data, 2698–2700
 - consistency models, 2926
 - geometry, 2897
 - optimal architecture, 2906
 - visualization, 2699
 - mixing coefficient, 146
 - mixing time
 - island model, 2127
 - mixing-layer depth, 4225
 - mixture constraint
 - projection, 892
 - mixture density network, 159
 - exercise, 162
 - Mixture of Experts
 - comparison, 1114
 - mixture of experts, 1539–1543
 - capacity-compute, 1542
 - connection to CL, 1203
 - definition, 1541
 - expert parallelism, 5135
 - GPT-4, 4926
 - layer, 1541
 - load balancing, 1542, 1557, 2152
 - medical segmentation, 3730
 - routing, 5138
 - task routing, 3730
 - video diffusion, 750
 - mixture weights
 - BMA, 4229
 - Mixture-of-Experts, 116, 620–628
 - advantages for tables, 2818–2820
 - capacity factor, 625–628
 - definition, 625
 - computational savings, 621
 - definition, 620

- definition for tables, 2816
- dense vs sparse, 621
- exercise, 627
- expert utilisation, 624
- layer, 620–628
- layer frequency, 626
- load balancing, 622–624
 - gradient, 623
 - optimum, 623
- load balancing loss, 622
- motivation, 2816
- noise-level specialisation, 624–628
- parameter interference, 2819
- practical configuration, 622
- relation to Engram, 1095
- routing, 621–628
- specialisation exercise, 627
- tabular, 2816–2820
- top- k routing, 621
- training objective, 626
- Z-loss, 626
- mixture-of-experts
 - architecture search, 1947–1949
 - communication, 4977
 - configuration, 1948
 - load balancing, 1948
 - NAS synergy, 1948
 - routing entropy, 1949
- MixUp
 - data augmentation, 4457
- Mixup, *see* regularization, Mixup
 - vicinal risk minimisation, 1555
- Mizar, 2944, 3003, 3035
 - and AI, 3003
 - key features, 3003
 - limitations, 3003
 - logical foundations, 3035
- Mizar language, 3066
- Mizar Mathematical Library, 2944, 3003, 3035
- ML engineering
 - distributed systems, 4927
- ML research
 - distributed knowledge gap, 4927
- MLP
 - architecture comparison, 4235
 - compute-bound, 4937
 - key-value exercise, 4067
 - key-value memory, 4060
 - tensor parallelism, 5045
- MLP as key-value memory, 4119
- MLP blocks
 - key-value memory, 4017
 - key-value memory exercises, 4019
 - role, 4016–4017
- MLP redundancy
 - circuit resilience, 4121
 - exercise, 4123
 - video transformer, 4119–4121
- MLSF
 - architecture, 4730–4731
 - cross-lingual transfer, 4734
 - definition, 4730
- MLTT, *see* Martin-Löf type theory
- MLWP
 - model comparison, 4198
 - taxonomy, 4198
- MLWP benchmarks
 - paradox, 4302
- MMD, *see* maximum mean discrepancy
 - aggregation exercise, 4608
- MMD loss
 - MADA aggregation, 4602
- MMPose, 4729
- MMSE estimator, 331
 - optimality, 331
- MnasNet, 1893
- MNIST, 5, 22
- mobile keyboard
 - cross-device FL, 4463

- mobility coefficient, 3358
- modality gap, 2463–2467, 2560–2565
 - challenge, 2638
 - cone effect, 2464
 - formal definition, 2463
 - lower bound, 2464
 - measurement exercise, 2466
 - reduction strategies, 2486
- modality prioritisation, 4081
 - empirical findings, 4082
- modality-adaptive architectures, 1960–1963
 - definition, 1962
 - formal definition, 1962
- mode assignment, 2718
- mode collapse, 25, 260, 281, 2025, 3590, 3744, 4444
 - amplification, 4572
 - and disengagement, 2070
 - as coevolutionary disengagement, 2252
 - clinical implications, 3590
 - definition, 281, 3590, 3820
 - discrete columns, 2722
 - EEG GAN, 3943
 - evaluation, 2705
 - exercise, 286
 - federated, 2883
 - federated GAN, 4487
 - federated GANs, 4453
 - GAN, 2758
 - GOGGLE vs CTGAN, 2757
 - minibatch standard deviation, 352
 - non-IID amplification, 4571
 - PacGAN defence, 2716
 - pix2pix, 3581
 - structural guarantee, 2716
 - tabular data, 2714
- mode coverage
 - diffusion, 2759
- diffusion models, 2769
 - privacy exercise, 4505
 - privacy tradeoff, 4502
 - segmentation, 3495
- mode covering, 398
- mode dropping
 - KL divergence, 2852
- mode-covering, 74, 171
- mode-seeking, 74, 171
- mode-specific normalisation, 2714, 2718–2721
 - algorithm, 2719
 - encoding, 2719
 - example, 2720
 - information preservation, 2720
 - preview, 2700
- mode-specific normalization
 - exercise, 2916
- model arithmetic
 - task vectors, 4168
- model cards
 - epilepsy AI, 3923
 - federated, 4652
- model comparison
 - ELBO, 199
- model compression
 - accessibility, 4816
- model diversity, 2710
- model FLOP utilisation
 - definition, 4930
- model FLOPs utilisation, 5142
- model heterogeneity
 - definition, 4583
 - federated learning, 4454, 4583
- model interpretability, 2755
- model memory footprint
 - breakdown, 4929
- Model Output Statistics, 4221–4232
 - exercise, 4412
 - formal definition, 4223

- model parallelism, 659
 - definition, 4929
 - early history, 4931
 - pipeline, 5050–5060
 - tensor, 5041–5050
- model privacy, 1571–1842
 - architecture dependence, 1972
- model quality
 - and imputation, 2785
- model scale
 - historical growth, 4925
- model selection
 - GMM, 155
 - structured data, 2904
- model self-knowledge, 4171
- model size
 - memory wall, 4925
- model state
 - memory check, 4940
- model uncertainty, 4290
 - stochastic physics, 4291
- model-to-data paradigm, 4449
- modelling capacity
 - classical methods, 3530
- modular forms, 3039
- modus tollens, 2969
- MoE, *see* Mixture-of-Experts
- MoE-cGAN, 3587
 - architecture, 3587
 - quality assurance, 3588
- MOGREPS-G, 4196
- moisture equation, 4205
- molecular design, 2696
- molecular dynamics, 2210
- molecular simulation
 - atmosphere, 4209
- molecule generation
 - GFlowNet, 2873
- moment generating function, 36
- moment retrieval, 2421–2427
- moments accountant, 1671, 2738, 3941
- momentum conservation, 3320
- momentum equation
 - atmospheric, 4203
- Monge problem, 3864
- Monge, Gaspard, 290, 3864
 - historical, 3876
- monosemantic features, 4051–4057
 - taxonomy figure, 4055
- monosemanticity
 - vision, 4086
- monoT5, 2489, 2590
 - example, 2489
 - vs BERT, 2595
- monotone convergence theorem, 5228
- monotone game, 2065
- Monte Carlo
 - single sample, 395
- Monte Carlo dropout, 3229, 3411
- Monte Carlo forecasting
 - theoretical skill, 4286
- Monte Carlo sampling
 - weather, 4202
- Monte Carlo Tree Search
 - theorem proving, 3042
- Monte Carlo tree search
 - Lean proof search, 3126–3129
 - theorem proving, 3099–3112
 - UCT formula, 3114
- Montreal Neurological Institute, 3776
- Moore, J Strother, 3004
- Moreau decomposition, 5275
- Morlet wavelet
 - EEG, 3832
- MOS, 22, *see* Mean Opinion Score, 4221–4232
 - evolution timeline, 4241
 - history, 4231
 - limitations, 4229–4230
 - operational pipeline, 4227

- operational vs developmental, [4227](#)
- original formulation, [4222](#)
- voice cloning, [4754](#)
- MOS (Mean Opinion Score), [4887–4888](#)
- mossy fibre sprouting, [3892](#)
- most general unifier, [3020](#)
- MoTiF
 - tropical cyclones, [4380](#)
- motion blindness
 - MLLMs, [4832](#)
- motion blur
 - endoscopy, [3514](#)
- motion control, [675–683](#)
 - dense flow, [679](#)
 - from language, [679](#)
 - optical flow, [678](#)
 - sparse trajectories, [678](#)
 - trajectories, [678](#)
 - types, [680](#)
- motion diffusion, [742](#)
- motion magnitude score, [740](#)
- MotionCtrl, [681](#)
- motor accessibility
 - quadriplegia
 - smart home scenario, [4871](#)
 - smart home, [4871](#)
 - voice interfaces, [4864](#)
- motor disabilities, [4690](#)
- motor imagery EEG
 - CVAE, [3824](#)
- Mouret, Jean-Baptiste, [2083, 2284](#)
- Moûsai, [846–847](#)
 - efficiency, [847](#)
- mouthings
 - sign language, [4736](#)
- Movie Gen
 - system design, [688](#)
 - text encoding, [686](#)
- movie recommendation
 - federated VAE, [4514](#)
- moving boundary problems, [3352–3357](#)
 - neural network challenges, [3354](#)
- MPI
 - collective operations, [4970](#)
 - execution model, [4960](#)
 - GPU limitations, [4971](#)
 - gradient averaging pseudocode, [5171](#)
 - history, [4960](#)
 - tutorial, [4959–4971](#)
- MPI-1, [4960](#)
- MPI-2
 - one-sided communication, [4960](#)
- MPI-3
 - non-blocking collectives, [4960](#)
- MPI-4
 - large count, [4960](#)
- MPI_Accumulate, [4966](#)
- MPI_Allgather, [4964](#)
- MPI_Allreduce, [4964](#)
 - example, [4968](#)
- MPI_ANY_SOURCE, [4962](#)
- MPI_ANY_TAG, [4962](#)
- MPI_Bcast, [4964](#)
- MPI_Comm_rank, [4961](#)
- MPI_Comm_size, [4961](#)
- MPI_Finalize, [4961](#)
- MPI_Get, [4966](#)
- MPI_Init, [4961](#)
- MPI_Irecv, [4962](#)
- MPI_Isend, [4962](#)
- MPI_Put, [4966](#)
- MPI_Recv, [4961](#)
- MPI_Reduce, [4964](#)
- MPI_Scatter, [4964](#)
- MPI_Send, [4961](#)
- MPI_Wait, [4962](#)
- MPICH, [4960](#)
- MPP, *see* Multiple Physics Pretraining
 - masking, [3254](#)

- scaling laws, 3255
- tokenisation, 3254
- zero-shot transfer, 3329
- MPP (Multiple Physics Pretraining), 3329, 3436
- MPTP, 3003
- MRAL (Mathematical Research Autonomy Levels), 3154
 - detailed, 3173
- MRI
 - epilepsy, 3776
 - fusion with EEG, 3934
 - high-resolution epilepsy, 3780
 - lesion detection
 - generative, 3955
 - outdated
 - epilepsy, 3900
 - subtle lesions, 3781
- MRI gradient artefacts, 3850
- MRI synthesis
 - federated, 4610
- MRI-negative epilepsy, 3780
- MRI-negative FCD, 3781
 - detection sensitivity, 3859
- MRI-to-PET synthesis, 3857–3860
 - clinical validation, 3859
- MRI-to-SPECT synthesis, 3857–3860
- MRL, *see* Matryoshka Representation Learning
- MRMS
 - Multi-Radar Multi-Sensor, 4240
- MRR (mean reciprocal rank), 2360
- MS MARCO, 2597–2598
 - bias, 2597
 - leaderboard progression, 2598
 - sparse judgments, 2597
- MSCOCO benchmark, 2417
- MSE, *see* mean squared error
 - minimisation trap, 4298
 - optimal predictor, 4297, 4298
 - spectral implications, 4360
- MTBF
 - distributed training, 5145
- MTEB, 2599–2600
 - leaderboard, 2600
 - score decomposition, 2608
- MTEB benchmark, 2387, 2388
- MTLD, 4763
 - definition, 4764
 - exercise, 4768
- multi-agent
 - theorem proving, 3090
- multi-agent collaboration
 - science, 1400–1406
- multi-agent debate
 - failure modes, 1316
- multi-agent policy gradient, 1378
- multi-agent systems, 1267–1463
 - 4-tuple, 1301
 - alignment, 1433
 - biological parallels, 1443
 - connections, 1432–1438
 - continual learning, 1435
 - definition, 1301
 - economies, 1442
 - exercises, 1445–1462
 - formalisms, 1301–1308
 - future, 1459
 - game theory, 1354–1363
 - grand unification, 1437
 - history, 1438–1440
 - information-theoretic capacity, 1428
 - memory, 1432
 - open problems, 1440
 - performance plateau, 1305
 - practitioner’s guide, 1445
 - reasoning, 1434
 - reasoning power, 1435
 - regression model, 1307
 - scaling laws, 1430

- unified framework, 1426–1432
- vision, 1444
- multi-annotator modelling, 3568–3570
 - confusion matrix, 3569
 - probabilistic segmentation, 3570
 - scalability, 3569
- multi-annotator segmentation, 3760
- multi-armed bandit
 - for LLM selection, 2149
 - model selection, 2149
 - resource allocation, 2173
 - tool selection, 1282, 1446
- multi-atlas label fusion, 3529
- multi-block coevolution, 2112
- multi-channel time series, 2802
- multi-diffusion
 - patch-based, 4355
- multi-domain EEG
 - information redundancy, 3830
- multi-domain EEG synthesis, 3819
- multi-expert segmentation, 3701
- multi-head attention
 - decomposition, 4012–4020
 - global workspace, 4149
 - tensor parallelism, 5045
- Multi-head Latent Attention, 115
- multi-hop reasoning
 - MDM vs ARM, 4110
- multi-information
 - exercise, 3500
 - segmentation, 3492
- multi-input operator learning, 3242
- multi-level optimisation
 - convergence, 3708
 - GenSeg, 3707
- multi-level retrieval, 2539
- multi-mesh icosahedron, 4266
- multi-modal formalization, 3078
- multi-modal generation
 - audio-visual, 744
- multi-modal imaging, 3483
- multi-objective evaluation, 2117
- multi-objective NAS, 1888
- multi-objective optimisation, 2283
 - exercise, 2294
- multi-objective optimization
 - Pareto coevolution, 2077
- multi-organ segmentation, 3503
 - abdomen, 3512
- multi-PDE benchmark, 3325
- multi-period discriminator, 783
 - coverage, 783
- multi-resolution STFT loss, 785
 - definition, 785
 - parameter choices, 786
 - properties, 854
 - vs. mel-spectrogram loss, 786
- multi-scale architecture, 3243
- multi-scale discriminator, 784
- multi-scale encoder
 - MetNet, 4240
- multi-scale expertise, 1226
- multi-scale representation
 - epilepsy, 3895
- multi-scale structure, 2788
- multi-stage retrieval, 2372
- multi-stage training, 652–661
 - definition, 652
 - exercise, 660
 - overview, 652
- multi-step generation, 3967, 3993–3994
 - interpretability challenge, 3966
- multi-step peer review agent, 2184
 - detailed example, 2185
- multi-step reasoning
 - with retrieval, 2536–2538
- multi-table data, 2697–2698
 - trend, 2711
- multi-table generation, 2829, 2924
 - challenge, 2829

- multi-task learning
 - FengWu, [4278](#)
- multi-task loss
 - FengWu, [4277](#)
- multi-track generation, [889–895](#)
 - approaches, [889](#)
 - comparison, [890](#)
 - definition, [889](#)
 - joint, [889](#)
 - sequential, [890](#)
- multi-vector retrieval
 - exercise, [2636](#)
- multicollinearity
 - MOS, [4224](#)
- multidimensional transformer,
 - [3884–3885](#)
- multigrid
 - analogy with diffusion, [3266](#)
- multigrid methods
 - connection to PiT, [3252](#)
- multilayer perceptron, [91](#)
 - shallow, [4232](#)
- multilingual sign language framework,
 - [4730](#)
- multimodal accessibility, [4875–4885](#)
 - formal definition, [4879](#)
 - hub architecture, [4880](#)
 - promise, [4875](#)
 - real-time translation, [4880](#)
 - translation pipelines, [4876](#)
- multimodal AI
 - epilepsy, [3903–3908](#)
- multimodal alignment, [2458](#)
- multimodal continual learning,
 - [1224–1229](#)
- multimodal distributions, [2693](#), [2718](#)
- multimodal embeddings, [2390](#),
 - [2457–2467](#)
 - comparison, [2465](#)
 - Gemini, [2461](#)
 - native, [2461](#)
- multimodal foundation model
 - epilepsy, [3934–3935](#)
 - open problem
 - epilepsy, [3935](#)
- multimodal fusion
 - early vs. late, [2420](#)
 - expressiveness, [2424](#)
 - for video, [2423–2427](#)
 - practical considerations, [2425](#)
 - strategies, [2424](#)
- multimodal generation
 - federated, [4663](#)
 - formal statement, [755](#)
 - neurodivergent, [4792](#)
 - unified audio, [948](#)
 - unified model, [755](#)
- multimodal interpretability, [4157–4158](#)
- multimodal LLMs
 - geospatial, [4717](#)
- multimodal prompting
 - benchmarks, [3736](#)
- multimodal RAG, [2538–2545](#)
 - architecture, [2543](#)
 - audio generation, [2541–2542](#)
 - generality, [2543](#)
 - open problems, [2545](#)
 - speech, [2540–2541](#)
 - video, [2542–2543](#)
 - visual documents, [2538–2540](#)
- multimodal retrieval, [2631](#)
 - benchmarks, [2600–2602](#)
- multimodal search
 - cold start, [2479](#)
 - hybrid queries, [2479](#)
 - product, [2463](#)
- multimodal VAE
 - epilepsy, [3935](#)
- multinomial diffusion, [2758](#), [2760](#)
 - closed form, [2762](#)

- example, 2763
- exercise, 2773, 2917
- forward process, 2762
- multiphysics
 - compositional generation, 3471
- multiphysics systems, 3371–3377
 - coupling challenges, 3371
 - universal compositionality, 3377
- multiple negatives ranking loss, 2392
- Multiple Physics Pretraining, 3254–3256
- multiple sclerosis
 - lesion segmentation, 3511
 - segmentation, 3480
- multiple synthesis, 2890
- multiple synthetic datasets, 2690
- multiply-accumulate
 - TPU, 4937
- MultiSimDiff, 3372–3377, 3471
 - nuclear thermal coupling, 3373
 - training, 3373
- multivariate time series, 2802
- μ SAM, 3647–3648
 - 3D embryo tracking, 3648
- museum accessibility
 - example, 4714
 - participatory evaluation, 4895
- music features
 - hierarchy, 2442
- music generation, 16, 765–959
 - copyright, 937
 - flow matching, 864–873
 - future directions, 945–949
 - long-form, 881–888
- music information retrieval, 2438–2447
 - pipeline, 2446
- music recommendation, 2441–2445
- music representation learning, 2443–2447
 - neural features, 2444
- music similarity, 2441–2445
- exercises, 2447
- Music Transformer, 906–908
- music-text retrieval, 2445
- musical form, 887–888
- musical structure
 - hierarchy, 881
 - long-form, 946
 - timescale hierarchy, 881
- musical structure metrics, 924–926
 - domain specificity, 926
 - self-similarity, 925
- musicality metrics, 923–927
- MusicFlow, 867–873
- MusicFX, 17, 911
- MusicGen, 767, 812–814, 2443
 - autoregressive connection, 943
 - classifier-free guidance, 814
 - codebook patterns, 951
 - diversity, 927
 - generation algorithm, 813
 - multi-modal conditioning, 822
 - token rates, 814
- MusicLDM, 839–841
 - augmentation strategies, 841
 - beat-synchronous mixup, 840
 - metrical structure preservation, 840
 - training data, 841
- MusicLM, 911–912, 2444
- MusicWeaver, 883
- must-link constraint, 2754
- must-not-link constraint, 2754
- mutation
 - semantic, 2137
- mutation operators
 - limitations, 2095–2097
- mutual information, 84
 - CLIP bound, 2409
 - cross-modal bound, 3856
 - definition, 1588, 2838
 - difference, 2838

- EEG-fMRI, 3856
- estimation, 2853
 - neural, 3856
- forward process, 1609
- preservation, 2853
- synthetic PET, 3862
- training data, 1588
- MYCIN, 3962
- N-gram hash, 1092
- n -gram model, 430–433
 - bigram, 430
 - definition, 430
- NADE, 433–435
 - algorithm, 435
 - Deep NADE, 435
 - definition, 434
 - weight sharing, 434
- naive encoding
 - failure, 2699
- Name Mover Heads, 4030
 - walkthrough, 4034
- nanowell array
 - cell tracking, 3662
- napari
 - μ SAM integration, 3648
- narrative consistency, 522
- NAS
 - evolutionary analogs, 2255
 - evolutionary lineage, 2254–2255
- NAS-Bench, 1986
- NAS-Bench-101, 1934
- NAS-Bench-201, 1934
- Nash equilibrium, 310, 5270
 - coevolutionary, 2065–2067
 - definition, 2065
 - exercise, 1451
 - GAN, 277, 3943
 - Prisoner’s Dilemma, 325
- NASNet, 1893, 2254
- NASWOT, 1931
- National Institutes of Health (NIH)
 - AI4AD, 4800
- National Meteorological Center, 4287
- National Weather Service
 - US, 4227
- natural gradient, 197, 248–251, 2043
 - definition, 249
 - diagonal Gaussian, 267
 - exponential families, 250
 - parameterization invariance, 250
 - practical methods, 251
 - SVI, 216
- natural language
 - ambiguity in proofs, 2936
 - evaluation difficulty, 2933
 - mathematical, 3039
 - proof limitations, 3136
 - tabular encoding, 2805
- natural language processing, 17
- natural language prompts, 3520
- natural numbers
 - in Lean, 2966
 - inductive definition, 2985
- natural parameter, 203
- natural selection, 2019
 - as algorithm, 2027–2030
- NaturalSpeech 2, 4747
 - architecture, 4748
- Navier–Stokes
 - distillation, 3391
 - flow matching, 3386
 - physics-informed diffusion, 3279
- Navier–Stokes equations, 3197–3199, 3325–3352
 - 2D vs 3D, 3428
 - atmospheric application, 4203
 - benchmark results, 3332
 - biofluid segmentation, 3731
 - compressible, 3327, 3416

- diffusion models, 3330
- existence and smoothness, 3428
- FNO, 3222
- GNOT, 3244
- incompressibility constraint, 3326
- incompressible, 3326
- Millennium Prize, 3427–3431
- millennium problem, 3198, 3731
- momentum equation, 3326
- overview, 3325
- singularity, 3473
- Transformer models, 3327
- navigation
 - blind users, 4810
- navigation pipeline
 - performance, 4916
- NC¹, 1036
- NCAM guidelines, 4877
- NCCL, 5036
 - architecture, 4972
 - asynchronous communication, 5140
 - comparison table, 4976
 - double binary tree, 5036
 - hybrid algorithm, 4972, 5036
 - NVIDIA Collective
 - Communications Library, 4971–4980
 - profiling, 4978
 - PyTorch integration, 4974
 - reinitialisation, 5151
 - ring all-reduce, 4972
 - specifications, 4977
 - topology auto-detection, 4943
 - tree algorithm, 4972
 - tuning, 5037
- nccl-tests
 - profiling, 4978
- NCCL_ALGO, 5037
- NCCL_BUFFSIZE, 5037
- NCEP/NCAR reanalysis, 4238
- nDCG, 2587
 - definition, 2604
 - example calculation, 2360
 - exercise, 2632
 - pairwise approximation, 2594
 - vs MAP, 2608
- nDCG (normalised discounted cumulative gain), 2360
- near-orthogonal features, 3971
- near-orthogonal vectors, 4042
- near-singularity detection, 3430
- nearest-point projection, 2778
- NEAT, 2047–2057, 2283
 - as ancestor of NAS, 2254
 - complexification, 2050
 - crossover, 2048
 - encoding, 2048
 - genome, 2048
 - mutation, 2048
 - speciation, 2049
- needle-in-a-haystack, 2097
- negative momentum, 321
 - definition, 321
 - exercises, 325
- negative prompts
 - Bayesian interpretation, 669
 - video, 668
- negative sampling
 - in recommendation, 2578
- negative transfer
 - EEG, 3932
- Nelson–Oppen method, 3023
- NeMo
 - diffusion, 5086–5087
- neocortex, 1176
- nested intervals theorem, 5193
- nested learning, 1216–1219
 - connection to CLS, 1218
 - definition, 1216
- nested NWP models, 4351

- Net2Net, 1869, 1957–1959
 - Net2DeeperNet, 1958
 - Net2WiderNet, 1957
 - wider, 1870
- network architecture
 - TabDDPM, 2771
- network fabric
 - high-radix switches, 4928
- network geometry breakdown, 3863, 3866
- network hub
 - seizure interception, 3874
- network hypersynchrony, 3891–3896
- network latency, 4811
- network morphism, 1957
- network regime
 - crossover, 5131
- network surgery, 3913
- network topology
 - adaptive, 5157
 - algorithm fitness, 4950
 - comparison, 4947
 - definition, 4943
 - distributed training, 4942–4951
 - epilepsy, 3912
 - families, 4944
 - formal definition, 4943
 - importance, 4943
- Neumann series, 3342
 - neural operator, 3342
- Neumann solution, 3354
- neural architecture
 - cooperative coevolution, 2064
- neural architecture search, 1845–1880, 2005
 - adaptive networks, 1868–1880
 - agentic AI, 1976
 - alignment, 1974
 - alternative RL methods, 1916
 - as privacy tool, 1974
 - bilevel optimisation, 1875–1880, 1886–1891
 - biological inspiration, 1845–1857
 - biological pruning, 1849
 - brain-inspired challenge, 1879
 - compositional safety, 1978
 - computational cost, 1915, 1986
 - computational waste, 1915
 - connection to NEAT, 2052
 - connections and synthesis, 1969–1981
 - constraints, 1853–1857
 - continual learning, 1959, 1970
 - contributions, 1934
 - controller policy, 1912
 - controller RNN, 1912
 - convergence guarantees, 1982
 - cost reduction, 2012
 - differentiable, 1920–1926
 - diffusion models, 1935–1943
 - EEG prediction, 3931
 - evolution, 1933
 - evolution analogy, 1879
 - evolutionary-generative, 1953
 - exercises, 1992–2003
 - bilevel optimization, 1994
 - convergence experiment, 2002
 - DARTS computation, 1993
 - dynamic programming, 1996
 - forgetting bound, 1998
 - generative model, 1996
 - information bottleneck, 1993
 - NP-hardness, 1999
 - Pareto frontier, 1995
 - privacy-aware, 2000
 - pruning equivalence, 1997
 - search space size, 1992
 - self-adapting LLM, 2001
 - sleep-inspired, 2000
 - weight sharing, 1995

for retrieval, 2610
 forgetting, 1881–1891
 formal definition, 1875
 foundation models, 1934, 1987
 generative approaches comparison, 1954
 generative model generalization, 1982
 generative models as tools, 1949–1955
 GPT-NAS, 1946
 graph diffusion, 1937
 grow then prune, 1852
 grow vs. start simple, 1852
 hardware-aware, 1928–1930
 historical perspective, 1992
 historical timeline, 1989–1992
 history, 1880
 information theory, 1857–1868
 intellectual lineage, 2004
 large language models, 1943–1949
 layer-wise decomposition, 1927
 mathematics, 1897–1910
 metabolic bound, 1853
 mixture-of-experts, 1947
 modern methods, 1926–1935
 multi-objective, 1986
 NP-hardness, 1876
 one-shot algorithm, 1909
 open problems, 1981–1989
 parameter sharing, 1916–1919
 polynomial complexity, 1926
 practical challenges, 1985
 privacy, 1972
 provable improvement, 1985
 pruning, 1944
 reinforcement learning, 1911–1916
 reproducibility, 1985
 reproducibility crisis, 1934
 reward signal, 1912

RL pipeline, 2009
 RL-based algorithm, 1913
 safety concerns, 1977
 sample complexity, 1914, 1982
 search space, 1846, 1892–1897
 search space extensions, 1915
 self-evolving AI, 1987
 surrogate-assisted, 2229
 survey, 1933
 theoretical foundations, 1982
 three pillars, 1876
 timeline, 1933
 training-free, 1930–1932
 diffusion models, 1940
 trajectory, 1933
 transferability, 1986
 unified diagram, 2015
 unified view, 1979
 Zoph and Le, 1911
 neural audio codec, 787–802
 as tokenizer, 795
 comparison, 795
 frozen decoder, 799
 generation pipeline, 799
 quality ceiling, 947
 training, 797
 training data bias, 796
 training objective, 797
 neural autonomy
 increasing, 3044
 neural circuit verification, 4158
 neural circuits
 as collective computation, 1271
 neural codec
 autoregressive modelling, 809–816
 expressivity conjecture, 830
 neural coding, 1857, 3969
 neural collaborative filtering, 2576
 neural field
 latent diffusion, 3392

- neural function evaluations (NFEs),
3289
- neural image retrieval, 2406–2417
- neural mass model, 3892
- neural memory
 - for continual learning, 1211–1219
- neural network, 6, 91
 - 1990s meteorology, 4233
 - early weather applications,
4232–4243
 - GRU, 1029
 - LSTM, 100
- neural network tricks, 1471–1558
 - composition, 1546–1549
 - design, 1558
 - interactions, 1558
- neural ODE, 460
- neural operators, 3210–3232
 - Bayesian, 3411
 - computational cost, 3458
 - Darcy benchmark, 3297
 - data requirements, 3230
 - discretisation dependence, 3230
 - GNOT, 3242
 - heat equation, 3308
 - kernel integral, 3452
 - limitations, 3227–3231
 - Transformer-based, 3233–3264, 3361
- neural OT
 - neuromodulation, 3937
- neural PDE solvers
 - uncertainty quantification, 3409
- neural population control, 3873
- neural processing unit, 4815–4816
 - application processor, 4815
- neural quantum states
 - history, 3370
- neural ranking models, 2589–2592
- neural spline flow, 458–459
 - monotonicity, 458
 - rational-quadratic, 458
- neural tangent kernel
 - spectral bias, 3457
- neural theorem proving, 2954,
3080–3095
 - comparison, 3044–3046, 3093
 - comparison table, 3093
 - early approaches, 3031–3046
 - history, 2956–2957
 - pipeline, 3093
- Neural Turing Machine, 1090–1092
- neural vocoder, 779–787
 - history, 780
- neural wavefunctions, 3366
- neural-symbolic integration
 - retrieval, 2572–2574
- NeuralGCM, 4274–4284
 - climate simulation, 4281
 - hybrid model, 4280
 - learned physics, 4281
- neuro-symbolic
 - AlphaGeometry, 3113
 - division of labour, 3122
 - mathematics, 2954
- NeuroBOLT, 3851
 - diffusion backbone, 3851
 - EEG encoder, 3851
- neurodevelopmental conditions, 4689
- neurodivergence, 4690
- neurodivergent learners
 - generative AI, 4789
- neurodiversity
 - definition, 4789
- NeuroDream, 1964–1966
 - mathematical formulation, 1965
 - NAS algorithm, 1967
- neuroevolution, 1880, 2047–2057, 2283
 - Atari, 2054
 - Baldwin effect, 2036
 - open problems, 2053

- sparse rewards, [2053](#)
- NeuroFlowNet, [3854–3856](#)
- neurogenesis, [1869](#)
 - artificial, [1869](#)
- neuroimaging
 - diffusion models, [3840–3847](#)
 - segmentation, [3511–3512](#)
- neuromodulation
 - optimal transport, [3936](#)
- neuromorphic computing
 - accessibility, [4816](#)
- neuron
 - polysemantic, [4039](#)
- neuron merging, [1966](#)
- neurons
 - as features, [3972](#)
- neuroplasticity, [1847](#)
- neuropsychological evaluation
 - missing, [3900](#)
- neuropsychology
 - pre-surgical, [3780](#)
- neuroscience
 - brain development, [1845–1857](#)
 - of retrieval, [2340–2348](#)
- neurosurgery
 - segmentation, [3480](#)
- neurosymbolic
 - broader principle, [2833](#)
 - constraint discovery, [2926](#)
 - correctness theorem, [2830](#)
 - key idea, [2825](#)
 - motivation, [2816](#)
 - tabular generation, [2820–2834](#)
- neurosymbolic generation
 - definition, [2821](#)
- neutral network
 - definition, [2033](#)
 - escape from local optima, [2033](#)
- neutral networks, [2032–2034](#)
- neutral theory, [2033](#)
- new row synthesis metric, [2849](#)
- Newell, Allen, [2939](#)
- Neyman–Pearson lemma, [1581](#)
 - audio detection, [930](#)
 - membership inference, [1581](#)
- NGR, *see* Non-homogeneous Gaussian Regression, [4228–4229](#)
 - comparison, [4241](#)
- NICE, [454](#)
- niche construction, [2244–2246](#)
 - and unbounded complexity, [2244](#)
 - biological examples, [2244](#)
 - definition, [2244](#)
 - exercise, [2249](#)
- niching
 - exercise, [2288](#)
- NIfTI
 - segmentation format, [3484](#)
- NIH ChestX-ray
 - MGDS attack, [4596](#)
- Nipkow, Tobias, [3001](#)
- NK model, [2022–2024](#)
 - definition, [2022](#)
 - ruggedness, [2023](#)
- nnU-Net, [3487, 3538–3539](#)
 - architecture selection, [3538](#)
 - challenge results, [3538](#)
 - design philosophy, [3538](#)
 - ensembling, [3538](#)
 - postprocessing, [3538](#)
 - preprocessing, [3538](#)
 - training configuration, [3538](#)
- no free lunch
 - architectures, [1983](#)
 - continual learning, [1181](#)
 - epilepsy AI, [3937](#)
 - medical segmentation, [3729](#)
- No Free Lunch theorem, [2247–2249](#)
 - and open-ended evolution, [2248](#)
 - exercise, [2249](#)

- formal statement, [2247](#)
- interpretation, [2247](#)
- structured problems, [2248](#)
- No Free Lunch theorems, [2274–2275](#), [2282](#)
 - exercise, [2295](#)
- no-free-lunch
 - ideal CL, [1237](#)
 - tabular generation, [2908](#)
- no-free-lunch theorem
 - tabular generation, [2827](#)
- NOAA
 - Project EAGLE, [4385](#)
- NOAA GEFS, [4196](#)
- NOAA GFS, [4218](#)
- NOAA Meteorological Development Laboratory, [4231](#)
- Nobel Prize, [14](#)
- NODDI dataset, [3852](#)
 - description, [3852](#)
- Noether's theorem, [3437](#)
- noise
 - spatiotemporal independence, [541](#)
- noise aggregation, [1654](#)
 - Jacobian connection, [1656](#)
 - local curvature, [1655](#)
- noise blending, [698](#)
- noise consistency
 - paired generation, [3702](#)
- noise prediction
 - DDPM, [2763](#)
- noise prediction loss, [4311](#)
- noise ramp, [704](#)
- noise schedule, [479](#), [495](#), [2760](#)
 - adaptation across stages, [659](#)
 - audio, [849–862](#)
 - audio-specific considerations, [850](#)
 - comparison example, [1631](#)
 - cosine, [849](#), [2760](#)
 - cosine advantage, [849](#)
 - examples, [632](#)
 - exercise, [634](#), [863](#)
 - image schedules for video, [628](#)
 - linear, [849](#), [2760](#)
 - linear vs cosine, [1631](#)
 - optimal, [630](#)
 - optimal privacy-preserving, [1635](#)
 - optimal video, [630](#)
 - privacy implications, [1630](#)
 - privacy knob, [1631](#)
 - privacy-aware design, [1634](#)
 - resolution-adapted, [632](#)
 - shift intuition, [630](#)
 - shifted cosine, [634](#)
 - shifted SNR, [629](#)
 - sigmoid, [849](#)
 - signal-to-noise ratio, [3745](#)
 - SNR shift, [629–635](#)
 - temporal, [704](#)
 - video, [628–635](#), [758](#)
 - visualisation, [633](#)
 - vulnerability, [1626–1636](#)
 - vulnerability comparison, [1634](#)
- noise-as-probe technique, [1653](#)
- noise-based attacks, [1649–1659](#)
- Noise2Music, [845–846](#)
- noisy channel, [1857](#)
- Nomic Embed, [2389](#)
- non-autoregressive generation
 - audio, [869](#)
- non-blocking collective operations, [4987](#)
- non-blocking communication
 - MPI, [4962–4964](#)
 - timeline figure, [4964](#)
- non-chronological backtracking, [3018](#)
- non-consensual synthesis, [939](#)
- non-convexity
 - federated unlearning, [4607](#)
- non-expansive mapping, [5274](#)
- non-Gaussian forecast distribution, [4321](#)

- Non-homogeneous Gaussian
 - Regression
 - NGR, [4229](#)
- non-i.i.d. data, [5116](#)
- non-IID data, [2883](#), [4568–4579](#)
 - Dirichlet partition
 - exercise, [4674](#)
 - federated learning, [4571](#)
 - formal definition, [4571](#)
 - generative models, [4571](#)
- non-iid data, [4448](#)
- non-local means, [330](#)
- non-manual markers, [4725](#)
 - grammatical role, [4726](#)
 - in lecture translation, [4732](#)
 - linguistic function, [4726](#)
 - mouthings, [4736](#)
- non-maximum suppression
 - StarDist, [3647](#)
- non-monotone games, [2066](#)
- non-saturating loss, [280](#)
 - definition, [280](#)
 - exercise, [285](#)
- non-separable functions, [2064](#)
- non-stationarity, [2697](#), [2788](#)
 - EEG, [3796](#), [3880](#)
- nonlinear approximation, [3225](#)
- nonlinear mapping
 - weather forecasting, [4232](#)
- nonlinear methods
 - environmental science, [4236](#)
- nonlinearity
 - atmospheric, [4229](#)
- Norell, Ulf, [3002](#)
- norm, [5212](#)
 - ℓ^p , [5212](#)
- norm_num, [3032](#)
- norm_num tactic, [2989](#)
- normal form
 - polynomial, [2989](#)
- normal-form game, [1355](#)
- Normal-Normal model, [208](#)
- NormalFloat, [1538](#)
- normalisation, [96](#)
 - adaptive instance, [353](#)
 - continuous features, [2771](#)
 - in generative models, [96](#)
 - spatially-adaptive, [355](#)
- normalising flow, [58](#)
- normalising flows
 - EEG scoring
 - exercise, [3789](#)
 - epilepsy applications, [3786](#)
 - exact likelihood, [4460](#)
 - federated, [4460](#)
 - seizure forecasting, [3786](#)
 - voice synthesis, [4749](#)
 - weather, [4198](#)
- normalization, [1488–1495](#)
 - DeepNorm, [1492](#)
 - definition, [1492](#)
 - stability, [1492](#)
 - group normalisation, [1489](#)
 - definition, [1490](#)
 - instance normalisation, [1490](#)
 - definition, [1490](#)
 - pre-norm vs post-norm, [1493](#)
 - QK-norm, [1491](#)
 - definition, [1491](#)
- normalizing flow, [7](#), [44](#), [449–469](#)
 - applications, [465–466](#)
 - composition, [451](#)
 - definition, [450](#)
 - variational inference, [465](#)
- normalizing flows, [236–241](#)
 - composition, [237](#)
 - definition, [236](#)
 - EEG synthesis, [3854](#)
 - flow matching, [750](#)
 - modern architectures, [239](#)

- planar flow, [238](#)
- radial flow, [239](#)
- universal approximation, [240](#)
- normative modelling
 - definition, [3846](#)
 - diffusion prior, [3841](#)
 - inside-out detection, [3844](#)
 - paradigm shift, [3911](#)
 - theory, [3846](#)
- normed space, [5212](#)
- notation
 - overloading, [3065](#)
- notations
 - custom, [2992](#)
- NOTEARS
 - EEG causal discovery, [3936](#)
- Nothing About Us Without Us,
 - [4701–4704](#), [4906](#)
- exercise, [4706](#)
- novel scenarios, [722](#)
- novelty
 - unbounded, [2238](#)
- novelty bonus, [2174](#)
- novelty rejection
 - embedding-based, [2198](#)
- novelty rejection filter, [2147](#)
- novelty reward, [3105](#)
- novelty search, [2081–2083](#), [2283](#)
 - deceptive problems, [2082](#)
 - definition, [2082](#)
 - in scientific discovery, [2217](#)
 - philosophical insight, [2083](#)
 - rejection filter, [2147](#)
 - theorem proving, [3109](#)
 - vs. fitness-based search, [2094](#)
 - weighted sampling, [2146](#)
- novelty-weighted sampling, [2146](#)
- nowcasting
 - cascaded diffusion, [4322](#)
 - decision making, [4371](#)
 - difficulty, [4362](#)
 - formal definition, [4361](#)
 - generative models advantage, [4372](#)
 - method comparison, [4372](#)
 - pipeline, [4371](#)
 - pipeline figure, [4436](#)
 - precipitation, [4361](#)
 - problem definition, [4361](#)
- NowcastNet, [4365](#)
 - advection embedding exercise, [4417](#)
 - advection operator, [4366](#)
 - formal definition, [4366](#)
 - inductive bias, [4372](#)
 - operational, [4388](#)
- NP
 - analogy with theorem proving, [2934](#)
 - generation vs verification, [2205](#)
- NP-hard, [166](#)
- NP-hardness
 - NAS, [1876](#)
 - neural architecture search, [1999](#)
 - parallelism search, [5155](#)
- NSGA-II, [2283](#)
 - exercise, [2294](#)
- NSNO, [3342](#)
- NTK-aware interpolation, [612](#)
- NTK-aware scaling, [1082–1083](#)
- nuclear pleomorphism, [3504](#)
- nuclear reactor simulation, [3373](#)
- nucleus sampling
 - tabular, [2810](#)
- nucleus segmentation, [3515](#)
- number of function evaluations
 - audio, [866](#)
- number theory
 - in Lean, [2976](#)
- numerical computation
 - verified, [2989](#)
- numerical precision

- language models, [2814](#)
- numerical simulation, [15](#)
- numerical stability
 - distributed training, [5105](#)
 - low precision, [5158](#)
- Numerical Weather Prediction, [4181](#)
- numerical weather prediction, [13](#), [4181–4202](#), [4212–4221](#)
 - data assimilation cost, [4374](#)
 - definition, [4212](#)
 - ENIAC, [4184](#)
 - history, [4210](#)
 - post-processing, [4221–4232](#)
- NUWA-XL, [702](#)
- NVIDIA
 - GPU evolution, [4935](#)
- NVIDIA DGX
 - architecture, [4948–4949](#)
 - placement exercise, [4951](#)
- NVLink, [4948](#), [5115](#)
 - architecture, [4935](#)
 - architecture diagram, [4928](#)
 - bandwidth, [4928](#), [4971](#), [5137](#)
 - communication hiding, [4989](#)
 - exercise, [4941](#)
 - H100, [4972](#)
 - ring topology, [4945](#)
 - tensor parallelism decision, [4940](#)
- NVLink 2.0, [4935](#)
- NVLink 3.0, [4935](#)
- NVLink 4.0, [4935](#)
- NVLink-C2C, [4935](#)
- NVMe offloading, [5068](#)
- NVSwitch, [4948](#)
 - architecture, [4935](#)
 - bisection bandwidth, [4948](#)
 - full-bisection bandwidth, [4936](#)
 - H100, [4935](#)
- NWP, *see* Numerical Weather Prediction, *see* numerical weather prediction
 - weather prediction
 - AI relationship, [4220](#)
 - computational cost, [4189–4192](#)
 - cost exercise, [4192](#)
 - diffusion model interface, [4327](#)
 - resolution scaling, [4190](#)
 - skill improvement, [4190](#)
- NWP pipeline, [4218](#)
- NYC Bar Association
 - AI disability report, [4848](#)
- Nyquist bound
 - temporal, [527](#)
- Nyquist frequency, [769](#)
- Nyquist theorem
 - video, [527](#)
- o-series
 - architecture, [3051](#)
 - performance, [3051](#)
- o1 (OpenAI), [3147](#)
- o1 model, [3051](#)
- o3 (OpenAI model), [3136](#)
- o3 (OpenAI), [3147](#)
- o3 model, [3051](#)
- o4-mini (OpenAI), [3147](#)
- o4-mini model, [3051](#)
- object detection
 - accessibility limitations, [4698](#)
 - YOLO, [4698](#)
- object-centric representation, [754](#)
- objective self-discovery, [2271](#)
- observation likelihood
 - DA, [4323](#)
- obtain tactic, [2995](#)
- “it is obvious that”, [2936](#)
- Occam’s razor
 - architecture search, [1885](#)
- occlusion traps
 - hallucination, [4831](#)
- OCR

- document understanding, 4698
- history, 4693
- HMM-based, 4698
- limitations, 4698
- reading order, 4698
- text protection, 4775
- Odd Order Theorem
 - formalization, 3000
- OFA, 1930
- offline architecture restructuring,
 - 1966–1967
- offline navigation
 - blind users, 4818
- offline operation
 - accessibility, 4814
- OGDA, *see* optimistic gradient
 - descent-ascent
- Okapi BM25, *see* BM25
- Olah, C., 3969, 3973
- Old Faithful, 145
- olean files, 2998
- Olshausen and Field, 4044–4045
- olympiad mathematics, 3033
- omega tactic, 2988–2989
- Omega test, 2989, 3023
- on-device speech synthesis, 4463
- Once-for-All, 1930
- one-hot encoding, 2700, 2771
 - limitations, 2795
- one-shot federated learning, 4636–4646
 - definition, 4636
- one-shot model
 - definition, 1904
- one-shot NAS, 1904–1910
 - algorithm, 1909
- one-sided communication
 - definition, 4966
 - MPI, 4966–4968
- OneMax problem, 2040
- online continual learning, 1177
- online scheduling, 1382
- online softmax
 - ring attention, 5055
- opacity problem, 3962
- open ball, 5195
- open cover, 5202
- open problems
 - AI for mathematics, 3133
 - calibrated extreme prediction, 4402
 - constraint-aware generation, 2713
 - differentiable NWP, 4406
 - kilometer-scale ensemble, 4403
 - mathematical AI, 3184
 - medical segmentation, 3727–3739
 - multi-scale modeling, 4406
 - online learning, 4404
 - PDE-based simulation, 3427–3442
 - physical conservation, 4403
 - seamless weather-climate, 4404
 - uncertainty quantification, 4405
 - universal metric, 2713
 - weather prediction, 4402–4407
- open questions
 - PDE simulation, 3442–3445
- open reasoning models, 3052
- open set, 5196
- open-domain question answering, 2523
- open-ended evolution, 2026, 2194,
 - 2237–2249, 2301
 - AlphaEvolve, 2108–2118
 - as holy grail, 2239
 - biological context, 2237
 - challenge exercise, 2295
 - complexity barrier, 2243
 - connection to LLM evolution, 2132
 - criterion, 2267–2269
 - conjecture, 2268
 - research directions, 2268
 - definition, 2238–2239
 - design, 2132

- drivers of discovery, 2219
- formal definition, 2238
- in DRQ, 2165
- LLM-based, 2246–2247
- manifesto, 2284
- metrics
 - exercise, 2293
- novelty vs. complexity, 2269
- open problems, 2267–2278
- POET, 2242
- requirements, 2246
- theorem proving, 3109
- open-ended search, 2019
- Open-Sora
 - training curriculum, 655
- open-source
 - mathematical AI, 3137
- open-source evolutionary systems, 2118–2131
- open-vocabulary segmentation, 3520, 3695–3697
 - definition, 3695
 - formal definition, 3695
- OpenAI, 3146–3150
 - mathematical AI approach, 3136
- OpenAI o-series, 3051–3052
- OpenASL dataset, 4728
- OpenDiT, 5086–5087
- OpenEvolve, 2119–2121
 - algorithm, 2119
 - architecture, 2119
 - multi-language support, 2120
 - results, 2120
- OpenMPI, 4960
- OpenNEURO
 - epilepsy dataset, 3859
- OpenStreetMap
 - accessibility, 4717
- operational AI weather
 - global overview, 4390
- operational deployment
 - cost-benefit, 4418
- operational forecast
 - Hurricane Milton case study, 4389
- operational forecasting
 - bottleneck, 4189
 - NOAA, 4231
 - validation
 - AI, 4386
- operational weather forecasting
 - AI, 4383–4394
- operator learning, 3210–3232
 - approximation rates, 3224
 - error analysis, 3227
 - problem formulation, 3211
 - theory, 3223
- optic cup segmentation, 3601
- optical coherence tomography
 - segmentation, 3481
- optical flow, 528–533
 - accessibility, 4898
 - as constraint, 528
 - definition, 528, 4363
 - radar extrapolation, 4363
 - temporal consistency, 3508
 - TrajGRU, 4240
- optical tomography, 3406
- Optima
 - exercise, 1449
- Optima framework, 1315–1316
 - generate, 1315
 - rank, 1315
 - select, 1315
 - train, 1315
- optimal transport, 261
- optimal critic
 - gradient norm, 296
- optimal difficulty
 - curriculum, 3108
- optimal intervention

- network, 3913
- optimal memory architecture, 1114
- optimal transport, 79, 367–381
 - advanced brain models, 3869–3879
 - advantages, 637
 - audio generation, 941
 - audio latent space, 871–873
 - basics, 3864
 - brain networks, 3862–3868
 - conditional path, 952
 - conditional paths, 868
 - definition, 369
 - dual, 371
 - EEG computation, 3946
 - EEG pipeline, 3914
 - entropic, 372
 - flow matching, 747, 4324
 - global federated, 4641–4642
 - history, 3876, 3877
 - Kantorovich, 369
 - linear program, 371
 - local federated, 4640–4641
 - metrics exercise, 3928
 - mini-batch, 380
 - neuromodulation, 3936–3937
 - paradigm shift, 3912
 - path, 636
 - path curvature, 868
 - problem, 80
 - Slater’s condition, 5278
 - topological data analysis, 3950
 - unbalanced, 381
 - Wasserstein distance, 3819
- optimality condition
 - subgradient, 5254
- optimistic gradient descent-ascent, 319–320
 - algorithm, 319
 - convergence, 320
 - definition, 319
 - exercises, 324
 - idea, 319
 - interpretation, 319
- optimization
 - federated LLMs, 4558–4568
- optimizer
 - distributed, 5022
- optimizer state
 - partitioning, 5061
- oracle pattern, 2990
- ORB, 2399
- orbit (dynamical systems), 1035
- OrCam MyEye, 4826
- orchestration, 1344–1354
 - configuration space, 1350
 - of model families, 3694
 - pattern comparison, 1353
 - token usage, 1349
- orchestrator, 1345
 - definition, 1345
 - meta-policy, 1346
- orchestrator-worker pattern, 1346–1347
 - empirical results, 1346
- order of a group, 2996
- order statistics, 1729
 - density, 1730
- ordered paramodulation, 3020
- ordinal columns, 2693
- ordinary least squares
 - bias correction, 4225
 - MOS exercise, 4231
- organ size disparity, 3512
- organs at risk, 3479
- Ornstein–Uhlenbeck process, 2770
- Ornstein–Uhlenbeck process, 530
- orographic lifting, 4222
- orthogonal projection, 2777
- orthopaedic surgery
 - segmentation, 3480
- OSI-FL, 4638–4639

- OT attention, 3870
- OT neuromodulation
 - open problem, 3937
- OT-guided stimulation
 - definition, 3936
- Otsimo, 4806
- Otsu's method, 3524
 - between-class variance, 3524
 - formulation, 3524
- out-of-distribution detection
 - PDEs, 3438
- outcome supervision, 1146
- outcome-based rewards, 3051
- output space
 - structured, 3967
- output structure, 3503–3505
 - loss function implications, 3505
- output variance
 - generative models, 3992
- OV circuit, 4005–4006, 4165
 - copy head exercises, 4011
 - definition, 4006
 - low rank, 4006
- overfitting
 - early neural networks, 4236
 - in evolutionary search, 2212
 - membership inference, 1591
 - MOS, 4223
 - score matching, 1614
- overlap length
 - and spectral bias, 2795
 - selection, 2802
- overlap parameter, 698
- overlap-add, 772
 - audio generation, 884
 - coherence bound, 885
 - exercise, 896
- overlap-and-blend, 2798–2799
 - definition, 2798
- overlapping segments, 2791
- Overlong Reward Shaping
 - definition, 1152
 - exercise, 1164
- oversampling, 3805
- P vs NP, 2951
- P-DARTS, 1925
- P-glycoprotein, 3778
- P/poly, 1036
- PAB, 649
- PAC-Bayes bounds, 1865
 - architecture selection, 1866
 - NAS challenge, 1868
- PacGAN, 2714
 - discriminator, 2716
- PackNet, 1199–1201, 1959
 - as evolutionary pruning, 2250
 - capacity, 1200
 - definition, 1199
- PACS
 - data volume, 4449
- PACT, 2956
- PACT (proving and co-training), 3044
- padding
 - middle, 2808
- paediatric oncology
 - federated learning example, 4456
- Page, Larry, 2628
- PageRank, 2628
- pairwise correlation, 2700
- pairwise correlations
 - evaluation, 2834
- pairwise metrics, 2838–2840
- pairwise ranking, 2586–2587
- pairwise reranking
 - definition, 2490
- PaLM
 - training, 5135
- Palmer, Tim, 4290

- PAN (Projective Adversarial Network),
 - 3576–3578
 - computational cost, 3578
 - multi-organ, 3578
- pancreas segmentation, 3576
 - PAN, 3576
- Pang et al. attack, 1642
- Pangu-Weather, 3440, 4181, 4195,
 - 4254–4264
 - foundation model, 4241
 - Hurricane Lee, 4260
 - lead-time-specific models, 4257
 - performance benchmarks, 4259
 - smoothing bias, 4261
 - spectral damping, 4299
 - training, 4258
- PanNuke dataset, 3505
- Panoptic Quality, 3505
- panoptic segmentation, 3479, 3503–3505
 - computational pathology, 3505
 - definition, 3505
- parabolic PDE, 3303–3311
- paradigm shift
 - GANs to diffusion, 2758
 - search to generation, 3038
- parallel attention and FFN, 114
- parallel computing
 - GPU, 4934
- parallel pattern, 951
- parallel type-checking, 2998
- parallelism
 - combination strategies, 5058
 - comparison table, 5056
 - hierarchical generation, 701
- parallelism–depth trade-off, 1064
- parameter count
 - growth, 4925
 - memory requirements, 4925
- parameter efficiency
 - GMS framework, 3707
- parameter interference, 2819
- parameter plasticity, 1881
- parameter prefetching, 5065
 - FSDP, 4988
 - implementation, 4996
- parameter server
 - architecture, 4931
 - reduce-broadcast, 5032
- parameter sharing
 - architecture search, 1916–1919
 - channels, 3881
- parameter transmission vs. gradient
 - transmission, 4451
- parameter-efficient adaptation
 - definition, 3636
 - exercise, 3641
 - medical imaging, 3636–3637
- parameter-efficient fine-tuning,
 - 1126–1129, 1791, 4552
 - medical segmentation, 3716
 - video, 761
- parameterisation
 - equivalence, 640
 - sub-grid processes, 4209
 - v-prediction, 639
- parameters
 - partitioning, 5063
- parametric knowledge, 4068
- parametric synthesis, 767
- paramodulation, 3020
- parasitism
 - digital, 2240
- Pareto coevolution, 2077
- Pareto front
 - exercise, 2294
- Pareto frontier
 - architecture search, 2011
 - continual learning, 1238
 - definition, 2140
 - privacy and utility, 2707

- privacy–utility, 2845
- prompt optimisation, 2139
- tabular generation, 2827
- Pareto optimality
 - dynamic networks, 1956
 - hardware-aware NAS, 1995
 - in NAS, 1888
 - privacy-utility, 2253
- parity function, 1036
 - exercise, 1066
- Parke–Taylor formula, 3169
- Parseval’s theorem
 - STFT, 774
- parsing
 - generated text, 2811
- partial differential equations, 3191–3450
 - classification, 3195
- partial F-test, 4224
- partial factorisation, 544
- partial participation
 - federated learning, 4453
- partial volume effect, 3524
- participatory design
 - disability, 4701
 - exercise, 4706
- participatory evaluation, 4894–4895, 4920
- particle swarm optimisation
 - exercise, 1453
- partition
 - by cosets, 2996
- partition function, 132, 501
 - intractability, 132
- partition of unity, 2798
- patch discriminator, 413
- patch-based processing, 3507
- PatchGAN, 347, 3581
 - 3D, 3858
 - as texture loss, 347
 - exercise, 359
 - receptive field, 348
- PatchGAN discriminator, 3817
- patchification, 590–592
 - convolutional interpretation, 592
 - definition, 590
 - patch size trade-off, 592
- path patching, 3974–3981, 4121, 4137
 - definition, 3974
- path-connected, 5210
- pathologies of coevolution, 2067–2079
- pathology
 - generative divergence, 3911
- pathology map, 3911
- pathwise gradient, 228
- patient H.M., 2340
- patient-adaptive GPT, 3888
 - (, 3887
 - exercise, 3898
- patient-specific evaluation, 3921
- patient-specific fine-tuning, 3887
- patient-specific model
 - EEG, 3931
- pattern completion, 2340–2341
 - hippocampal, 2341
- pattern matching, 2985–2987
 - autoformalization, 3066
 - semantic-free, 4024
- Pauli exclusion principle, 3365
- PauliNet, 3366
- Paulson, Lawrence, 2945, 3001
- PC algorithm
 - time series, 3936
- PC-DARTS, 1925
- PCI DSS, 2887
- PCIe
 - bandwidth comparison, 4928
 - comparison with NVLink, 4936
- PCIe bus
 - communication hiding limitation, 4989

- PDE classification
 - elliptic, [3450](#)
 - hyperbolic, [3450](#)
 - parabolic, [3450](#)
- PDE conditioning, [3253](#)
 - domain-wise, [3253](#)
 - point-wise, [3253](#)
- PDE foundation model, [3322](#), [3472](#)
 - definition, [3435](#)
- PDE foundation models, [3253–3256](#), [3434](#)
 - synthesis, [3256](#)
- PDE residual
 - as guidance, [3275](#)
 - bounds for diffusion samples, [3291](#)
 - in diffusion training, [3277](#)
- PDE-constrained inverse problems, [3397–3427](#)
- PDE-constrained optimisation
 - neural control, [3873](#)
- PDE-Diffusion, [3331](#)
- PDEBench, [3416](#)
 - diffusion-reaction, [3306](#)
 - FNO baseline reproduction, [3468](#)
 - implementation exercise, [3426](#)
 - shallow water, [3323](#)
 - Transformer baseline, [3465](#)
- PDEformer, [3258–3259](#), [3436](#)
- PDFed, [5130](#)
 - overhead, [5130](#)
- PDM, [2415](#)
- Peano axioms, [2944](#), [2966](#), [2985](#)
- PEAQ, [918](#)
- PEFT, *see* parameter-efficient fine-tuning
- penalised logistic regression
 - SAPEMOS, [4228](#)
- Penfield, Wilder, [3776](#)
- Pentium FDIV bug, [3004](#)
- per-sample gradients, [1677](#)
- per-token loss, [1714](#)
- Perceiver IO
 - Aurora, [4378](#)
 - complexity analysis, [4382](#)
- percentile estimation
 - ensemble size, [4318](#)
- perceptual audio quality, [917–919](#)
 - limitations for generation, [919](#)
 - PEAQ, [918](#)
 - PESQ, [917](#)
 - ViSQOL, [918](#)
- perceptual distortion
 - minimisation, [4785](#)
- perceptual evaluation
 - frontier, [947](#)
- perceptual features
 - temporal consistency, [707](#)
- perceptual hashing, [4497](#)
- perceptual loss, [413](#)
 - exercises, [340](#)
 - lip synthesis, [4739](#)
- perceptual quality, [522](#)
- Perelman, Grigori, [3112](#)
- perfect reward signal, [3097](#)
- performance verification, [2213](#)
- performance paradox
 - weather forecasting, [4395](#)
- periodic features
 - distance, [2923](#)
- periodicity, [2789](#)
 - audio, [783](#)
- permeability field, [3297](#)
- permutation calibration, [2496](#)
- permutation equivariance, [2752](#)
 - proof, [3896](#)
 - set transformer, [3881](#)
- permutation invariance
 - graphs, [2696](#)
 - streamline set, [3890](#)
 - tabular, [2807](#)
- permuted MNIST, [1184](#)

- perplexity, 20, 432–433
 - as membership signal, 1713
 - definition, 432, 1713
 - federated language models, 4458
- persistence diagram
 - diagonal matching, 3865
 - metric space, 3875
 - Wasserstein distance, 3865
- persistence diagrams
 - EEG, 3950
- persistent homology, 3514
 - epilepsy, 3864
- personalisation
 - DreamBooth, 692
 - federated generative models, 4454
 - practical considerations, 694
 - regularisation, 695
 - summary, 696
 - training duration, 695
 - video diffusion, 689–696
- personalisation–generalisation trade-off, 4467
- personalised FedVAE, 4512–4513
 - definition, 4512
- personalised retrieval, 2415
- personality erasure
 - TTS, 4746
- personalized content
 - federated, 4621–4635
 - federated generation, 4628–4629
- personalized federated diffusion, 4645
- personalized federated learning
 - PEFT, 4565
- personalized medicine
 - epilepsy, 3931–3932
- perturbation curve, 3966
- perturbation-based methods, 3983
- PESQ, 917
- PET synthesis
 - federated, 4610
 - MRI-to-PET, 4611
- PETAL, 1754–1762
 - information loss, 1755
 - limitations, 1755
 - membership score, 1754
 - score decomposition, 1755
- Petals, 5124
 - block design, 5125
- Petrov–Galerkin method, 3239
- PFGM, *see* Poisson Flow Generative Models
- pharmaceutical
 - federated drug discovery, 4458
- pharmacoresistance
 - mechanisms, 3778
- phase artifacts, 935
- phase breaks, 2790
- phase change
 - in-context learning, 4026
 - moving interface, 3352
 - scaling analysis, 4028
- phase locking value, 3800
- phase reconstruction, 773
- phase spectrogram, 773
- phase synchrony, 3864
- phase transition, 132
 - superposition, 4043
- phase-contrast microscopy, 3509, 3662
- phase-field models, 3357–3364
- pHash-KT, 4497
- phenotypic convergence, 2159
 - definition, 2160
- phenytoin, 3775
- Phi (IIT), 4148
- philosophy
 - AI introspection, 4147
 - neural PDE solvers, 3445–3449
- philosophy of explanation, 3963
- philosophy of mind
 - AI connection, 3963

- PHiSeg, 3563–3566
 - brain lesion segmentation, 3565
 - comparison with Probabilistic U-Net, 3565
 - sampling exercise, 3571
- phoneme alignment, 4751
 - forced, 4751
- photonic crystals, 3350
- photonics
 - generative models, 3350
- Phys-Instruct, 3289, 3389
- physical accessibility, 4864–4875
- physical consistency
 - diffusion models, 4328
 - downscaling warning, 4360
 - problem, 3283
 - score-based generation, 3282
- physical guarantees, 3437–3438
- physical law
 - limits on evolution, 2264
- physical plausibility, 522
 - desideratum, 523
 - open problem, 753
- physical-layer authentication
 - federated, 4623
- physics guarantee, 3391
- physics–learning duality, 3449
- physics-attention, 3245
 - block, 3246
 - mechanism, 3246
- physics-aware generation, 718–726
- physics-constrained generation, 753–754
 - formal statement, 753
- physics-constrained machine learning, 4409
- physics-embedded neural network, 4365
- physics-free learning
 - precipitation, 4241
- physics-guided diffusion, 719–726
 - algorithm, 723
 - computational cost, 720
 - definition, 719
 - exercise, 726
 - guidance schedule, 724
- physics-informed DeepONet, 3216
- physics-informed diffusion model, 3467
- physics-informed diffusion models,
 - 3277–3295
 - residual reduction, 3279
 - theory, 3279
 - training objective, 3277
 - vs. guidance, 3294
 - weighting schedule, 3278
- physics-informed distillation, 3354, 3390
- physics-informed neural networks,
 - 2210, 3203–3210, 3417
 - 2D heat equation, 3460
 - benchmarking, 3418
 - Burgers’ equation, 3451
 - causality, 3207
 - failure modes, 3206
 - failure on discontinuous IC, 3307
 - failure on shocks, 3314
 - gradient pathologies, 3207
 - heat equation, 3307, 3451
 - loss balancing, 3206
 - segmentation, 3731
 - shock resolution, 3315
 - Transformer backbone, 3260
 - wave equation, 3451
- physics-informed segmentation, 3731
- PhysioNet
 - seizure challenge, 3923
- Pi type, 2972
- Pi-type, 2985
- piano roll, 905
- PIDDM, 3354, 3390
 - distillation strategy, 3355
 - porous medium equation, 3356
- PIDE, 3013

-
- PIDMs, *see* physics-informed diffusion models
 - Piggyback, 1199–1201
 - definition, 1200
 - Pinakes, 2625
 - Pinecone, 2379
 - PINNacle, 3418
 - PINNs, *see* physics-informed neural networks
 - advection equation, 3464
 - causal training, 3457
 - spectral bias, 3457
 - PINNsFormer, 3260–3263, 3315
 - ablation, 3262
 - architecture, 3261
 - motivation, 3260
 - Pseudo Sequence Generator, 3260
 - wavelet activation, 3261
 - Pinsker’s inequality, 79
 - PINTO, 3361–3364
 - Cahn–Hilliard application, 3362
 - generalisation, 3362
 - pipeline boundaries
 - activation storage, 5101
 - pipeline bubble, 4994, 5135
 - bounded, 5141
 - bubble fraction formula, 4995
 - exercise, 5105
 - micro-batch, 5140
 - pipeline bubbles
 - diffusion model, 5175
 - pipeline parallelism, 5050–5060
 - autograd challenge, 5098
 - batch size, 5058
 - bubble fraction, 5173
 - communication hiding, 4994–4999
 - computation-communication overlap, 4984
 - definition, 5050
 - diffusion models, 5085
 - fault-tolerant, 5125
 - formal definition, 5050
 - InfiniBand placement, 4949
 - pipeline schedule
 - 1F1B, 5051
 - comparison, 5052
 - GPipe, 5051
 - interleaved, 5052
 - pipeline scheduling
 - diffusion models, 5167
 - pipeline speedup, 2796
 - pipeline stage
 - activation communication, 4994
 - layer count, 5141
 - pipelined-parallel architecture, 2788
 - pipelined-parallel generation, 2791
 - PiT, 3250
 - attention definition, 3252
 - attention structure, 3251
 - connection to numerical methods, 3252
 - hierarchical attention, 3251
 - mechanism, 3251
 - positional bias, 3251
 - pitch
 - periodic structure, 783
 - pitch class, 813
 - Pitts, W., 3969
 - Pix2Pix, 12
 - pix2pix, 346–348, 3580–3582
 - architecture diagram, 3581
 - epilepsy, 3857
 - exercise, 359
 - framework, 3580
 - L1 vs L2 loss, 347
 - loss function, 347, 3580
 - PatchGAN, 347
 - Pix2Video, 716
 - pixel reconstruction
 - DLG, 4592

- pixel-space diffusion, 534
 - computational cost, 557
- pixel-wise uncertainty
 - limitations, 3558
 - underestimates correlations, 3558
- PixelCNN, 7, 415, 439–441
 - blind spot, 439
 - horizontal stack, 440
 - vertical stack, 440
- PixelRNN, 437–439
 - Diagonal BiLSTM, 439
 - Row LSTM, 438
- pixels
 - continuity, 4086
- PixelSNAIL, 444
- Plücker coordinates, 676–678
 - advantages, 676
 - epipolar constraint, 682
 - exercise, 759
 - mathematical properties, 682
 - zoom, 683
- Plackett-Luce model, 2588
 - consistency, 2588
- Plain Language
 - dyslexia, 4793
- plan-then-diffuse, 731
- planar flow, 238
 - Jacobian determinant, 238, 266
- planetary boundary layer, 4210
- planning
 - as search, 1290–1294
 - formal definition, 1290
 - SMT-based, 3023
- planning target volume, 3482
- plasticity, 1181
 - architecture, 1882
- plasticity rate, 1851
- plate notation, 128
 - exercise, 142
- Plato
 - anamnesis, 2333–2334
- Platonic representation hypothesis, 2625
- plug-and-play methods, 3402
- pluralistic alignment, 4135–4137
- POD-DeepONet, 3217
- POET, 2242–2243, 2284
- point annotations, 3520
- point prompts, 3519
- point-to-point communication
 - definition, 4961
 - MPI, 4961
- PointNet
 - streamline encoding, 3890
- pointwise reranking
 - definition, 2488
- Poisson distribution, 31
- Poisson equation, 3295–3303
 - DeepONet implementation, 3462
 - definition, 3296
 - Green’s function representation, 3234
 - Transformer operators, 3301
 - well-posedness, 3296
- Poisson Flow Generative Models, 3309–3311
 - connection to diffusion, 3310
 - definition, 3309
- policy
 - architecture search, 1912
- policy gradient, 1046–1047
 - architecture search, 1913
 - derivation, 1914
 - multi-agent, 1458
 - NAS formulation, 1913
 - REINFORCE, 1046
- policy network
 - AlphaGo, 3114
 - pre-training, 3118
 - theorem proving, 3100
- policy paradigm, 3038

- policy recommendations
 - accessibility AI, 4906–4908
- Polu, Stanislas, 3036
- polynomial arithmetic
 - automation, 2989
- polynomial factorisation, 3031
- polynomial ideals, 3033
- polynomial preconditioning, 4987
- polyp detection
 - open-vocabulary, 3696
- polyp segmentation, 3514, 3611
 - synthetic data, 3582
- polyrith tactic, 2989–2990
- polysemanticity, 3969, 4038–4046
 - causes, 4039
 - InceptionV1 example, 3969
- polysemy, 2335
- POMDP
 - agent formulation, 1445
 - for agents, 1279
- pooling, 98
- popularity bias
 - in LLM recommendations, 2580
- population genetics, 2279
- population-based training
 - forgetting protection, 2250
- porous medium equation, 3356
- portfolio simulation, 2889
- position bias
 - in LLM recommendations, 2580
 - in reranking, 2496
- position interpolation, 1081–1082
 - exercise, 1110
- position-induced attention, 3251
- Position-induced Transformer, *see* PiT
- positional encoding, 102, 1505–1510
 - 3D, 607–613
 - definition, 608
 - ALiBi, 1507
 - extrapolation, 1507
 - exercise, 612
 - for diffusion, 2764
 - generalisation, 611–613
 - learnable vs fixed, 612
 - learned, 1506
 - motivation, 607
 - NTK-aware scaling, 1508
 - definition, 1508
 - rotary (RoPE), 104
 - base frequency, 105
 - definition, 104
 - sinusoidal
 - 3D, 608
 - inner product, 609
 - limitations, 609
 - YaRN, 1509
 - definition, 1509
- positional encodings
 - for PDE domains, 3235
- positional structure
 - PDE attention, 3250
- positron emission tomography, 3506
 - segmentation, 3481
 - synthesis, 3857–3860
- post-processing
 - NWP, 4221–4232
- post-processing attacks, 936
- post-processing immunity
 - differential privacy, 4456
- post-processing property, 2737
- post-processing vs forecasting
 - distinction, 4233
- post-stroke rehabilitation
 - language, 4761
- post-training
 - introspection, 4144
- post-training attack
 - federated learning, 4595
- posterior
 - intractability, 388

- posterior collapse, [260](#)
- posterior collapse, [397–398](#), [4515](#)
 - definition, [397](#)
 - exercise, [400](#)
 - federated VAEs, [4515](#)
 - segmentation, [3495](#)
- posterior distillation
 - FedVAE, [4465](#)
- posterior distribution, [46](#), [166](#)
 - importance, [169](#)
- posterior ensemble
 - iEEG, [3855](#)
- posterior predictive check, [259](#)
- posterior sampling, [3401](#)
 - source separation, [890](#)
- posterior variance
 - diffusion, [488](#)
- postictal EEG, [3794](#)
- potential function, [131](#)
- potential game, [316](#)
- potential-based reward shaping, [3105](#)
- Potter, Mitchell, [2062](#), [2282](#)
- power budget
 - implantable device, [3931](#)
- power iteration, [298](#)
 - for spectral norm, [298](#)
- power spectral density, [2794](#), [3794](#)
- power spectrogram, [773](#)
- power spectrum
 - atmospheric fields, [4352](#)
 - recovery by diffusion, [4354](#)
- PPO, [1047](#), [1133–1137](#)
 - clipped objective, [3102](#)
 - critic initialization, [1137](#)
 - for language models, [3053](#)
 - mathematical reasoning, [3102–3104](#)
 - per-token reward, [1135](#)
 - PPO-ptx, [1135](#)
 - pseudocode for theorem proving, [3111](#)
- PQ, *see* product quantization
- pragmatic accessibility, [4698](#)
- pre-ictal detection
 - worked example, [3868](#)
- pre-ictal state
 - cross-channel synchrony, [3879](#)
 - detection, [3866](#)
- pre-market validation, [3723](#)
- pre-norm
 - gradient flow, [1552](#)
- pre-surgical evaluation, [3773](#)
 - Phase 1, [3779](#)
 - Phase 2, [3780](#)
- pre-surgical planning
 - connectivity mapping, [3893](#)
- precipitation
 - diffusion nowcasting, [4322](#)
 - forecast blurring, [4195](#)
 - multi-scale structure, [4322](#)
- precipitation forecast
 - 12-hour, [4240](#)
- precipitation nowcasting
 - ConvLSTM, [4238](#)
- precipitation occurrence
 - Bayesian network, [4235](#)
- precipitation probability
 - MOS, [4224](#)
- precipitation probability of occurrence
 - PoP, [4225](#)
- precision, [3545](#)
 - k , [2534](#), [2604](#)
 - k , [2359](#)
- precision agriculture
 - AI weather forecasting, [4390](#)
- precision and recall, [22](#)
 - generative models, [284](#)
- Predetermined Change Control Plan, [3939](#)
- predicate logic, [3016](#)
- predicates, [3020](#)

- predictability, [4183–4202](#)
 - beyond horizon, [4188](#)
 - convective, [4362](#)
 - theoretical limit, [4187](#)
 - weather, [4394–4410](#)
- predictability horizon, [4184](#)
 - theoretical limit, [4187](#)
- predictand
 - definition, [4223](#)
- predictands
 - MOS, [4224](#)
- prediction evolution
 - across layers, [4014](#)
- predictive DBS, [3916](#)
- predictive distribution
 - variational, [198](#)
- predictive policy improvement, [1372](#)
- predictive retrieval, [2623](#)
 - definition, [2623](#)
- predictive text
 - AAC, [4760](#)
 - motor impairment, [4701](#)
- predictor interactions
 - MOS, [4230](#)
- predictor selection, [4223](#)
- predictors
 - NWP, [4222](#)
- preemptible instances, [5145](#)
- preemptible training, [5149](#)
- preemption handler, [5150](#)
- preference data, [1129](#)
- prefix conditioning, [2799](#)
- prefix tuning
 - federated, [4460](#)
- preictal EEG, [3794](#)
- Premarket Approval
 - AI, [3939](#)
- PREMIA, [1762–1771](#)
 - alignment process, [1762](#)
 - alignment vulnerabilities, [1766](#)
- attack framework, [1764–1771](#)
- DPO leakage, [1763](#)
- implications for alignment safety, [1768](#)
- membership signal, [1764](#)
- optimal weights, [1765](#)
- pipeline, [1765](#)
- preference pair membership
 - inference, [1763](#)
- RLHF vs DPO, [1767](#)
- RLHF vs DPO signal comparison, [1768](#)
- score function, [1765](#)
- premise retrieval
 - pipeline, [3094](#)
- premise selection, [2956](#), [3035](#)
 - neural, [3035](#)
- prenatal neural development, [1846](#)
- Presburger arithmetic, [2989](#)
- pressure levels
 - 13, [4312](#)
- pretraining
 - codomain attention, [3255](#)
 - cross-domain, [3375](#)
 - introspective capability, [4143](#)
 - multi-physics, [3254](#)
- previous token head
 - positional embedding analysis, [4027](#)
- previous token heads, [4021](#)
 - in IOI circuit, [4031](#)
 - walkthrough, [4034](#)
- primality testing
 - verified, [2989](#)
- prime number
 - in Lean, [2976](#)
- Prime Number Theorem
 - formalization, [3001](#)
- primitive equations, [4202–4221](#)
 - conservation, [4207](#)

- definition, [4206](#)
- derivation exercise, [4410](#)
- NeuralGCM, [4280](#)
- principal component analysis, [386–387](#)
- Principia Mathematica, [2944](#), [3079](#)
 - automated proofs, [2940](#)
- prior, [388](#)
- Priority Flow Control, [4968](#)
- Priority Learning Channel, [4731](#)
 - reward design, [4734](#)
- PRISM
 - comparison, [4602](#)
 - mask search, [4600](#)
 - mask transmission, [4600](#)
 - server initialisation, [4600](#)
 - sparsity exercise, [4608](#)
- PRISM defense, [4599–4603](#)
 - architecture figure, [4602](#)
- Prisma framework, [4085–4095](#)
- Prisma toolkit, [4087–4088](#)
 - architectures, [4087](#)
 - evaluation, [4087](#)
- prismatic fuel element, [3374](#)
- Prisoner's Dilemma, [2281](#)
 - coevolutionary forgetting, [2072](#)
- privacy, [19](#)
 - accessibility data, [4704](#)
 - centralisation failure, [4448](#)
 - comparison of mechanisms, [4521–4523](#)
 - comparison table, [4522](#)
 - data sharing, [2690](#)
 - EEG data, [3940–3941](#)
 - evaluation, [2705](#)
 - evolutionary perspective, [2253–2254](#)
 - federated generative models, [4590](#)
 - federated learning, [5116](#)
 - gradient leakage, [4453](#)
 - healthcare, [2885](#)
 - healthcare data, [2687](#)
 - leakage in federated learning, [2883](#)
 - LLM-based generation, [2829](#)
 - LLM-based methods, [2859](#)
 - medical data, [4814](#)
 - missing evaluation, [2837](#)
 - neuroimaging data, [3781](#)
 - noise schedule, [1630](#)
 - re-identification risk exercise, [3830](#)
 - regulatory constraints, [5119](#)
 - retrieval systems, [2612](#)
 - synthetic data, [4661](#)
 - definition, [4662](#)
 - synthetic data as shield, [3827](#)
 - synthetic EEG, [3826–3828](#)
 - synthetic medical data, [3590](#)
 - tabular data, [2898–2904](#)
 - VAE defense, [4517–4526](#)
- privacy accounting, [1673–1674](#)
 - comparison, [1674](#)
- privacy amplification
 - example, [1673](#)
 - federated learning, [4454](#)
 - subsampling, [1673](#)
 - through subsampling, [4464](#)
- privacy amplification by subsampling, [4464](#)
- privacy analysis
 - EEG generative model, [3951](#)
- privacy attack
 - federated VAE
 - research challenge, [4681](#)
- privacy attacks
 - federated learning, [4589–4599](#)
- privacy defense
 - attack matrix table, [4607](#)
 - comparison, [4602–4603](#)
 - comparison table, [4603](#)
- privacy defenses
 - federated learning, [4599–4608](#), [4609](#)

- privacy guarantee
 - definition, 2706
 - generative model, 3940
- privacy metrics, 2844–2846
- privacy paradox
 - generative AI, 4444
- privacy–utility trade-off, 2845
- privacy–utility tradeoff, 2738, 4496
- privacy-preserving data sharing, 2889–2891
- privacy-preserving embeddings, 2480
- privacy-preserving machine learning, 4443–4467
- privacy-preserving retrieval, 2620–2621
- privacy-preserving synthesis, 4456
- privacy-preserving training, 3717
- privacy-quality tradeoff, 1632–1636
 - exercises, 1636
 - formal statement, 1632
- privacy-utility tradeoff, 1600, 2705–2708
 - communication, 4668
 - conjecture, 1689
 - theoretical limits, 2906
 - visualization, 2707
- privacy-utility-fairness trilemma, 1829
 - empirical evidence, 1829
- private information retrieval, 2620
- probabilistic benchmark
 - weather, 4397
- probabilistic calibration, 4193–4202, 4228
 - definition, 4193
 - exercise, 4243
- probabilistic context-free grammar (PCFG), 4127
- probabilistic forecast, 4189
 - SAPEMOS, 4228
- probabilistic forecasting
 - value, 4371
- probabilistic graphical model, 125
 - generative models, 139
- probabilistic method, 3039
- probabilistic outputs
 - universal framework, 3712
- probabilistic retrieval, 2626
- probabilistic segmentation, 3557–3571
 - framework, 3494
 - model definition, 3557
 - problem formulation, 3557
- Probabilistic U-Net, 3560–3570
 - combining module, 3560
 - ELBO, 3747
 - implementation challenge, 3571
 - KL divergence, 3570
 - LIDC-IDRI, 3563
 - posterior network, 3560
 - prior network, 3560
- probabilistic U-Net, 3493
- probabilistic weather forecasting
 - Bayesian, 4235
- probability, 29
- probability distribution
 - categorical distribution, 32
 - multinomial distribution, 32
 - probability density function, 30
 - probability mass function, 29
- probability flow ODE, 645, 2771
- Probability of Detection
 - POD
 - nowcasting, 4239
- probability of improvement, 2227
- probability of precipitation
 - PoP
 - definition, 4225
- probe noise, 1653
 - signal amplification, 1654
- probing
 - deception detection, 4139–4140
 - deceptive alignment, 4134
- probing classifier

- video transformers, [4119](#)
- probit regression
 - precipitation, [4225](#)
- problem finding
 - vs problem solving, [3180](#)
- problem formulation
 - weather forecasting, [4243](#)
- problem-solution coevolution,
 - [2219–2220](#)
 - definition, [2219](#)
- procedural memory, [1175](#)
 - in agents, [1282](#)
- process group
 - DDP, [5023](#)
 - NCCL, [5023](#)
- process groups
 - PyTorch, [5100](#)
- process reward models, [3051](#)
- process supervision, [1146](#)
- product quantization, [2378–2379](#)
 - codebook learning, [2378](#)
 - definition, [2378](#)
- product type, [2947](#)
- program evolution
 - sample-efficient, [2144–2153](#)
- program extraction, [3000](#)
- program synthesis
 - connection to theorem proving,
 - [2947](#)
- progress curve, [3059](#)
 - mathematical benchmarks, [3061](#)
- progress signal
 - definition, [3089](#)
- progressive DARTS, [1925](#)
- progressive distillation, [649](#), [3729](#)
 - algorithm, [878](#)
 - audio, [878–880](#)
 - definition, [649](#)
 - error accumulation, [880](#)
 - number of rounds, [879](#)
 - schedule example, [879](#)
- progressive growing, [351–352](#), [1870](#),
 - [1957–1959](#)
 - exercise, [360](#)
 - fade-in, [351](#)
 - minibatch standard deviation, [352](#)
 - stability, [351](#)
 - training schedule, [351](#)
- progressive neural networks, [1198–1199](#),
 - [1959](#), [1971](#)
 - definition, [1198](#)
 - growth, [1199](#)
- progressive shrinking, [1909](#), [1957](#)
- progressive training
 - resolution curriculum, [655](#)
- progressive widening, [3129](#)
- projection
 - feature fixing, [2778](#)
 - manifold, [2778](#)
- projective geometry, [3123](#)
- Prometheus
 - evaluation paradigm, [4890–4891](#)
 - LLM-as-a-judge, [4878–4879](#)
- prompt engineering
 - constraints, [2822](#)
 - evolutionary, [2134](#)
 - for evolution, [2105](#)
 - medical SAM, [3637–3639](#)
 - medical segmentation, [3749](#)
 - negative prompt paradox, [3638](#)
- prompt enhancement, [683–689](#)
 - definition, [684](#)
 - encoding considerations, [686](#)
 - evaluation, [688](#)
 - information-theoretic view, [685](#)
 - LLM rewriting, [684](#)
- prompt evolution, [2133–2144](#)
- prompt generation
 - automated, [3697–3698](#)
- prompt sensitivity, [2129](#)

- prompt-mask distribution, 3699
- Prompt2LangGloss, 4731
- PROMPT2SIGN dataset, 4729–4732
 - description, 4730
- promptable segmentation, 3618
 - ambiguity, 3620
 - definition, 3619
 - exercise, 3642
 - paradigm, 3619
- Promptagator, 2505–2506
 - round-trip filter, 2505
 - scientific literature, 2506
- prompting
 - standard failures, 3047
- proof
 - multiple valid proofs, 2935
 - understanding, 2933
- proof as game, 3035
- proof assistants, 2934, 3080–3095
 - autoformalization, 3064
 - integration, 3076
- proof automation, 2987–2999
- proof by contradiction, 2935, 2981
- proof checker
 - as reward signal, 2934
 - kernel, 2943
- proof elegance, 3028
- proof gaps, 2936
- proof irrelevance, 2966
- proof kernel, 3080
 - as reward oracle, 3096
- proof progress prediction, 3088
- proof reconstruction, 3027
- proof search, 3040–3043
 - as graph exploration, 3040
 - as RL, 2961
 - automated, 3085
 - combinatorial, 3016
 - neural guided, 2956
- proof search tree, 3127
- proof sketches, 3039
- proof state, 2987, 3081
 - as MDP state, 3096
 - deduplication, 3129
 - definition, 3116
- proof strategy
 - human-guided, 3025
- proof styles, 2935
- proof transfer, 3008–3009
- proof verification
 - feedback, 3090
- ProofGPT, 3068
- Prop (universe), 2966
- proper orthogonal decomposition, 3215, 3226
- proper scoring rule, 3566, 4229
 - evaluation suite, 4420
 - exercise, 4415
- proportional sampling, 2866
- propositional equality, 2981
- propositional extensionality, 2974
- propositional logic, 3017–3018
- propositions as types, 2946, 2965, 2984
- PROSE-FD, 3322, 3436
- prostate segmentation
 - inter-observer variability, 3484
- protanopia, 4770
- Protected Health Information, 4446
- protein design, 14, 2875
 - generative models, 2893
- protein engineering, 2893
- protein folding, 14
- protein structure prediction, 2893
- provenance
 - EEG generation, 3933
 - frontier, 949
- proximal gradient method, 5276
- proximal initialisation, 1664
 - query complexity, 1664
- proximal operator, 5274

- of ℓ_1 norm, 5279
- proximal policy optimisation, 1047
- pruning, 1539–1543
 - as NAS, 1997
 - definition, 1540
 - iterative, 1959
 - iterative magnitude pruning, 1541
 - magnitude, 1540
 - structured, 1540, 4817
 - unstructured, 1540, 4817
 - wearable AI, 4816
- PSD, *see* power spectral density
- pseudo sequence generator, 3260
 - definition, 3261
- pseudo-healthy synthesis, 3841–3842, 3911
 - BraTS, 3920
 - epilepsy, 3955
 - evaluation exercise, 3928
 - exercises, 3848
 - key idea, 3841
 - pipeline figure, 3843
- pseudo-labels, 3520
- pseudo-likelihood, 1746
 - definition, 1746
- pseudo-prompt, 3696
- pseudo-rehearsal
 - EEG, 3933
- PSNR, 22, 332, 561
 - fMRI synthesis, 3854
 - limitations, 3924
 - MRI evaluation, 3923
- psychoacoustic mapping
 - exercise, 4715
- psychoacoustics
 - accessibility, 4708
- publish–subscribe systems, 1276
- PubMedCLIP, 3520
- PUCT algorithm
 - modifications for proofs, 3129
- PUCT formula, 3119
- Pulay forces, 3369
- pulmonary nodule detection, 3510
- pulse sequence
 - MRI, 3481
- punctuated equilibrium, 2032–2034
- pure literal elimination, 3018
- push-forward distribution, 57
 - notation, 57
- Putnam competition, 3150
- Putnam, Hilary, 3018
- Pyramid Attention Broadcast, 649
- PySTEPS, 4363
- Pythagorean theorem
 - for vectors, 305
- Python
 - as genome representation, 2182
- PyTorch
 - distributed autograd, 5100
 - DistributedDataParallel, 4974–4976, 4988
 - FSDP, 5065–5069
 - Gloo, 4976
 - gradient bucketing, 4991–4999
 - torch.distributed.checkpoint, 5148
- PyTorch DDP, 5023
 - bucket_cap_mb, 4991
- PyTorch Elastic
 - algorithm, 5149
 - rendezvous, 5149
- PyTorch FSDP, 4927
 - backward_prefetch, 4997
 - forward_prefetch, 4997
- Q-composition, 4006
 - definition, 4008
- Q-Former, 2413
 - definition, 2413
- QD-VAE, 2089
- QED Manifesto, 3079

- QiankunNet, 3366–3371
 - chemical accuracy, 3368
- QK circuit, 4004–4005, 4165
 - definition, 4005
 - low rank, 4005
 - rank exercises, 4011
- QLoRA, 694, 1127, 1538
 - definition, 694
 - membership inference, 1792
 - memory savings, 694
- quadratic bottleneck, 548
 - proposition, 549
- quality of life
 - accessibility AI, 4905
- quality skew, 4570
- quality vs. diversity, 1369
- quality-diversity, 2079–2094, 2284, 2301
 - and generative models, 2088–2094
 - and LLMs, 2089
 - at scale, 2284
 - exercise, 2290
 - generative model exercise, 2090
 - history, 2093
 - in Core War, 2157
 - information-theoretic view, 2094
 - open problems, 2090
 - paradigm, 2080–2081
 - philosophy, 2080
 - problem definition, 2081
 - theorem proving, 3109
- quality-privacy trade-off
 - EEG, 3941
- quality-speed trade-off, 877
- quantifier elimination, 2989
- quantifiers, 3020
- quantisation
 - activations, 5077
 - audio, 770
 - EEG model, 3931
- quantisation error
 - RVQ, 950
 - visual vocabulary, 2405
- quantitative phase imaging, 3663
- quantity skew, 4570
 - definition, 4571
- quantization, 1535–1539
 - AWQ, 1537
 - definition, 1535
 - GPTQ, 1537
 - INT8, 4816, 4817
 - post-training, 1536, 4817
 - QLoRA, 1538
 - uniform, 1536
 - zero-point, 1536
- quantum mechanics
 - neural methods, 3364–3371
- quasi-identifiers, 2889
- quasi-periodicity, 2789
- quasiconvex function, 5256
- query efficiency, 1664
- query expansion
 - exercise, 2637
- query generation, 2529
- query-by-committee, 2225
- query-by-humming, 2442–2443
 - exercises, 2446
- question answering
 - MDM vs ARM, 4110
- Quiroga, R., 3969
- quotient group, 2997
- quotient types, 2974
- Qwen, 114
 - 2.5, 116
- R-Drop, *see* regularization, R-Drop
- R-GAP attack, 4592
- Rényi differential privacy, 2738
 - composition, 1674
- Rabe, Markus, 3036
- rack

- data centre unit, 4927
- radar artifacts, 4363
- radar image
 - prediction, 4238
- radar nowcasting
 - deep learning, 4238
- Radford, Alec, 2630
- radial flow, 239
- radiation parameterisation, 4209
- radiation pneumonitis, 3483
- radiation therapy
 - planning, 3553
- radiation therapy planning, 3508
- RadioDiff-k, 3345
- radiologist
 - eye movement, 3688
- radiologist shortage, 3484
- radiotherapy
 - planning example, 3482
 - segmentation, 3477
- RAG, *see* retrieval-augmented
 - generation, *see*
 - retrieval-augmented generation, 2629
 - evaluation metric
 - challenge, 2639
 - exercise, 1113, 2634
 - explosion, 2630
 - multimodal, 3907
 - privacy risks, 2613
- RAG evaluation, 2533–2535, 2602–2604
 - component attribution, 2607
 - component-wise vs end-to-end, 2602
 - faithfulness failure, 2603
 - formal framework, 2603
 - pipeline design, 2608
 - quality decomposition, 2603
- RAGAS, 2533, 2602–2604
 - example, 2534
- RainScaleGAN, 4355
 - exercise, 4360
 - loss function, 4356
- raised cosine
 - properties, 2792
 - spectral properties, 2794
- raised cosine blending, 2792
- Ramanujan Machine, 3176
- RAMPNET, 4719
 - definition, 4720
 - error analysis exercise, 4724
 - precision recall, 4911
- random forests
 - decision trees, 3529
 - ensemble prediction, 3529
 - medical segmentation, 3528–3530
 - segmentation, 3487
- random Fourier features, 2840
- random hyperplane hashing, 2376
- random measurement matrix, 3821
- random network
 - PRISM, 4599
- random seed compression, 4479
- random splitting
 - inflation, 3922
- random variable, 29
 - definition, 29
- range constraints, 2703, 2777, 2825
- rank
 - MPI, 4960
- rank histogram, 4194
 - ensemble calibration, 4315
 - exercise, 4201
- RankBoost, 2586
- Rankine–Hugoniot condition, 3312
- RankSVM, 2586, 2628
- RANSAC, 2399
- Rao–Blackwellization, 222
- rapid intensification
 - definition, 4399

-
- RapidParc, 3890–3891
 - rare categories
 - diffusion coverage, 2769
 - rare disease
 - federated augmentation, 4616–4618
 - medical imaging, 4445
 - rare events, 2690
 - rare subtype
 - federated augmentation, 4457
 - raster scan, 438
 - rate coding, 1857
 - capacity, 1858
 - rate-distortion
 - watermark capacity, 954
 - watermarking, 931
 - rate-distortion function
 - retrieval, 2556
 - rate-distortion theory, 1859, 2473
 - EEG, 3931
 - neural tradeoff, 1859
 - retrieval, 2556–2565
 - curve, 2677
 - rationalisation, 3963
 - Ravuri et al. (2021), 4364
 - Ray, Thomas, 2240
 - RBF kernel
 - MMD, 4602
 - RCVAE
 - Bi-LSTM, 3826
 - RDMA, 4565
 - federated learning, 4479
 - figure, 4968
 - Remote Direct Memory Access, 4966–4968
 - WAN, 5132
 - Re-AudioLDM, 2453, 2541
 - Re-Con-GAN, 3585–3587
 - architecture, 3586
 - clinical deployment, 3587
 - re-noising, 2780
 - ReAct, 1280
 - ReAct agent, 2181
 - reaction-diffusion, *see* diffusion-reaction equations
 - real analysis
 - formalization, 2992–2995
 - in Lean, 2977
 - real-time
 - applications, 3434
 - definition, 3433
 - real-time audio generation, 945–946
 - formal statement, 945
 - real-time factor, 874
 - definition, 874
 - hardware dependence, 874
 - real-time processing
 - accessibility, 4810–4820
 - CVD, 4779–4789
 - real-time segmentation, 3514, 3728
 - real-time simulation, 3433–3434
 - real-time synthesis, 4755
 - real-time video generation, 751
 - conjecture, 751
 - realism
 - measuring, 2834
 - RealNVP, 239, 4460
 - multi-scale architecture, 454
 - reanalysis, 4217
 - definition, 4217
 - reasoning, 1027–1071
 - as continual learning, 1243
 - historical timeline, 1066
 - open problems, 1064
 - synthesis, 1062–1065
 - reasoning model, 3147
 - reasoning models, 3050–3064
 - recall, 3545
 - image-level, 3842
 - k , 2534, 2604
 - k , 2359

- RECAP, [2453](#), [2542](#)
- receptive field
 - GNN, [2751](#)
 - U-Net, [3547](#)
 - U-Net vs. DiT, [842](#)
- Rechenberg, Ingo, [2041](#), [2280](#)
- Reciprocal Rank Fusion, [2372](#)
- RECIST criteria, [3480](#)
- recognition network, [230](#)
- recommendation, [2574–2585](#)
 - as retrieval, [2574–2585](#)
 - content as regularizer, [2579](#)
 - debiasing LLM, [2584](#)
 - diversity, [2584](#)
 - formal framework, [2575](#)
 - LLM bias analysis, [2580](#)
 - LLM-based, [2579–2585](#)
 - retrieval connection, [2575](#)
- recommendation system
 - federated, [4510](#)
 - movie, [4514–4515](#)
- reconstruction error
 - anomaly detection, [3839](#)
- reconstruction loss
 - binary cross-entropy, [392](#)
 - exercise, [400](#)
 - MSE, [392](#)
 - per-column, [2754](#)
 - tabular GAN, [2735](#)
- rectified flow, [635–644](#), [3385](#)
 - advection equation exercise, [3294](#)
 - audio, [871](#)
 - convergence, [638](#), [871](#)
 - definition, [638](#)
 - PDE operators, [3287–3295](#)
 - definition, [3287](#)
 - practical considerations, [639](#)
 - straighter trajectories, [637](#)
- rectified flow-based operator, [3385](#)
- rectified linear unit
 - vanishing gradient, [4242](#)
- recurrence, [1028–1035](#)
 - iterative computation, [1029](#)
- recurrent dynamics
 - atmosphere, [4237](#)
- recurrent neural network, [100](#)
- recurrent state
 - conditioning, [727](#)
- recurrent VAE
 - temporal EEG, [3826](#)
- recursion, [1027–1071](#)
- recursive halving-doubling, [4946](#), [5035](#)
 - cost, [4954](#)
- recursive meta-evolution, [2201](#)
- recursive reasoning, [1056–1062](#)
 - exercise, [1069](#)
- recursive self-improvement, [1987](#), [2187](#), [2258–2260](#)
 - definition, [2259](#)
 - diagram, [2015](#)
 - formal definition, [2187](#)
 - practical concerns, [2260](#)
 - simulation, [2267](#)
 - thermodynamic limits, [2259](#)
 - through NAS, [1987](#)
- Red Queen dynamics
 - convergence, [2031](#)
 - definition, [2030](#)
- Red Queen effect, [2067–2069](#), [2220](#), [2252](#)
 - exercise, [2289](#)
- Red Queen hypothesis, [1373](#), [2030–2032](#), [2154–2170](#)
 - biological origins, [2154](#)
- RED-DiffEq, [3404](#)
 - algorithm, [3404](#)
- red-teaming
 - evolutionary, [2166](#)
- Redcode, [2155](#)
- reduce-scatter, [4972](#)
 - exercise, [5040](#)

- ring all-reduce, 4988
- ZeRO-3, 5064
- reduced basis methods, 3215, 3226
- reduced Gaussian grid, 4214
- redundant computation
 - fault tolerance, 5150
- Reed-Solomon coding, 5151
- reference counting
 - Lean 4, 2998
- reference policy, 3055
- referential integrity, 2698
 - constraints, 2703
- ReFL, 1158
 - objective, 1158
- reflection tokens, 2527
- reflective coevolution, 2133–2144
- Reflexion, 1052
- reflow procedure, 871
 - music generation, 873
- REFUGE dataset
 - MGDS, 4596
- refutation, 2955
- refutation completeness, 3017
- region growing, 3524–3525
 - algorithm, 3525
 - leakage problem, 3525
- region-of-interest loss, 3858
- regional MOS, 4227
- register file
 - GPU, 4935
- register spilling
 - performance, 4935
- regression coefficients
 - MOS example, 4227
- regression to mean
 - exercise, 4201
- regression to the mean, 4298
 - neural operators, 3227
- regression-based downscaling, 4351
- regret bound
 - UCB allocation, 2177
 - unified exploration, 2233
- regularisation
 - privacy, 1615
 - privacy benefit, 1683
 - score matching, 1682
- regularised score matching, 1683
 - algorithm, 1684
- regularization, 1516–1521
 - CutMix, 1520
 - dropout, 1517
 - Bayesian interpretation, 1518
 - ensemble interpretation, 1517
 - DropPath, 1518
 - label smoothing, 1519
 - calibration, 1519
 - Mixup, 1520
 - R-Drop, 1521
 - stochastic depth, 1518
- regulation
 - data privacy, 4445
 - synthetic data, 2911
- regulatory approval
 - generative models, 3722
 - medical AI, 3722
- regulatory considerations
 - medical AI, 3713
- regulatory convergence
 - global, 4447
- regulatory pathway
 - neurology AI, 3939
- regulatory traceability
 - EEG, 3933–3934
- reidentification, 2889
- REINFORCE, 219, 1046
 - architecture search, 1913
 - baseline, 1914
 - variance reduction, 1046
- reinforcement learning
 - architecture search, 1911–1916

- comparison table, 3109
- contrastive feedback, 2496
- difficulty adjustment, 4793
- for reasoning, 1044–1051
- for theorem proving, 2961–2962
 - diagram, 2962
- GRPO, 3053–3058
- language models, 3053
- limitations for LLM optimisation, 2133
- mathematical reasoning, 3095–3112
- reasoning models, 3050
- reward signal, 2934
- sign language production, 4731, 4734
- summary for mathematics, 3109
- theorem proving, 3036, 3112
- reinforcement learning from human feedback
 - disability, 4850
- rejection sampling, 54, 2703
- relational data, 2697–2698
 - CTGAN limitation, 2728
- relative attention
 - music, 906
 - periodic structure, 907
- relaxation gap, 1901
- relevance filtering, 3025
- reliability
 - log-scale analysis, 5146
- ReLU, 2715, 5253
 - dying neuron, 1553
 - dying neurons, 1502
 - Lipschitz constant, 305
 - vanishing gradient solution, 4238
- renewable energy
 - variability, 4182
 - weather downscaling, 4358
- Renyi divergence, 1773
 - definition, 1774
- repair module
 - efficiency, 2831
- reparameterisation trick, 40, 61, 2753, 4507
 - definition, 61
 - exercise, 400
 - gradient, 394
 - Gumbel-Softmax, 1902
 - VAE, 394–395
 - video, 568
- reparameterization trick, 224–230
 - definition, 225
 - discrete variables, 227
 - Gaussian example, 265
 - gradient estimator, 225
 - multivariate Gaussian, 228
 - VAE, 256
- repeated games, 1357
- replay
 - buffer size, 1197
 - for LLMs, 1196–1198
 - paradox, 1196
- replay buffer
 - FengWu training, 4278
 - theorem proving, 3086, 3119
 - weather forecasting, 4277
- replay-based methods, 1193–1198
- representation
 - fixed vs. open-ended, 2243
- representation bridging, 4092
- representation engineering, 4073
- representation laundering, 4141
- representation learning, 386
- representational capacity, 2072
- representational harm
 - disability, 4843
- reproducibility
 - acceleration, 3927
 - benchmark evaluation, 4240
 - distributed training, 5023

- EEG AI, 3952
- epilepsy AI, 3918–3929
- evaluation, 2849
- LLM-evolutionary search, 2129
- tabular generation, 2907
- reproducibility checklist, 3925
- reproducibility crisis
 - epilepsy AI, 3922
- reproducible research, 2690
- ReProver, 2957, 3082–3084
 - results, 3083
- reranker, 2373
- reranking, 2487–2499
 - accuracy-efficiency tradeoff, 2499
 - complexity theorem, 2498
 - computational complexity, 2497–2499
 - empirical comparison, 2498
 - in recommendation, 2575
 - listwise, 2490–2493
 - pairwise, 2490
 - pointwise, 2488–2489
 - reliability, 2499
 - setwise, 2493–2495
- Rerender-A-Video, 717
- resampling ablation, 3974–3981
 - definition, 3975
- research challenge
 - differentiable data assimilation, 4421
 - evaluation suite, 4420
 - foundation model proposal, 4421
 - generative downscaler, 4420
 - minimal MLWP, 4419
- research challenges
 - distributed training, 5166
 - epilepsy generative AI, 3953–3957
 - implementation, 5178
 - medical image segmentation, 3762–3771
 - weather prediction, 4419
- research frontier
 - accessibility AI, 4908
 - distributed training, 5166
- research-level mathematics, 3060
- resective surgery
 - failure rate, 3912
- reservoir sampling, 1193
 - guarantee, 1194
- residual
 - stochastic, 4353
- residual flow, 459–460
 - invertibility, 459
- residual map, 3911
 - exercise, 3928
 - lesion localisation, 3841
- residual MLP, 2764
- residual network, 98
- residual quantisation, 2511
- residual stream, 3973, 3974, 4002–4003, 4138
 - bilinear structure, 4003
 - communication bus, 4002
 - definition, 4002
 - direct vs indirect effects, 4013
 - harmful content, 4139
 - integrated information, 4148
 - video transformer, 4119
- residual vector quantisation, 788–802
 - approximation bound, 788
 - bitrate, 801
 - definition, 788
 - effective codebook size, 790
 - error bound, 950
 - gradient estimation, 797
 - orthogonality, 789
 - voice cloning, 4749
 - VQ-VAE connection, 942
- ReSLM, 2540
- resolution

- method, 2955
- principle, 3017–3021
- theorem proving, 2940
- resolution generalisation, 611
- resolution independence, 3431
- resolution-based proving, 3017–3021
- resolved variable, 3018
- resolvent, 3018
- resource allocation
 - evolutionary computation, 2198
- resource-constrained devices
 - federated learning, 4523–4524
- responsibility, 150
- responsible AI
 - evolutionary red-teaming, 2167
- restarts
 - SAT solving, 3018
- resting-state fMRI
 - epilepsy, 3894
- retention threshold
 - MOS, 4224
- retinal fundus photography, 3506
- retinal vessel segmentation, 3513, 3535
- retinitis pigmentosa, 4690
- retraining
 - cost, 2775
 - cost comparison, 4606
- retrieval, 2642
 - architecture search, 2610
 - as fundamental operation, 2616
 - as generation, 2333
 - as reconstruction, 2337–2338
 - biological, 2344
 - consolidation with prior chapters, 2608–2616
 - continual learning, 2612
 - convergence with generation, 2641
 - dense, 2364–2374
 - diffusion features, 2610
 - embodied agents, 2621
 - evaluation metrics
 - exercise, 2635
 - evolutionary methods, 2611
 - exercises, 2631–2640
 - fairness
 - exercise, 2637
 - for alignment, 2614
 - forward pass interpretation, 2683
 - frontier problems, 2623–2625
 - fundamental limits, 2564
 - GAN features, 2609
 - generative models, 2333
 - generative models as, 2608
 - grand unification, 2615–2616
 - diagram, 2682
 - history, 2519, 2625–2631
 - classical era, 2626
 - dense retrieval era, 2629
 - generative era, 2630
 - learning to rank era, 2628
 - manual era, 2625
 - in evolutionary algorithms, 2612
 - information theory
 - exercise, 2634
 - information-theoretic bounds, 2554–2565
 - key takeaways, 2641
 - memory augmentation, 2545–2554
 - memory connection, 2613
 - multi-hop reasoning, 2620
 - news problem, 2617
 - novel modality
 - challenge, 2637
 - open problems, 2616–2625
 - philosophical implications, 2642
 - philosophy, 2333–2340
 - real-time updating, 2617–2618
 - reconstruction thesis, 2338
 - scaling limits, 2559
 - structured data, 2901–2904

- summary, 2640–2642
- synthetic augmentation, 2901
- synthetic data, 2499–2508
- thermodynamic limits, 2623
- timeline, 2683
- top 10 open problems, 2624
- ultimate insight, 2642
- universality conjecture, 2624
- VAE formulation, 2609
- VAE latent space, 2609
- verification, 2565–2574
- view of learning, 2615
- retrieval benchmarks, 2596–2608
 - limitations, 2605–2607
- retrieval evaluation, 2596–2608
- retrieval information curve, 2561
- retrieval metrics, 2604–2605
- retrieval problem
 - definition, 2336
 - formal definition, 2336–2337
- retrieval query
 - quality, 2530
- retrieval recall
 - information-theoretic bound, 2558
- retrieval-augmented description, 2476
- retrieval-augmented fine-tuning
 - exercise, 2636
- retrieval-augmented generation, 1096–1098, 2520–2531
 - architecture, 2520
 - as memory system, 2552
 - audio, 2453–2457, 2541
 - definition, 2454
 - exercises, 2456
 - clinical reasoning, 3682–3685
 - cognitive architecture, 2553
 - definition, 3682
 - design exercise, 3693
 - end-to-end training, 2522
 - error analysis, 2562–2565
 - error example, 2563
 - error propagation, 2563
 - example, 2523
 - factuality reranking, 4837
 - formal definition, 2521
 - Fusion-in-Decoder, 2526
 - hallucination reduction, 3683
 - history, 2524
 - knowledge freshness, 3684
 - medical segmentation, 3680–3692, 3755
 - medical VQA, 3907–3908
 - memory capacity, 2554
 - MIA vulnerability, 1824
 - multimodal, 2538
 - original framework, 2520–2524
 - privacy defenses, 1825
 - RAG-Sequence, 2521–2522
 - inference, 2523
 - RAG-Token, 2521–2522
 - RETRO, 2524
 - segmentation decisions, 3684
 - Self-RAG, 2527
 - theorem proving, 2535, 3081
- retrieve-once reasoning, 2537
- retrieve-then-read, 2529
- retrieve-then-rerank, 2488
- RETRO, 2524–2526, 2630
 - chunked cross-attention, 2525
 - computational cost, 2525
 - scaling properties, 2525
- RetroMAE, 2387
- reverse engineering
 - neural networks, 3971
- reverse KL, 170
- reverse process, 477
 - categorical, 2763
 - continuous, 2763
 - memorisation, 1610
 - video, 546

- reverse SDE, 4340
 - Hilbert space, 3271
- reverse-time SDE, 2776, 3269
- reward feedback learning, 1158
- reward function
 - structured generation, 2865
 - theorem proving, 3096
- reward functions
 - scientific discovery, 2217–2219
- reward hacking, 1132, 4134
 - absence in theorem proving, 3096
 - example, 1136
- reward model, 1050, 1129–1133
 - architecture, 1132
 - identifiability, 1131, 1162
 - K-way rankings, 1131
 - proxy, 2934
 - training, 1130
- reward shaping, 1295–1301
 - combination strategies, 3107
 - invariance, 1297
 - potential-based, 3111
 - theorem proving, 3104–3107
- Reynolds decomposition, 4353
- Reynolds number, 3197, 3326
 - microscale computation, 3453
- ReZero, 1498
- RFdiffusion, 14
- RFO, *see* rectified flow, PDE operators,
 - see* rectified flow-based operator
- Richardson weather generator, 4289
- Richardson’s forecast factory, 4210
- Richardson, Lewis Fry, 4184, 4210
- Riemann Hypothesis, 2951
- Riemann problem, 3317
 - neural operators, 3318
- Riesz representation theorem, 5215
- Riffusion, 823–825
 - equivalence to image diffusion, 824
 - example, 824
- right to be forgotten
 - federated learning, 4603–4609
- right to explanation, 3962
- right to human alternatives, 4908
- rigid body constraints, 720
- ring all-reduce, 5033
 - algorithm, 4972–4980
 - bandwidth optimality, 4947, 4955–4956
 - bandwidth optimality proof, 5035
 - benchmarking, 5179
 - communication cost, 4946
 - cost analysis, 4946–4947
 - definition, 4972
 - exercise, 5040
 - figure, 4974
 - NCCL 1.x, 4972
 - step-by-step trace, 5170
 - topology, 4945
- Ring Attention, 1088–1090, *see* attention,
 - ring
 - exercise, 1111
- ring attention, 5011, 5055–5060
 - communication, 1554
 - communication cost, 5059, 5176
 - communication lower bound, 5156
 - definition, 5011
- ring equalities
 - automation, 2989
- ring tactic, 2989
- ring topology, 4944–4945
 - gradient aggregation, 4931
- Ring-AllReduce
 - Baidu, 4931
 - bandwidth, 4930
- Risch algorithm, 3032
 - example, 3032
- RKHS, 2840
- RLCF, 2496–2497

- reward function, 2496
- training, 2497
- RLHF, 1123–1167, 1762–1771, 2934
 - architecture search, 1888
 - as alignment mechanism, 2262
 - as coevolution, 2261
 - bias reduction exercise, 4853
 - bias reduction proposition, 4850
 - concealment hypothesis, 4138–4139
 - disability inclusion, 4850, 4906, 4910
 - introspective capability, 4144
 - KL regularization, 258
 - KL constraint, 1133
 - limitations, 4133
 - mathematical reasoning, 3136
 - optimal policy, 1133
 - PPO, 1133–1137
 - retrieval interpretation, 2615
- RMSE
 - connectivity, 3924
 - fMRI synthesis, 3854
 - Wasserstein trajectory, 3924
- RMSE reduction
 - ANN vs MOS, 4236
- RMSNORM, 113
- RNS
 - NeuroPace, 3918
- Robbins–Monro conditions, 264
- Robbins–Monro conditions, 216
- Robertson, Stephen, 2627
- Robinson, Alan, 2955
- Robinson, J. Alan, 3017
- robot locomotion
 - repertoire, 2086
- robotic assistive devices, 4867
 - safety constraints, 4868
- robustness
 - through diversity, 2080
- ROC curve
 - epilepsy triage, 3949
- RoCE
 - RDMA over Converged Ethernet, 4966
 - RDMA over Ethernet, 4928
 - specifications, 4968
- rock-paper-scissors
 - cycling, 2068
 - exercise, 2067
- Rocq, *see* Coq
- rollout
 - error accumulation, 4313
- rollout drift
 - 12-hour steps, 4313
- Rombach et al.
 - latent diffusion, 5082
- ROME
 - rank-one update, 4167
- ROME algorithm, 4058–4067
 - correctness, 4062
 - covariance computation, 4062
 - derivation, 4060
 - example, 4063
 - key vector, 4061
 - pipeline, 4063
 - rank-one proof exercise, 4067
 - rank-one update, 4062
 - value vector, 4061
- ROME hypothesis, 4059
- roofline analysis, 555
- roofline model
 - definition, 4936
 - exercise, 4941
 - figure, 4938
 - performance bound, 4937
- RoPE, *see* positional encoding, rotary, *see* Rotary Position Embedding, 2389
 - exercise, 120
 - NTK scaling, 1553
 - video diffusion, 750

- RoPE (rotary position embedding), 4116
- Rosenbrock function, 2044
- Rosetta Stone
 - autoformalization analogy, 2960
- Rosin, Christopher, 2282
- Rotary Position Embedding
 - 1D, 610
 - 3D, 610–613
 - definition, 610
 - relative position, 611
 - exercise, 613
 - extrapolation, 611
 - NTK-aware interpolation, 612
 - relative position, 610
- rotation matrix, 104
- ROUGE, 22
- round-trip error
 - denoising loss connection, 1662
 - latent diffusion, 1663
 - membership statistic, 1662
- router, 1203
- ROVER
 - optical flow
 - nowcasting baseline, 4239
- row-parallel linear, 5043–5050
 - definition, 5044
- RQ-VAE, 2511–2513
 - reconstruction error, 2512
- RRF, 2372
- RTMPose, 4729
- Rubin’s combining rules, 2890
- Rubin, Donald, 2690
- rule-based systems
 - accessibility, 4701
 - fixed scripts, 4701
- Rumelhart, D. E., 3962
- Runway Gen-3, 12
- Russell, Bertrand, 2944
- Russian roulette estimator, 460
- RVQ, *see* residual vector quantisation
 - token structure, 796
- Ryan, Conor, 2282
- S-Inhibition Heads, 4030
 - walkthrough, 4034
- s-step Krylov methods, 4987
- s-step methods
 - numerical stability, 4987
 - optimal s exercise, 4990
- S4 (structured state space), 4124
- S6MOD, 1215–1216
 - definition, 1215
 - temporal frequencies, 1216
- SA-1B dataset, 3621
- saddle point, 310, 5268
 - exercises, 323
- SAE
 - noise filtering, 4091
 - regularisation theorem, 4091
- SAE reconstruction
 - practical conditions, 4091
 - regularisation effect, 4089–4092
 - regularisation exercise, 4095
- safe evolutionary search, 2324
- safe head function, 2972
- SafeDiffCon, 3317, 3414
 - definition, 3317
- safety
 - accessibility AI, 4829–4841
 - dementia AI, 4807
 - evolutionary systems, 2256–2267
 - retrieval filtering, 2614
- safety guarantees
 - neurostimulation, 3918
- safety-critical accessibility, 4903
- safety-critical tools
 - offline requirement, 4908
- SageMath, 3031
- Sakallah, Kareem, 3018
- saliency maps, 3985

-
- Salimans, Tim, [2284](#)
 - Salmon, W. C., [3964](#)
 - Salton, Gerard, [2349](#), [2627](#)
 - SAM, [3618–3627](#)
 - 3D adaptation, [3632–3633](#)
 - ambiguity, [3620](#)
 - architecture, [3620–3623](#)
 - diagram, [3622](#)
 - automated prompting, [3697](#)
 - bidirectional propagation, [3637](#)
 - box prompts, [3624](#)
 - cVAE
 - training details, [3700](#)
 - cVAE taming, [3699](#)
 - decoder, [3696](#)
 - democratisation, [3618](#)
 - exercise, [3641](#)
 - fine-tuning for microscopy, [3647](#)
 - image encoder, [3620](#)
 - mask decoder, [3621](#)
 - mask prompts, [3624](#)
 - medical adaptations, [3627–3641](#)
 - medical imaging, [3625–3627](#)
 - negative prompts
 - insight, [3638](#)
 - paradox, [3638](#)
 - open-vocabulary extension, [3695](#)
 - paradigm, [3619](#)
 - performance
 - exercise, [3642](#)
 - organ complexity, [3626](#)
 - point prompts, [3623](#)
 - prompt encoder, [3620](#)
 - prompt engineering, [3637–3639](#)
 - prompt quality, [3624](#)
 - prompt sensitivity, [3698–3701](#)
 - taming proposition, [3700](#)
 - prompt sensitivity as feature, [3699](#)
 - prompt types, [3623–3624](#)
 - pseudo-prompts, [3696](#)
 - SA-1B, [3621](#)
 - slice selection, [3638](#)
 - strengths and limitations, [3625–3627](#)
 - text prompts, [3624](#)
 - timeline, [3627](#)
 - train once prompt anywhere, [3619](#)
 - zero-shot
 - failure, [3625](#)
 - success, [3625](#)
 - warning, [3625](#)
 - SAM 2, [3627–3641](#)
 - 3D abdominal CT, [3632](#)
 - 3D propagation
 - diagram, [3634](#)
 - architecture, [3628–3629](#)
 - streaming memory, [3628–3629](#)
 - definition, [3628](#)
 - diagram, [3629](#)
 - text-guided, [3680](#)
 - SAM 3, [3629–3630](#)
 - combined prompts, [3630](#)
 - exemplar prompts, [3630](#)
 - SAM-Med2D, [3519](#)
 - SAM-Med3D, [3519](#)
 - SAM2
 - architecture, [3904](#)
 - cell tracking, [3658–3659](#)
 - medical imaging, [3904](#)
 - sam4celltracking, [3658](#)
 - SAMA, [1693](#)
 - aggregated score, [1693](#)
 - algorithm, [1694](#)
 - exercise, [1702](#)
 - masked prediction score, [1693](#)
 - medical text, [1696](#)
 - performance, [1696](#)
 - scaling behaviour, [1696](#)
 - threshold calibration, [1694](#)
 - SaMD

- Software as a Medical Device, 3940
- SaMIA, 1746–1753
 - advantage over joint likelihood, 1749
 - bidirectional memorization, 1750
 - computational cost, 1751
 - connection to masked language models, 1750
 - importance sampling estimator, 1748
 - membership score, 1748
 - Monte Carlo estimation, 1747
 - practical considerations, 1751
 - pseudo-likelihood, 1746
 - sampling-based estimator, 1747
 - variance reduction, 1748–1749
 - when to use, 1752
- sample complexity
 - audio vs. image, 855
 - evolutionary, 2274
 - federated rare disease, 4617
 - lower bound proof, 1914
 - RL-NAS, 1914
 - ShinkaEvolve, 2151
 - tabular generation, 2906
- sample Dice, 3496
- sample efficiency
 - evolutionary exploration, 2236
 - flow matching, 870
 - in program evolution, 2145
 - PPO vs GRPO, 3104
 - RL limitations, 2134
- sample rate
 - speech vs. music, 770
- sample size
 - generative segmentation, 3500
- sampling, 51
 - inverse CDF, 52
 - tabular generation, 2810–2812
- sampling acceleration, 645–652
 - combined example, 651
 - combined strategies, 650
- sampling cost
 - video, 547
- sampling mechanism
 - privacy key, 4598
- sampling methods, 137
- sampling rate, 769
- sampling speed
 - diffusion vs GAN, 2768
- sampling theorem, 527
- sandboxing, 1299
- SAPEMOS, 4227
 - exercise, 4232
 - precipitation example, 4228
- SAT
 - definition, 3017
- SAT Competition, 3018
- SAT solvers, 2940
 - in mathematics, 2955
- SAT solving, 3017–3018
- satellite observations, 4185
- satisfiability modulo theories, 3021–3024
- satisfiability problem, 3017
- SBERT, *see* Sentence-BERT
- SCAFFOLD, 4472
 - federated training, 4613
- scalability
 - classical methods, 3531
 - GOGGLE, 2757
 - structured data, 2708
 - video generation, 523
- scalable oversight, 4146–4147, 4175
 - definition, 4146
- scale space, 2398
- Scale-Invariant Feature Transform, *see* SIFT
- scaling
 - distributed necessity, 4926

- interpretability, 4153
- scaling analysis
 - video diffusion, 551
- scaling laws
 - agent systems, 1431
 - architecture, 1963
 - architecture search, 1984
 - conjectured for PDEs, 3435
 - continual learning, 1245
 - derivation, 3255
 - empirical evidence, 1431
 - empirical observations, 658
 - evolutionary search, 2129, 2297
 - exercise, 4933
 - exponents, 1461
 - fitting, 1456
 - generative discovery, 3469
 - mathematical reasoning, 3046
 - memorization, 1704
 - model size, 4925
 - PDE foundation models, 3255, 3435
 - retrieval vs. parameters, 2526
 - retrieval-parameter trade-off, 2526
 - test-time compute, 3120
 - turn count, 1304, 1448
 - universality, 1441
 - video diffusion, 658
- scalogram
 - EEG, 3832
- scalp EEG
 - deep brain inference, 3854
 - intracranial translation, 3825
 - to iEEG, 3854
- scan-to-diagnosis pipeline, 3704
 - stages, 3704
- ScaNN, 2382–2383
- scanner artefacts, 3726
- scattering amplitude, 3168
 - background, 3168
- scene perception
 - BLV, 4707
- scene-to-audio synthesis
 - definition, 4708
- Scene2Audio, 4708
 - implementation, 4919
 - psychoacoustic mapping, 4911
- scheduled sampling, 444
- scheduling, 1380–1390
 - makespan, 1455
 - makespan minimisation, 1381
 - SMT, 3024
- schema
 - definition, 2038
- schema complexity, 2708
- schema reversals
 - hallucination, 4831
- Schema Theorem, 2280
 - exercise, 2288
- schema theorem, 2038–2040
 - limitations, 2039
 - statement, 2038
- Schizas, C.N., 4235
- Schmidt et al. (2025)
 - downscaling, 4357
- Schrödinger bridge, 381
- Schrödinger equation, 3364–3371
 - time-independent, 3364
- Schrems II
 - data transfer implications, 4449
- Schur complement, 40
- Schwartz class, 5231
- Schwefel, Hans-Paul, 2041
- scientific applications, 13
- scientific discovery, 1400–1406, 3447
 - agent pipeline, 1458
 - broader landscape, 2893
 - synthetic data, 2891–2896
- scientific machine learning, 3191–3450
- scientific verification, 2569–2574
- score

- function space, 3379
- score composition, 3372, 5092
 - definition, 5093
 - distributed, 5159
 - inference-time composition, 5094
 - quality vs inference cost, 5096
 - variational bound, 5093
 - with constraints, 3377
- score correction, 2781–2787
- score distillation, 2392
- score estimation
 - error propagation, 3291
- score function, 85, 220, 500, 501, 2771, 2776
 - approximation theory, 3291
 - definition, 85, 502, 3269, 4340
 - distributed memory, 1614
 - exercise, 2773
 - Gaussian field, 3293
 - Gaussian random field, 3453
 - Hilbert space, 3271
 - memorisation, 1611
 - noise identity, 504
 - noise prediction equivalence, 4309
 - SDE, 3833
 - source separation, 955
 - zero expectation, 265
- score function estimator, 219–224
- score function identity, 219
- score incompatibility, 2373
- score matching, 85, 506
 - data parallelism, 5089
 - denoising, 510, 2759
 - differentiable physics, 3408
 - EEG, 3833
 - exercise, 3395
 - function space, 3272, 3380
 - function spaces, 3267–3295
 - hybrid spaces, 2897
 - Hyvärinen, 86
 - identity, 87
 - implicit, 507
 - independence across samples, 5082
 - overfitting, 1614
 - practical implementation, 517
 - privacy attack, 4594
 - regularisation, 1682
 - sliced, 512
 - well-posedness, 3273
- Score Matching via Differentiable Physics, 3308
- score network
 - architecture search, 1941
 - computational graph, 1936
- score routing, 5128
 - exercise, 5134
- score-based data assimilation, 4343
 - definition, 4343
- score-based diffusion
 - function space, 3383
- score-based generation, 2771
- score-based generative models
 - PDE solutions, 3463
- score-based models, 330
 - continuous-time SDE, 4339–4340
 - inverse problems, 3300
 - PDEs, 3264–3295
 - weather, 4197
- scratchpad, 1039
 - definition, 1039
- screen readers
 - historical limitations, 4689
 - history, 4693
 - time cost of inaccessibility, 4697
- screen-reader friendliness, 4888–4889
- screening
 - segmentation role, 3479
- screening regression, 4222
- scribble annotations, 3520
- scribble prompts, 3519

- SDE
 - reverse-time, 2776
- SDEdit
 - video, 712–714
 - definition, 712
 - noise level selection, 714
 - trade-off, 713
- SDMetrics, 2849–2850
- SDXL
 - distributed training, 5085
- sea surface temperature
 - data assimilation, 3282
- seamless prediction
 - weather-climate, 4421
- search
 - theorem proving, 3088
- search algorithms
 - theorem proving, 3040–3043
- search and learning
 - shared architecture, 3137
- search space, 1892–1897
 - bias, 1894
 - cardinality, 1898
 - combinatorics, 1898
 - comparison, 1894
 - connectivity, 1892
 - constraints, 1892
 - formal definition, 1892
 - generative models, 1894
 - lower bound, 1897
 - hierarchy, 2007
 - operations, 1892
 - proofs, 2934
 - redundancy, 1899
 - richness tradeoff, 1896
 - taxonomy, 1892
- Searle, John, 4147
- seasonal harmonic
 - MOS predictor, 4226
- SecMI, 1626–1636
 - algorithm, 1627
 - exercises, 1635
 - intermediate timestep vulnerability, 1628
 - method, 1626
 - noise space vs image space, 1627
 - optimal timestep, 1629
 - practical configuration, 1635
 - step-wise estimation error, 1626
 - timestep vulnerability, 1627
- secure aggregation, 4451
- secure multi-party computation, 4522
 - comparison with VAEFL, 4521
 - cross-silo applicability, 4463
 - federated learning, 4599
- Seedat, Nabeel, 2775
- SEEDS, 4193, 4318–4329
 - 10000 members, 4319
 - architecture, 4318
 - cyclone forecasting, 4319
 - ensemble emulation, 4318
 - training objective, 4320
- SEG metric, 3643
 - exercise, 3665
- SegAN, 3573–3576, 3744
 - architecture, 3573
 - architecture diagram, 3575
 - boundary error reduction, 3574
 - brain tumour segmentation, 3575
 - exercise, 3594
 - multi-scale loss, 3574
 - training stability, 3576
- Segment Anything Model, 3493, 3618–3627
 - annotation interface, 3518
 - microscopy adaptation, 3647
 - prompt sensitivity, 3749
 - text-guided, 3767
- segmentation
 - continuous relaxation, 3597

- democratisation, 3618
- loss functions, 3485
- segmentation mask, 3478
- segmentation metrics, 3495, *see*
 - evaluation metrics
- segmentation taxonomy, 3501–3523
 - annotation interface, 3518–3521
 - clinical target, 3509–3515
 - exercise, 3522
 - four axes, 3501
 - limitations, 3521
 - motivation, 3501
 - output structure, 3503–3505
 - spatial dimensionality, 3506–3508
 - tree diagram, 3502
- segmentation uncertainty
 - annotation interface, 3521
- SegVol, 3639–3640
 - architecture, 3639
 - text prompts, 3639
 - training data, 3639
 - vs slice-by-slice, 3640
 - zoom-in mechanism, 3640
- seismic inversion, 3404
- seizure
 - classification, 3776
 - definition, 3774
 - geometric interpretation, 3863
 - onset types, 3776
 - phase transition, 3865
- seizure control
 - DBS, 3873
- seizure detection
 - CHB-MIT benchmark, 3834
 - class imbalance, 3803–3813
 - diffusion models, 3831–3840
- seizure mechanism
 - causal inference, 3935
- seizure onset localisation
 - EEG-fMRI, 3954
- seizure onset zone, 3775
 - attention, 3882
- seizure prediction
 - accuracy-latency tradeoff, 3889
 - EPILEPSIAE, 3920
 - generative predictor
 - definition, 3930
 - implantable devices, 3931
 - open problem, 3931
 - OT-based, 3866
 - RCVAE, 3826
 - real-time, 3930
 - transformers, 3879–3889
 - wearable system, 3953
- seizure prediction horizon, 3921
- seizure propagation
 - BOT model, 3871
 - branched OT, 3870
 - network, 3863
 - tractography, 3891–3896
- seizure propagation pathway
 - prediction, 3893
- seizure propagation transformer, 3894
- seizure states
 - formal definition, 3795
- seL4, 3001
- selection, 2019–2022
 - pressure analysis, 2292
 - stalling, 2030
- selection pressure, 2028
- selective attention, 4149
- selective checkpointing, 5072–5073
 - trade-off, 5073
- selective forgetting, 1884
 - decision tree, 2006
- selective gates
 - exercise, 4132
- selective recoloring, 4780
 - optimality, 4785
 - practical implications, 4786

- proposition, 4785
- selective state spaces, 1104
- self-adaptation, 2122
- self-adapting networks, 1955–1960
- self-attention
 - color perception, 4779
 - EEG
 - channel, 3946
 - long-range dependencies, 3234
- self-awareness
 - in AI, 3963
- self-configuring segmentation, 3538
- self-consistency, 1042–1043
 - definition, 1042
 - error reduction, 1042
 - exercise, 1068
- self-correction
 - emergent, 3052
- self-critique, 2528
- self-explanation
 - feature activations, 4145
- self-improving AI
 - medical segmentation, 3734
 - optimal stopping, 3736
- self-improving retrieval, 2624
 - challenge, 2639
- self-improving systems, 2200
- self-knowledge
 - models, 4142–4152
- self-model circuit, 4143
- self-modification
 - safety analysis, 2293
- self-modifying systems
 - safety, 1968
- self-organising map
 - SOM
 - meteorology, 4236
- self-paced learning, 2229
- self-play
 - AlphaGo, 3114

- AlphaZero training loop, 3115
- conjecture quality, 3111
- for conjectures, 3179
- mathematical reasoning, 3107
- theorem proving, 3101, 3118
- self-prompting pipeline, 3697–3698
 - definition, 3697
 - exercise, 3713
- Self-RAG, 2527–2529, 2630
 - inference algorithm, 2528
 - reflection tokens, 2528
 - training, 2528
- Self-Refine, 1053
 - exercise, 1070
 - fixed-point interpretation, 1053
- self-refinement
 - iterative, 3698
- self-reflection, 1051–1056
- self-replay, 1197
- self-similarity matrix, 887, 925
 - as guidance, 887
- self-supervised learning
 - EEG exercise, 3898
 - exercises, 2438
 - speech, 2431–2434
- self-supervised pretraining
 - EEG, 3887–3888
- self-verification, 2571–2574
 - retrieval quality dependence, 2572
- Selsam, Daniel, 3033
- SELU, *see* activation functions, SELU
 - fixed point, 1553
- semantic verification, 2213
- semantic gradient, 2101–2137
 - definition, 2102
- semantic ID, 2581
 - collision probability, 2582
 - granularity analysis, 2585
- semantic IDs, 2511–2513
 - definition, 2512

- semantic memory, [2341–2342](#)
 - definition, [2342](#)
 - in agents, [1281](#)
- semantic parsing
 - mathematics, [3066](#)
- semantic retrieval
 - philosophy, [2336](#)
- semantic segmentation, [3503–3504](#)
 - definition, [3478](#), [3503](#)
 - organ segmentation, [3503](#)
- semantic variation operator, [2097](#)
- semantic variation operators, [2095–2108](#)
- semi-amortized inference, [232](#)
- semi-decidability, [3017](#)
- semi-implicit scheme, [4215](#)
- semi-Lagrangian scheme, [4213](#), [4214](#)
- semi-supervised learning
 - EEG, [3820](#)
- Semon, Richard, [2343](#)
- sensitivity, [3545](#), [3809](#)
 - definition, [3810](#)
 - EEG, [3923](#)
 - seizure prediction, [3882](#)
- sensitivity axiom, [3987](#)
- sensor data, [2801](#)
- sensory processing, [4690](#)
- Sentence-BERT, [2385–2386](#)
 - training, [2386](#)
- sentiment analysis
 - debiasing example, [4851](#)
 - disability bias, [4843](#)
- separability
 - and cooperative coevolution, [2062](#)
- seq2seq
 - EEG report generation, [3902](#)
- sequence
 - 1/n convergence, [2978](#)
 - convergence in Lean, [2977](#)
- sequence parallelism, [5013](#), [5050–5060](#), [5135–5144](#)
- causal masking, [5013](#)
 - definition, [5054](#), [5137](#)
 - formal definition, [5055](#), [5138](#)
 - ring attention, [5055](#)
- sequence representation
 - proofs, [3037](#)
- sequential compactness, [2995](#), [5203](#)
- sequential computation
 - via token generation, [3048](#)
- sequential dependency, [2689](#)
- sequential forward screening
 - algorithm, [4224](#)
- sequential screening regression, [4223](#)
- SerAPI, [3013](#)
- serialization
 - exercise, [2921](#)
 - tabular, [2804](#)
 - tabular data, [2804–2806](#)
- server rack
 - structure, [4927](#)
- Set Attention Block
 - definition, [3881](#)
 - permutation equivariance, [3896](#)
- set convolution
 - weather observations, [4375](#)
- set theory
 - vs. type theory, [2945](#)
- Set Transformer
 - EEG, [3880–3884](#)
- set transformer
 - attention mechanism, [3881](#)
- set.mm
 - coverage, [3004](#)
- set.mm database, [3004](#)
- SetConv encoder, [4375](#)
 - density channel, [4382](#)
- setwise reranking, [2493–2495](#)
 - algorithm, [2493](#)
 - complexity, [2494](#)
 - definition, [2493](#)

- legal search, [2494](#)
- severe weather prediction
 - ANN, [4234](#)
- SEVIR benchmark, [4323](#)
- SFNO, *see* Spherical Fourier Neural Operator
- SFT
 - exercise, [1161](#)
- shadow AI, [1825–1826](#)
 - definition, [1825](#)
 - healthcare example, [1825](#)
 - key risk, [1826](#)
 - vulnerability, [1825](#)
- shadow model attack
 - MIA, [4596](#)
- shadow models, [1595–1607](#)
 - attack algorithm, [1602](#)
 - CIFAR-10 example, [1603](#)
 - computational cost, [1604](#), [1725](#)
 - convergence, [1603](#)
 - definition, [1601](#)
 - historical context, [1669](#)
 - paradigm, [1601](#)
- shallow atmosphere approximation, [4208](#)
- shallow network
 - limitations, [4237–4238](#)
- shallow water equations, [3319–3325](#), [3416](#)
 - applications, [3320](#)
 - conservation, [3324](#)
 - definition, [3320](#)
 - hyperbolicity, [3320](#)
 - rotating sphere, [3322](#)
 - system, [3320](#)
- Shanahan, Murray, [4150](#)
- Shannon channel, [1309](#)
- Shannon entropy, [67](#)
- Shannon–Nyquist theorem, [770](#)
- SHAP, [3964](#), [3966](#), [3970](#), [3981–3983](#)
 - comparison figure, [3986](#)
 - exercise, [3990](#)
 - KernelSHAP, [3983](#)
 - key idea, [3983](#)
 - VAE attribution exercise, [4001](#)
- Shapley value, [1361](#)
 - computation, [1376](#)
 - exercise, [1451](#)
- Shapley values, [3981–3983](#)
 - axioms, [3982](#)
 - computation, [4164](#)
 - definition, [3982](#)
- shared memory
 - GPU, [4934](#)
- shared noise
 - temporal consistency, [710](#)
- Shazam algorithm, [2438–2441](#)
- Shi et al. (2017), [4370](#)
- Shi et al. 2025, [4111–4113](#)
- Shi, Xingjian, [4238](#)
 - 2017 paper, [4239](#)
- shifted window attention, [605](#)
 - definition, [606](#)
- ShinkaEvolve, [2144–2153](#), [2286](#)
 - algorithm, [2150](#)
 - architecture, [2151](#)
 - bandit-based selection, [2149](#)
 - full algorithm, [2150](#)
 - novelty filtering, [2198](#)
 - novelty-weighted sampling, [2146](#)
 - rejection sampling, [2147](#)
 - results, [2151](#)
 - sample complexity, [2151](#)
 - sample efficiency problem, [2145](#)
- shock waves, [3312](#)
 - PINNsFormer, [3262](#)
- Shokri, Reza, [1577](#)
- short convolution
 - Mamba, [4126](#)
- short-term load forecasting

- STLF, [4233](#)
- Short-Time Fourier Transform, [771–802](#)
 - compression ratio, [774](#)
 - definition, [771](#)
 - linearity, [773](#)
 - parameter choices, [771](#)
 - resolution tradeoff, [800](#)
- short-time Fourier transform
 - EEG, [3885](#)
- short-time Fourier transform (STFT), [3798](#)
 - definition, [3798](#)
- Shostak method, [3023](#)
- shrinkage problem, [4049–4051](#)
 - practical impact, [4050](#)
- Shwartz-Ziv and Tishby, [1862](#)
- Siamese diffusion, [3605–3606](#)
 - architecture, [3605](#)
 - morphological consistency, [3606](#)
 - noise consistency, [3605](#)
- Siamese network, [2385](#)
- Siamese-diffusion, [3702–3703](#)
 - architecture, [3702](#)
 - consistency loss, [3703](#)
 - cross-attention coupling, [3703](#)
- Siena dataset, [3920](#)
- Siena EEG dataset, [3837](#)
- Siena Scalp EEG dataset, [3800](#)
 - description, [3801](#)
- SIFT, [2397–2400](#)
 - descriptor, [2398](#)
 - exercises, [2400](#)
 - invariance properties, [2399](#)
 - keypoint localisation, [2398](#)
 - orientation assignment, [2398](#)
- SigLIP, [2411–2412](#)
 - batch independence, [2412](#)
 - loss definition, [2412](#)
 - loss function, [2412](#)
- Sigma type, [2972](#)
- Sigma-type, [2985](#)
- sigmoid belief network, [433](#)
- sigmoid loss
 - for contrastive learning, [2412](#)
- sigmoid saturation
 - backpropagation, [4237](#)
- sign language, [4690](#)
 - avatar rendering, [4729](#)
 - ethical analysis exercise, [4744](#)
 - evaluation by signers, [4743](#)
 - fluency illusion, [4741](#)
 - generation, [4694](#)
 - gloss, [4912](#)
 - historical note, [4733](#)
 - interpreting history, [4733](#)
 - linguistic diversity, [4899](#)
 - linguistic structure, [4725](#)
 - manual parameters, [4726](#)
 - real-time recognition, [4810](#)
 - video generation, [4701](#)
- sign language production, [4725–4735](#)
 - challenge, [4725](#)
 - cost-quality tradeoff, [4742](#)
 - definition, [4727](#)
 - formal definition, [4727](#)
 - GenASL pipeline, [4728](#)
 - keypoint representation, [4733](#)
 - latency, [4732](#)
 - lecture translation, [4732](#)
 - multilingual, [4729, 4734](#)
 - native signer evaluation, [4743](#)
- signal-to-noise ratio, [494](#)
 - audio latent, [848](#)
 - definition, [628](#)
 - IWAE, [242](#)
 - mismatch, [629](#)
 - resolution-adapted, [632](#)
 - shifted schedule, [629](#)
 - temporal uniformity, [542](#)
 - video, [541, 628–635, 758](#)

- signing space, 4726
- SignLLM, 4729–4732
- SignSGD, 4477
- silent data corruption
 - GPU, 5152
- SiLU, 113
- SiLU activation, 2764
- silver medal
 - AI equivalent, 3112
- SIMD
 - GPU architecture, 4934
- SimDiffPDE, 3332
- SimGDA, *see* gradient descent ascent,
 - simultaneous
- similarity
 - Aristotelian, 2334
- similarity metrics
 - music, 938
- Simon, Herbert, 2939
- simp lemmas
 - database, 2988
- simp only, 2988
- simp tactic, 2987–2988
- simple type theory, 3001
 - in HOL, 3003
- Simplex
 - in SMT, 3023
- simplicity
 - TabDDPM, 2772
- simplification
 - automated, 2987
- Sims, Karl, 2059, 2281
- SIMT
 - definition, 4934
 - warp execution, 4934
- simulated annealing
 - connection to NAS, 1903
- simulation, 2690
- simulation-based verification,
 - 2209–2211
 - definition, 2209
- simultaneous EEG-fMRI, 3850
- sine-cosine encoding, 2795
- single point of failure
 - accessibility systems, 4814
- single-minus amplitude, 3168
- single-path one-shot, 1906
- single-player game
 - theorem proving, 3116
- single-step generation
 - audio, 877
 - definition, 877
- singular vector perturbations, 4288
 - ECMWF, 4291
- Sinkhorn algorithm, 367–381, 3656, 3868
 - convergence exercise, 383
 - convergence rate, 376
 - derivation, 373–377
 - EEG, 3946
 - exercise, 3878
 - exercise, 382
 - factorisation, 375
 - iterations, 375
 - log-domain, 380
 - matrix scaling, 375
 - numerical example, 377
 - numerical stability exercise, 383
 - pseudocode, 376
- Sinkhorn divergence, 377–378
 - definition, 377
 - exercise, 383
 - properties, 377
 - sample complexity exercise, 383
- Sinkhorn generative modelling, 378–380
 - algorithm, 379
 - exercise, 382
 - loss function, 378
- sinusoidal embedding
 - diffusion, 4312
- sinusoidal encoding, 2764

- sinusoidal positional encoding, 3887
- SISCOM, 3780, 3857
- situational disability, 4691
- Sivic, Josef, 2401
- skeletal motion, 742
- sketch-based image retrieval, 2416
- skill module, 1203
- skill score
 - weather forecasting, 4398
- skip connection, 98
 - dominance in DARTS, 1924
- skip connections, 1495–1501
 - DenseNet, 1498
 - gradient flow, 3741
 - hyper-connections, 1499
 - mHC, 1499
 - pre-activation, 1496
 - precursor in FCN, 3532
 - preserving spatial resolution, 3534
 - ReZero, 1498
 - unraveled view, 1496
- Skolem functions, 3020
- Skolemization, 3020
- Slater condition, 5262
 - strong duality, 5263
- Slater determinant, 3365–3371
 - antisymmetry, 3455
- Sledgehammer, 2945, 3001, 3025–3027
- sleep consolidation, 1871–1874
 - computational overhead, 1874
- sleep-inspired optimization, 1964–1969
- SLERP, 583
- sliding window attention, 1086
- sliding window reranking, 2491–2493
 - algorithm, 2492
 - convergence, 2492
- SLIM-Diff, 3844–3846
 - consistency remark, 3846
 - exercises, 3849
 - location conditioning, 3845
 - results, 3845
 - training strategy, 3845
- slope discontinuity, 2790
- SLoRA, 4554
- slow loop
 - normative model, 3916
- small world graph, 2379
- small-noise injection, 1654
- smart city
 - federated generative models, 4624–4625
- smart glasses
 - accessibility, 4826
- smart grid
 - federated load profile synthesis, 4625
- smart home
 - LLM orchestration, 4871
- SMART system, 2349, 2627
- SMDP, *see* Score Matching via Differentiable Physics, 3408
- SMILES
 - generation, 2891
- smoothing, 431
 - backoff, 431
 - interpolation, 431
 - Kneser–Ney, 431
 - Laplace, 431
- smoothness class, 5231
- SMOTE, 3805
 - formula, 3806
 - limitations, 3806
- SMPC
 - federated learning, 4602
- SMT, 3021–3024
 - definition, 3022
 - theories, 3023
- SMT Competition, 3023
- smt method, 3027
- SMT solvers, 2940

- in mathematics, 2955
- Snake activation, 793
 - analysis, 801
 - motivation, 793
 - properties, 794, 958
- snakes, *see* active contours
 - energy functional, 3742
- SNLI-VE benchmark, 4889
- Sobolev norm, 3420
- Sobolev space, 5233
- social communication
 - generative AI, 4797
- social mirror (GenAI), 4791
- social model of disability, 4691
- social networks
 - synthetic, 2696
- Socrates, 2333
- Socratic tutoring
 - LLM, 4805
- soft conditioning, 2783
- soft guidance
 - constraints, 2822
- soft prompts, 4555
- soft-thresholding, 5275
- softmax, 2715
- softmax output
 - discretised, 4241
- softmax relaxation, 1900
- software testing, 2690
- software verification, 3023
- solution operator, 3211
 - definition, 3211
- Sonderby, Casper Kaae, 4240–4241
- Song and Chissom
 - fuzzy time series, 4234
- Song et al. 2023, 3851
- song structure, 887–888
 - current limitations, 888
 - recurrence, 888
- SOP
 - seizure occurrence period, 3922
- Sora, 3, 12, 722–726, 4115
 - architecture, 5143
 - architecture overview, 539
 - parallelism, 5143
 - prompt enhancement, 684
 - system design, 687
 - world simulation capabilities, 722
- Sort (universe), 2966
- sorting networks, 2281, 3132
 - coevolution, 2058
- sound event retrieval, 2451–2453
 - definition, 2452
- soundness, 3016
- SoundStream, 790–791
 - architecture, 790
 - bitrate, 791
- source separation, 889–895
 - algorithm, 892
 - as inverse generation, 890–895
 - Bayesian, 955
 - connection to vision, 892
 - diffusion-based, 891
 - exercise, 897
 - posterior sampling, 890
- space-filling curves
 - spatial locality, 3904
- spaced repetition, 1174
- SPADE, 355–357
 - Cityscapes example, 357
 - exercise, 360
 - generalisation of BatchNorm, 356
 - layer definition, 356
- spaghetti plot, 4431
- sparse attention, 1085–1090
 - complexity, 1554
 - long video, 748
 - pattern, 1085
 - transformer efficiency, 4824
 - VLM, 4082

- wearable AI, 4824
- sparse autoencoder, 4046–4058
 - architecture, 4047
 - architecture figure, 4052
 - definition, 4047
 - dictionary size, 4053–4055
 - discovered features, 4051–4057
 - exercises, 4057, 4058
 - gated variant, 4050–4051
 - GPT-2 example, 4055–4057
 - L1 tradeoff, 4167
 - loss function, 4048
 - microscope analogy, 4053
 - pre-encoder bias, 4048
 - scaling, 4053–4055
 - shrinkage, 4049–4051
 - toy transformer, 4172
 - training objective, 4048
 - transcoder variant, 4051–4058
- sparse autoencoders, 3969, 3971, 3973
 - causal faithfulness, 4103
 - density exercise, 4094
 - diffusion models, 4098–4102
 - example, 4099
 - feature density, 4086
 - timestep-stratified, 4099
 - vision, 4085–4095
- sparse coding
 - biological, 3969
 - differences from superposition, 4045
 - visual cortex, 4044–4045
- sparse coding hypothesis, 3969
- sparse distributed representations, 2346–2347
- sparse memory, 1208–1209
 - forgetting reduction, 1208
- sparse memory layer, 1208
- sparse models
 - contrast, 4932
- sparse observations
 - developing world, 4227
- sparse retrieval, 2365
- sparse reward
 - theorem proving, 3097
- sparse rewards, 2133
 - theorem proving, 3116
- sparsity
 - adjacency matrix, 2749
 - Ampere Tensor Cores, 4935
- SPASS, 3020–3021
- SpateGAN-ERA5, 4357
- spatial compression, 534
- spatial conditioning
 - neuroimaging diffusion, 3841
- spatial correlation
 - MOS, 4230
 - uncertainty, 3558
- spatial dimensionality, 3506–3508
 - illustration, 3509
- spatial downscaling
 - NWP-to-station, 4226
- spatial features
 - EEG, 3885
- spatial locality, 2689
- spatial pooling
 - MOS, 4228
- spatial proximity
 - attention, 3251
- spatial relationships
 - object detection, 4698
- spatial structure
 - no exploitation
 - shallow MLP, 4237
 - preservation
 - ConvLSTM, 4239
 - shallow network limitation, 4237
- spatio-temporal GCN, 3869
- spatio-temporal segmentation, 3508
- spatiotemporal attention, 598–607

- computational cost, 599
- full, 598
 - definition, 599
- infeasibility, 599
- spatiotemporal coherence
 - downscaling, 4357
- spatiotemporal compression, 3393
- spatiotemporal data manifold, 526–533
- spatiotemporal diffusion
 - 4D cardiac, 4614
- spatiotemporal diffusion model
 - architecture, 4614
- spatiotemporal downscaling
 - definition, 4358
- spatiotemporal factorisation, 572
 - definition, 572
 - initialisation, 573
- spatiotemporal patchification, 4116
- spatiotemporal prediction
 - ConvLSTM, 4238
- spatiotemporal segmentation, 3727
- speaker encoder
 - exercise, 4757
- speaker similarity, 4755
- speaker verification, 158, 939
- specialisation
 - generative models, 2710
- speciation, 2049
- specific humidity, 4205
- specificity, 3545, 3809, 4064
 - definition, 3810
 - EEG, 3923
- SPECT
 - epilepsy, 3780, 3857
 - ictal blood flow, 3857
- spectral accuracy
 - turbulence, 3340
- spectral analysis
 - denoising errors, 1649
- spectral artefacts, 2790
- spectral bias, 3206
 - Burgers equation, 3314
 - MLP, 3260
 - tanh activations, 3457
 - wavelet mitigation, 3262
 - wavelet remedy, 3261
- spectral consistency, 2794–2795
 - definition, 2794
 - generative weather models, 4302
- spectral energy
 - latent space, 585
- spectral gap
 - reanalysis, 4396
- spectral inheritance theorem, 4396
- spectral loss, 785
- spectral methods, 4213
 - NWP, 4213
 - weather forecasting, 4244–4254
- spectral metrics, 3421
- spectral norm, 298
- spectral normalisation, 281, 298, 459
 - computation, 305
 - definition, 298
 - discriminator, 784
 - Lipschitz bound, 298
 - not needed, 2769
 - privacy, 1683
- spectral peaks, 2439
- spectral power
 - denoising error, 1650
- spectral preservation, 2794
- spectral signature
 - diffusion artifacts, 934
- spectrogram, 2439, 3798
 - phase reconstruction error, 828
 - resolution exercise, 832
 - resolution for generation, 827
 - seizure onset example, 3800
 - time-frequency trade-off, 826
- spectrogram diffusion

- definition, 825
- mathematics, 825
- sound effects, 827
- spectrogram generation, 823–830
 - limitations, 827–830
 - codec comparison, 829
- spectrogram watermarking, 931–932
 - definition, 931
- specular reflections, 3514
- speculative decoding, 1543–1546
 - algorithm, 1543
 - correctness, 1544
 - definition, 1543
 - Medusa, 1545
 - speedup, 1544
 - speedup computation, 1557
- speech disabilities, 4745–4758
- speech RAG, 2540
- speech recognition
 - dysarthria, 4689
- speech retrieval, 2427–2438
- speech-to-speech
 - dysarthric, 4755
- speech-to-text
 - Whisper, 4694
- speed
 - automated segmentation, 3488
- SPH
 - seizure prediction horizon, 3922
- sphere packing, 3132
- Spherical FNO, 3223
- Spherical Fourier Neural Operator, 3321
 - definition, 3321
- spherical harmonics, 3321
- spike-and-wave discharges, 3884
- spinal cord injury, 4690
- spinodal decomposition, 3359
- spinor-helicity formalism, 3168
- SPLADE, 2370–2371
- SPLADE++, 2371
- SplineDist, 3510
- split inference, 4822–4823
- split learning, 4454
 - diffusion models, 5129
- split-brain patients, 3963
- SPMD
 - MPI, 4960
- spoken document retrieval, 2427–2438
 - architectures, 2435
 - error propagation, 2437
 - hybrid, 2436
- spoken term detection, 2430–2431
 - DTW-based, 2431
- spot instance
 - training strategy, 5150
- spot instances
 - preemption, 5149
- SPPT
 - limitations, 4292
- spread-skill ratio
 - calibration, 4315
 - DEffusion, 4321
- spread-skill relationship, 4295
 - definition, 4295
- SPV-MIA, 1738–1746
 - approximation error, 1744
 - comparison with other methods, 1743
 - connection to LiRA, 1739
 - decoding strategy, 1742
 - mathematical framework, 1740
 - multilingual considerations, 1745
 - pipeline, 1741
 - prompt design, 1740–1746
 - score function, 1739
 - score separation, 1740
 - self-calibration principle, 1738
 - self-prompted reference text, 1739
 - temperature effects, 1742
 - temperature-variance trade-off,

-
- 1743
 - squeeze theorem, 2978
 - SRAM
 - GPU L1 cache, 4934
 - SSIM, 22, 561
 - evaluation, 3928
 - fMRI synthesis, 3854
 - medical image quality, 4458
 - MRI evaluation, 3923
 - SSReflect, 3000
 - stability, 1181
 - GRPO vs PPO, 3103
 - stability metrics, 3422
 - stability–plasticity dilemma, 1172, 1180–1182
 - definition, 1181
 - stability-plasticity
 - landscape, 1238
 - stability-plasticity dilemma, 1881
 - architectural resolution, 1970
 - history, 1882
 - Stable Audio, 767, 843–845
 - generation parameters, 844
 - timing conditioning, 843
 - waveform VAE, 843
 - Stable Diffusion, 8, 10
 - alignment, 1157
 - architecture, 5082
 - differential privacy, 1678
 - distributed training, 5028
 - throughput example, 5029
 - stable diffusion
 - accessibility, 4703
 - Stable Diffusion XL
 - parameter count, 4932
 - training, 4925
 - Stable Video Diffusion, 673
 - guidance weights, 663
 - stain normalisation, 3515
 - stain transfer, 3515
 - staining domain shift, 3515
 - stale gradients, 5022
 - staleness
 - convergence penalty, 4582
 - standardisation
 - evaluation, 2849
 - Stanley, Kenneth, 2047, 2081, 2283
 - STAPLE algorithm, 3530, 3552, 3760
 - neural extensions, 3569
 - STaR, 1045–1046
 - algorithm, 1045
 - as approximate RL, 1046
 - STAR-C
 - behavioural symptoms, 4801
 - dementia caregiving, 4801
 - sundowning episode, 4806
 - star-convex polygon, 3646
 - StarDist, 3510, 3646–3647
 - state space
 - atmospheric, 4188
 - state space model
 - linear complexity, 3904
 - recurrence
 - exercise, 3909
 - state-space model, 1102–1106
 - convolution, 1103
 - exercise, 1113
 - definition, 1103
 - exercise, 1112
 - memory capacity, 1105
 - state-space model (SSM), 4124–4133
 - state-space models, 1962
 - station-specific forecasting, 4223
 - stationarity assumption
 - MOS, 4226
 - statistical downscaling, 4351
 - diffusion, 4324
 - statistical estimation, 44
 - NWP correction, 4222
 - statistical fidelity, 2704

- statistical physics, 181
 - connection to GFlowNets, 2869
- statistical separation, 1618
- statistical tests, 2850–2852
- statistical verification, 2211–2213
 - protocol, 2211
- statistics
 - physically relevant, 3422
- status epilepticus
 - treatment latency, 3913
- steering vector, 4139
- steering vectors
 - bias, 4112
 - exercises, 4114
- Stefan condition, 3353
 - derivation, 3356
- Stefan problem, 3352–3357, 3454
 - classical formulation, 3352
- Stefan, Jožef, 3354
- steganography
 - EEG, 3934
 - model internals, 4141
- Stein discrepancy, 245
- Stein operator, 244
 - definition, 244
- Stein variational gradient descent, 201, 243–248
 - algorithm, 246
 - RBF kernel, 266
 - theory, 247
 - update rule, 246
 - vs MCMC, 247
- Stein’s identity, 244
 - exercise, 266
 - proof, 245
- step-by-step reasoning, 3048
- step-wise estimation error, 1626
- Stephen Hawking
 - voice, 4699
- stepping stones, 2081
 - evolutionary, 2242
- stereo-EEG, 3776, 3885
 - clinical use, 3780
- stereoelectroencephalography
 - propagation ground truth, 3894
 - synthesis, 3820
- STFT, *see* Short-Time Fourier Transform
 - frame count derivation, 949
 - window design, 3815
- stigmergy, 1272
- stitching function, 2792
- stitching mechanism, 2791
- stochastic differential equation
 - tabular, 2770–2771
- stochastic differential equations
 - EEG diffusion, 3833
 - function space, 3455
- stochastic downscaling
 - uncertainty, 4326
- stochastic ensemble
 - VAE, 4345
- stochastic generation, 3992–3993
 - formal obstacle, 3992
- stochastic gradient descent, 94
- stochastic interpolants, 643
 - definition, 643
- stochastic masking
 - privacy, 4599–4603
- stochastic masking defense
 - definition, 4600
- stochastic PDE, 3383
- stochastic physics parameterisation, 4291–4292
- stochastic rounding
 - distributed, 5159
- stochastic sampling, 3967
 - interpretability challenge, 3966
- stochastic segmentation, 3493
- stochastic variational inference, 213–219
 - algorithm, 215

- convergence, 217
- LDA, 217
- stochastic weather generator, 4289–4290
- stochastic-dynamic prediction, 4285
- stop-gradient, 411
- storage
 - biological, 2343
- storm coherence
 - counterexample, 4331
- StormDiT, 4367
 - architecture, 4367
 - formal definition, 4368
 - training, 4368
- straggler problem, 4581
 - bound, 4581
 - simulation, 4588
- stragglers
 - federated learning, 4451
- straight-through estimator, 229, 411, 575–577, 4600
- RVQ, 797
- Strassen’s algorithm, 3132
 - AlphaEvolve improvement, 2113
- stratified model allocation, 2123–2126
 - cost efficiency, 2124
- stratosphere, 4208
- streaming
 - dense retrieval, 2618
- streaming audio, 874
- streaming audio generation, 946
- streaming indexing, 2618
- streaming memory
 - definition, 3628
- Streaming Multiprocessor
 - definition, 4934
- StreamingT2V, 700
- streamlines
 - tractography, 3890
- StreetViewAI, 4719
- stress testing
 - synthetic data, 2788, 2889
- StrictMono, 2994
- stride, 97
- strong convexity, 5246
- strong lottery ticket
 - definition, 4600
 - privacy, 4599
- strong lottery ticket hypothesis, 4600
- structural bottleneck, 2753
- structural causal model
 - EEG, 3935
- structural connectivity
 - seizure spread, 3891–3896
- structural connectome
 - epilepsy model, 3893
 - simulation, 3897
- structural constraints
 - quantification, 523
- structural control, 537
- structural fidelity, 2834–2847
 - performance bound, 2924
- structural fidelity score, 2846–2847
- structural Hamming distance, 2862
- structural plasticity, 1881
- structural pruning
 - as NAS, 1944
 - performance guarantee, 1945
- structural recursion, 2986
- structural similarity
 - exercise, 2919
 - segmentation samples, 3496
 - SSIM, 4593
- structure functions, 3422
- structure prediction
 - unifying theme, 3137
- structure-conditioned generation, 887
- structured explanations, 4147
- structured data
 - advanced topics, 2876–2884
 - application comparison, 2894

- application taxonomy, 2885
- applications, 2884–2896
- challenges, 2698–2713
 - unified view, 2709
- chapter preview, 2712
- connections, 2896–2904
- definition, 2692
- diffusion, 2896–2904
- ethics, 2910–2912
- exercises, 2916–2929
- formal goal, 2711–2712
- foundation models, 2908–2912
- future directions, 2904–2912
- generation pipeline, 2691
- history, 2710–2713, 2912–2916
- importance, 2690–2691
- introduction, 2687–2691
- landscape, 2691–2698
- open challenge, 2894
- open problems, 2904–2912
- practical challenges, 2907–2910
- retrieval augmentation, 2901–2904
- score matching, 2897
- taxonomy, 2692
- theoretical foundations, 2906–2910
- timeline, 2710
- unified view, 2902–2904
- vs. images and text, 2688–2691
- structured data generation
 - formal problem, 2711
- structured data retrieval, 2602
- structured exploration, 3133
- structured generation
 - GFlowNet, 2865–2876
- structured variational inference, 189
- STT, *see* simple type theory
- stuff classes, 3505
- style mimicry, 938
- style transfer, 10, 334–336
 - audio, 893–895
 - constrained posterior, 893
 - definition, 10
 - diffusion models, 4101
 - exercises, 4104
 - exercises, 339
 - Gram matrix, 334
 - interpreting the loss, 335
 - loss function, 334
 - neural, 334
 - time series, 2800
- StyleGAN, 3, 10, 352–355
 - AdaIN, 353
 - disentanglement, 353
 - exercise, 360
 - mapping network, 353
 - style mixing, 354
 - style mixing matrix, 354
 - truncation trick, 354
- sub-grid processes, 4209
- sub-hundred regime
 - rare disease, 4616
- sub-manifold, 2777
- sub-seasonal to seasonal prediction, 4404
- sub-trajectory balance, 2870
- subdifferential, 5252
 - examples, 5253
 - of max function, 5277
- subgradient, 5252
 - existence, 5252
- subgrid parameterisations
 - learned, 4281
- subgroup, 2996
- subject identity preservation, 740
- subject mismatch
 - EEG, 3781
- sublevel set, 5256
 - convexity, 5256
- subsamped Gaussian mechanism, 1673
- subseasonal-to-seasonal prediction,

-
- 4341–4343
 - subsequence, 2993
 - formalization, 2994
 - substitution
 - as sole rule, 3004
 - logical, 3020
 - subway
 - accessibility, 4818
 - successful tactic reward, 3105
 - successor function, 2944
 - SUDEP, 3773
 - Sudoku
 - video models, 4116
 - sufficient statistic, 203
 - sum type, 2947
 - sum-product algorithm, 134, 136
 - Sumerian tablets, 2625
 - Sun et al. (2026), 4367
 - sundowning
 - AI coaching, 4806
 - Suno, 16, 912–913
 - training data, 913
 - super-resolution, 11, 331–333
 - atmosphere, 4324
 - connection to downscaling, 4350
 - definition, 11
 - exercises, 339
 - ill-posedness, 332
 - meteorological, 4349
 - methods, 332
 - problem formulation, 331
 - PSNR vs. perceptual quality, 332
 - turbulent flows, 3284
 - supercomputer
 - NWP, 4189
 - superensemble
 - generative, 4321
 - member weighting, 4321
 - physical coherence, 4322
 - superhuman mathematical reasoning, 3110
 - supernet, 1904–1910, 1916
 - definition, 1904
 - visualisation, 2008
 - superposition calculus, 3020
 - superposition hypothesis, 3969, 4038–4046
 - capacity, 4042–4046
 - capacity bound, 3971, 4042
 - core insight, 4039
 - definition, 4040
 - exercises, 4045, 4046
 - exponential capacity, 4043
 - formal definition, 4040
 - phase transition, 4043–4046
 - practical implications, 4044
 - toy model, 4040–4046
 - toy model figure, 4042
 - supervised fine-tuning, 1125–1129
 - objective, 1126
 - supervised learning
 - limitations for theorem proving, 3096
 - support function, 5259
 - supremum, 5191
 - SURF, 2399
 - surface delta function, 2782
 - surface measure, 2782
 - surgical concordance
 - tractography model, 3894
 - surgical navigation
 - segmentation, 3479
 - surgical outcome prediction
 - epilepsy, 3895
 - surgical planning
 - segmentation, 3479
 - tractography, 3890
 - surgical referral
 - drug-resistant epilepsy, 3900
 - surprise loss, 1212

- surprise metric, [1212](#)
- surrogate model, [2226](#)
 - in LLM evolution, [2129](#)
- surrogate-assisted evolution, [2226–2229](#)
- surrogate-assisted evolutionary
 - algorithm, [2227](#)
 - convergence, [2228](#)
- surveillance
 - privacy-by-design
 - federated, [4625](#)
- survival probability
 - distributed training, [5146](#)
 - example, [5146](#)
- survival rate, [2096](#)
 - semantic operators, [2098](#)
- Sutskever, Ilya, [3036](#)
- SVD
 - optimality, [2357](#)
- SVGD, *see* Stein variational gradient descent
 - attractive force, [246](#)
 - repulsive force, [246](#)
- SVI
 - latent Dirichlet allocation, [264](#)
- SVM
 - accessibility, [4701](#)
- swarm intelligence, [1270–1277](#), [1390–1400](#)
 - bees, [1271](#)
- SWARM parallelism, [5125–5126](#)
- SwiGLU, [113](#), [1061](#), *see* gating, SwiGLU
- Swin Transformer
 - connection to weather, [4254](#)
 - CVD correction, [4780](#)
 - shifted windows, [3543](#)
- Swin-UNet, [3541](#)
 - architecture, [3543](#)
- SwinTransformer
 - FuXi, [4274](#)
- SwinTransformer V2, [4275](#)
- SwinVRNN, [4345–4346](#)
 - comparison, [4346](#)
 - definition, [4345](#)
 - ELBO, [4345](#)
- Swish, [113](#)
- sycophancy, [1298–1301](#), [1316](#)
 - debate degradation, [1317](#)
 - formal characterisation, [1298](#)
 - mechanism, [1317](#)
- Syed, A., [3976](#)
- SyGuS, [3022](#)
- symbol manipulation, [3016](#)
- symbolic AI
 - vs. connectionist AI, [3447](#)
- symbolic computation, [3031–3033](#)
- symbolic differentiation, [3031](#)
- symbolic execution, [3023](#)
- symbolic integration, [3031](#)
- symbolic music
 - discrete diffusion, [944](#)
- symbolic music diffusion, [908](#)
 - accompaniment, [910](#)
 - exercise, [928](#)
 - sparsity, [909](#)
- symbolic music generation, [904–910](#)
 - diffusion, [908–910](#)
 - MIDI, [904](#)
 - sequence length, [906](#)
 - transformer, [906–908](#)
- symbolic regression, [2047](#)
- symbolicism, [2833](#)
- Symphony, [1338](#)
- symplectic structure, [326](#)
- symptomatogenic zone, [3775](#)
- SymPy, [3031](#)
- synaptic consolidation, [1176](#), [1177](#)
- synaptic flow, [1931](#)
- synaptic homeostasis hypothesis, [1871](#), [1964](#)
 - detail, [1872](#)

- Synaptic Intelligence, 1188–1190
 - comparison with EWC, 1190
 - definition, 1189
- synaptic plasticity
 - rate, 1851
- synaptic pruning, 1845, 1849–1853
 - timeline, 1851
- synaptogenesis, 1846
 - postnatal, 1847
- sync loss, 745
- synchronous federated learning, 4451
- synchronous training, 5022
- synergy
 - generative and foundation models, 3694
- synesthetic retrieval, 2476, 2483
- SynFlow, 1931
- SynLBD, 2711
- synoptic-scale forecasting, 4222
- syntactic verification, 2213
- syntactic role
 - attention, 4145
- syntax extensions, 2992
- syntax-guided synthesis, 3022
- synthesis
 - of mathematical AI approaches, 3181
 - structured data, 2896–2904
- synthetic 4D cardiac CT
 - corpus, 4615
- synthetic control arms, 4618
- synthetic data
 - anatomically consistent, 3702
 - as privacy surrogate, 4456
 - autoformalization, 3071–3072
 - autoformalization exercise, 3078
 - clinical trials, 4618
 - conformal prediction, 2879
 - coverage analysis, 3123
 - definition, 2687
 - detection, 2843–2844
 - evaluation, 2704–2705, 2834
 - evaluation exercise, 2895
 - evaluation frameworks, 2849–2865
 - geometry, 3122–3123
 - legal ownership, 4649–4650
 - medical imaging, 3492
 - motivation, 2687–2691
 - non-IID reduction, 4576
 - privacy, 2844–2846, 3940
 - privacy guarantees, 3765
 - quality
 - federated, 4665
 - quality axes, 2706
 - quality metric, 2879
 - uncertainty, 2877
 - verification
 - decentralised, 4665
- synthetic data evaluation, 2740
- synthetic data fidelity
 - definition, 2704
- synthetic data generation, 2499–2508
 - evaluation trap, 2507
 - federated, 4456
 - risks, 2506–2508
- Synthetic Data Vault, 2849–2850
- synthetic digital twin, 4763
 - validation, 4763
- synthetic EEG
 - deepfake risk, 3933–3934
 - patient-specific pipeline, 3817
 - privacy analysis, 3830
 - weaponization, 3933
- synthetic medical images
 - risks, 3593
- synthetic microscopy, 3660–3662
- synthetic neural data
 - privacy shield, 3827
- synthetic pathology generation, 3610
- synthetic PET

- clinical utility, [3857](#)
- FCD detection, [3859](#)
- limitations, [3860](#)
- synthetic pretraining
 - privacy, [1692](#)
- synthetic proof generation, [3092](#)
- synthetic proofs
 - generation, [3071](#)
- synthetic time series, [2788–2803](#)
- SynthID, [931–932](#)
- System 2 intervention, [2173](#)
- system design
 - exercise, [2773](#)
- System Usability Scale (SUS), [4886–4887](#)
- systematic bias
 - NWP, [4222](#)
- systematic error
 - NWP, [4225](#)
- systems reply
 - Chinese Room, [4147](#)
- systems research
 - call to action, [5167](#)
- systole, [3513](#)
- systolic array
 - architecture, [4937](#)
 - arithmetic intensity, [4939](#)
 - definition, [4937](#)
- Szegedy, Christian, [2960, 3034](#)
- Sørensen–Dice coefficient, *see* Dice coefficient
- T-GATE, [649](#)
- T5
 - frozen text encoder, [4996](#)
 - text encoder, [615](#)
- T5 encoder
 - video diffusion, [685](#)
- Tabby, [2815–2834](#)
 - architecture, [2816–2818](#)
 - training algorithm, [2817](#)
- TabDDPM, [2758–2774](#)
 - advantages, [2769](#)
 - architecture, [2764–2765](#)
 - combined forward process, [2765](#)
 - comparison with GANs, [2772](#)
 - comparison with LLMs, [2804](#)
 - continuous-time, [2770–2771](#)
 - ELBO, [2765–2767](#)
 - practical considerations, [2771–2772](#)
 - reverse process, [2763–2768](#)
 - sampling, [2768](#)
 - training, [2767–2768](#)
 - worked example, [2774–2775](#)
- TabStruct, [2834–2847](#)
 - worked example, [2847](#)
- TabuLa, [2804–2815](#)
 - advantages and limitations, [2813–2815](#)
 - architecture, [2807–2809](#)
 - training algorithm, [2808](#)
- tabula rasa learning, [3114](#)
- tabular data, [2692–2696](#)
 - as text, [2804](#)
 - challenges, [2714](#)
 - column relationships, [2745](#)
 - consistency models, [2876](#)
 - example, [2694](#)
 - federated GAN, [4491](#)
 - formal definition, [2692](#)
 - geometry, [2695](#)
- tabular GAN, [2713–2729](#)
 - CTAB-GAN+, [2729–2743](#)
 - memorization, [2898](#)
 - open problems, [2743](#)
- tabular generation
 - advanced LLM, [2815–2834](#)
 - comparison, [2815, 2858–2861](#)
 - diffusion-based, [2758–2774](#)
 - diversification era, [2915](#)
 - early era, [2912](#)

- foundation era, 2915
- graph-based, 2745–2758
- history, 2826–2834, 2912–2916
- LLM comparison, 2828
- LLM revolution, 2826–2834
- LLM-based, 2804–2815
- standardization era, 2912
- timeline, 2710–2713, 2826
- tabular serialization
 - definition, 2805
- tabular-to-image encoding, 2729
 - definition, 2730
- tactic
 - apply, 2970
 - as MDP action, 3096
 - calc, 2971
 - cases, 2971
 - constructor, 2971
 - definition, 2969
 - exact, 2970
 - have, 2971
 - induction, 2971
 - intro, 2970
 - omega, 2970
 - ring, 2970
 - rw, 2970
 - simp, 2970
- tactic generation, 2961, 3083
 - pipeline, 3094
 - variable length, 3128
- tactic language, 2206
- tactic prediction, 3081
- tactic suggestion, 3085
- TacticM monad, 2991
- tactics
 - as actions, 3035
 - custom, 2991
 - Lean 4, 2987–2999
 - proof commands, 2944
- tail risk, 4194
 - weather forecasting, 4405
- Talagrand diagram, 4194
- talking head synthesis, 744–745
- tanh activation, 2715
- Tao, Terence, 3112
 - on Erdos problems, 3163
- Tarski–Grothendieck set theory, 3003, 3035
- task decomposition, 1290
- task prioritisation
 - LLM, 3914
- task sequence, 1177
- task similarity, 1182
- task vector
 - definition, 4078
 - properties, 4078
- task vectors, 4078–4084
 - colour task example, 4080
 - cross-modal alignment score
 - exercise, 4084
 - cross-modal patching, 4079
 - cross-modal transfer, 4168
 - figure, 4081
 - layer-dependent alignment, 4079
 - LLM to VLM transfer, 4080
 - modality invariance, 4079
 - orthogonality
 - exercise, 4084
 - patching effect, 4080
- task-coupled generation, 3707–3711
 - definition, 3707
- task-incremental learning, 1177
- taxicab metric, 5195
- taxonomy of generative models, 6
- TBPTL, *see* Truncated Backpropagation Through Loops
- TC^0 , 1036, 3047
 - exercise, 1066
- teacher forcing, 429, 443–444
 - definition, 443

- teacher-student
 - architecture consolidation, 1966
- teacher-student framework
 - FedDT, 4582
- teeth
 - in face models, 4737
- teleprinter
 - TTY, 4693
- temperature
 - diversity in evolution, 2121
 - in contrastive learning, 2369
- temperature annealing
 - convergence, 1904
 - in NAS, 1903
- temperature evolution
 - atmosphere, 4205
- temperature forecasting
 - MOS, 4224
- temperature gradient
 - downscaling, 4355
- temperature parameter
 - in contrastive learning, 2410
- temperature scaling
 - tabular, 2810
 - trade-off, 2810
- temperature-load relationship
 - nonlinear, 4233
- tempered posterior, 664
- template-based formalization, 3066
- Temple University EEG Corpus, 3781
 - bias, 3941
- Temple University Hospital EEG
 - Corpus, 3887
- Temple University Hospital EEG corpus,
 - 3902
- temporal alignment
 - optimality, 900
- temporal attention, 536
 - window, 697
- temporal autocorrelation, 529–533
 - decay, 529
 - definition, 529
 - noise schedule, 630
 - optimal schedule, 630
- temporal caching
 - error bound, 759
- temporal coding, 1857
- temporal coherence, 522, 2788, 2792
 - definition, 530
 - desideratum, 523
 - formal definition, 530
 - measurement, 706–711
 - requirements, 522
 - subsampling, 531
- temporal compression, 534
 - bound, 568
- temporal concept schedule, 4097
 - exercises, 4103
 - practical computation, 4098
- temporal consistency, 706–711, 3508
 - attention window, 708
 - bounds, 740
 - CLIP-based metric, 738
 - degradation, 708
 - diversity trade-off, 710
 - exercise, 711
 - identity, 707
 - improvement methods, 709
 - long-range, 752
 - metrics, 706, 738
 - score, 706
 - shared noise, 710
- temporal constraints, 2703
- temporal convolution, 536
 - EEG, 3851
- temporal data
 - challenges, 2708
- temporal dependencies, 2788
- temporal discriminator, 4364
- temporal dynamics

- fuzzy, [4234](#)
 - shallow MLP limitation, [4237](#)
- temporal feature caching, [647](#)
 - definition, [647](#)
- temporal features
 - naive encoding, [2795](#)
- temporal grounding, [2421–2427](#)
 - instructional videos, [2422](#)
 - loss function, [2422](#)
 - methods, [2421](#)
- temporal inpainting, [886](#)
- temporal lobe epilepsy, [3777](#)
 - SEEG, [3820](#)
 - surgery, [3776](#)
- temporal lobe FCD, [3860](#)
- temporal localisation, [932](#)
- temporal micro-structure, [935](#)
- temporal receptive field, [536](#)
- temporal resolution, [2618](#)
- temporal retrieval, [2417–2427](#)
- temporal split
 - requirement, [3919](#)
- temporal splitting
 - exercise, [3927](#)
- temporal structure
 - evaluation, [2864](#)
- temporal transformer
 - ordering, [2420](#)
- temporally correlated noise, [710](#)
 - exercise, [711](#)
- temporary disability, [4691](#)
- Tendsto, [2994](#)
- Tensor Cores
 - introduction, [4935](#)
- tensor parallelism, [5041–5050](#)
 - all-reduce, [5137](#)
 - architectural implications, [4927](#)
 - backward pass, [5049](#)
 - bandwidth exercise, [4941](#)
 - communication hiding, [4997](#)
 - communication volume, [5047](#)
 - definition, [5043](#)
 - degree 8, [5142](#)
 - DiT, [5084–5085](#)
 - formal definition, [5043](#)
 - limits, [5049](#)
 - memory analysis, [5088](#)
 - memory savings, [5173](#)
 - NCCL, [4977](#)
 - NVLink placement, [4949](#)
 - NVLink requirement, [4936](#)
 - overlap, [5050](#)
 - transformers, [4931](#)
- tensor programs
 - distributed AD, [5163](#)
- term frequency, [2350](#)
- termination checking, [2986](#)
- ternary compression, [4478](#), [4582](#)
- ternary quantisation, [4582](#)
 - definition, [4583](#)
- ternary quantization, [4477](#)
- terrain effects
 - NWP, [4222](#)
- Terzopoulos, Demetri, [3525](#)
- Tesseract
 - OCR, [4698](#)
- test data generation, [2690](#)
- test generation, [3023](#)
- test-time compute, [3147](#)
 - AlphaProof, [3119](#)
 - definition, [3147](#)
- test-time compute scaling, [3051](#)
- testing
 - limitations, [2934](#)
- testing effect, [1175](#)
- text
 - sequential structure, [2688](#)
- text conditioning
 - diffusion, [5083](#)
 - video, [537](#)

- text contrast
 - Daltonization, [4778](#)
- text embeddings, [2385–2396](#)
- text simplification, [4914](#)
 - cognitive accessibility, [4694](#)
 - reading level, [4701](#)
- text-guided segmentation, [3520–3521](#), [3680–3692](#), [3755](#)
- text-to-audio retrieval, [2469](#)
- text-to-image
 - federated, [4535–4536](#)
- text-to-image retrieval, [2467](#)
- text-to-speech
 - concatenative, [4698](#)
 - generic, *see* voice cloning
 - HMM limitations, [4698](#)
 - HMM synthesis, [4698](#)
 - voice cloning, [4689](#)
- text-to-video, [683–689](#)
 - dual encoding, [685](#)
 - memory requirements, [4932](#)
 - pipeline diagram, [687](#)
 - pipeline overview, [683](#)
 - prompt rewriting, [684](#)
 - system considerations, [687](#)
- texture synthesis, [334–336](#)
 - methods, [334](#)
 - problem formulation, [334](#)
- TF-IDF, [2350–2351](#), [2627](#)
 - connection to language models, [2355](#)
 - example, [2351](#)
 - scoring function, [2351](#)
- TGSAM-2, [3680–3682](#)
 - endoscopic polyp segmentation, [3681](#)
- theorem prover
 - Lean 4, [2964–2983](#)
- theorem provers, [2205–2209](#)
 - comparative survey, [2999–3015](#)
 - comparison, [3005](#)
- theorem proving
 - agent-based, [1458](#)
 - AlphaProof, [3112–3131](#)
 - as game playing, [3115](#)
 - classical, [3015–3031](#)
 - example, [2536](#)
 - practical formalization, [2984–2999](#)
 - recall requirement, [2568](#)
 - reinforcement learning, [3095–3112](#)
 - retrieval, [2565–2574](#)
 - diagram, [2679](#)
 - retrieval for, [2566–2574](#)
 - retrieval-augmented proof search, [2567](#)
 - search complexity, [2567](#)
 - with retrieval, [2535](#)
- TheoremLlama, [3091–3092](#)
 - fine-tuning, [3092](#)
 - informal-formal alignment, [3092](#)
 - results, [3092](#)
- theory conflict clause, [3024](#)
- theory solvers, [3022](#)
- thermodynamic energy equation, [4205](#)
- thermodynamic limits
 - computation, [4812](#)
- thermodynamics
 - limits on evolution, [2264](#)
- thermoelasticity, [3376](#)
 - exercise, [3376](#)
- thermosphere, [4208](#)
- theta waves, [3793](#)
- thickness
 - 1000–500 hPa, [4225](#)
 - predictor, [4226](#)
- thin structure segmentation, [3513](#)
- thing classes, [3505](#)
- thinking time, [3051](#)
- thinking tokens, [3051](#)
- Thompson sampling, [2180](#)

- thorax segmentation, 3510–3511
- thought communication, 1310–1314
 - generative model, 1311
 - identifiability, 1312, 1448
- thought retrieval, 2481
- thread
 - GPU concurrency, 4934
- threat model
 - black-box, 1660
 - white-box, 1637
- threat models, 1582
 - black-box, 1583
 - gray-box, 1583
 - label-only, 1584
 - white-box, 1583
- Threat-Focused Guidance
 - TFG, 4594
- three walls
 - memory, compute, communication, 4929
- three-electrode seizure prediction, 3883
- three-stage evolution, 4082
- thresholding, 3524–3525
 - adaptive, 3525
 - adjacency matrix, 2748
 - global, 3524
- throughput
 - distributed training, 5058
- Thurston, William, 3142
- Tierra, 2240–2241
 - complexity trajectory, 2240
- TIGER, 2513–2514, 2581–2583
 - music recommendation, 2513
- TIGGE, 4196
- Tikhonov regularisation, 3400
- tiling
 - memory optimisation, 4935
- timbre transfer, 893–895
 - definition, 893
 - diffusion inversion, 894
- time constraints
 - IMO, 3033
- time embedding, 2764
- time integration
 - NWP, 4215
- time series
 - challenges, 2788–2789
 - conditional generation, 2799–2800
 - generation, 2696–2697
 - generation challenges, 2788
 - infilling, 2799
 - length, 2789
 - multi-scale, 2788
 - non-stationary, 2788
 - segment concatenation, 2789
 - ubiquity, 2788
- time series generation, 2788–2803
 - comparison, 2801
- time-frequency coupling
 - pre-ictal, 3885
- time-frequency features
 - transformer, 3884
- time-frequency representations, 3798
- time-lapse microscopy
 - cell tracking example, 3517
 - segmentation, 3508
- time-series generation
 - federated
 - WaveStitch, 4629
- time-stepping
 - diffusion models, 3281
- time-to-action
 - triage, 3925
- TimeGAN
 - SEEG, 3820–3821
- timeout
 - in proof search, 3101
- timestep vulnerability, 1627
 - intermediate, 1628
- timestep-dependent architecture, 1940

- timestep-stratified SAE, [4099](#)
- timing conditioning, [843](#)
 - definition, [843](#)
 - exercise, [957](#)
- Tiny Recursive Model, [1060](#)
 - convergence, [1060](#)
 - exercise, [1069](#)
- TinyML
 - accessibility, [4824](#)
 - definition, [4824](#)
- Tishby, Naftali, [1860](#)
- TissueNet dataset, [3644](#)
- Titans, [1212–1213](#)
 - architecture, [1212](#)
 - memory capacity, [1213](#)
 - memory module, [1212](#)
- TOFU benchmark, [1800](#)
 - definition, [1800](#)
 - gradient ascent example, [1801](#)
- token dependencies, [3993](#)
- token dropping
 - MoE, [5142](#)
- token economics, [1349–1354](#)
 - performance prediction, [1349](#)
- token embedding
 - W_E , [4003](#)
- token pattern
 - audio codecs, [796](#)
 - coarse-to-fine, [796](#)
 - delay, [796](#)
 - flat, [796](#)
 - tradeoff, [796](#)
- token-then-render, [731](#)
- TokenFlow, [715](#)
- tokenisation
 - as universal interface, [1961](#)
 - audio, [795](#)
 - physics-aware, [3245](#)
- tokenization
 - of proofs, [3037](#)
- Tonegawa, Susumu, [2342](#)
- Tononi, Giulio, [1872](#), [1964](#), [4148](#)
- tool integration
 - AI and provers, [3013](#)
- tool use, [1282–1284](#)
- tool-augmented agent, [1282](#)
- top- k routing, *see* Mixture-of-Experts, routing
- top- k sampling
 - tabular, [2810](#)
- Top-K sparsification, [4476](#), [5009](#)
 - definition, [5009](#)
 - error feedback, [5010](#)
- topological data analysis
 - epilepsy, [3950](#)
- topological phase diagram, [3865–3866](#)
- topological space, [5216](#)
- topology
 - adaptive, [1460](#)
 - algorithm co-design, [4942](#)
 - key idea, [4950](#)
 - message complexity, [1449](#)
 - vascular segmentation, [3513](#)
- topology-aware generation, [3514](#)
- torchrun
 - elastic, [5148](#)
- torus network
 - bisection bandwidth, [5170](#)
- torus topology, [4945](#)
- total correlation
 - disentanglement
 - EEG, [3943](#)
 - mask averaging, [3557](#)
- total variation
 - denoising, [329](#)
 - exercises, [338](#)
 - temporal, [531](#)
- total variation distance, [78](#), [2837](#)
- TotalSegmentator, [3489](#), [3503](#)
- Toth, Zoltan, [4287](#)

- tournament evaluation, 2158
- toxic language
 - disability terminology, 4843
- toy model of superposition, 4040–4046
- TPTP format, 3003
- TPU
 - AlphaProof, 3113
 - architecture, 4937
 - matrix unit size, 4939
 - weather forecasting, 4314
- TPU Pod
 - torus interconnect, 4945
- TPU v4
 - specification, 4939
- TPU v5
 - specification, 4939
- TPU v5p
 - ICI advantage, 4939
- TRA metric, 3644
 - exercise, 3665
- trace, 5223
- track accuracy
 - hurricane, 4399
- tracking graph, 3655
- tracking-by-detection, 3652, 3752
- tracking-by-segmentation, 3652
- tractable density, 7
- tractable likelihood
 - normalizing flows, 3855
- tractography
 - seizure propagation exercise, 3897
 - seizure propagation figure, 3893
 - transformers, 3889–3896
- tractography parcellation, 3890
- TractSeg atlas, 3891
- traffic flow
 - federated synthesis, 4625
- train analogy
 - alpha-beta model, 4951
- train on synthetic, test on real, 2705, 2841–2843
- train-on-synthetic paradigm, 3817
- training, 9
 - algorithm, 858
 - audio diffusion, 853–862
 - budget estimation, 859
 - composite loss, 853
 - data challenges, 855–862
 - energy-based cropping, 856
 - neural networks, 94
 - practical considerations, 859
 - stability, 857–862
- training bottleneck
 - three walls, 4925
- training compute
 - example, 4940
- training cost
 - video diffusion, 554
- training data
 - disability bias, 4902
 - disability weighting, 4907
 - from formal libraries, 3013
 - PDE operators, 3230
 - representation bias, 4703
 - scale, 4926
 - scarcity in mathematics, 2963
- training dynamics
 - factual knowledge, 4071
 - phase change, 4026
- training loss
 - diffusion model, 4311
- training memory
 - breakdown, 5042
- training objective, 6
- training oscillation, 2758
- training sample
 - class imbalance, 4234
- training set size
 - early ML meteorology, 4236
- training simplicity

- diffusion vs GAN, 2768
- training stability
 - diffusion, 2769
 - GOGGLE vs CTGAN, 2757
- training system design
 - exercise, 5080
- training time
 - example, 4940
 - single GPU, 4926
- training-by-sampling, 2714, 2721
 - algorithm, 2722
 - coverage proof, 2918
 - exercise, 2916
 - oversampling bound, 2724
 - why it works, 2722
- training-free NAS, 1930–1932
 - error bound, 1931
- trajectory balance, 2869
 - condition, 2869
- trajectory constraints, 720
- trajectory curvature, 637
- trajectory reward, 1295
- TrajGRU, 4239–4370
 - inductive bias, 4372
 - update rule, 4370
- transcoder, 4051–4058
 - definition, 4051
 - exercises, 4058
- transcript-based editing, 4872
- transfer learning
 - bound, 654
 - EEG classification, 3818
 - epilepsy pathology, 3895
 - evolutionary, 2275–2276
 - research directions, 2276
 - implications, 655
 - PDEs, 3256
 - tabular data, 2908
 - video, 654
- transfer ratio, 2276
- Transformer, 101, 2629
 - as PINN backbone, 3260
 - circuit complexity, 1037
 - convergence, 3044
 - elliptic PDEs, 3301
 - for theorem proving, 3036
 - in TC^0 , 1037
 - limitations, 1037
 - million-point, 3321
 - modern components, 113
 - motivation for PDEs, 3234
 - neural operator, 3465
 - parallel block, 114
 - PDE solvers, 3233–3264
 - comparison, 3263
 - pre-norm vs. post-norm, 104
 - tabular applications, 2804
- transformer, 4124
 - architectural evolution, 1943
 - architecture search, 1943
 - bias auditing, 4138
 - channel selection
 - iEEG, 3946
 - comparison
 - seizure prediction, 3888
 - diffusion backbone, 5083
 - EEG encoding, 3851
 - elastic, 1945
 - epilepsy applications, 3786
 - geometry, 3122
 - medical segmentation, 3541–3543, 3603
 - memory analysis, 5173
 - modality conquest, 1961
 - searchable components, 1943
 - seizure prediction, 3879–3889
 - comparison table, 3889
 - strengths and limitations, 3543
 - universal substrate, 1961
- transformer circuits, 4002–4012

- composition, 4022
- induction heads, 4020–4028
- transformer output decomposition, 4009–4010
- theorem, 4009
- transformer revolution, 3046
- Transformers
 - multiphysics operators, 3375
 - quantum chemistry, 3366
- transformers
 - computational limitations, 3047
 - mathematical reasoning, 3046–3064
- transit map
 - accessibility, 4771
- transition
 - shallow to deep learning, 4241
- translational equivariance
 - ConvLSTM, 4239
- Transolver, 3245–3327
 - aggregation, 3246
 - airfoil, 3328
 - attention, 3246
 - complex geometries, 3247
 - complexity, 3247
 - distribution, 3246
 - motivation, 3245
 - physics-attention, 3246
 - slice assignment, 3246
- transparency
 - definition, 3964
 - formal definition, 3964
- transport plan, 369
 - exercise, 382
- transposed convolution, 97, 3532
- TransUNet, 3487, 3541
 - architecture, 3541
- travel stages
 - BLV navigation, 4718
- treatment planning
 - segmentation role, 3479
- TREC, 2349, 2627
- tree aggregation
 - privacy, 4627
- tree all-reduce, 4973–4974, 5032
 - benchmarking, 5179
 - NCCL 2.x, 4972
- Tree of Thoughts, 1043–1044
 - BFS algorithm, 1044
 - definition, 1044
 - exercise, 1070
- tree-structured networks, 3036
- triangular map, 451
- trillion-parameter models, 4153
- TripleGAN
 - 99.7 percent accuracy, 3819
 - consistency loss analysis, 3830
 - EEG, 3819
 - three-domain coupling, 3819
- triplet loss, 2392
- tritanopia, 4770
- TRM, *see* Tiny Recursive Model
- tropical cyclone
 - path distribution, 4319
 - track versus intensity, 4399
- tropical cyclone intensity
 - ANN, 4234
- tropical cyclones
 - ensemble forecasting, 4315
 - multi-source forecasting, 4380
- troposphere, 4208
- TRTR baseline, 2741
- TRTS, 2842
- true but unprovable statements, 2939
- Truncated Backpropagation Through
 - Loops, 1060, 1061
- truncated sampling
 - fluid dynamics, 3388
- truncation trick, 354
 - exercise, 360
- trust

- calibration
 - assistive AI, [4861](#)
 - formal verification, [3085](#)
- trust in computation, [3033](#)
- trusted computing base, [2992](#), [3011](#)
- trusted kernel, [2943](#)
- trustworthiness
 - neural PDE solvers, [3437](#)
- Trybulec, Andrzej, [2944](#), [3003](#)
- Tseitin transformation, [3017](#)
- TSTR, [2841–2843](#)
 - augmentation, [2842](#)
 - definition, [2841](#)
 - multiple models, [2842](#)
- TSTR protocol, [2741](#)
- TTUR, *see* two-timescale learning
- TUH
 - preprocessing, [3920](#)
- TUH EEG Corpus, [3800](#)
 - description, [3801](#)
 - pretraining, [3887–3888](#)
- TUH EEG corpus, [3920](#)
- Tulving, Endel, [2341](#)
- tumour board, [3553](#)
 - definition, [3553](#)
- tumour core, [3552](#)
- tumour margins
 - infiltrative, [3549](#)
- tumour volume
 - monitoring, [3553](#)
- tumour-infiltrating lymphocytes, [3480](#)
- tumour-related epilepsy, [3920](#)
- turbo the snail
 - AlphaProof solution, [3126](#)
- turbo the snail (IMO 2024), [3113](#)
- turbulence, [3333–3352](#)
 - L^2 loss failure, [3340](#)
 - computational challenge, [3333](#)
 - CoNFILd, [3393](#)
 - diffusion models, [3336](#)
 - ECHO compression, [3394](#)
 - energy cascade, [3334](#)
 - equilibrium distribution, [3337](#)
 - GAN with spectrum constraint, [3336](#)
 - GANs, [3336](#)
 - generative models, [3333–3352](#)
 - Kolmogorov microscale, [3453](#)
 - Kolmogorov theory, [3197](#)
 - mode collapse, [3341](#)
 - operator-generative hybrid, [3339](#)
 - real-time, [3433](#)
 - scaling laws, [3469](#)
 - super-resolution, [3339](#)
 - diffusion, [3339](#)
 - surrogate comparison, [3466](#)
 - uncertainty quantification, [3413](#)
- turbulence simulation, [15](#)
- Turing completeness
 - of ADAS search space, [2182](#)
- Turing machine simulation, [3049](#)
- Turing patterns, [3305](#)
- Turing, Alan, [3016](#)
 - morphogenesis, [3305](#)
- turn-count power law, [1304–1308](#)
 - theorem, [1304](#)
- Turpin, M., [3971](#)
- Tutel, [5143](#)
- TVD, *see* total variation distance
- twCRPS
 - extreme events, [4316](#)
- Tweedie estimate, [2780](#)
- Tweedie estimate
 - and conditional score, [2782](#)
 - Harpoon, [2780](#)
- Tweedie formula
 - fluid fields, [3388](#)
 - function space, [3381](#)
- Tweedie’s formula, [502](#)
- two-alternative forced choice, [741](#)

- two-stage neural network
 - load forecasting, 4234
- two-timescale learning, 321
 - definition, 321
 - exercises, 326
 - intuition, 321
- two-tower model
 - algorithm, 2583
 - for recommendation, 2577
 - generalization bound, 2577
- Type (universe), 2966
- type safety
 - mathematical objects, 2985
- type theory, 2946–2953
- type-level computation, 2984
- typeclass, 2973
 - inference failure, 2981
- U-FNO, 3223
- U-Net, 99, 3487, 3532
 - 1D vs. 2D parameter count, 847
 - 2D, 3506
 - architecture, 596
 - architecture limitations, 1935
 - audio diffusion, 842
 - backbone, 3560
 - bias circuits, 4112
 - bottleneck, 3533
 - channel dimensions, 3533
 - comparison with DiT, 596
 - contracting path, 3533
 - copy and crop, 3534
 - data augmentation, 3535
 - encoder equations, 3533
 - expanding path, 3533
 - federated diffusion, 4459
 - federated synthetic pre-training, 4619
 - GAN generator, 3816
 - loss function, 3534
 - one-dimensional, 3816
 - output layer, 3534
 - parameter count, 3741
 - PDE solver, 3417
 - receptive field, 3547, 3741
 - retinal vessels, 3535
 - strengths, 1941
 - vs FNO, 3306
 - vs. DiT, 1941
 - weight map, 3535
- U-shaped relationship
 - temperature, 4233
- U-shaped scaling law, 1094
 - exercise, 1111
- U-Transformer
 - FuXi, 4275
- UAU-Net, 3685
- UCB formula, 3042
- UCB score
 - exploration constant, 3110
 - theorem proving, 3100
- UCF-101, 22
- Udio, 913–914
- UHB FCD dataset, 3920
- UKMO UM, 4218
- ultrasound
 - automated diagnosis, 3908
 - cine segmentation, 3508
 - segmentation, 3481
 - VQA, 3907
- uncanny valley, 523
- uncertainty
 - abstention, 4074
 - and diversity, 2081
 - discarded by discriminative models, 3490
 - sources in synthetic data, 2878
- uncertainty calibration
 - clinical decision support, 3769
 - distribution shift, 3712

- uncertainty estimation
 - clinical value, [3686](#)
 - semantic characterisation, [3686](#)
 - vision-language models, [3685–3686](#)
- uncertainty principle
 - time-frequency, [774](#)
- uncertainty propagation
 - Bayesian network, [4235](#)
- uncertainty quantification, [2877–2884](#), [3409–3427](#)
 - 4D volumes, [3727](#)
 - motivation, [3410](#)
 - neural operators, [3229](#)
 - NWP, [4219](#)
 - synthesis, [3860](#)
- uncertainty sampling, [2224](#)
- uncertainty-based task weighting, [4278](#)
- uncertainty-guided synthesis, [2882](#)
- uncinate fasciculus, [3891](#)
- unconditional diffusion prior, [4321](#)
- undecidability, [3016](#)
- underdamped Langevin dynamics, [1679](#)
- underflow
 - FP16 gradients, [5074](#)
- undersampling, [3805](#)
- understanding
 - computation vs., [4147](#)
 - neural networks, [3445](#)
- unembedding matrix
 - W_U , [4003](#)
- UNet
 - pipeline parallelism, [4996](#)
- unification, [3020–3021](#)
- unified architectures, [1960–1963](#)
- unified audio generation, [948](#)
 - formal statement, [948](#)
- unified evolutionary view, [2255–2256](#)
- unified exploration framework, [2232](#)
- unified framework
 - medical segmentation, [3711](#)
- unifier, [3020](#)
- uniform distribution, [33](#)
- uniform time-step sampling, [2767](#)
- uninterpreted functions, [3023](#)
- union bound, [42](#)
- unique factorisation, [2935](#)
- uniqueness constraints, [2825](#)
- UniSim, [723](#)
- Unisolver, [3253–3256](#), [3436](#)
 - architecture, [3253](#)
 - universality, [3253](#)
 - zero-shot, [3254](#)
- unit propagation, [3018](#)
- univalence axiom, [3002](#)
- universal acid, [2022](#)
- universal approximation
 - Fourier Neural Operator, [3224](#)
 - GMM, [156](#), [162](#)
 - operators, [3212](#), [3458](#)
- universal approximation theorem, [93](#)
 - limitation, [4237](#)
 - meteorology, [4232](#)
- universal design, [4691](#), [4882–4883](#)
 - generative AI enhancement, [4882](#)
 - generative models, [4904](#)
- universal embedding hypothesis, [2486](#)
- universal embedding model, [2390](#)
 - exercises, [2396](#)
- universal formal language, [3016](#)
- Universal Grammar, [1174](#)
- universal medical segmentation,
 - [3711–3713](#)
 - challenges, [3712](#)
 - deployment warning, [3713](#)
 - formal specification, [3712](#)
 - four pillars, [3711](#)
- universal quantification
 - as Pi-type, [2985](#)
- Universal Reasoning Model, [1061](#)
- universal retrieval, [2616–2617](#)

- definition, 2616
- information-theoretic limit, 2617
- research directions, 2617
- universal segmentation model, 3729
- Universal Transformer, 1057
 - Turing completeness, 1057
- universality conjecture, 2624
- universe
 - Lean, 2966
- universe error, 2981
- universes
 - type theory, 2948
- University of Cyprus
 - meteorological ANN, 4235
- unknown onset seizure, 3777
- unlearning
 - vs retraining, 4604
- unlearning verification
 - baseline MIA, 4605
 - criterion, 4606
 - membership inference, 4605–4606
 - post-unlearning MIA, 4605
 - retrain baseline, 4605
- unpaired image translation, 3582
- unreasonable effectiveness, 3447
- unrolled optimisation, 321
 - definition, 322
 - exercises, 326
- unsafe AI generation
 - disability, 4844
- unused parameters
 - DDP, 5025
- upper confidence bound, 2227
- upper confidence bound for trees, 3114
- upsampling, 97
 - vocoder, 780
- Urban, Josef, 2956, 3034
- URM, *see* Universal Reasoning Model
- user study
 - accessibility AI, 4918
- User-Centered Interpretability, 3965
- user-centred evaluation, 4886–4887
- utility metrics, 2840–2841
- utility-privacy tradeoff, 2890
- v -parameterisation
 - audio, 850
- V-ASR loss, 4739
 - feedback loop, 4740
- V-composition, 4006
 - definition, 4008
 - exercises, 4011
- V-Net, 3537–3538
 - architecture, 3537
 - Dice loss, 3537
 - residual connections, 3537
- v -prediction, 639–644
 - advantages, 640
- v -structure, 129, 141
- V1
 - sparse coding, 4044
- V100
 - specification, 4935
 - Tensor Cores, 4935
- VAE, 7
 - atmospheric compression, 4319
 - EEG summary, 3828
 - ELBO, 3823–3824
 - encoder
 - frozen, 5082
 - epilepsy applications, 3785
 - federated, 4459
 - ELBO, 4465
 - for prompt generation, 3698
 - gradient leakage, 4453
 - latent code opacity, 3994
 - latent space interpretability, 3994
 - limitations, 2758
 - PDE fields, 3393
 - reconstruction loss, 4341

- reconstruction-regularisation
 - tradeoff, [3829](#)
- tabular generation exercise, [2922](#)
- voice cloning, [4749](#)
- weather latent space, [4341](#)
- VAE-cGAN
 - 76 percent IED accuracy, [3825](#)
 - scalp-to-iEEG, [3825](#)
- VAE-HGB
 - architecture, [4622](#)
 - IoT anomaly detection, [4622–4623](#)
- VAEFL, [4517–4526](#)
 - defense mechanism, [4519](#)
 - encryption analogy, [4524](#)
 - reconstruction error bound, [4520–4521](#)
- VAEs
 - vs. other families, [2709](#)
- valid prediction time, [3422](#)
- validity checking, [2811](#)
- validity rate
 - confidence bound, [2832](#)
- VALL-E, [4747](#)
 - architecture, [4748](#)
 - duration modelling, [4757](#)
- valley cold-pooling, [4226](#)
- value alignment
 - challenges, [2262](#)
 - coevolution conjecture, [2262](#)
 - coevolution hypothesis, [2261–2263](#)
- value function
 - approximation error, [3054](#)
 - challenges in theorem proving, [3103](#)
 - learned, [3042](#)
- value incommensurability, [4135–4137](#)
- value network
 - AlphaGo, [3114](#)
 - pre-training, [3118](#)
 - proof search, [3128](#)
 - theorem proving, [3100](#)
 - training, [3101](#)
- Vampire (prover), [3020–3021](#)
- Vampire (theorem prover), [2955](#)
- van der Schaar, Mihaela, [2775](#), [2834](#)
- Van Valen, Leigh, [2030](#), [2154](#)
- vanishing gradient, [100](#)
 - early neural networks, [4237–4238](#)
 - exercise, [1066](#), [4243](#)
 - revisited, [1034](#)
 - RNN gradient product, [1034](#)
 - sigmoid activation, [4237](#)
- vanishing gradients
 - GAN, [280](#)
- Vanroose, Wim, [4986](#)
- VAR (Visual Autoregressive Modelling), [444](#)
- variable elimination, [135](#)
 - ordering, [143](#)
- variable weighting
 - ERA5, [4311](#)
- variable-length generation
 - timing, [957](#)
- variable-size output, [2696](#)
- variance, [35](#)
- variance dip, [2799](#)
- variance exploding, [495](#)
- variance preservation, [2799](#)
- variance preserving, [495](#)
- variance reduction, [221–224](#)
 - baseline, [1914](#)
- variance schedule, [479](#)
- variance-preserving path, [637](#)
- variation, [2019–2022](#)
- variation operator
 - classical, [2095](#)
 - semantic, [2097](#)
- variational autoencoder, [385–401](#)
 - and quality-diversity, [2089](#)
 - architecture, [391–392](#)
 - as retrieval system, [2609](#)

- definition, 391
- EEG, 3823–3829
 - ELBO, 3947
- ELBO, 4506
- exercise, 401
- federated, 4506–4517
- from GMM, 163
- graphical model, 144
- implementation, 396
- PDE fields, 3393
- reparameterisation trick, 4506
- sampling, 396
- training, 395–396
- variants, 403
- vs. GAN, 398
- variational autoencoders
 - VI perspective, 256
 - weather, 4198
- variational data assimilation, 4191, 4216
- variational diffusion model, 256, 479, 496
- variational EM, 209–213
 - algorithm, 210
 - convergence, 210
 - E-step, 210
 - latent factor model, 264
 - M-step, 210
- variational Gaussian mixture, 2718–2721
 - exercise, 2916
 - fitting, 2718
- variational inference, 137, 165–203
 - alternative divergences, 174
 - audio autoencoders, 940
 - connection to GFlowNets, 2871
 - diagnostics, 259
 - exercises, 261–269
 - importance weighted, 241
 - mean-field, 183
 - message passing, 251
 - method selection, 259
 - modern generative models, 255–258
 - motivation, 166
 - natural gradient, 248–251
 - normalizing flows, 236–241
 - open problems, 261
 - optimization equivalence, 268
 - particle-based, 243
 - pitfalls, 260
 - practical guide, 259–261
 - reverse KL, 172
 - video VAE, 747
- variational message passing, 197
- variational principle
 - exact, 252
- variogram
 - precipitation, 4356
- variogram score, 4336
 - definition, 4336
- vascular segmentation, 3513–3514, 3731
- Vaswani, Ashish, 2629
- VBench, 737
 - aesthetic quality, 738
 - appearance style, 738
 - background consistency, 738
 - colour, 738
 - decomposition, 737
 - dynamic degree, 738
 - human action, 738
 - imaging quality, 738
 - motion smoothness, 738
 - multiple objects, 738
 - object class, 738
 - overall consistency, 738
 - scene, 738
 - spatial relationship, 738
 - subject consistency, 738
 - temporal flickering, 738
 - temporal style, 738
- VBVR-Bench, 4116

- VCG mechanism, 1359
- VCTK, 4754
- VDocRAG, 2538
- Vector (dependent type), 2972
- Vector of Locally Aggregated Descriptors, *see* VLAD
- vector quantisation, 408
 - audio codecs, 787, 942
 - codebook, 409
 - compression, 575
 - definition, 787
 - EEG tokens, 3888
 - exercise, 589
 - loss, 576
 - training, 575
 - video, 574–579
 - definition, 575
- vector space
 - in Lean, 2979
- vector space model, 2349–2350, 2627
- vectors
 - length-indexed, 2984
- vehicle aerodynamics
 - Transolver, 3247
- velocity field
 - audio, 864
- Veo, 12, 4115
- verbal reinforcement learning, 1052
- verifiability
 - hierarchy, 2934
 - mathematical proofs, 2933
- verification
 - agent safety, 1441
 - co-evolved verifiers, 2270
 - complexity, 2211
 - coverage bound, 2570
 - evolutionary systems, 2269–2270
 - exercise, 2291
 - experimental vs formal, 3139
 - information-theoretic limits,
 - 2570–2574
 - mechanical, 3080
 - of AI mathematics, 3177
 - persistent, 3044
 - pipeline, 2678
 - retrieval-based, 2531–2574
 - software, 3024
 - with retrieval, 2535–2536
- verification advantage, 3182
- verification gap, 2269, 2936
 - definition, 2937
- verification hierarchy, 2942
- verification loop
 - mathematical research, 3134
- verification oracle, 2204
- verification pyramid, 2213–2215
- verification systems, 2203–2215
 - challenge, 2204–2205
- verification targets
 - ECMWF ENS, 4308
- verified multimodal agent, 2185
- vertical federated learning, 4451
- VGG features
 - perceptual loss, 3586
- VGG16
 - EEG transfer learning, 3818
- VGM, *see* variational Gaussian mixture
- vicinal risk minimisation, 1520
- video
 - continuous representation, 526
- video autoencoder, 534, 556–562
 - architecture, 560
 - causal, 562–574
 - evaluation metrics, 561
 - reconstruction quality, 562
 - spatial-only, 557
 - definition, 558
 - spatiotemporal, 558
 - definition, 558
 - trade-offs, 558

- video compression, 551
 - as pillar, 534
- Video ControlNet, 679
 - architecture diagram, 680
 - definition, 679
 - multiple controls, 683
- video corpus moment retrieval, 2421–2427
 - algorithm, 2423
 - definition, 2421
- video DDPM
 - sampling algorithm, 546
 - training algorithm, 544
- video diffusion, 521–556
 - adaptation, 689–696
 - applications, 741–746
 - architecture, 590–598
 - as world model, 721
 - attention, 598
 - audio connection, 943
 - chapter overview, 524
 - chapter summary, 763
 - computational challenges summary, 555
 - computational cost, 557
 - computational efficiency, 751
 - compute budgets, 658
 - conditioning, 613–620
 - connections to other chapters, 746–750
 - controllable generation, 675–683
 - exercises, 756–762
 - forward process, 756
 - from image diffusion, 540–548
 - guidance, 661–669
 - hybrid parallelism, 5142
 - Mixture-of-Experts, 620
 - motivation, 521
 - multi-condition guidance, 662
 - open problems, 751–756
 - positional encoding, 607
 - practical training, 658
 - training at scale, 652–661
 - summary, 661
 - world model, 762
- video diffusion model, 763
- video diffusion models, 3508
- video editing, 712–718
 - advanced methods, 715
 - attention injection, 714
 - comparison of methods, 716
 - diffusion-based, 712–718
 - exercise, 717
 - inversion bottleneck, 717
 - SDEdit, 712
- video frame interpolation, 584
- video generation, 12, 521–556
 - control, 537–540
 - desiderata, 523
 - evaluation, 733–741
 - future directions, 751–756
 - history, 524
 - image-conditioned, 669–675
 - inference speed, 645–652
 - interpretability, 4115–4124
 - long-form, 697–706
 - method comparison, 539
 - physics, 718
 - pillar interaction, 540
 - scale comparison, 522
 - tensor parallelism, 5048
 - three pillars, 533–540
- video generation pipeline
 - definition, 534
- Video Google, 2401–2402
 - system architecture, 2402
- video RAG, 2542
 - challenges, 2543
- video retrieval, 2417–2427
 - exercises, 2425

- open problems, 2426
- problem definition, 2418
- temporal challenge, 2418
- temporal reasoning, 2419
- training data, 2423
- video segmentation
 - medical, 3507
- video signal
 - continuous, 526
- video tensor, 521
 - definition, 527
- video tokenisation, 574–579
- video-conditioned music generation, 898
- video-EEG monitoring, 3779
- video-text retrieval, 2470
 - benchmarks, 2601
 - performance gap, 2601
 - temporal sensitivity, 2470
- video-to-music generation, 897–900
 - exercise, 929
- VideoGPT, 579
- VideoPoet, 728
- VideoViT, 4117–4118
- virtual creatures, 2059, 2276–2277, 2281
- virtual observables, 3277
- virtual try-on, 746
- virtual weights, 4006
 - definition, 4008
 - exercises, 4011
- viseme
 - ambiguity, 4736
 - classification exercise, 4743
 - definition, 4735
 - formal definition, 4735
 - generation, 4917
- viseme generation, 4735–4745
- vision impairment
 - prevalence, 4690
- Vision Transformer, 2408
 - connection to weather, 4254
 - CVD correction, 4779
 - transfer to weather, 4262
- vision transformer
 - EEG analogy, 3886
- Vision Transformer (ViT), 1961
- vision transformers
 - CLS token, 4088
 - interpretability, 4085–4095
- vision-language
 - historical convergence, 2425
- vision-language model
 - federated, 4638–4639
- vision-language models, 1771–1780, 3666–3680
 - accessibility, 4689, 4883
 - architecture, 1772
 - architecture types, 3673–3678
 - benchmarks, 3686–3688
 - clinical workflow, 3668
 - CLIP, 2408–2411
 - definition, 3668
 - embodied AI, 3691
 - encoder-based alignment, 3673
 - encoder-decoder, 3674
 - epilepsy, 3903–3908
 - future directions, 3691
 - hallucination, 3689–3691, 3755
 - historical development, 3692
 - hybrid architectures, 3678
 - image description, 4704
 - interactive dialogue, 3691
 - interpretability, 3995, 4076–4085
 - language transforms segmentation, 3666
 - language-centric bias, 4076
 - formal treatment, 4077
 - motivation, 3667
 - ambiguity, 3667
 - explainability, 3667

- rare conditions, 3667
 - rigidity, 3667
 - multimodal attention, 3674
 - objective complementarity, 3680
 - pretraining objectives, 3678–3680
 - self-supervision, 3691
 - zero-shot medical, 3767
- ViSQOL, 918
- visual entailment, 4888–4889
 - accessibility, 4834
 - formal definition, 4834
 - threshold calibration, 4840
- visual features
 - continuity, 4086
- visual fine-tuning
 - weight perturbation, 4081
- visual quality, 523
- visual question answering
 - knowledge-based, 2539
 - medical, 3907–3908
 - multi-level, 2539
- visual representations
 - linguistic anchoring, 4083–4084
- visual search, 2397–2406
- visual vocabulary, 413, 2400–2406
 - definition, 2400
- visual word, 2400
- visual-language pretraining
 - definition, 4710
 - museum accessibility, 4710
- VisualSpeaker, 4736
 - Gaussian splatting, 4738
 - transformer decoder, 4739
- ViT-H, 3620
- VITS, 4747
 - architecture, 4747
- VLAD, 2402
 - as special case of Fisher vector, 2404
 - definition, 2402
 - relationship to BoVW, 2403
- VLCL benchmark, 1225
- VLM, *see* vision-language models
 - accessibility, 4694
 - explainability, 3995
 - semantic interface, 3711
- VLM hallucination
 - benchmarks, 4889
- VLM probing, 4081–4085
- VLMs
 - accessibility connection, 4703
- vocabulary mismatch, 2364
- vocal identity
 - dignity, 4757
- vocal synthesis
 - quality frontier, 947
- vocoder, *see* neural vocoder
 - definition, 779
- vocoding
 - definition, 780
- vocoding problem, 780
- VOGN, 251
- voice banking, 4699, 4753
 - proactive, 4900
 - timing, 4753
- voice cloning, 4745–4758
 - accessibility, 4694
 - ALS, 4699
 - ALS patients, 4810
 - ethical concerns, 939
 - few-shot, 4901
 - low-resource, 4753
 - models, 4747
 - MOS WER tradeoffs, 4913
 - pipeline, 4754
 - privacy, 4756
- voice control
 - generative AI, 4864
- Voicebox, 4747
 - architecture, 4749
- volumetric segmentation, 3506

- abdominal CT, [3632](#)
- bidirectional propagation, [3637](#)
- propagation strategy, [3634](#)
- SAM, [3632–3633](#)
- SegVol, [3639–3640](#)
- slice selection, [3638](#)
- Von Neumann minimax theorem, [5269](#)
- von Neumann, John, [311](#), [2239](#), [4210](#)
 - self-replicating automata, [2239](#)
 - self-replicating machines, [1989](#)
- Voronoi partitioning, [2381](#)
- vortex stretching, [3428](#)
- voxel anisotropy, [3507](#)
- VP-SDE, [3833](#)
 - definition, [3833](#)
 - exercises, [3847](#)
- VQ-GAN, [412](#)
- VQ-VAE, [256](#), [409](#)
 - audio codec connection, [942](#)
 - EEG compression, [3833](#)
 - EEG diffusion, [3832](#)
 - fMRI encoding, [3851](#)
 - video pipeline, [578](#)
- VQ-VAE-2, [413](#)
- VQA, *see* visual question answering
- VRAM requirements, [3716](#)
- VS Code
 - Lean extension, [3084](#)
- VSIDS heuristic, [3018](#)
-
- w infinity symmetry, [3170](#)
- wagon-wheel effect, [527](#)
- wake-sleep cycle
 - architecture optimization, [2014](#)
- wake-sleep phases, [1964](#)
- Walker, Matthew, [1964](#)
- WAN-optimised collectives, [5132](#)
- Wan-Video
 - VAE, [572](#)
- Wang et al. (2026)
 - energy downscaling, [4358](#)
- Wang, K., [3973](#)
- warmup
 - exercise, [5030](#)
 - linear, [5027](#)
- warp
 - definition, [4934](#)
- warp divergence
 - efficiency penalty, [4934](#)
- warp field
 - spatiotemporal, [4240](#)
- warping error, [740](#)
- Wasserstein distance, [79](#), [2853–2854](#)
 - as OT, [370](#)
 - colour histogram alignment, [3588](#)
 - connectivity, [3913](#)
 - connectivity exercise, [3928](#)
 - CTGAN, [2725](#)
 - definition, [80](#), [3588](#), [3819](#), [3864](#)
 - EEG, [3946](#)
 - EEG synthesis, [3819–3820](#)
 - FAD, [953](#)
 - finite diagrams
 - exercise, [3877](#)
 - FVD, [735](#)
 - Gaussian, [81](#), [916](#)
 - image quality, [3588](#)
 - medical images, [3744](#)
 - metric property, [81](#), [3875](#)
 - one dimension, [81](#), [304](#)
 - parallel lines example, [289](#)
 - persistence diagrams, [3865](#), [3950](#)
 - metric, [3875](#)
 - point mass example, [304](#)
 - recap, [290](#)
 - seizure prediction, [3866](#)
 - sliced, [512](#)
 - time series of diagrams, [3866](#)
 - W_1 , [290](#)
 - W_2 , [306](#)

-
- Wasserstein GAN, 289–307
 - algorithm, 299
 - comparison, 306
 - comparison with Sinkhorn, 380
 - convergence, 302
 - discrete data, 306
 - epilepsy, 3861
 - exercise, 383
 - federated, 4488
 - influence, 303
 - loss monitoring, 301
 - MNIST example, 302
 - objective, 294
 - practical considerations, 302
 - training dynamics, 300
 - value function, 294
 - vs. original GAN, 294
 - weight clipping algorithm, 299
 - Wasserstein gradient flow, 247
 - Wasserstein space
 - EEG neuromodulation, 3936
 - water vapour, 4205
 - water-filling, 5267
 - watermark capacity
 - exercise, 954
 - information-theoretic bound, 931
 - watermarking
 - green list, 1819
 - interaction with MIA, 1820
 - LLM, 1819
 - open problem
 - EEG, 3934
 - synthetic EEG, 3934
 - unforgeable, 949
 - verifiable, 948
 - Wav2Vec 2.0
 - lip synthesis, 4738
 - wav2vec 2.0, 2432
 - WavCaps, 807
 - wave equation
 - PINN loss, 3451
 - wave phenomena, 3341–3352
 - wavelet activation, 3315
 - wavelet activation function, 3261
 - wavelet shrinkage, 329
 - wavelet transform, 3798
 - WaveNet, 7, 16, 441–443, 767
 - autoregressive connection, 943
 - speech synthesis, 443
 - WaveStitch, 2788–2803, 4629–4630
 - algorithm, 2796–2803
 - applications, 2800–2803
 - architecture, 2790–2791
 - base model, 2802
 - comparison, 2801
 - conditional generation, 2799–2800
 - cyclic encoding, 2795–2796
 - efficiency, 2796
 - energy, 2801
 - finance, 2800
 - full algorithm, 2797
 - healthcare, 2800
 - implementation, 2801–2803
 - IoT, 2801
 - mathematical analysis, 2798–2799
 - overview, 2788
 - segment scheme, 2791
 - spectral consistency, 2794–2795
 - stitching, 2791
 - variance preservation, 2799
 - WavLM, 2433–2434
 - denoising objective, 2433
 - WavRAG, 2541
 - WCAG, 4690–4697
 - 97% non-compliance, 4692
 - AI analogue, 4907
 - audit exercise, 4697
 - definition, 4691
 - four principles, 4691
 - non-compliance statistics, 4692

- POUR principles, [4691](#)
- weak convergence, [84](#)
 - Wasserstein, [291](#)
- weak formulation, [3238](#)
- weak solutions
 - Leray, [3429](#)
- weak topology, [5217](#)
- weakly supervised segmentation,
 - [3519–3520](#)
- wearable devices
 - form factors, [4826](#)
- wearable EEG, [3882](#)
 - deployment constraints, [3889](#)
 - seizure prediction example, [3883](#)
- weather
 - economic value, [4182](#)
- Weather and Forecasting journal, [4234](#)
- weather compensation
 - load forecasting, [4233](#)
- weather compensation module, [4234](#)
- weather forecasting
 - diffusion models, [3280](#), [3282](#)
 - economic importance, [4182](#)
 - economic value, [4418](#)
 - historical development, [4184](#)
 - history, [4182–4192](#)
 - operational deployment, [4383–4394](#)
 - probabilistic, [4285](#)
- weather prediction, [13](#), [3440–3442](#),
 - [4181–4202](#)
 - AI revolution, [3441](#)
 - as generative modelling, [4190](#)
 - classification, [4198](#)
 - computational cost, [4193](#)
 - three pillars, [4193](#)
- weather regime
 - MOS limitation, [4230](#)
 - predictability, [4229](#)
- weather vs. climate, [3442](#)
- WeatherBench
 - benchmark, [4240](#)
- WeatherBench 2, [4395](#)
- WeatherNext 2, [4334–4336](#)
 - 15-day forecasts, [4336](#)
 - architecture, [4334](#)
 - heatwave example, [4336](#)
 - performance, [4335](#)
 - speed, [4335](#)
 - speedup exercise, [4338](#)
- Weaviate, [2379](#)
- web accessibility
 - compliance rates, [4843](#)
- web search
 - as retrieval, [2337](#)
- WebAIM Million, [4692](#)
 - 97% non-compliance, [4696](#)
- WeBank
 - federated learning, [4451](#)
- Wei, Jason, [3048](#)
- weight clipping, [295](#)
 - definition, [295](#)
 - limitations, [304](#)
- weight coupling, [1905](#)
- weight decay, [46](#)
 - privacy, [1683](#)
- weight map
 - touching objects, [3535](#)
- weight matrix
 - partitioning, [5041–5060](#)
- weight overwriting, [1183](#)
- weight sharing, [1904–1910](#)
 - approximation bound, [1995](#)
 - approximation error, [1918](#)
 - coupling problem, [1905](#)
 - diversity tradeoff, [1908](#)
 - error bound, [1907](#)
 - exercise, [1879](#)
 - fidelity, [1905](#)
 - methods comparison, [1907](#)
 - ranking preservation, [1919](#)

- weight tying, 1546
 - definition, 1546
- Weiss, Pierre-Ernest, 183
- well-posedness
 - reverse SDE, 3384
- WER
 - voice cloning, 4754
- WER (Word Error Rate), 4887–4888
- West Virginia University Medicine, 3899
- Western nations
 - representation, 4138
- WGAN
 - federated, 4488
- WGAN-GP, *see* gradient penalty, 2725
 - algorithm, 300
 - convergence, 2726
 - EEG, 3819–3820
 - federated exercise, 4494
 - gradient penalty derivation, 3829
 - implementation, 306
 - objective, 297
- wheelchair users
 - AI representations, 4844
- Whisper, 2434–2435
 - audio support, 4792
 - for retrieval, 2435
- white matter
 - parcellation, 3889–3896
- white matter bundles
 - parcellation, 3890
- white matter connectivity
 - seizure propagation, 3890
- white matter lesions, 3511
- white matter parcellation
 - global context transformer, 3890–3891
- white-box access, 1583
- white-box attacks
 - advantage bound, 1643
 - adversary definition, 1637
 - challenge problem, 1648
 - comparison, 1645
 - current state, 1690
 - diffusion models, 1636–1649
 - diminishing returns, 1647
 - exercises, 1647
 - Fisher information, 1639
 - gradient alignment, 1639
 - gradient decomposition, 1641
 - gradient norm, 1638
 - influence functions, 1644
 - layer-wise decomposition, 1643
 - Pang et al., 1642
 - practical scenarios, 1638
 - subsumes loss attacks, 1647
 - threat model, 1637
 - timestep selection, 1642
 - training epochs, 1644
 - vs. black-box, 1665
- white-box detection, 930
- Whitehead, Alfred North, 2944
- whitening
 - for retrieval, 2406
- whole tumour, 3552
- whole-proof generation, 2962
- whole-slide image, 3515
- whole-slide imaging, 3480
 - resolution, 3481
- wide-area network
 - distributed training, 5159
- Wiener process, 2770, 2776
 - cylindrical, 3378
 - on Hilbert spaces, 3270
- Wiesel, T. N., 3969
- Wigner, Eugene, 3447
- WikiMIA benchmark, 1733–1738
 - definition, 1733
 - results, 1734
- WikiText, 22
- Wilby, Robert, 4289

- Wilcox, Christian, [3036](#)
- Wilcoxon signed-rank test, [921](#)
- Wilks, Daniel, [4289](#)
- Wilson-Cowan model, [3892](#)
 - network exercise, [3897](#)
 - perturbed connectome, [3895](#)
- wind direction
 - MOS, [4225](#)
- wind energy
 - ensemble forecasting, [4358](#)
- wind forecasting
 - MOS, [4224](#)
- wind power
 - forecasting, [4316](#)
- wind speed
 - MOS, [4225](#)
- window attention, [605–607](#)
 - 3D definition, [606](#)
 - receptive field, [606](#)
 - shifted, [606](#)
- window function, [771](#)
- windowed generation, [2801](#)
- wiring optimisation, [1854](#)
 - minimisation, [1855](#)
 - spectral placement, [1856](#)
- Witkin, Andrew, [3525](#)
- Wittgenstein, Ludwig, [2335–2336](#)
- WLOG, [2936](#)
- Wolpert, David, [2282](#)
- word error rate, [4754](#), [4913](#)
- word sense disambiguation, [2335](#)
- word2vec, [2366](#), [2628](#)
- WordNet
 - accessibility applications, [4877](#)
 - concept graph
 - exercise, [4715](#)
- workflow orchestration, [3913](#)
- workflow policy, [3914](#)
- working memory, [4150](#)
 - in agents, [1281](#)
- working memory hypothesis, [3048](#)
- World Health Organization
 - disability statistics, [4689](#)
- world knowledge
 - in language models, [2805](#)
- world model, [718–726](#), [1292–1294](#)
 - compounding error, [1292](#)
 - definition, [718](#), [1292](#)
 - evaluation, [725](#)
 - formal definition, [718](#)
 - frontier systems, [722](#)
 - generalisation depth, [724](#)
 - implicit, [721](#)
 - theoretical basis, [721](#)
 - limitations, [722](#)
 - planning exercise, [762](#)
 - simulation vs. understanding, [724](#)
 - video diffusion, [749](#)
- world simulator
 - video model as, [4115](#)
- Wright, Sewall, [2022](#)
- WSD schedule, *see* learning rate
 - schedule, WSD
 - properties, [1555](#)
- Wu et al. attack, [1654](#)
- X-ray synthesis
 - federated, [4610](#)
- x_0 -parameterisation
 - audio, [850](#)
- x_0 -prediction, [490](#)
- XAI, [3964](#)
 - classical, [3966](#)
 - classical methods, [3981–3990](#)
 - failure for generative models,
 - [3990–4002](#)
 - failure modes
 - table, [3997](#)
 - taxonomy, [3997](#)
 - gap figure, [3998](#)

- historical critique, 4000
- Xavier initialisation
 - derivation, 1550
- XCON, 3962
- XGBoost
 - for detection, 2844
- XLNet
 - comparison, 2807
- XOR problem, 118, 2052
- XP-2 dataset, 3852
 - description, 3854
- Xu, Lei, 2714
- Yang et al. 2024, 3852
- YaRN, *see* positional encoding, YaRN, 1083–1509
 - exercise, 1110
 - frequency zones, 1554
- Yeom’s bound, 1592
- Yices, 3022–3023
- YouTube
 - recommendation system, 2577
- Yule’s K, 4763
 - definition, 4764
 - exercise, 4768
- Z-loss, 626
- Z3, 2955, 3022–3023
- ZeRO, 5060–5069
 - communication analysis exercise, 5069
 - communication overhead, 5064
 - DeepSpeed, 4932
 - exercise, 5069, 5114
 - memory analysis, 5174
 - memory comparison figure, 5183
 - memory reduction, 5137
 - optimizer state partitioning, 5111
 - prefetching, 5065
 - Stage 1, 5061
 - memory formula, 5062
 - Stage 2, 5062
 - memory formula, 5062
 - Stage 3, 5063
 - communication pattern, 5064
 - memory formula, 5063
 - VAE training, 5167
 - worked example, 5067
- zero ablation, 3975
 - out-of-distribution, 3980
- Zero Redundancy Optimizer, 5060–5069
- ZeRO-3
 - parameter partitioning, 4932
- zero-avoiding, 171
- zero-convolution
 - training stability, 680
- zero-cost proxy, 1930–1932
 - combination, 1932
 - definition, 1930
 - diffusion models, 1940
 - error bound, 1931
 - examples, 1930
- zero-forcing, 171
- zero-forgetting
 - expansion, 1958
- ZeRO-Infinity, 5068
- zero-inflated distribution
 - precipitation, 4241
- zero-shot classification, 2409
 - as retrieval, 2452
- zero-shot conditioning, 2775
- zero-shot cross-modal transfer, 2471–2472
 - exercise, 2486
 - transfer loss, 2472
- zero-shot downscaling, 4358
- zero-shot generation
 - formal definition, 2910
 - tabular, 2910
- zero-shot inference
 - PDEs, 3254

- zero-shot learning
 - voice cloning, [4745](#)
- zero-shot segmentation
 - medical imaging, [3493](#)
 - text-guided, [3767](#)
- zero-shot super-resolution, [3221](#), [3456](#)
- zero-shot tracking, [3658](#)
- zero-shot voice cloning
 - definition, [4752](#)
- zero-sum game, [310](#)
- zeroth-order gradient estimation
 - definition, [4559](#)
- zeroth-order optimisation, [2056](#)
- zeroth-order optimization, [2179](#), [4559](#)
- convergence, [4563](#)
- convergence rate, [4563](#)
- variance, [4567](#)
- ZFC, [2939](#)
- ZFC set theory, [3004](#)
- Zhai et al. attack, [1653](#)
- Zhang et al. (2023), [4365](#)
- Zhengzhou flood
 - NowcastNet, [4367](#)
- Zhu, Ligeng
 - gradient inversion, [4453](#)
- Zisserman, Andrew, [2401](#)
- Zoph, Barret, [1911](#)