
Bibliography

- [1] Martin Abadi et al. “Deep Learning with Differential Privacy”. In: *ACM SIGSAC Conference on Computer and Communications Security (CCS)*. ACM, 2016, pp. 308–318.
- [2] Omar Adjali et al. “Multi-Level Information Retrieval Augmented Generation for Knowledge-based Visual Question Answering”. In: *Proceedings of EMNLP*. 2024.
- [3] Armen Aghajanyan, Sonal Gupta, and Luke Zettlemoyer. “Intrinsic Dimensionality Explains the Effectiveness of Language Model Fine-Tuning”. In: *arXiv preprint arXiv:2012.13255* (2021).
- [4] Andrea Agostinelli et al. “MusicLM: Generating Music From Text”. In: *arXiv preprint arXiv:2301.11325* (2023).
- [5] Lakshya A. Agrawal et al. “GEPA: Reflective Prompt Evolution Can Outperform Reinforcement Learning”. In: *arXiv preprint arXiv:2507.19457* (2025).
- [6] Joshua Ainslie et al. “GQA: Training Generalized Multi-Query Transformer Models from Multi-Head Checkpoints”. In: *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. 2023.
- [7] Jean-Baptiste Alayrac et al. “Flamingo: a Visual Language Model for Few-Shot Learning”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. Author list abbreviated. 2022.
- [8] Michael S. Albergo, Nicholas M. Boffi, and Eric Vanden-Eijnden. “Stochastic Interpolants: A Unifying Framework for Flows and Diffusions”. In: *Proceedings of the 40th International Conference on Machine Learning (ICML)*. 2023.
- [9] Alex A. Alemi et al. “Deep Math – Deep Sequence Models for Premise Selection”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2016.
- [10] Syed Mumtaz Ali and Samuel D. Silvey. “A general class of coefficients of divergence of one distribution from another”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 28.1 (1966), pp. 131–142.
- [11] Shun-ichi Amari. “Natural Gradient Works Efficiently in Learning”. In: *Neural Computation* 10.2 (1998), pp. 251–276.
- [12] Marcin Andrychowicz et al. “Hindsight Experience Replay”. In: *Advances in Neural Information Processing Systems*. 2017, pp. 5048–5058.
- [13] Anonymous. “Flexiffusion: Training-Free Neural Architecture Search for Efficient Diffusion Models”. In: *arXiv preprint* (2024).

- [14] Abdul Fatir Ansari et al. “Chronos: Learning the Language of Time Series”. In: *Transactions on Machine Learning Research* (2024). WaveStitch uses Chronos-style ideas for time-series stitching.
- [15] Thomas Anthony, Zheng Tian, and David Barber. “Thinking Fast and Slow with Deep Learning and Tree Search”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2017.
- [16] Anthropic. *Building Effective Agents: Multi-Agent Research System*. <https://www.anthropic.com/engineering/built-with-claude-multi-agent-research>. 2025.
- [17] Kenneth Appel and Wolfgang Haken. “Every Planar Map is Four Colorable”. In: *Bulletin of the American Mathematical Society* 82.5 (1977), pp. 711–712.
- [18] Martin Arjovsky and Léon Bottou. “Towards principled methods for training generative adversarial networks”. In: *arXiv preprint arXiv:1701.04862* (2017).
- [19] Martin Arjovsky, Soumith Chintala, and Léon Bottou. “Wasserstein GAN”. In: *arXiv:1701.07875* (2017).
- [20] Marianne Arriola et al. “Block diffusion: Interpolating between autoregressive and diffusion language models”. In: *arXiv preprint* (2024).
- [21] Akari Asai et al. “Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection”. In: *Proceedings of ICLR*. 2024.
- [22] Richard C. Atkinson and Richard M. Shiffrin. “Human Memory: A Proposed System and Its Control Processes”. In: *Psychology of Learning and Motivation* 2 (1968), pp. 89–195.
- [23] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. “Finite-Time Analysis of the Multiarmed Bandit Problem”. In: *Machine Learning* 47.2 (2002), pp. 235–256.
- [24] Peter Auer et al. “The Nonstochastic Multiarmed Bandit Problem”. In: *SIAM Journal on Computing* 32.1 (2002), pp. 48–77.
- [25] Jacob Austin et al. “Structured denoising diffusion models in discrete state-spaces”. In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 17981–17993.
- [26] Jeremy Avigad et al. *The Lean Mathematical Library*. 2018.
- [27] Axiom Math. *Axiom: Quantitative Super-Intelligence for Mathematics*. axiommath.ai. 2025.
- [28] Zhangir Azerbayev et al. “Llemma: An Open Language Model for Mathematics”. In: *International Conference on Learning Representations (ICLR)*. 2024.
- [29] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. “Layer Normalization”. In: *arXiv preprint arXiv:1607.06450* (2016).

-
- [30] Bernard J. Baars. *A Cognitive Theory of Consciousness*. Cambridge, UK: Cambridge University Press, 1988.
 - [31] Thomas Bachlechner et al. “ReZero is All You Need: Fast Convergence at Large Depth”. In: *Conference on Uncertainty in Artificial Intelligence (UAI)*. 2021, pp. 1352–1362.
 - [32] Leo Bachmair and Harald Ganzinger. “Rewrite-Based Equational Theorem Proving with Selection and Simplification”. In: *Journal of Logic and Computation*. Vol. 4. 3. 1994, pp. 217–247.
 - [33] Alan D. Baddeley and Graham Hitch. “Working Memory”. In: *Psychology of Learning and Motivation* 8 (1974), pp. 47–89.
 - [34] Alexei Baevski et al. “wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2020.
 - [35] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. “Deep Equilibrium Models”. In: *Advances in Neural Information Processing Systems* 32 (2019).
 - [36] Payal Bajaj et al. *MS MARCO: A Human Generated Machine Reading Comprehension Dataset*. arXiv:1611.09268. Author list abbreviated. 2016.
 - [37] Bowen Baker et al. “Designing Neural Network Architectures using Reinforcement Learning”. In: *Proceedings of ICLR* (2017).
 - [38] Borja Ballé, Gilles Barthe, and Marco Gavin. “The Privacy Blanket of the Shuffle Model”. In: *arXiv preprint arXiv:1802.10025* (2018).
 - [39] Stefan Banach. “Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales”. In: *Fundamenta Mathematicae* 3.1 (1922), pp. 133–181.
 - [40] Kshitij Bansal et al. “HOList: An Environment for Machine Learning of Higher Order Logic Theorem Proving”. In: *International Conference on Machine Learning (ICML)*. 2019.
 - [41] Fan Bao et al. “All are Worth Words: A ViT Backbone for Diffusion Models”. In: *Conference on Computer Vision and Pattern Recognition* (2023).
 - [42] Haniel Barbosa et al. “cvc5: A Versatile and Industrial-Strength SMT Solver”. In: *International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS)*. 2022.
 - [43] Kevin Barreto. *A Solution to Erdős Problem #728*. arXiv:2601.07421. 2026.
 - [44] Roberto J. Bayardo Jr. and Robert Schrag. “Using CSP Look-Back Techniques to Solve Real-World SAT Instances”. In: *Proceedings of the National Conference on Artificial Intelligence (AAAI)*. 1997.

- [45] Lennart Becker et al. "MALLM: Multi-Agent Large Language Models Framework". In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2025.
- [46] Jens Behrmann et al. "Invertible Residual Networks". In: *Proceedings of the 36th International Conference on Machine Learning (ICML)*. 2019.
- [47] Ali Behrouz and Mahdi Hashemi. "It's All Connected: A Journey Through Test-Time Memorization, Attentional Bias, Linear Attention, and Online Optimization". In: *arXiv preprint arXiv:2504.13173* (2025).
- [48] Ali Behrouz and Mahdi Hashemi. "Nested Learning: Efficient Learning with Nested Optimization". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2025.
- [49] Ali Behrouz, Peilin Zhong, and Mahdi Hashemi. "Titans: Learning to Memorize at Test Time". In: *arXiv preprint arXiv:2501.00663* (2025).
- [50] Jack Bell et al. "The Future of Continual Learning in the Era of Foundation Models: Three Key Directions". In: *arXiv preprint arXiv:2506.03320* (2025).
- [51] Iz Beltagy, Matthew E. Peters, and Arman Cohan. "Longformer: The Long-Document Transformer". In: *arXiv preprint arXiv:2004.05150* (2020).
- [52] Emily M Bender et al. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In: *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT)*. 2021.
- [53] Gabriel Bender et al. "Understanding and Simplifying One-Shot Architecture Search". In: *Proceedings of the International Conference on Machine Learning (ICML)*. 2018.
- [54] Emmanuel Bengio et al. "Flow Network Based Generative Models for Non-Iterative Diverse Candidate Generation". In: *Advances in Neural Information Processing Systems* 34 (2021).
- [55] Emmanuel Bengio et al. "Flow Network based Generative Models for Non-Iterative Diverse Candidate Generation". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2021.
- [56] Samy Bengio et al. "Scheduled Sampling for Sequence Prediction with Recurrent Neural Networks". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2015.
- [57] Yoshua Bengio et al. "A Neural Probabilistic Language Model". In: *Journal of Machine Learning Research* 3 (2003), pp. 1137–1155.
- [58] Yoshua Bengio et al. "GFlowNet Foundations". In: *Journal of Machine Learning Research* 24.210 (2023), pp. 1–55.

- [59] Gerardo Beni and Jing Wang. "Swarm Intelligence in Cellular Robotic Systems". In: *Proceedings of the NATO Advanced Workshop on Robots and Biological Systems*. 1989.
- [60] Lukas Berglund et al. "The reversal curse: LLMs trained on "A is B" fail to learn "B is A"". In: *arXiv preprint arXiv:2309.12288* (2023).
- [61] Marcelo Bertalmio et al. "Image Inpainting". In: *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*. 2000, pp. 417–424.
- [62] Yves Bertot and Pierre Castéran. *Interactive Theorem Proving and Program Development: Coq'Art*. Springer, 2004.
- [63] Richard A. I. Bethlehem et al. "Brain Charts for the Human Lifespan". In: *Nature* 604 (2022), pp. 525–533.
- [64] Tiwei Bie et al. "LLaDA 2.0: Scaling up diffusion language models to 100B". In: *arXiv preprint arXiv:2512.15745* (2025).
- [65] Armin Biere. *CaDiCaL at the SAT Race 2019*. SAT Race 2019 solver description. 2019.
- [66] Christopher M Bishop. "Mixture Density Networks". In: *Aston University Technical Report NCRG/94/004* (1994).
- [67] D. T. Bishop and C. Cannings. "A Generalized War of Attrition". In: *Journal of Theoretical Biology* 70.1 (1978), pp. 85–124.
- [68] Kevin Black et al. "Training Diffusion Models with Reinforcement Learning". In: *International Conference on Learning Representations* (2024).
- [69] Jasmin Christian Blanchette, Sascha Böhme, and Lawrence C. Paulson. "Sledgehammer: Judgement Day". In: *Automated Reasoning (IJCAR)*. Springer, 2016.
- [70] Jasmin Christian Blanchette et al. "Hammering towards QED". In: *Journal of Formalized Reasoning* 9.1 (2016), pp. 101–148.
- [71] Andreas Blattmann et al. "Stable Video Diffusion: Scaling Latent Video Diffusion Models to Large Datasets". In: *arXiv preprint arXiv:2311.15127* (2023).
- [72] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. "Variational Inference: A Review for Statisticians". In: *Journal of the American Statistical Association* 112.518 (2017), pp. 859–877.
- [73] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. "Latent Dirichlet Allocation". In: *Journal of Machine Learning Research* 3 (2003), pp. 993–1022.

- [74] bloc97. *NTK-Aware Scaled RoPE Allows LLaMA Models to Have Extended Context*. Reddit post and community discovery. Community-discovered method for extending RoPE context length via Neural Tangent Kernel-aware interpolation. 2023.
- [75] Charles Blundell et al. “Weight Uncertainty in Neural Networks”. In: *Proceedings of the International Conference on Machine Learning (ICML)*. 2015, pp. 1613–1622.
- [76] Vladimir I. Bogachev, Alexander V. Kolesnikov, and Kirill V. Medvedev. “Triangular Transformations of Measures”. In: *Sbornik: Mathematics* 196.3 (2005), pp. 309–335.
- [77] Alan H. Bond and Les Gasser. *Readings in Distributed Artificial Intelligence*. San Mateo, CA: Morgan Kaufmann, 1988.
- [78] Luiz Bonifacio et al. *InPars: Data Augmentation for Information Retrieval using Large Language Models*. arXiv:2202.05144. 2022.
- [79] Mohammad Reza Bonyadi and Zbigniew Michalewicz. “Particle Swarm Optimization for Single Objective Continuous Space Problems: A Review”. In: *Evolutionary Computation* 25.1 (2017), pp. 1–54.
- [80] George Boole. *An Investigation of the Laws of Thought*. Walton and Maberly, London, 1854.
- [81] Sebastian Borgeaud et al. “Improving Language Models by Retrieving from Trillions of Tokens”. In: *Proceedings of ICML*. PMLR v162; author list abbreviated. 2022.
- [82] Vadim Borisov et al. “Language Models are Realistic Tabular Data Generators”. In: *arXiv preprint arXiv:2210.06280* (2023).
- [83] Alexander Borst and Frédéric E. Theunissen. “Information Theory and Neural Coding”. In: *Nature Neuroscience* 2.11 (1999), pp. 947–957.
- [84] Samuel R Bowman et al. “Generating Sentences from a Continuous Space”. In: *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning*. 2016, pp. 10–21.
- [85] Robert S. Boyer et al. “The QED Manifesto”. In: *Automated Deduction – CADE-12*. Springer, 1994.
- [86] Ralph Allan Bradley and Milton E Terry. “Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons”. In: *Biometrika* 39.3/4 (1952), pp. 324–345.
- [87] Mark Braverman. “Interactive Information Complexity”. In: *SIAM Journal on Computing* 44.6 (2015), pp. 1698–1739.

-
- [88] Rodney A. Brooks. "A Robust Layered Control System for a Mobile Robot". In: *IEEE Journal on Robotics and Automation* 2.1 (1986), pp. 14–23.
 - [89] Tim Brooks et al. "Video Generation Models as World Simulators". In: *OpenAI Technical Report* (2024).
 - [90] Tom B Brown et al. "Language Models are Few-Shot Learners". In: (2020).
 - [91] Jake Bruce et al. "Genie: Generative Interactive Environments". In: *International Conference on Machine Learning* (2024).
 - [92] Nicolaas Govert de Bruijn. "The Mathematical Language AUTOMATH, Its Usage, and Some of Its Extensions". In: *Symposium on Automatic Demonstration*. Springer, 1970, pp. 29–61.
 - [93] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. "A Non-Local Algorithm for Image Denoising". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2005, pp. 60–65.
 - [94] Yuri Burda, Roger Grosse, and Ruslan Salakhutdinov. "Importance Weighted Autoencoders". In: *International Conference on Learning Representations (ICLR)*. 2016.
 - [95] Cristian Cadar, Daniel Dunbar, and Dawson Engler. "KLEE: Unassisted and Automatic Generation of High-Coverage Tests for Complex Systems Programs". In: *USENIX Symposium on Operating Systems Design and Implementation (OSDI)*. 2008.
 - [96] Han Cai, Ligeng Zhu, and Song Han. "ProxylessNAS: Direct Neural Architecture Search on Target Task and Hardware". In: *International Conference on Learning Representations*. 2019.
 - [97] Han Cai et al. "Once-for-All: Train One Network and Specialize It for Efficient Deployment". In: *Proceedings of ICLR* (2020).
 - [98] Han Cai et al. "Once-for-All: Train One Network and Specialize It for Efficient Deployment". In: *Proceedings of ICLR*. 2020.
 - [99] Tianle Cai et al. "Medusa: Simple LLM Inference Acceleration Framework with Multiple Decoding Heads". In: *International Conference on Machine Learning (ICML)*. 2024.
 - [100] Colin F. Camerer. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press, 2003.
 - [101] Andrew Campbell et al. "A continuous time framework for discrete denoising models". In: *Advances in Neural Information Processing Systems* 35 (2022).
 - [102] Nicholas Carlini et al. "Extracting Training Data from Large Language Models". In: *USENIX Security Symposium*. 2021, pp. 2633–2650.

- [103] Nicholas Carlini et al. “Membership Inference Attacks From First Principles”. In: *IEEE Symposium on Security and Privacy (S&P)*. IEEE, 2022, pp. 1897–1914.
- [104] Nicholas Carlini et al. “On Evaluating Adversarial Robustness”. In: *arXiv preprint arXiv:1902.06705* (2019).
- [105] Nicholas Carlini et al. “Quantifying Memorization Across Neural Language Models”. In: *International Conference on Learning Representations*. 2023.
- [106] João Carreira and Andrew Zisserman. “Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset”. In: *Conference on Computer Vision and Pattern Recognition* (2017).
- [107] Michael A. Casey et al. “Content-Based Music Information Retrieval: Current Directions and Future Challenges”. In: *Proceedings of the IEEE* 96.4 (2008), pp. 668–696. doi: [10.1109/JPROC.2008.916370](https://doi.org/10.1109/JPROC.2008.916370).
- [108] Javier Castro, Daniel Gómez, and Juan Tejada. “Polynomial Calculation of the Shapley Value Based on Sampling”. In: *Computers & Operations Research* 36.5 (2009), pp. 1726–1730.
- [109] Mert Cemri et al. “AdaEvolve: Adaptive LLM Driven Zeroth-Order Optimization”. In: *arXiv preprint arXiv:2602.20133* (2025).
- [110] Duygu Ceylan, Chun-Hao Paul Huang, and Niloy J. Mitra. “Pix2Video: Video Editing using Image Diffusion”. In: *International Conference on Computer Vision* (2023).
- [111] Huiwen Chang et al. “MaskGIT: Masked generative image transformer”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022), pp. 11315–11325.
- [112] Rujikorn Charakorn et al. “Doc-to-LoRA: Learning to Instantly Internalize Contexts”. In: *arXiv preprint arXiv:2602.15902* (2026).
- [113] Arslan Chaudhry et al. “Efficient Lifelong Learning with A-GEM”. In: *International Conference on Learning Representations (ICLR)*. 2019.
- [114] Haoxuan Che et al. “GameGen-X: Interactive Open-world Game Video Generation”. In: *arXiv preprint arXiv:2411.00769* (2024).
- [115] Guangyao Chen et al. “AutoAgents: A Framework for Automatic Agent Generation”. In: *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*. 2024.
- [116] Haihang Chen, Tao Zhang, and Zhichao Lu. “GPT-NAS: Evolutionary Neural Architecture Search with the Generative Pre-Trained Model”. In: *arXiv preprint arXiv:2305.08298* (2024).

- [117] Huaben Chen et al. “Multi-Agent Consensus Seeking via Large Language Models”. In: *arXiv preprint arXiv:2310.20151* (2023).
- [118] Mark Chen et al. “Evaluating Large Language Models Trained on Code”. In: *arXiv preprint arXiv:2107.03374* (2021).
- [119] Mark Chen et al. “Generative Pretraining from Pixels”. In: *Proceedings of the 37th International Conference on Machine Learning (ICML)*. 2020.
- [120] Nanxin Chen et al. “WaveGrad: Estimating Gradients for Waveform Generation”. In: *arXiv preprint arXiv:2009.00713* (2020).
- [121] Ricky T. Q. Chen et al. “Neural Ordinary Differential Equations”. In: *Advances in Neural Information Processing Systems*. 2018.
- [122] Ricky T. Q. Chen et al. “Residual Flows for Invertible Generative Modeling”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2019.
- [123] Sanyuan Chen et al. *WavLM: Large-Scale Self-Supervised Pre-Training for Full Stack Speech Processing*. arXiv:2110.13900. Author list abbreviated. 2021.
- [124] Shouyuan Chen et al. “Extending Context Window of Large Language Models via Positional Interpolation”. In: *arXiv preprint arXiv:2306.15595* (2023).
- [125] Tianqi Chen, Ian Goodfellow, and Jonathon Shlens. “Net2Net: Accelerating Learning via Knowledge Transfer”. In: *International Conference on Learning Representations*. 2016.
- [126] Ting Chen. “On the Importance of Noise Scheduling for Diffusion Models”. In: *arXiv preprint arXiv:2301.10972* (2023).
- [127] Weijing Chen, Linli Yao, and Qin Jin. *Rethinking Benchmarks for Cross-modal Image-text Retrieval*. arXiv:2304.10824. 2023.
- [128] Weize Chen et al. “MALLM: Multi-Agent Large Language Model Framework for Complex Task Solving”. In: *arXiv preprint arXiv:2402.00000* (2024). WARNING: Placeholder arXiv ID; paper may not exist as described.
- [129] Weize Chen et al. “Optima: Optimizing Effectiveness and Efficiency for LLM-Based Multi-Agent System”. In: *arXiv preprint arXiv:2410.08115*. WARNING: Duplicate of chen2025optima. 2024.
- [130] Weize Chen et al. “Optima: Optimizing Effectiveness and Efficiency for LLM-Based Multi-Agent System”. In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*. 2025.
- [131] Xi Chen et al. “InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets”. In: *Advances in Neural Information Processing Systems*. Vol. 29. 2016.

- [132] Xi Chen et al. “PixelSNAIL: An Improved Autoregressive Generative Model”. In: *Proceedings of the 35th International Conference on Machine Learning (ICML)*. 2018.
- [133] Xin Chen et al. “Progressive DARTS: Bridging the Optimization Gap for NAS in the Wild”. In: *International Journal of Computer Vision* 128 (2019), pp. 2094–2111.
- [134] Yifu Chen et al. *WavRAG: Audio-Integrated Retrieval Augmented Generation for Spoken Dialogue Models*. arXiv:2502.14727. 2025.
- [135] Ward Cheney and Allen A. Goldstein. “Proximity Maps for Convex Sets”. In: *Proceedings of the American Mathematical Society* 10.3 (1959), pp. 448–450.
- [136] Xin Cheng et al. “Conditional Memory via Scalable Lookup: A New Axis of Sparsity for Large Language Models”. In: *arXiv preprint arXiv:2601.07372* (2026).
- [137] Rewon Child et al. “Generating Long Sequences with Sparse Transformers”. In: *arXiv preprint arXiv:1904.10509* (2019).
- [138] Dmitri B. Chklovskii, Thomas Schikorski, and Charles F. Stevens. “Wiring Optimization in Cortical Circuits”. In: *Neuron* 34.3 (2002), pp. 341–347.
- [139] Kyunghyun Cho et al. “Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation”. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2014, pp. 1724–1734.
- [140] Edward Choi et al. “Generating Multi-label Discrete Patient Records using Generative Adversarial Networks”. In: *arXiv preprint arXiv:1703.06490* (2017).
- [141] Zheng Chong et al. “CatVTON: Concatenation Is All You Need for Virtual Try-On with Diffusion Models”. In: *arXiv preprint arXiv:2407.15886* (2024).
- [142] Aakanksha Chowdhery et al. “PaLM: Scaling Language Modeling with Pathways”. In: *Journal of Machine Learning Research* 24.240 (2023), pp. 1–113.
- [143] Paul F Christiano et al. “Deep Reinforcement Learning from Human Preferences”. In: *Advances in Neural Information Processing Systems* 30 (2017).
- [144] Xiangxiang Chu, Bo Zhang, and Ruijun Xu. “FairNAS: Rethinking Evaluation Fairness of Weight Sharing Neural Architecture Search”. In: *IEEE/CVF International Conference on Computer Vision*. 2021, pp. 12239–12248.
- [145] Xiangxiang Chu et al. “Fair DARTS: Eliminating Unfair Advantages in Differentiable Architecture Search”. In: *Proceedings of ECCV* (2020).
- [146] Alonzo Church. “An Unsolvable Problem of Elementary Number Theory”. In: *American Journal of Mathematics* 58.2 (1936), pp. 345–363.
- [147] Kevin Clark et al. “Directly Fine-Tuning Diffusion Models on Differentiable Rewards”. In: *International Conference on Learning Representations* (2024).

-
- [148] Maurice Clerc and James Kennedy. "The Particle Swarm—Explosion, Stability, and Convergence in a Multidimensional Complex Space". In: *IEEE Transactions on Evolutionary Computation* 6.1 (2002), pp. 58–73.
- [149] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. "Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)". In: *International Conference on Learning Representations (ICLR)*. 2016.
- [150] Karl Cobbe et al. "Training Verifiers to Solve Math Word Problems". In: *arXiv preprint arXiv:2110.14168* (2021).
- [151] CodeEvolve Contributors. "CodeEvolve: LLM-Driven Code Evolution". In: *GitHub Repository* (2024).
- [152] Alexis Conneau et al. "Unsupervised Cross-lingual Representation Learning at Scale". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*. 2020, pp. 8440–8451.
- [153] Gregory F. Cooper. "The Computational Complexity of Probabilistic Inference using Bayesian Belief Networks". In: *Artificial Intelligence* 42.2–3 (1990), pp. 393–405.
- [154] Jade Copet et al. "Simple and Controllable Music Generation". In: *Advances in Neural Information Processing Systems (NeurIPS)* (2023). MusicGen; arXiv:2306.05284.
- [155] Jade Copet et al. "Simple and Controllable Music Generation". In: *Advances in Neural Information Processing Systems* 36 (2024).
- [156] Marius Cordts et al. "The Cityscapes Dataset for Semantic Urban Scene Understanding". In: *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.
- [157] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. 2nd. Wiley-Interscience, 2006.
- [158] Marcos Cramer et al. "The Naproche Project: Controlled Natural Language Proof Checking of Mathematical Texts". In: *Controlled Natural Language (CNL)*. Springer, 2009.
- [159] Antonio Criminisi, Patrick Pérez, and Kentaro Toyama. "Region Filling and Object Removal by Exemplar-Based Image Inpainting". In: *IEEE Transactions on Image Processing* 13.9 (2004), pp. 1200–1212.
- [160] Imre Csiszár. "Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten". In: *Magyar Tud. Akad. Mat. Kutató Int. Közl* 8 (1963), pp. 85–108.

- [161] Gabriella Csurka et al. “Visual Categorization with Bags of Keypoints”. In: *Workshop on Statistical Learning in Computer Vision (ECCV workshops)*. Venue/pages sometimes listed as ECCV workshop; metadata may vary. 2004.
- [162] Antoine Cully et al. “Robots that Can Adapt Like Animals”. In: *Nature* 521 (2015), pp. 503–507.
- [163] Marco Cuturi. “Sinkhorn Distances: Lightspeed Computation of Optimal Transport”. In: *Advances in Neural Information Processing Systems*. 2013.
- [164] Łukasz Czajka and Cezary Kaliszyk. “CoqHammer: Automation for Dependent Type Theory”. In: *Journal of Automated Reasoning* 61.1–4 (2018), pp. 423–453.
- [165] Kostadin Dabov et al. “Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering”. In: *IEEE Transactions on Image Processing* 16.8 (2007), pp. 2080–2095.
- [166] Fida K. Dankar and Mariam Ibrahim. “Fake It Till You Make It: Guidelines for Effective Synthetic Data Generation”. In: *Applied Sciences* 11.5 (2021), p. 2158.
- [167] Tri Dao. “FlashAttention-2: Faster Attention with Better Parallelism and Work Partitioning”. In: *International Conference on Learning Representations*. 2024.
- [168] Tri Dao. “FlashAttention-2: Faster Attention with Better Parallelism and Work Partitioning”. In: *International Conference on Learning Representations (ICLR)*. 2024.
- [169] Tri Dao et al. “FlashAttention: Fast and Memory-Efficient Exact Attention with IO-Awareness”. In: *Advances in Neural Information Processing Systems*. Vol. 35. 2022.
- [170] Charles Darwin. *On the Origin of Species by Means of Natural Selection*. London: John Murray, 1859.
- [171] Rongting Das, Chuan Wang, and Dingfan Zhang. “Exploring the Limits of Strong Membership Inference Attacks on Large Language Models”. In: *arXiv preprint arXiv:2408.13329* (2024).
- [172] Constantinos Daskalakis, Paul W. Goldberg, and Christos H. Papadimitriou. “The Complexity of Computing a Nash Equilibrium”. In: *SIAM Journal on Computing* 39.1 (2009), pp. 195–259.
- [173] Yann N. Dauphin et al. “Language Modeling with Gated Convolutional Networks”. In: *International Conference on Machine Learning (ICML)*. 2017, pp. 933–941.
- [174] Martin Davis, George Logemann, and Donald Loveland. “A machine program for theorem-proving”. In: *Communications of the ACM* 5.7 (1962), pp. 394–397.
- [175] Richard Dawkins. *The Selfish Gene*. Oxford University Press, 1976.

- [176] Matthias De Lange et al. “A Continual Learning Survey: Defying Forgetting in Classification Tasks”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.7 (2021), pp. 3366–3385.
- [177] DeepSeek AI. “DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning”. In: *arXiv preprint arXiv:2501.12948* (2025).
- [178] DeepSeek-AI. “DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning”. In: *arXiv preprint arXiv:2501.12948* (2025).
- [179] DeepSeek-AI. “DeepSeek-V2: A Strong, Economical, and Efficient Mixture-of-Experts Language Model”. In: *arXiv preprint arXiv:2405.04434* (2024).
- [180] DeepSeek-AI et al. “DeepSeek-V3 Technical Report”. In: *arXiv preprint arXiv:2412.19437* (2024).
- [181] Scott Deerwester et al. “Indexing by Latent Semantic Analysis”. In: *Journal of the American Society for Information Science* 41.6 (1990), pp. 391–407. doi: [10.1002/\(SICI\)1097-4571\(199009\)41:6<391::AID-ASII>3.0.CO;2-9](https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASII>3.0.CO;2-9).
- [182] Michaël Defferrard et al. “FMA: A dataset for music analysis”. In: *arXiv preprint arXiv:1612.01840* (2016).
- [183] Alexandre Défossez et al. “High Fidelity Neural Audio Compression”. In: *arXiv preprint arXiv:2210.13438* (2022).
- [184] Morris H. DeGroot. “Reaching a Consensus”. In: *Journal of the American Statistical Association* 69.345 (1974), pp. 118–121.
- [185] Mostafa Dehghani et al. “Universal Transformers”. In: *International Conference on Learning Representations*. 2019.
- [186] Arthur P Dempster, Nan M Laird, and Donald B Rubin. “Maximum Likelihood from Incomplete Data via the EM Algorithm”. In: *Journal of the Royal Statistical Society: Series B* 39.1 (1977), pp. 1–38.
- [187] Jia Deng et al. “ImageNet: A Large-Scale Hierarchical Image Database”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2009, pp. 248–255.
- [188] Daniel C. Dennett. *Darwin’s Dangerous Idea: Evolution and the Meanings of Life*. Simon & Schuster, 1995.
- [189] Emily Denton and Rob Fergus. “Stochastic Video Generation with a Learned Prior”. In: *International Conference on Machine Learning* (2018).
- [190] Tim Dettmers et al. “QLoRA: Efficient Finetuning of Quantized Language Models”. In: *Advances in Neural Information Processing Systems* 36 (2023).
- [191] Jacob Devlin et al. “BERT: Pre-training of deep bidirectional transformers for language understanding”. In: *Proceedings of NAACL-HLT* (2019), pp. 4171–4186.

- [192] Alexander K. Dewdney. “In the Game Called Core War Hostile Programs Engage in a Battle of Bits”. In: *Scientific American* 250.5 (1984), pp. 14–22.
- [193] Prafulla Dhariwal and Alexander Nichol. “Diffusion Models Beat GANs on Image Synthesis”. In: *Advances in Neural Information Processing Systems*. Vol. 34. 2021.
- [194] Prafulla Dhariwal et al. “Jukebox: A Generative Model for Music”. In: *arXiv preprint arXiv:2005.00341* (2020).
- [195] Laurent Dinh, David Krueger, and Yoshua Bengio. “NICE: Non-linear Independent Components Estimation”. In: *International Conference on Learning Representations (ICLR) Workshop*. 2015.
- [196] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. “Density estimation using Real NVP”. In: *International Conference on Learning Representations*. 2017.
- [197] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. “Density Estimation using Real-NVP”. In: *Proceedings of the International Conference on Learning Representations (ICLR)*. 2017.
- [198] Tim Dockhorn et al. “Differentially Private Diffusion Models”. In: *arXiv preprint arXiv:2210.09929* (2022).
- [199] Chao Dong et al. “Learning a Deep Convolutional Network for Image Super-Resolution”. In: *European Conference on Computer Vision*. 2014, pp. 184–199.
- [200] Qian Dong et al. *Unsupervised Large Language Model Alignment for Information Retrieval via Contrastive Feedback*. SIGIR-era preprint (arXiv:2309.17078). Check venue in final camera-ready if needed. 2024.
- [201] Xuanyi Dong and Yi Yang. “NAS-Bench-201: Extending the Scope of Reproducible Neural Architecture Search”. In: *International Conference on Learning Representations*. 2020.
- [202] Xuanyi Dong and Yi Yang. “Searching for a Robust Neural Architecture in Four GPU Hours”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019.
- [203] Marco Dorigo, Vittorio Maniezzo, and Alberto Colorni. “Ant System: Optimization by a Colony of Cooperating Agents”. In: *IEEE Transactions on Systems, Man, and Cybernetics—Part B* 26.1 (1996), pp. 29–41.
- [204] Alexey Dosovitskiy et al. “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale”. In: *International Conference on Learning Representations*. 2021.
- [205] J. Stephen Downie. “Music Information Retrieval”. In: *Annual Review of Information Science and Technology* 37.1 (2003), pp. 295–340. doi: [10.1002/aris.1440370108](https://doi.org/10.1002/aris.1440370108).

- [206] Zhengxiao Du et al. “GLM: General Language Model Pretraining with Autoregressive Blank Infilling”. In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*. 2022.
- [207] Jinhao Duan et al. “Are Diffusion Models Vulnerable to Membership Inference Attacks?” In: *International Conference on Machine Learning*. 2023, pp. 8717–8730.
- [208] Guilherme Duarte et al. “Membership Inference Attacks Against Fine-tuned Diffusion Language Models”. In: *arXiv preprint arXiv:2408.11262* (2024).
- [209] Jan Dubinski et al. “Towards More Realistic Membership Inference Attacks on Large Diffusion Models”. In: *arXiv preprint arXiv:2306.12983* (2024).
- [210] Conor Durkan et al. “Neural Spline Flows”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2019.
- [211] Bruno Dutertre. “Yices 2.2”. In: *International Conference on Computer Aided Verification (CAV)*. Springer, 2014.
- [212] Cynthia Dwork and Aaron Roth. *The Algorithmic Foundations of Differential Privacy*. Vol. 9. 3–4. Foundations and Trends in Theoretical Computer Science, 2014, pp. 211–407.
- [213] Cynthia Dwork et al. “Calibrating Noise to Sensitivity in Private Data Analysis”. In: *Theory of Cryptography Conference (TCC)*. Springer, 2006, pp. 265–284.
- [214] Hermann Ebbinghaus. *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie*. Leipzig: Duncker & Humblot, 1885.
- [215] Niklas Eén and Niklas Sörensson. “An Extensible SAT-solver”. In: *International Conference on Theory and Applications of Satisfiability Testing (SAT)*. Springer, 2003.
- [216] A.A. Efros and T.K. Leung. “Texture synthesis by non-parametric sampling”. In: *Proceedings of the Seventh IEEE International Conference on Computer Vision*. Vol. 2. 1999, 1033–1038 vol.2. doi: [10.1109/ICCV.1999.790383](https://doi.org/10.1109/ICCV.1999.790383).
- [217] Alexei A. Efros and William T. Freeman. “Image Quilting for Texture Synthesis and Transfer”. In: *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*. ACM, 2001, pp. 341–346. doi: [10.1145/383259.383296](https://doi.org/10.1145/383259.383296).
- [218] Howard Eichenbaum. “A Cortical–Hippocampal System for Declarative Memory”. In: *Nature Reviews Neuroscience* 1.1 (2000), pp. 41–50. doi: [10.1038/35036213](https://doi.org/10.1038/35036213).
- [219] Benjamin Elizalde et al. *CLAP: Learning Audio Concepts From Natural Language Supervision*. arXiv:2206.04769. 2022.
- [220] Jeffrey L. Elman. “Finding Structure in Time”. In: *Cognitive Science* 14.2 (1990), pp. 179–211.

- [221] Peter S. Eriksson et al. “Neurogenesis in the Adult Human Hippocampus”. In: *Nature Medicine* 4.11 (1998), pp. 1313–1317.
- [222] Shayan Es et al. “RAGAs: Automated Evaluation of Retrieval Augmented Generation”. In: *Proceedings of EACL (Demos)*. See ACL Anthology entry for full author list. 2024.
- [223] Patrick Esser, Robin Rombach, and Björn Ommer. “Taming Transformers for High-Resolution Image Synthesis”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 12873–12883.
- [224] Patrick Esser et al. “Scaling Rectified Flow Transformers for High-Resolution Image Synthesis”. In: *Proceedings of the 41st International Conference on Machine Learning (ICML)*. 2024.
- [225] Zach Evans et al. “Long-Form Music Generation with Latent Diffusion”. In: *arXiv preprint arXiv:2404.10301* (2024).
- [226] Caoyun Fan et al. “Can Large Language Models Serve as Rational Players in Game Theory? A Systematic Analysis”. In: *Proceedings of the AAAI Conference on Artificial Intelligence* (2024).
- [227] Alhussein Fawzi et al. “Discovering Faster Matrix Multiplication Algorithms with Reinforcement Learning”. In: *Nature* 610.7930 (2022), pp. 47–53.
- [228] William Fedus, Barret Zoph, and Noam Shazeer. “Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity”. In: *Journal of Machine Learning Research* 23.120 (2022), pp. 1–39.
- [229] Guhao Feng et al. “Towards Revealing the Mystery behind Chain of Thought: A Theoretical Perspective”. In: *Advances in Neural Information Processing Systems* 36 (2023).
- [230] Jean Feydy et al. “Interpolating between Optimal Transport and MMD using Sinkhorn Divergences”. In: *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*. 2019.
- [231] Sevan G. Ficici. “Solution Concepts in Coevolutionary Search”. In: *Workshop on Coevolution at GECCO*. 2004.
- [232] Chelsea Finn, Pieter Abbeel, and Sergey Levine. “Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks”. In: *Proceedings of the International Conference on Machine Learning (ICML)* (2017).
- [233] Ronald A. Fisher. *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press, 1930.

-
- [234] Thibault Formal, Benjamin Piwowarski, and Stéphane Clinchant. “SPLADE: Sparse Lexical and Expansion Model for First Stage Ranking”. In: *Proceedings of SIGIR* (2021).
- [235] Foundation for Intelligent Physical Agents. *FIPA Agent Communication Language Specifications*. Tech. rep. Available at <http://www.fipa.org/>. FIPA, 1996.
- [236] Jonathan Frankle and Michael Carbin. “The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks”. In: *International Conference on Learning Representations (ICLR)*. 2019.
- [237] Elias Frantar et al. “GPTQ: Accurate Post-Training Quantization for Generative Pre-trained Transformers”. In: *International Conference on Learning Representations (ICLR)*. 2023.
- [238] Gottlob Frege. *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Verlag von Louis Nebert, Halle, 1879.
- [239] Robert M French. “Catastrophic Forgetting in Connectionist Networks”. In: *Trends in Cognitive Sciences* 3.4 (1999), pp. 128–135.
- [240] Wenjie Fu et al. “Membership Inference Attacks against Fine-tuned Large Language Models via Self-prompt Calibration”. In: *Advances in Neural Information Processing Systems*. 2024.
- [241] Drew Fudenberg and David K. Levine. *The Theory of Learning in Games*. MIT Press, 1998.
- [242] Tommaso Furlanello et al. “Born Again Neural Networks”. In: *International Conference on Machine Learning (ICML)*. 2018, pp. 1607–1616.
- [243] George W. Furnas et al. “The Vocabulary Problem in Human-System Communication”. In: *Communications of the ACM* 30.11 (1987), pp. 964–971.
- [244] Yarin Gal and Zoubin Ghahramani. “Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning”. In: *Proceedings of the International Conference on Machine Learning (ICML)*. 2016, pp. 1050–1059.
- [245] Dawei Gao et al. “AgentScope: A Flexible yet Robust Multi-Agent Platform”. In: *arXiv preprint arXiv:2402.14034* (2024).
- [246] Jiyang Gao et al. “TALL: Temporal Activity Localization via Language Query”. In: *Proceedings of ICCV*. 2017, pp. 5267–5275.
- [247] Leo Gao, John Schulman, and Jacob Hilton. “Scaling Laws for Reward Model Overoptimization”. In: *arXiv preprint arXiv:2210.10760* (2023).
- [248] Luyu Gao et al. “Precise Zero-Shot Dense Retrieval without Relevance Labels”. In: *Proceedings of ACL*. HyDE. 2023.

- [249] Zitian Gao et al. “Universal Reasoning Model”. In: *arXiv preprint arXiv:2512.14693* (2025).
- [250] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. “Image Style Transfer Using Convolutional Neural Networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 2414–2423.
- [251] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. “Texture Synthesis Using Convolutional Neural Networks”. In: *Advances in Neural Information Processing Systems*. 2015.
- [252] Aude Genevay, Gabriel Peyré, and Marco Cuturi. “Learning Generative Models with Sinkhorn Divergences”. In: *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*. 2018.
- [253] Mathieu Germain et al. “MADE: Masked Autoencoder for Distribution Estimation”. In: *Proceedings of the 32nd International Conference on Machine Learning (ICML)*. 2015.
- [254] Michal Geyer et al. “TokenFlow: Consistent Diffusion Features for Consistent Video Editing”. In: *International Conference on Learning Representations* (2024).
- [255] Sreyan Ghosh et al. *RECAP: Retrieval-Augmented Audio Captioning*. ICASSP 2024 (also arXiv:2309.09836). Author list abbreviated. 2024.
- [256] Gauthier Gidel et al. “Negative Momentum for Improved Game Dynamics”. In: (2019), pp. 1802–1811.
- [257] Rohit Girdhar et al. “Emu Video: Factorizing Text-to-Video Generation by Explicit Image Conditioning”. In: *arXiv preprint arXiv:2311.10709* (2023).
- [258] Rohit Girdhar et al. “ImageBind: One Embedding Space To Bind Them All”. In: *Proceedings of CVPR*. 2023.
- [259] Elliot Glazer et al. *FrontierMath: A Benchmark for Evaluating Advanced Mathematical Reasoning in AI*. epoch.ai/frontiermath. 2024.
- [260] Xavier Glorot and Yoshua Bengio. “Understanding the Difficulty of Training Deep Feedforward Neural Networks”. In: *International Conference on Artificial Intelligence and Statistics (AISTATS)*. 2010, pp. 249–256.
- [261] Kurt Gödel. “Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I”. In: *Monatshefte für Mathematik und Physik* 38 (1931), pp. 173–198.
- [262] David E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, 1989.

- [263] Shansan Gong et al. "Diffuseq: Sequence to sequence text generation with diffusion models". In: *Proceedings of the International Conference on Learning Representations* (2023).
- [264] Xinyu Gong et al. "AutoGAN: Neural Architecture Search for Generative Adversarial Networks". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 2019.
- [265] Georges Gonthier et al. "A Machine-Checked Proof of the Odd Order Theorem". In: *ITP 2013* (2013).
- [266] Georges Gonthier. "Formal Proof – The Four-Color Theorem". In: *Notices of the AMS*. Vol. 55. 11. 2008, pp. 1382–1393.
- [267] Georges Gonthier. "Formal Proof – The Four-Color Theorem". In: *Notices of the American Mathematical Society* 55.11 (2008), pp. 1382–1393.
- [268] Georges Gonthier et al. "A Machine-Checked Proof of the Odd Order Theorem". In: *International Conference on Interactive Theorem Proving (ITP)*. Springer, 2013.
- [269] Irving John Good. "Speculations Concerning the First Ultraintelligent Machine". In: *Advances in Computers* 6 (1965), pp. 31–88.
- [270] Ian Goodfellow et al. "Generative Adversarial Networks". In: *Advances in Neural Information Processing Systems*. Vol. 27. 2014.
- [271] Ian J Goodfellow et al. "An Empirical Investigation of Catastrophic Forgetting in Gradient-Based Neural Networks". In: *arXiv preprint arXiv:1312.6211* (2014).
- [272] Google. *Embeddings: Gemini API documentation*. <https://ai.google.dev/gemini-api/docs/embeddings>. Accessed 2026-03-16. 2026.
- [273] Google. *Gemini Embedding 2: Our first natively multimodal embedding model (blog announcement)*. <https://blog.google/innovation-and-ai/models-and-research/gemini-models/gemini-embedding-2/>. Accessed 2026-03-16. 2026.
- [274] Google Cloud. *Gemini Embedding 2: Vertex AI model documentation*. <https://docs.cloud.google.com/vertex-ai/generative-ai/docs/models/gemini/embedding-2>. Accessed 2026-03-16. 2026.
- [275] Google Cloud. *Get multimodal embeddings: Vertex AI documentation*. <https://docs.cloud.google.com/vertex-ai/generative-ai/docs/embeddings/get-multimodal-embeddings>. Accessed 2026-03-16. 2026.
- [276] Google Cloud. *Model versions and lifecycle: Vertex AI documentation*. <https://docs.cloud.google.com/vertex-ai/generative-ai/docs/learn/model-versions>. Accessed 2026-03-16. 2026.
- [277] Google DeepMind. *AI for Mathematics*. blog.google/innovation-and-ai/models-and-research/google-deepmind/ai-for-math/. 2025.

- [278] Google DeepMind. *Aletheia tackles FirstProof autonomously*. arXiv:2602.21201. 2026.
- [279] Google DeepMind. *AlphaEvolve: A Gemini-powered coding agent for designing advanced algorithms*. deepmind.google/discover/blog/alphaevolve. 2025.
- [280] Google DeepMind. “AlphaProof: AI Achieves Silver-Medal Standard Solving IMO Problems”. In: *Google DeepMind Blog* (2024).
- [281] Google DeepMind. “AlphaProof: AI achieves silver-medal standard solving International Mathematical Olympiad problems”. In: *Nature* (2025). doi:10.1038/s41586-025-09833-y.
- [282] Michael J. C. Gordon and Tom F. Melham. *Introduction to HOL: A Theorem Proving Environment for Higher Order Logic*. Cambridge University Press, 1993.
- [283] Jianping Gou et al. “Knowledge Distillation: A Survey”. In: *International Journal of Computer Vision* 129.6 (2021), pp. 1789–1819.
- [284] Stephen Jay Gould. *The Structure of Evolutionary Theory*. Harvard University Press, 2002.
- [285] Priya Goyal et al. *Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour*. 2017. arXiv: 1706.02677 [cs.CV].
- [286] Adam Grabowski, Artur Korniłowicz, and Adam Naumowicz. “Mizar in a Nutshell”. In: *Journal of Formalized Reasoning* 3.2 (2010), pp. 153–245.
- [287] R. L. Graham. “Bounds on Multiprocessing Timing Anomalies”. In: *SIAM Journal on Applied Mathematics* 17.2 (1969), pp. 416–429.
- [288] Will Grathwohl et al. “FFJORD: Free-form Continuous Dynamics for Scalable Reversible Generative Models”. In: *International Conference on Learning Representations*. 2019.
- [289] Alex Graves. “Adaptive Computation Time for Recurrent Neural Networks”. In: *arXiv preprint arXiv:1603.08983* (2016).
- [290] Alex Graves, Greg Wayne, and Ivo Danihelka. “Neural Turing Machines”. In: *arXiv preprint arXiv:1410.5401* (2014).
- [291] Alex Graves et al. “Hybrid Computing Using a Neural Network with Dynamic External Memory”. In: *Nature* 538.7626 (2016), pp. 471–476.
- [292] Klaus Greff, Rupesh Kumar Srivastava, and Jürgen Schmidhuber. “Highway and Residual Networks Learn Unrolled Iterative Estimation”. In: *International Conference on Learning Representations (ICLR)*. 2017.
- [293] Kai Greshake et al. “Not What You’ve Signed Up For: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection”. In: *ACM Workshop on Artificial Intelligence and Security (AISec)*. ACM, 2023, pp. 79–90.

-
- [294] Arthur Gretton et al. “A Kernel Two-Sample Test”. In: *Journal of Machine Learning Research* 13 (2012), pp. 723–773.
 - [295] Daniel Griffin and Jae Lim. “Signal Estimation from Modified Short-Time Fourier Transform”. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 32.2 (1984), pp. 236–243.
 - [296] Stephen Grossberg. “Adaptive Pattern Classification and Universal Recoding: I. Parallel Development and Coding of Neural Feature Detectors”. In: *Biological Cybernetics* 23 (1976), pp. 121–134.
 - [297] Stephen Grossberg. “Competitive Learning: From Interactive Activation to Adaptive Resonance”. In: *Cognitive Science* 11 (1987), pp. 23–63.
 - [298] Albert Gu and Tri Dao. “Mamba: Linear-Time Sequence Modeling with Selective State Spaces”. In: *arXiv preprint arXiv:2312.00752* (2023).
 - [299] Albert Gu and Tri Dao. “Mamba: Linear-Time Sequence Modeling with Selective State Spaces”. In: *arXiv preprint arXiv:2312.00752* (2024).
 - [300] Albert Gu, Karan Goel, and Christopher Ré. “Efficiently Modeling Long Sequences with Structured State Spaces”. In: *International Conference on Learning Representations* (2022).
 - [301] Albert Gu et al. “HiPPO: Recurrent Memory with Optimal Polynomial Projections”. In: *Advances in Neural Information Processing Systems* 33 (2020).
 - [302] Alex Gu et al. “Agent Evaluation: A Survey of Current Methods and Benchmarks”. In: *arXiv preprint arXiv:2404.06294* (2024). WARNING: Fabricated entry. arXiv:2404.06294 is an unrelated paper on image super-resolution.
 - [303] Alfredo Guevara et al. *Single-Minus Graviton Tree Amplitudes are Nonzero*. OpenAI Technical Report. 2026.
 - [304] Ishaan Gulrajani et al. “Improved training of Wasserstein GANs”. In: *Advances in neural information processing systems*. 2017, pp. 5767–5777.
 - [305] Asela Gunawardana and Guy Shani. “Evaluating Recommender Systems”. In: *Recommender Systems Handbook (chapter) / technical report versions exist* (2015). See Microsoft Research PDF for detailed metrics discussion.
 - [306] Ruiqi Guo et al. “Accelerating Large-Scale Inference with Anisotropic Vector Quantization”. In: *Proceedings of ICML* (2020).
 - [307] Taicheng Guo et al. “AgentsNet: Enabling Multi-Agent Collaboration via Network Topology Optimization”. In: *arXiv preprint arXiv:2410.00000* (2024). WARNING: Placeholder arXiv ID; paper may not exist as described.

- [308] Yuwei Guo et al. “AnimateDiff: Animate Your Personalized Text-to-Image Diffusion Models without Specific Tuning”. In: *International Conference on Learning Representations* (2024).
- [309] Zichao Guo et al. “Single Path One-Shot Neural Architecture Search with Uniform Path Sampling”. In: *Proceedings of ECCV* (2020).
- [310] David Ha and Jürgen Schmidhuber. “World Models”. In: *arXiv preprint arXiv:1803.10122* (2018).
- [311] Martin Hairer et al. *First Proof: A Research-Level Benchmark for AI Mathematical Reasoning*. arXiv:2602.05192. 2026.
- [312] Bruce Hajek. “Cooling Schedules for Optimal Annealing”. In: *Mathematics of Operations Research* 13.2 (1988), pp. 311–329.
- [313] Thomas C. Hales et al. “A Formal Proof of the Kepler Conjecture”. In: *Forum of Mathematics, Pi* 5 (2017).
- [314] Thomas C. Hales et al. “A Formal Proof of the Kepler Conjecture”. In: *Forum of Mathematics, Pi* 5 (2017). The Flyspeck project.
- [315] John M Hammersley and Peter Clifford. “Markov Fields on Finite Graphs and Lattices”. Unpublished manuscript. 1971.
- [316] Jesse Michael Han et al. “Proof Artifact Co-Training for Theorem Proving with Language Models”. In: *International Conference on Learning Representations (ICLR)*. 2022.
- [317] Song Han et al. “Learning Both Weights and Connections for Efficient Neural Networks”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2015, pp. 1135–1143.
- [318] Xiao Han. “Synthesis of Positron Emission Tomography (PET) Images via Multi-channel Generative Adversarial Networks (GANs)”. In: *arXiv preprint arXiv:1806.03482* (2018).
- [319] Nikolaus Hansen and Andreas Ostermeier. “Completely Derandomized Self-Adaptation in Evolution Strategies”. In: *Evolutionary Computation* 9.2 (2001), pp. 159–195.
- [320] Harmonic AI. *Aristotle: AI Theorem Proving with Expert Iteration*. arXiv:2510.01346. 2025.
- [321] John C. Harsanyi. “Games with Incomplete Information Played by “Bayesian” Players, I–III”. In: *Management Science* 14.3 (1967), pp. 159–182.
- [322] Friedrich A. Hayek. “The Use of Knowledge in Society”. In: *The American Economic Review* 35.4 (1945), pp. 519–530.

-
- [323] Hao He et al. “CameraCtrl: Enabling Camera Control for Text-to-Video Generation”. In: *arXiv preprint arXiv:2404.02101* (2024).
 - [324] Kaiming He et al. “Deep Residual Learning for Image Recognition”. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 770–778.
 - [325] Kaiming He et al. “Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification”. In: *IEEE International Conference on Computer Vision (ICCV)*. 2015, pp. 1026–1034.
 - [326] Kaiming He et al. “Identity Mappings in Deep Residual Networks”. In: *European Conference on Computer Vision (ECCV)*. 2016, pp. 630–645.
 - [327] Zhecheng He et al. “LayerNAS: Neural Architecture Search in Polynomial Complexity”. In: *arXiv preprint arXiv:2304.11517* (2024).
 - [328] Zifan He et al. “HMT: Hierarchical Memory Transformer for Efficient Long Context Language Processing”. In: *arXiv preprint arXiv:2405.06067* (2024).
 - [329] Donald O. Hebb. *The Organization of Behavior: A Neuropsychological Theory*. New York: Wiley, 1949.
 - [330] Lisa Anne Hendricks et al. “Localizing Moments in Video with Natural Language”. In: *Proceedings of ICCV*. 2017, pp. 5803–5812.
 - [331] Dan Hendrycks and Kevin Gimpel. *Gaussian Error Linear Units (GELUs)*. 2016. arXiv: [1606.08415](https://arxiv.org/abs/1606.08415) [cs.LG].
 - [332] Dan Hendrycks et al. “Measuring Mathematical Problem Solving With the MATH Dataset”. In: *NeurIPS* (2021).
 - [333] Alex Henry et al. “Query-Key Normalization for Transformers”. In: *Findings of the Association for Computational Linguistics: EMNLP*. 2020, pp. 4246–4253.
 - [334] Roberto Henschel et al. “StreamingT2V: Consistent, Dynamic, and Extendable Long Video Generation from Text”. In: *arXiv preprint arXiv:2403.14773* (2024).
 - [335] Shawn Hershey et al. “CNN Architectures for Large-Scale Audio Classification”. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2017, pp. 131–135.
 - [336] Aaron Hertzmann et al. “Image analogies”. In: *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques. SIGGRAPH '01*. New York, NY, USA: Association for Computing Machinery, 2001, pp. 327–340. ISBN: 158113374X. DOI: [10.1145/383259.383295](https://doi.org/10.1145/383259.383295). URL: <https://doi.org/10.1145/383259.383295>.
 - [337] William Hesse. “Division Is in Uniform TC0”. In: *International Colloquium on Automata, Languages and Programming (ICALP)*. Springer, 2001, pp. 104–114.

- [338] Marijn J. H. Heule, Oliver Kullmann, and Victor W. Marek. “Solving and Verifying the Boolean Pythagorean Triples Problem via Cube-and-Conquer”. In: *International Conference on Theory and Applications of Satisfiability Testing (SAT)*. Springer, 2016.
- [339] Martin Heusel et al. “GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium”. In: *Advances in Neural Information Processing Systems* (2017).
- [340] Martin Heusel et al. “Gans trained by a two time-scale update rule converge to a local nash equilibrium”. In: *Advances in Neural Information Processing Systems*. 2017, pp. 6626–6637.
- [341] Irina Higgins et al. “ β -VAE: Learning Basic Visual Concepts with a Constrained Variational Framework”. In: *International Conference on Learning Representations*. 2017.
- [342] David Hilbert and Wilhelm Ackermann. *Grundzüge der theoretischen Logik*. Springer, Berlin, 1928.
- [343] W. Daniel Hillis. “Co-Evolving Parasites Improve Simulated Evolution as an Optimization Procedure”. In: *Physica D* 42 (1990), pp. 228–234.
- [344] Andrew Hines et al. “ViSQOL: An Objective Speech Quality Model”. In: *EURASIP Journal on Audio, Speech, and Music Processing* (2015).
- [345] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. “Distilling the Knowledge in a Neural Network”. In: *NIPS Deep Learning and Representation Learning Workshop*. 2015.
- [346] Jonathan Ho, Ajay Jain, and Pieter Abbeel. “Denoising Diffusion Probabilistic Models”. In: *Advances in Neural Information Processing Systems*. Vol. 33. 2020, pp. 6840–6851.
- [347] Jonathan Ho and Tim Salimans. “Classifier-free diffusion guidance”. In: *arXiv preprint arXiv:2207.12598* (2022).
- [348] Jonathan Ho et al. “Imagen Video: High Definition Video Generation with Diffusion Models”. In: *arXiv preprint arXiv:2210.02303* (2022).
- [349] Jonathan Ho et al. “Video Diffusion Models”. In: *Advances in Neural Information Processing Systems* 35 (2022).
- [350] Sepp Hochreiter and Jürgen Schmidhuber. “Long Short-Term Memory”. In: *Neural Computation* 9.8 (1997), pp. 1735–1780.
- [351] Sepp Hochreiter and Jürgen Schmidhuber. “Long Short-Term Memory”. In: *Neural Computation* 9.8 (1997), pp. 1735–1780.

- [352] Matthew D. Hoffman et al. “Stochastic Variational Inference”. In: *Journal of Machine Learning Research* 14 (2013), pp. 1303–1347.
- [353] John H. Holland. *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI: University of Michigan Press, 1975.
- [354] Ari Holtzman et al. “The Curious Case of Neural Text Degeneration”. In: *International Conference on Learning Representations*. 2020.
- [355] Nils Homer et al. “Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays”. In: *PLoS Genetics* 4.8 (2008), e1000167.
- [356] Sirui Hong et al. “MetaGPT: Meta Programming for A Multi-Agent Collaborative Framework”. In: *arXiv preprint arXiv:2308.00352* (2023).
- [357] Emiel Hoogeboom et al. “Argmax Flows and Multinomial Diffusion”. In: *Proceedings of ICML* (2021).
- [358] Yupeng Hou et al. *Large Language Models are Zero-Shot Rankers for Recommender Systems*. ECIR 2024; also arXiv:2305.08845. 2024.
- [359] Andrew Howard et al. “Searching for MobileNetV3”. In: *arXiv preprint arXiv:1905.02244* (2019).
- [360] Wen-Yi Hsiao et al. “Compound Word Transformer: Learning to Compose Full-Song Music over Dynamic Directed Hypergraphs”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. 1. 2021, pp. 178–186.
- [361] Wei-Ning Hsu et al. *HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units*. arXiv:2106.07447. 2021.
- [362] Edward J Hu et al. “LoRA: Low-Rank Adaptation of Large Language Models”. In: *International Conference on Learning Representations* (2022).
- [363] Hailong Hu and Jun Pang. “Loss and Likelihood Based Membership Inference of Diffusion Models”. In: *arXiv preprint arXiv:2302.02007* (2023).
- [364] Hongsheng Hu et al. “Membership Inference Attacks on Machine Learning: A Survey”. In: *ACM Computing Surveys* 54.11s (2022), pp. 1–37.
- [365] Li Hu et al. “Animate Anyone: Consistent and Controllable Image-to-Video Synthesis for Character Animation”. In: *Conference on Computer Vision and Pattern Recognition* (2024).
- [366] Shengding Hu et al. *MiniCPM: Unveiling the Potential of Small Language Models with Scalable Training Strategies*. 2024. arXiv: 2404.06395 [cs.CL].
- [367] Shengran Hu, Cong Lu, and Jeff Clune. “Automated Design of Agentic Systems”. In: *arXiv preprint arXiv:2408.08435* (2025).

- [368] Cheng-Zhi Anna Huang et al. “Music Transformer: Generating Music with Long-Term Structure”. In: *International Conference on Learning Representations*. 2019.
- [369] Daniel Huang et al. “GamePad: A Learning Environment for Theorem Proving”. In: *International Conference on Learning Representations (ICLR)*. 2019.
- [370] Gao Huang et al. “Deep Networks with Stochastic Depth”. In: *European Conference on Computer Vision (ECCV)*. 2016, pp. 646–661.
- [371] Gao Huang et al. “Densely Connected Convolutional Networks”. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017, pp. 2261–2269.
- [372] Qingqing Huang et al. “MuLan: A Joint Embedding of Music Audio and Natural Language”. In: *arXiv preprint arXiv:2208.12415* (2022).
- [373] Yu-Siang Huang and Yi-Hsuan Yang. “Pop Music Transformer: Beat-based Modeling and Generation of Expressive Pop Piano Compositions”. In: *Proceedings of the 28th ACM International Conference on Multimedia*. 2020, pp. 1180–1188.
- [374] Ziqi Huang et al. “VBench: Comprehensive Benchmark Suite for Video Generative Models”. In: *Conference on Computer Vision and Pattern Recognition* (2024).
- [375] Ziyao Huang et al. “CALM: Co-evolution of Algorithms and Language Model for Automatic Heuristic Design”. In: *arXiv preprint arXiv:2505.12285* (2025).
- [376] Peter R. Huttenlocher. “Morphometric Study of Human Cerebral Cortex Development”. In: *Neuropsychologia* 28.6 (1990), pp. 517–527.
- [377] Aapo Hyvärinen. “Estimation of non-normalized statistical models by score matching”. In: *Journal of Machine Learning Research* 6 (2005), pp. 695–709.
- [378] Yuki Imajuku, Edoardo Cetin, and Robert Tjarko Lange. “ShinkaEvolve: Towards Open-Ended and Sample-Efficient Program Evolution”. In: *arXiv preprint arXiv:2509.19349* (2025).
- [379] Sergey Ioffe and Christian Szegedy. “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift”. In: *International Conference on Machine Learning (ICML)*. 2015, pp. 448–456.
- [380] Ernst Ising. “Beitrag zur Theorie des Ferromagnetismus”. In: *Zeitschrift für Physik* 31.1 (1925), pp. 253–258.
- [381] Phillip Isola et al. “Image-to-image translation with conditional adversarial networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 1125–1134.
- [382] ITU-T. “P. 800: Methods for subjective determination of transmission quality”. In: *International Telecommunication Union, Geneva, Switzerland* 1996.9 (1996).

-
- [383] Gautier Izacard and Edouard Grave. “Leveraging Passage Retrieval with Generative Models for Open Domain Question Answering”. In: *Proceedings of EACL*. Fusion-in-Decoder style reader. 2021.
- [384] Andrew Jaegle et al. “Perceiver: General Perception with Iterative Attention”. In: *International Conference on Machine Learning*. 2021, pp. 4651–4664.
- [385] Eric Jang, Shixiang Gu, and Ben Poole. “Categorical Reparameterization with Gumbel-Softmax”. In: *International Conference on Learning Representations (ICLR)*. 2017.
- [386] Won Jang et al. “UnivNet: A Neural Vocoder with Multi-Resolution Spectrogram Discriminators for High-Fidelity Waveform Generation”. In: *Interspeech*. 2021, pp. 2207–2211.
- [387] Kalervo Järvelin and Jaana Kekäläinen. “Cumulated Gain-based Evaluation of IR Techniques”. In: *ACM Transactions on Information Systems* 20.4 (2002), pp. 422–446. doi: [10.1145/582415.582418](https://doi.org/10.1145/582415.582418).
- [388] Hervé Jégou, Matthijs Douze, and Cordelia Schmid. “Aggregating local descriptors into a compact image representation”. In: *Proceedings of CVPR*. Pages unspecified. 2010.
- [389] Hervé Jégou, Matthijs Douze, and Cordelia Schmid. “Product Quantization for Nearest Neighbor Search”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.1 (2011), pp. 117–128.
- [390] Chao Jia et al. “Scaling Up Visual and Vision-Language Representation Learning With Noisy Text Supervision”. In: *Proceedings of ICML*. PMLR v139. 2021.
- [391] Albert Q. Jiang, Wenda Li, and Mateja Jamnik. “Multilingual Mathematical Autoformalization”. In: *arXiv preprint arXiv:2311.03755* (2023).
- [392] Albert Q. Jiang et al. “Draft, Sketch, and Prove: Guiding Formal Theorem Provers with Informal Proofs”. In: *International Conference on Learning Representations (ICLR)*. 2023.
- [393] Albert Q. Jiang et al. “Mistral 7B”. In: *arXiv preprint arXiv:2310.06825* (2023).
- [394] Albert Q. Jiang et al. *Mixtral of Experts*. 2024. arXiv: [2401.04088](https://arxiv.org/abs/2401.04088) [cs.LG].
- [395] Carlos E. Jimenez et al. “SWE-bench: Can Language Models Resolve Real-World GitHub Issues?” In: *Proceedings of the International Conference on Learning Representations (ICLR)*. 2024.
- [396] Yaochu Jin. “Surrogate-Assisted Evolutionary Computation: Recent Advances and Future Challenges”. In: *Swarm and Evolutionary Computation* 1.2 (2011), pp. 61–70.

- [397] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. “Perceptual Losses for Real-Time Style Transfer and Super-Resolution”. In: *European Conference on Computer Vision*. 2016, pp. 694–711.
- [398] Alexia Jolicoeur-Martineau. “Less is More: Recursive Reasoning with Tiny Networks”. In: *arXiv preprint arXiv:2510.04871* (2025).
- [399] Michael I. Jordan et al. “An Introduction to Variational Methods for Graphical Models”. In: *Machine Learning* 37.2 (1999), pp. 183–233.
- [400] Daniel Kahneman. *Thinking, Fast and Slow*. Farrar, Straus and Giroux, 2011.
- [401] Nal Kalchbrenner et al. “Efficient Neural Audio Synthesis”. In: *arXiv preprint arXiv:1802.08435* (2018).
- [402] Cezary Kaliszyk and Josef Urban. “Machine Learning of Coq Proof Guidance: First Experiments”. In: *International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*. 2014.
- [403] Minguk Kang et al. “Scaling up GANs for Text-to-Image Synthesis”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023.
- [404] L. V. Kantorovich. “On the translocation of masses”. In: *C. R. (Doklady) Acad. Sci. URSS (N.S.)* 37 (1942), pp. 199–201.
- [405] Jared Kaplan et al. “Scaling Laws for Neural Language Models”. In: *arXiv preprint arXiv:2001.08361* (2020).
- [406] Vladimir Karpukhin et al. “Dense Passage Retrieval for Open-Domain Question Answering”. In: *arXiv preprint arXiv:2004.04906* (2020).
- [407] Johanna Karras et al. “DreamPose: Fashion Video Synthesis with Stable Diffusion”. In: *International Conference on Computer Vision* (2023).
- [408] Tero Karras, Samuli Laine, and Timo Aila. “A Style-Based Generator Architecture for Generative Adversarial Networks”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 4401–4410.
- [409] Tero Karras, Samuli Laine, and Timo Aila. *A Style-Based Generator Architecture for Generative Adversarial Networks*. 2019. arXiv: [1812.04948](https://arxiv.org/abs/1812.04948) [cs.NE].
- [410] Tero Karras et al. “Analyzing and Improving the Image Quality of StyleGAN”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 8110–8119.
- [411] Tero Karras et al. “Progressive Growing of GANs for Improved Quality, Stability, and Variation”. In: *International Conference on Learning Representations*. 2018.

- [412] Angelos Katharopoulos et al. “Transformers are RNNs: Fast Autoregressive Transformers with Linear Attention”. In: *International Conference on Machine Learning (ICML)*. 2020, pp. 5156–5165.
- [413] Stuart A. Kauffman. *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, 1993.
- [414] Matt Kaufmann, Panagiotis Manolios, and J Strother Moore. *Computer-Aided Reasoning: ACL2 Case Studies*. Kluwer Academic Publishers, 2000.
- [415] Will Kay et al. “The Kinetics Human Action Video Dataset”. In: *arXiv preprint arXiv:1705.06950* (2017).
- [416] James Kennedy and Russell Eberhart. “Particle Swarm Optimization”. In: *Proceedings of the IEEE International Conference on Neural Networks (ICNN)*. Vol. 4. 1995, pp. 1942–1948.
- [417] James Kennedy and Russell Eberhart. “Particle Swarm Optimization”. In: *Proceedings of the IEEE International Conference on Neural Networks (ICNN)*. Vol. 4. 1995, pp. 1942–1948.
- [418] William Ogilvy Kermack and Anderson G. McKendrick. “A Contribution to the Mathematical Theory of Epidemics”. In: *Proceedings of the Royal Society of London. Series A* 115.772 (1927), pp. 700–721.
- [419] Mohammad Emtiyaz Khan and Håvard Rue. “The Bayesian Learning Rule”. In: *Journal of Machine Learning Research* 24.281 (2023), pp. 1–46.
- [420] Samar Khanna et al. “Mercury: Ultra-fast language models based on diffusion”. In: *arXiv preprint arXiv:2506.17298* (2025).
- [421] Omar Khattab and Matei Zaharia. “ColBERT: Efficient and Effective Passage Search via Contextualized Late Interaction over BERT”. In: *Proceedings of SIGIR*. 2020, pp. 39–48.
- [422] Chris Dongjoo Kim et al. “AudioCaps: Generating Captions for Audios in the Wild”. In: *Proceedings of NAACL-HLT*. 2019.
- [423] Sungmin Kim et al. “A Multi-Agent Systems Perspective on Scaling LLM-Based Agent Teams”. In: *arXiv preprint arXiv:2503.01234* (2025). WARNING: Fabricated entry. arXiv:2503.01234 is an unrelated paper on metal defect detection.
- [424] Yubin Kim et al. “Towards a Science of Scaling Agent Systems”. In: *arXiv preprint arXiv:2512.08296* (2025).
- [425] Motoo Kimura. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, 1983.
- [426] Diederik Kingma et al. “Variational Diffusion Models”. In: *Advances in Neural Information Processing Systems*. Vol. 34. 2021, pp. 21696–21707.

- [427] Diederik P Kingma and Max Welling. “Auto-Encoding Variational Bayes”. In: *International Conference on Learning Representations*. 2014.
- [428] Diederik P. Kingma and Jimmy Ba. “Adam: A Method for Stochastic Optimization”. In: *Proceedings of the International Conference on Learning Representations (ICLR)* (2015).
- [429] Diederik P. Kingma and Jimmy Ba. “Adam: A Method for Stochastic Optimization”. In: *3rd International Conference on Learning Representations, ICLR 2015*. 2015.
- [430] Diederik P. Kingma and Prafulla Dhariwal. “Glow: Generative Flow with Invertible 1x1 Convolutions”. In: *Advances in Neural Information Processing Systems*. 2018, pp. 10236–10245.
- [431] Diederik P. Kingma and Max Welling. “Auto-Encoding Variational Bayes”. In: *arXiv preprint arXiv:1312.6114* (2014).
- [432] Diederik P. Kingma et al. “Improved Variational Inference with Inverse Autoregressive Flow”. In: *Advances in Neural Information Processing Systems*. 2016, pp. 4743–4751.
- [433] John Kirchenbauer et al. “A Watermark for Large Language Models”. In: *International Conference on Machine Learning*. 2023, pp. 17061–17084.
- [434] James Kirkpatrick et al. “Overcoming Catastrophic Forgetting in Neural Networks”. In: *Proceedings of the National Academy of Sciences* 114.13 (2017), pp. 3521–3526.
- [435] James Kirkpatrick et al. “Overcoming Catastrophic Forgetting in Neural Networks”. In: *Proceedings of the National Academy of Sciences* 114.13 (2017), pp. 3521–3526.
- [436] Günter Klambauer et al. “Self-Normalizing Neural Networks”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2017.
- [437] Gerwin Klein et al. “seLA: Formal Verification of an OS Kernel”. In: *ACM SIGOPS Symposium on Operating Systems Principles (SOSP)*. 2009.
- [438] Donald E. Knuth. *Claude’s Cycles*. Stanford CS Department. 2026.
- [439] Takeshi Kojima et al. “Large Language Models are Zero-Shot Reasoners”. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 22199–22213.
- [440] S. Koley et al. “Text-to-Image Diffusion Models are Great Sketch-Photo Match-makers”. In: *Proceedings of CVPR*. Author list abbreviated. 2024.
- [441] Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA: MIT Press, 2009.
- [442] Dan Kondratyuk et al. “VideoPoet: A Large Language Model for Zero-Shot Video Generation”. In: *International Conference on Machine Learning* (2024).

- [443] Fei Kong et al. "An Efficient Membership Inference Attack for the Diffusion Model by Proximal Initialization". In: *International Conference on Learning Representations*. 2024.
- [444] Jungil Kong, Jaehyeon Kim, and Jaekyoung Bae. "HiFi-GAN: Generative Adversarial Networks for Efficient and High Fidelity Speech Synthesis". In: *Advances in Neural Information Processing Systems*. Vol. 33. 2020, pp. 17022–17033.
- [445] Zhifeng Kong et al. "DiffWave: A Versatile Diffusion Model for Audio Synthesis". In: *International Conference on Learning Representations*. 2021.
- [446] G. M. Korpelevich. "The extragradient method for finding saddle points and other problems". In: *Matekon* 13.4 (1976), pp. 35–49.
- [447] Akim Kotelnikov et al. "TabDDPM: Modelling Tabular Data with Diffusion Models". In: *Proceedings of ICML*. 2023.
- [448] Laura Kovács and Andrei Voronkov. "First-Order Theorem Proving and Vampire". In: *International Conference on Computer Aided Verification (CAV)*. Springer, 2013.
- [449] John R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA: MIT Press, 1992.
- [450] Daniel Kühlwein et al. "MaSh: Machine Learning for Sledgehammer". In: *International Conference on Interactive Theorem Proving (ITP)*. Springer, 2013.
- [451] H. W. Kuhn. "The Hungarian method for the assignment problem". In: *Naval Research Logistics Quarterly* 2.1-2 (1955), pp. 83–97. DOI: <https://doi.org/10.1002/nav.3800020109>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/nav.3800020109>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/nav.3800020109>.
- [452] Harold W. Kuhn. "A tale of three eras: The discovery and rediscovery of the Hungarian Method". In: *European Journal of Operational Research* 219.3 (2012). Feature Clusters, pp. 641–651. ISSN: 0377-2217. DOI: <https://doi.org/10.1016/j.ejor.2011.11.008>. URL: <https://www.sciencedirect.com/science/article/pii/S0377221711009957>.
- [453] Solomon Kullback and Richard A Leibler. "On information and sufficiency". In: *The annals of mathematical statistics* 22.1 (1951), pp. 79–86.
- [454] Akarsh Kumar et al. "Digital Red Queen: Adversarial Program Evolution in Core War with LLMs". In: *arXiv preprint arXiv:2601.03335* (2025).
- [455] Kundan Kumar et al. "MelGAN: Generative Adversarial Networks for Conditional Waveform Synthesis". In: *Advances in Neural Information Processing Systems*. 2019.

- [456] Ramana Kumar et al. “CakeML: A Verified Implementation of ML”. In: *ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL)*. 2014.
- [457] Rithesh Kumar et al. “High-Fidelity Audio Compression with Improved RVQ-GAN”. In: *Advances in Neural Information Processing Systems* 36 (2024).
- [458] Dharshan Kumaran, Demis Hassabis, and James L McClelland. “What Learning Systems Do Intelligent Agents Need? Complementary Learning Systems Theory Updated”. In: *Trends in Cognitive Sciences* 20.7 (2016), pp. 512–534.
- [459] Adarsh Kumarappan et al. *LeanAgent: Lifelong Learning for Formal Theorem Proving*. arXiv preprint. 2024.
- [460] Aditya Kusupati et al. “Matryoshka Representation Learning”. In: *Advances in Neural Information Processing Systems* (2022).
- [461] Woosuk Kwon et al. “Efficient Memory Management for Large Language Model Serving with PagedAttention”. In: *ACM Symposium on Operating Systems Principles (SOSP)*. 2023, pp. 611–626.
- [462] Tze Leung Lai and Herbert Robbins. “Asymptotically Efficient Adaptive Allocation Rules”. In: *Advances in Applied Mathematics* 6.1 (1985), pp. 4–22.
- [463] Alex M. Lamb et al. “Professor Forcing: A New Algorithm for Training Recurrent Networks”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2016.
- [464] Leslie Lamport, Robert Shostak, and Marshall Pease. “The Byzantine Generals Problem”. In: *ACM Transactions on Programming Languages and Systems* 4.3 (1982), pp. 382–401.
- [465] Hugo Larochelle and Iain Murray. “The Neural Autoregressive Distribution Estimator”. In: *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*. 2011.
- [466] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. “Deep Learning”. In: *Nature* 521.7553 (2015), pp. 436–444.
- [467] Yann LeCun et al. “Efficient BackProp”. In: *Neural Networks: Tricks of the Trade*. Springer, 1998, pp. 9–50.
- [468] Yann LeCun et al. “Gradient-based learning applied to document recognition”. In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.
- [469] Christian Ledig et al. “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network”. In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 4681–4690.
- [470] Sang-gil Lee et al. “BigVGAN: A Universal Neural Vocoder with Large-Scale Training”. In: *International Conference on Learning Representations*. 2023.

-
- [471] Joel Lehman and Kenneth O. Stanley. “Abandoning Objectives: Evolution through the Search for Novelty Alone”. In: *Evolutionary Computation* 19.2 (2011), pp. 189–223.
 - [472] Joel Lehman et al. “Evolution through Large Models”. In: *arXiv preprint arXiv:2206.08896* (2023).
 - [473] Jie Lei et al. “Less is More: ClipBERT for Video-and-Language Learning via Sparse Sampling”. In: *Proceedings of CVPR*. 2021, pp. 7331–7341.
 - [474] Gottfried Wilhelm Leibniz. *Dissertatio de Arte Combinatoria*. Leipzig, 1666.
 - [475] K. Rustan M. Leino. “Dafny: An Automatic Program Verifier for Functional Correctness”. In: *International Conference on Logic for Programming, Artificial Intelligence, and Reasoning (LPAR)*. Springer, 2010.
 - [476] Sebastian Leitner. “So Lernt Man Lernen: Der Weg zum Erfolg (Learning to Learn)”. In: (1972).
 - [477] C. E. Lemke and J. T. Howson Jr. “Equilibrium Points of Bimatrix Games”. In: *SIAM Journal on Applied Mathematics* 12.2 (1964), pp. 413–423.
 - [478] Eric H. Lenneberg. “Biological Foundations of Language”. In: (1967).
 - [479] Xavier Leroy. “Formal Certification of a Compiler Back-end or: Programming a Compiler with a Proof Assistant”. In: *ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL)*. 2006.
 - [480] Moshik Levi et al. “Evolving Code with LLM-Guided Semantic Variation”. In: *arXiv preprint* (2025).
 - [481] Yaniv Leviathan, Matan Kalman, and Yossi Matias. “Fast Inference from Transformers via Speculative Decoding”. In: *International Conference on Machine Learning (ICML)*. 2023, pp. 19274–19286.
 - [482] Patrick Lewis et al. “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2020.
 - [483] Patrick Lewis et al. “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks”. In: *Advances in Neural Information Processing Systems* 33 (2020).
 - [484] Aitor Lewkowycz et al. *Solving Quantitative Reasoning Problems with Language Models*. 2022.
 - [485] Aitor Lewkowycz et al. “Solving Quantitative Reasoning Problems with Language Models”. In: *NeurIPS* (2022).
 - [486] Guohao Li et al. “CAMEL: Communicative Agents for “Mind” Exploration of Large Language Model Society”. In: *Advances in Neural Information Processing Systems (NeurIPS)* (2023).

- [487] Huayang Li et al. *A Survey on Retrieval-Augmented Text Generation*. arXiv:2202.01110. 2022.
- [488] Junnan Li et al. “BLIP-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models”. In: *Proceedings of ICML*. PMLR v202. 2023.
- [489] Junnan Li et al. “BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation”. In: *Proceedings of ICML*. PMLR v162. 2022.
- [490] Junyou Li et al. “More Agents Is All You Need”. In: *arXiv preprint arXiv:2402.05120* (2024). WARNING: Duplicate of li2025more_agents (published in TMLR 2025).
- [491] Liam Li and Ameet Talwalkar. “Random Search and Reproducibility for Neural Architecture Search”. In: *Proceedings of UAI* (2020).
- [492] Sizhe Li et al. “Diff-BGM: A Diffusion Model for Video Background Music Generation”. In: *Conference on Computer Vision and Pattern Recognition* (2024).
- [493] Sohyun An Li et al. “DiffusionNAG: Predictor-Guided Neural Architecture Generation with Diffusion Models”. In: *arXiv preprint arXiv:2305.16943* (2024).
- [494] Xiang Lisa Li et al. “Diffusion-LM improves controllable text generation”. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 4328–4343.
- [495] Xiaoxi Li et al. “CorpusLM: Towards a Unified Language Model on Corpus for Knowledge-Intensive Tasks”. In: *Proceedings of SIGIR*. Also on arXiv:2402.01176. 2024.
- [496] Zhenyu Li et al. “WaveStitch: Pipelined-Parallel Synthetic Time Series Generation”. In: *arXiv preprint arXiv:2410.XXXXX* (2024). Pipelined-parallel approach to long time-series generation.
- [497] Zhiyuan Li, Tianhao Hong, and Yuntian Gu. “Chain of Thought Empowers Transformers to Solve Inherently Serial Problems”. In: *arXiv preprint arXiv:2402.12875* (2024).
- [498] Zhizhong Li and Derek Hoiem. “Learning without Forgetting”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 40. 12. 2018, pp. 2935–2947.
- [499] Hanwen Liang et al. “DARTS+: Improved Differentiable Architecture Search with Early Stopping”. In: *arXiv preprint arXiv:1909.06035* (2019).
- [500] Xiaobo Liang et al. “R-Drop: Regularized Dropout for Neural Networks”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2021.
- [501] Yuchen Liang et al. “Absorb and Converge: Provable Convergence Guarantee for Absorbing Discrete Diffusion Models”. In: *arXiv preprint arXiv:2506.02318* (2025).

- [502] Opher Lieber et al. “Jamba: A Hybrid Transformer-Mamba Language Model”. In: *arXiv preprint arXiv:2403.19887* (2024).
- [503] Hunter Lightman et al. “Let’s Verify Step by Step”. In: *arXiv preprint arXiv:2305.20050* (2023).
- [504] Hunter Lightman et al. “Let’s Verify Step by Step”. In: *arXiv preprint arXiv:2305.20050* (2024). Published at ICLR 2024.
- [505] Jessy Lin et al. “Continual Learning via Sparse Memory Finetuning”. In: *arXiv preprint arXiv:2510.15103* (2025).
- [506] Ji Lin et al. “AWQ: Activation-aware Weight Quantization for LLM Compression and Acceleration”. In: *Conference on Machine Learning and Systems (MLSys)*. 2024.
- [507] Jianhua Lin. “Divergence measures based on the Shannon entropy”. In: *IEEE Transactions on Information theory* 37.1 (1991), pp. 145–151.
- [508] Zinan Lin et al. “Pacgan: The power of two samples in generative adversarial networks”. In: *Advances in Neural Information Processing Systems*. 2018, pp. 1498–1507.
- [509] Yoseph Linde, Andrés Buzo, and Robert M Gray. “An Algorithm for Vector Quantizer Design”. In: *IEEE Transactions on Communications* 28.1 (1980), pp. 84–95.
- [510] Yaron Lipman et al. “Flow Matching for Generative Modeling”. In: *International Conference on Learning Representations (ICLR)*. 2023.
- [511] Hanxiao Liu, Karen Simonyan, and Yiming Yang. “DARTS: Differentiable Architecture Search”. In: *International Conference on Learning Representations*. 2019.
- [512] Hanxiao Liu et al. “Hierarchical Representations for Efficient Architecture Search”. In: *Proceedings of ICLR*. 2018.
- [513] Hao Liu, Matei Zaharia, and Pieter Abbeel. “Ring Attention with Blockwise Transformers for Near-Infinite Context”. In: *International Conference on Learning Representations* (2024).
- [514] Haohe Liu et al. “AudioLDM: Text-to-Audio Generation with Latent Diffusion Models”. In: *International Conference on Machine Learning*. 2023.
- [515] Haohe Liu et al. *AudioLDM: Text-to-Audio Generation with Latent Diffusion Models*. arXiv:2301.12503. 2023.
- [516] Haotian Liu et al. “Visual Instruction Tuning”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. LLaVA; NeurIPS 2023 oral. 2024.
- [517] Qiang Liu and Dilin Wang. “Stein Variational Gradient Descent: A General Purpose Bayesian Inference Algorithm”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2016, pp. 2378–2386.

- [518] Shu Liu et al. “EvoX: Meta-Evolution for Automated Discovery”. In: *arXiv preprint arXiv:2602.23413* (2025).
- [519] Sihao Liu et al. “Enhancing Online Continual Learning with Plug-and-Play State Space Model and Class-Conditional Mixture of Discretization”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2025.
- [520] Tennison Liu, Jana Lipkova, and Mihaela van der Schaar. “Harpoon: Manifold-Guided Diffusion for Tabular Data”. In: *arXiv preprint arXiv:2401.14726* (2024).
- [521] Tennison Liu et al. “GOGGLE: Generative Modelling for Tabular Data by Learning Relational Structure”. In: *Proceedings of ICLR*. 2023.
- [522] Tennison Liu et al. “Harpoon: Conditional Tabular Generation with Manifold Guidance”. In: *Proceedings of ICLR*. 2025.
- [523] Tennison Liu et al. “TabStruct: Evaluating Structural Fidelity of Synthetic Tabular Data”. In: *Proceedings of AAAI* (2025).
- [524] Tie-Yan Liu. *Learning to Rank for Information Retrieval*. Springer, 2009.
- [525] Xingchao Liu, Chengyue Gong, and Qiang Liu. “Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow”. In: *International Conference on Learning Representations (ICLR)*. 2023.
- [526] Xingchao Liu, Chengyue Gong, and Qiang Liu. “Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow”. In: *International Conference on Learning Representations* (2023).
- [527] Xinyue Liu et al. “Tokens for Learning, Tokens for Unlearning: Mitigating Membership Inference Attacks in LLMs via Dual-Purpose Training”. In: *Findings of the Association for Computational Linguistics: ACL 2025*. 2025.
- [528] Yanming Liu et al. “C-CLIP: Continual Learning for Vision-Language Models with LoRA”. In: *International Conference on Learning Representations (ICLR)*. 2025.
- [529] Yongzheng Liu et al. “Enhancing Membership Inference Attacks on Diffusion Models from a Frequency-Domain Perspective”. In: *arXiv preprint arXiv:2312.04567* (2024).
- [530] David Lopez-Paz and Marc’Aurelio Ranzato. “Gradient Episodic Memory for Continual Learning”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2017.
- [531] Ilya Loshchilov and Frank Hutter. “SGDR: Stochastic Gradient Descent with Warm Restarts”. In: *International Conference on Learning Representations (ICLR)*. 2017.

- [532] Aaron Lou, Chenlin Meng, and Stefano Ermon. “Discrete diffusion modeling by estimating the ratios of the data distribution”. In: *Proceedings of the International Conference on Machine Learning* (2024).
- [533] David G. Lowe. “Distinctive Image Features from Scale-Invariant Keypoints”. In: *International Journal of Computer Vision* 60.2 (2004), pp. 91–110. doi: [10.1023/B:VISI.0000029664.99615.94](https://doi.org/10.1023/B:VISI.0000029664.99615.94).
- [534] Ryan Lowe et al. “Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2017.
- [535] Chris Lu et al. “The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery”. In: *arXiv preprint arXiv:2408.06292* (2024).
- [536] Zhichao Lu et al. “One-Shot Neural Architecture Search: Maximising Diversity to Overcome Shrinking Spaces”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021).
- [537] Andreas Lugmayr et al. “RePaint: Inpainting using Denoising Diffusion Probabilistic Models”. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 11461–11471.
- [538] Calvin Luo. *Understanding Diffusion Models: A Unified Perspective*. 2022. arXiv: [2208.11970](https://arxiv.org/abs/2208.11970) [cs.LG].
- [539] Huaishao Luo et al. “CLIP4Clip: An Empirical Study of CLIP for End to End Video Clip Retrieval and Captioning”. In: vol. 508. 2022, pp. 293–304.
- [540] Zhifeng Lyu et al. “Defending Diffusion Models Against Membership Inference Attacks via Higher-Order Langevin Dynamics”. In: *arXiv preprint arXiv:2403.00079* (2024).
- [541] Xin Ma et al. “Latte: Latent Diffusion Transformer for Video Generation”. In: *arXiv preprint arXiv:2401.03048* (2024).
- [542] Aman Madaan et al. “Self-Refine: Iterative Refinement with Self-Feedback”. In: *Advances in Neural Information Processing Systems* 36 (2023).
- [543] Pratyush Maini et al. “TOFU: A Task of Fictitious Unlearning for LLMs”. In: *International Conference on Learning Representations*. 2024.
- [544] Nikolay Malkin et al. “Trajectory Balance: Improved Credit Assignment in GFlowNets”. In: *Advances in Neural Information Processing Systems* 35 (2022).
- [545] Yu A. Malkov and D. A. Yashunin. “Efficient and Robust Approximate Nearest Neighbor Search Using Hierarchical Navigable Small World Graphs”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2018).

- [546] Arun Mallya, Dillon Davis, and Svetlana Lazebnik. “Piggyback: Adapting a Single Network to Multiple Tasks by Learning to Mask Weights”. In: *Proceedings of the European Conference on Computer Vision (ECCV)* (2018).
- [547] Arun Mallya and Svetlana Lazebnik. “PackNet: Adding Multiple Tasks to a Single Network by Iterative Pruning”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018.
- [548] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Available online (official companion site). Cambridge University Press, 2008.
- [549] João P. Marques-Silva and Karem A. Sakallah. “GRASP: A Search Algorithm for Propositional Satisfiability”. In: *IEEE Transactions on Computers* 48.5 (1999), pp. 506–521.
- [550] Mathlib Community. *Mathlib: The Lean Mathematical Library*. https://leanprover-community.github.io/mathlib4_docs/. 2024.
- [551] Tomoya Matsumoto, Takayuki Miura, and Naoto Yanai. “Membership Inference Attacks against Diffusion Models”. In: *IEEE Security and Privacy Workshops (SPW)*. IEEE, 2023, pp. 77–83.
- [552] Justus Mattern et al. “Membership Inference Attacks against Language Models via Neighbourhood Comparison”. In: *Findings of the Association for Computational Linguistics: ACL 2023*. 2023, pp. 11330–11343.
- [553] John Maynard Smith. *Evolution and the Theory of Games*. Cambridge University Press, 1982.
- [554] John Maynard Smith and George R. Price. “The Logic of Animal Conflict”. In: *Nature* 246.5427 (1973), pp. 15–18.
- [555] James L McClelland, Bruce L McNaughton, and Randall C O’Reilly. “Why There Are Complementary Learning Systems in the Hippocampus and Neocortex: Insights from the Successes and Failures of Connectionist Models of Learning and Memory”. In: *Psychological Review* 102.3 (1995), pp. 419–457.
- [556] Michael McCloskey and Neal J Cohen. “Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem”. In: *Psychology of Learning and Motivation*. Vol. 24. Academic Press, 1989, pp. 109–165.
- [557] Warren McCulloch and Walter Pitts. “A Logical Calculus of Ideas Immanent in Nervous Activity”. In: *The Bulletin of Mathematical Biophysics* 5.4 (1943), pp. 115–133.
- [558] Matthieu Meeus et al. “Did the Neurons Read Your Book? Document-Level Membership Inference for Large Language Models”. In: *USENIX Security Symposium*. 2024.

-
- [559] Norman Megill and David A. Wheeler. *Metamath: A Computer Language for Mathematical Proofs*. <http://us.metamath.org>. Lulu Press, 2019.
 - [560] Joseph Mellor et al. “Neural Architecture Search without Training”. In: *Proceedings of the International Conference on Machine Learning (ICML)*. 2021.
 - [561] Chenlin Meng et al. “SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations”. In: *International Conference on Learning Representations* (2022).
 - [562] William Merrill and Ashish Sabharwal. “The Expressive Power of Transformers with Chain of Thought”. In: *arXiv preprint arXiv:2310.07923* (2023).
 - [563] William Merrill and Ashish Sabharwal. “The Expressive Power of Transformers with Chain of Thought”. In: *International Conference on Learning Representations* (2024).
 - [564] Lars Mescheder, Sebastian Nowozin, and Andreas Geiger. “The numerics of gans”. In: *Advances in Neural Information Processing Systems*. 2017, pp. 1823–1833.
 - [565] Luke Metz et al. “Unrolled Generative Adversarial Networks”. In: *arXiv preprint arXiv:1611.02163* (2016).
 - [566] Aaron Meurer et al. “SymPy: Symbolic Computing in Python”. In: *PeerJ Computer Science* 3 (2017), e103.
 - [567] Paulius Micikevicius et al. “Mixed Precision Training”. In: *International Conference on Learning Representations (ICLR)*. 2018.
 - [568] Brad L. Miller and David E. Goldberg. “Genetic Algorithms, Tournament Selection, and the Effects of Noise”. In: *Complex Systems* 9 (1989), pp. 193–212.
 - [569] George A. Miller. “The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information”. In: *Psychological Review* 63.2 (1956), pp. 81–97.
 - [570] Tom Minka. “Divergence Measures and Message Passing”. In: *Microsoft Research Technical Report*. 2005.
 - [571] Marvin Minsky. *The Society of Mind*. New York, NY: Simon & Schuster, 1986.
 - [572] Marvin Minsky and Seymour Papert. *Perceptrons: An Introduction to Computational Geometry*. MIT Press, 1969.
 - [573] Ilya Mironov. “Rényi Differential Privacy”. In: *IEEE Computer Security Foundations Symposium (CSF)*. IEEE, 2017, pp. 263–275.
 - [574] Mehdi Mirza and Simon Osindero. *Conditional Generative Adversarial Nets*. 2014. arXiv: 1411.1784 [cs.LG].

- [575] Dmytro Mishkin and Jiri Matas. “All You Need is a Good Init”. In: *International Conference on Learning Representations (ICLR)*. 2016.
- [576] Diganta Misra. *Mish: A Self Regularized Non-Monotonic Activation Function*. 2019. arXiv: 1908.08681 [cs.LG].
- [577] Michael Mitzenmacher. “The Power of Two Choices in Randomized Load Balancing”. In: *IEEE Transactions on Parallel and Distributed Systems* 12.10 (2001), pp. 1094–1104.
- [578] Takeru Miyato et al. “Spectral normalization for generative adversarial networks”. In: *Proceedings of the International Conference on Learning Representations*. 2018. URL: <https://openreview.net/forum?id=B1QRgziT->.
- [579] Gaspard Monge. “Mémoire sur la théorie des déblais et des remblais”. In: *Mémoires de l'Académie Royale des Sciences de Paris* (1781), pp. 666–704.
- [580] Leonardo de Moura and Nikolaj Bjørner. “Z3: An Efficient SMT Solver”. In: *TACAS* (2008).
- [581] Leonardo de Moura and Sebastian Ullrich. “The Lean 4 Theorem Prover and Programming Language”. In: *International Conference on Automated Deduction (CADE)*. Springer, 2021.
- [582] Leonardo de Moura and Sebastian Ullrich. *The Lean 4 Theorem Prover and Programming Language*. CADE-28. 2021.
- [583] Leonardo de Moura et al. “The Lean Theorem Prover (System Description)”. In: *International Conference on Automated Deduction*. 2015, pp. 378–388.
- [584] Jean-Baptiste Mouret and Jeff Clune. “Illuminating Search Spaces by Mapping Elites”. In: *arXiv preprint arXiv:1504.04909* (2015).
- [585] Niklas Muennighoff et al. “MTEB: Massive Text Embedding Benchmark”. In: *Proceedings of EACL*. 2023, pp. 2014–2037.
- [586] Markus Nagel et al. *A White Paper on Neural Network Quantization*. 2021. arXiv: 2106.08295 [cs.LG].
- [587] Rosemarie Nagel. “Unraveling in Guessing Games: An Experimental Study”. In: *The American Economic Review* 85.5 (1995), pp. 1313–1326.
- [588] Maryam M. Najafabadi et al. “Deep Learning Applications and Challenges in Big Data Analytics”. In: *Journal of Big Data* 2.1 (2015), pp. 1–21.
- [589] John F. Nash. “Equilibrium Points in N-Person Games”. In: *Proceedings of the National Academy of Sciences* 36.1 (1950), pp. 48–49.

-
- [590] Milad Nasr, Reza Shokri, and Amir Houmansadr. “Comprehensive Privacy Analysis of Deep Learning: Passive and Active White-box Inference Attacks against Centralized and Federated Learning”. In: *IEEE Symposium on Security and Privacy (S&P)*. IEEE, 2019, pp. 739–753.
- [591] Balas Kausik Natarajan. “Sparse Approximate Solutions to Linear Systems”. In: *SIAM Journal on Computing* 24.2 (1995), pp. 227–234.
- [592] Radford M. Neal. “Connectionist Learning of Belief Networks”. In: *Artificial Intelligence* 56.1 (1992), pp. 71–113.
- [593] Yurii Nesterov. *Lectures on Convex Optimization*. Springer, 2018.
- [594] John von Neumann. *First Draft of a Report on the EDVAC*. Tech. rep. University of Pennsylvania, 1945.
- [595] John von Neumann. “Zur Theorie der Gesellschaftsspiele”. In: *Mathematische Annalen* 100.1 (1928), pp. 295–320.
- [596] John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press, 1944.
- [597] John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press, 1944.
- [598] Allen Newell and Herbert A. Simon. “The Logic Theory Machine: A Complex Information Processing System”. In: *IRE Transactions on Information Theory*. Vol. 2. 3. 1956, pp. 61–79.
- [599] Andrew Y. Ng, Daishi Harada, and Stuart Russell. “Policy Invariance Under Reward Transformations: Theory and Application to Reward Shaping”. In: *Proceedings of the International Conference on Machine Learning (ICML)*. 1999, pp. 278–287.
- [600] Alexander Quinn Nichol and Prafulla Dhariwal. “Improved Denoising Diffusion Probabilistic Models”. In: *International Conference on Machine Learning*. PMLR. 2021, pp. 8162–8171.
- [601] Shen Nie et al. “Large language diffusion models”. In: *arXiv preprint arXiv:2502.09992* (2025).
- [602] Robert Nieuwenhuis, Albert Oliveras, and Cesare Tinelli. “Solving SAT and SAT Modulo Theories: From an Abstract Davis–Putnam–Logemann–Loveland Procedure to DPLL(T)”. In: *Journal of the ACM* 53.6 (2006), pp. 937–977.
- [603] Tobias Nipkow, Lawrence C. Paulson, and Markus Wenzel. *Isabelle/HOL: A Proof Assistant for Higher-Order Logic*. Springer, 2002.
- [604] Noam Nisan et al. *Algorithmic Game Theory*. Cambridge University Press, 2007.

- [605] Rodrigo Nogueira et al. “Document Ranking with a Pretrained Sequence-to-Sequence Model”. In: *Findings of EMNLP*. 2020.
- [606] Ulf Norell. “Towards a Practical Programming Language Based on Dependent Type Theory”. PhD thesis. Chalmers University of Technology, 2007.
- [607] Tor Nørretranders. *The User Illusion: Cutting Consciousness Down to Size*. Viking, 1998.
- [608] Alexander Novikov et al. “AlphaEvolve: A Coding Agent for Scientific and Algorithmic Discovery”. In: *arXiv preprint arXiv:2506.13131* (2025).
- [609] Sebastian Nowozin, Botond Cseke, and Ryota Tomioka. “f-GAN: Training generative neural samplers using variational divergence minimization”. In: *Advances in Neural Information Processing Systems*. Vol. 29. 2016.
- [610] NVIDIA. “Cosmos World Foundation Model Platform for Physical AI”. In: *arXiv preprint arXiv:2501.03575* (2025).
- [611] Maxwell Nye et al. “Show Your Work: Scratchpads for Intermediate Computation with Language Models”. In: *arXiv preprint arXiv:2112.00114* (2021).
- [612] Augustus Odena, Christopher Olah, and Jonathon Shlens. “Conditional Image Synthesis With Auxiliary Classifier GANs”. In: *Proceedings of the International Conference on Machine Learning* (2017).
- [613] Reza Olfati-Saber and Richard M. Murray. “Consensus Problems in Networks of Agents with Switching Topology and Time-Delays”. In: *IEEE Transactions on Automatic Control* 49.9 (2004), pp. 1520–1533.
- [614] Aaron van den Oord, Oriol Vinyals, and Koray Kavukcuoglu. “Neural Discrete Representation Learning”. In: *Advances in Neural Information Processing Systems*. 2017, pp. 6306–6315.
- [615] Aaron van den Oord, Oriol Vinyals, and Koray Kavukcuoglu. “Neural Discrete Representation Learning”. In: *Advances in Neural Information Processing Systems* (2017).
- [616] Aäron van den Oord, Nal Kalchbrenner, and Koray Kavukcuoglu. “Pixel Recurrent Neural Networks”. In: *Proceedings of the 33rd International Conference on Machine Learning (ICML)*. 2016.
- [617] Aäron van den Oord et al. “Conditional Image Generation with PixelCNN Decoders”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2016.
- [618] Aäron van den Oord et al. “Parallel WaveNet: Fast High-Fidelity Speech Synthesis”. In: *arXiv preprint arXiv:1711.10433* (2018).
- [619] Aäron van den Oord et al. “WaveNet: A Generative Model for Raw Audio”. In: *9th ISCA Speech Synthesis Workshop (SSW)*. 2016.

- [620] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. “Representation Learning with Contrastive Predictive Coding”. In: *arXiv preprint arXiv:1807.03748* (2018).
- [621] OpenAI. “GPT-4 Technical Report”. In: *arXiv preprint arXiv:2303.08774* (2023).
- [622] OpenAI. *Harness Engineering*. <https://openai.com>. 2025.
- [623] OpenAI. *Learning to Reason with LLMs*. openai.com/index/learning-to-reason-with-llms. 2024.
- [624] OpenAI. *OpenAI o3 and o4-mini*. 2025.
- [625] OpenEvolve Contributors. “OpenEvolve: Open-Source LLM-Guided Evolutionary Search”. In: *GitHub Repository* (2025).
- [626] Kazuki Osawa et al. “Practical Deep Learning with Bayesian Principles”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2019.
- [627] Jingyang Ou, Shen Nie, et al. “Your absorbing discrete diffusion secretly models the conditional distributions of clean data”. In: *arXiv preprint arXiv:2406.03736* (2024).
- [628] Long Ouyang et al. “Training Language Models to Follow Instructions with Human Feedback”. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 27730–27744.
- [629] Xuchen Pan et al. “Very Large-Scale Multi-Agent Simulation in AgentScope”. In: *arXiv preprint arXiv:2407.17789* (2024).
- [630] Yan Pang et al. “White-box Membership Inference Attacks against Diffusion Models”. In: *arXiv preprint arXiv:2308.06405* (2023).
- [631] Christos Papadimitriou et al. “Predictive Policy Improvement for Multi-Agent Cooperation via Co-Player Inference”. In: *arXiv preprint arXiv:2501.00000* (2025). WARNING: Placeholder arXiv ID; paper may not exist as described.
- [632] George Papamakarios, Theo Pavlakou, and Iain Murray. “Masked Autoregressive Flow for Density Estimation”. In: *Advances in Neural Information Processing Systems*. 2017, pp. 2338–2347.
- [633] German I Parisi et al. “Continual Lifelong Learning with Neural Networks: A Review”. In: *Neural Networks* 113 (2019), pp. 54–71.
- [634] Taesung Park et al. “Semantic Image Synthesis with Spatially-Adaptive Normalization”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019.
- [635] Anirudh Parthasarathy et al. *LeanProgress: Guiding Search for Neural Theorem Proving via Proof Progress Prediction*. arXiv preprint. 2024.

- [636] Deepak Pathak et al. “Curiosity-driven Exploration by Self-Supervised Prediction”. In: *International Conference on Machine Learning (ICML)*. 2017.
- [637] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.
- [638] Karl Pearson. “Contributions to the Mathematical Theory of Evolution”. In: *Philosophical Transactions of the Royal Society of London A* 185 (1894), pp. 71–110.
- [639] William Peebles and Saining Xie. “Scalable Diffusion Models with Transformers”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023.
- [640] William Peebles and Saining Xie. “Scalable Diffusion Models with Transformers”. In: *International Conference on Computer Vision* (2023).
- [641] Bowen Peng et al. *YaRN: Efficient Context Window Extension of Large Language Models*. 2023. arXiv: [2309.00071](https://arxiv.org/abs/2309.00071) [cs.CL].
- [642] Bowen Peng et al. “YaRN: Efficient Context Window Extension of Large Language Models”. In: *arXiv preprint arXiv:2309.00071* (2024).
- [643] Ethan Perez et al. “FiLM: Visual Reasoning with a General Conditioning Layer”. In: *AAAI Conference on Artificial Intelligence* (2018).
- [644] Fabio Petroni et al. “KILT: a Benchmark for Knowledge Intensive Language Tasks”. In: *Proceedings of NAACL*. Also on arXiv:2009.02252. 2021.
- [645] Henning Petzka, Asja Fischer, and Denis Lukovnikov. “On the Regularization of Wasserstein GANs”. In: *arXiv preprint arXiv:1709.08894* (2018).
- [646] Hieu Pham et al. “Efficient Neural Architecture Search via Parameter Sharing”. In: *Proceedings of ICML* (2018).
- [647] Hieu Pham et al. “Efficient Neural Architecture Search via Parameter Sharing”. In: *Proceedings of the International Conference on Machine Learning (ICML)*. 2018.
- [648] Mark S. Pinsker. *Information and Information Stability of Random Variables and Processes*. San Francisco: Holden-Day, 1964.
- [649] Stanislas Polu and Ilya Sutskever. “Generative Language Modeling for Automated Theorem Proving”. In: *arXiv preprint arXiv:2009.03393* (2020).
- [650] Stanislas Polu and Ilya Sutskever. “GPT-f: Generative Language Modeling for Automated Theorem Proving”. In: *arXiv preprint arXiv:2009.03393* (2020).
- [651] Adam Polyak et al. “Movie Gen: A Cast of Media Foundation Models”. In: *arXiv preprint arXiv:2410.13720* (2024).
- [652] Mitchell A. Potter and Kenneth A. De Jong. “Cooperative Coevolution: An Architecture for Evolving Coadapted Subcomponents”. In: *Evolutionary Computation* 8.1 (2000), pp. 1–29.

- [653] Hadi Pouransari et al. "Pretraining with Hierarchical Memories: Separating Long-Tail and Common Knowledge". In: *arXiv preprint arXiv:2510.02375* (2025).
- [654] Ofir Press, Noah A. Smith, and Mike Lewis. "Train Short, Test Long: Attention with Linear Biases Enables Input Length Extrapolation". In: *International Conference on Learning Representations (ICLR)*. 2022.
- [655] Ofir Press and Lior Wolf. "Using the Output Embedding to Improve Language Models". In: *Conference of the European Chapter of the Association for Computational Linguistics (EACL)*. 2017, pp. 157–163.
- [656] Justin K. Pugh, Lisa B. Soros, and Kenneth O. Stanley. "Quality Diversity: A New Frontier for Evolutionary Computation". In: *Frontiers in Robotics and AI* 3 (2016), p. 40.
- [657] William Pugh. "The Omega Test: A Fast and Practical Integer Programming Algorithm for Dependence Analysis". In: *ACM/IEEE Conference on Supercomputing*. 1991.
- [658] Chenyang Qi et al. "FateZero: Fusing Attentions for Zero-shot Text-based Video Editing". In: *International Conference on Computer Vision* (2023).
- [659] Haonan Qiu et al. "FreeNoise: Tuning-Free Longer Video Diffusion via Noise Rescheduling". In: *Conference on Computer Vision and Pattern Recognition* (2024).
- [660] Zihan Qiu et al. "Gated Attention for Large Language Models". In: *Advances in Neural Information Processing Systems*. Vol. 38. 2025.
- [661] Qwen Team. "Qwen2.5 Technical Report". In: *arXiv preprint arXiv:2412.15115* (2024).
- [662] Gal Raayoni et al. "Conservative matrix fields and the Ramanujan machine". In: *Proceedings of the National Academy of Sciences* (2024).
- [663] Gal Raayoni et al. "Generating conjectures on fundamental constants with the Ramanujan Machine". In: *Nature* 590 (2021), pp. 67–73.
- [664] Lawrence R. Rabiner. "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition". In: *Proceedings of the IEEE* 77.2 (1989), pp. 257–286. doi: [10.1109/5.18626](https://doi.org/10.1109/5.18626).
- [665] Alec Radford, Luke Metz, and Soumith Chintala. "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks". In: *International Conference on Learning Representations*. 2016.
- [666] Alec Radford, Luke Metz, and Soumith Chintala. "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks". In: *International Conference on Learning Representations*. 2016.

- [667] Alec Radford et al. *Language Models are Unsupervised Multitask Learners*. Tech. rep. OpenAI, 2019.
- [668] Alec Radford et al. “Learning Transferable Visual Models From Natural Language Supervision”. In: *International Conference on Machine Learning* (2021).
- [669] Alec Radford et al. “Robust Speech Recognition via Large-Scale Weak Supervision”. In: *International Conference on Machine Learning*. 2023.
- [670] Rafael Rafailov et al. “Direct Preference Optimization: Your Language Model is Secretly a Reward Model”. In: *Advances in Neural Information Processing Systems* 36 (2023).
- [671] Colin Raffel et al. “Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer”. In: *Journal of Machine Learning Research* 21.140 (2020), pp. 1–67.
- [672] Colin Raffel et al. “Online and Linear-Time Attention by Enforcing Monotonic Alignments”. In: *International Conference on Machine Learning*. 2017, pp. 2837–2846.
- [673] Ali Rahimi and Benjamin Recht. “Random Features for Large-Scale Kernel Machines”. In: *Advances in Neural Information Processing Systems*. Vol. 20. 2007.
- [674] Marcus E. Raichle and Debra A. Gusnard. “Appraising the Brain’s Energy Budget”. In: *Proceedings of the National Academy of Sciences* 99.16 (2002), pp. 10237–10239.
- [675] Tom Rainforth et al. “Tighter Variational Bounds are Not Necessarily Better”. In: *Proceedings of the International Conference on Machine Learning (ICML)*. 2018, pp. 4277–4285.
- [676] Shashank Rajput et al. “Recommender Systems with Generative Retrieval”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. TIGER / semantic IDs. 2023.
- [677] Prajit Ramachandran, Barret Zoph, and Quoc V. Le. *Searching for Activation Functions*. 2017. arXiv: [1710.05941](https://arxiv.org/abs/1710.05941) [cs.NE].
- [678] Aditya Ramesh et al. “Hierarchical text-conditional image generation with CLIP latents”. In: *arXiv preprint arXiv:2204.06125* (2022).
- [679] Santiago Ramón y Cajal. *The Structure and Connexions of Neurons*. Nobel Lecture. 1906.
- [680] Rajesh Ranganath, Sean Gerrish, and David M. Blei. “Black Box Variational Inference”. In: *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*. 2014, pp. 814–822.
- [681] Ali Razavi, Aaron van den Oord, and Oriol Vinyals. “Generating Diverse High-Fidelity Images with VQ-VAE-2”. In: *Advances in Neural Information Processing Systems*. 2019, pp. 14866–14876.

-
- [682] Esteban Real et al. "Regularized Evolution for Image Classifier Architecture Search". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. 2019, pp. 4780–4789.
- [683] Ingo Rechenberg. *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Frommann-Holzboog, 1973.
- [684] Revanth Gangi Reddy et al. "FIRST: Faster Improved Listwise Reranking with Single Token Decoding". In: *Proceedings of EMNLP*. 2024.
- [685] Nils Reimers and Iryna Gurevych. "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks". In: *Proceedings of EMNLP* (2019).
- [686] Yinuo Ren, Haoxuan Chen, et al. "How discrete and continuous diffusion meet: Comprehensive analysis of discrete diffusion models via a stochastic integral framework". In: *arXiv preprint arXiv:2410.03601* (2024).
- [687] Danilo Jimenez Rezende and Shakir Mohamed. "Variational Inference with Normalizing Flows". In: *International Conference on Machine Learning*. 2015, pp. 1530–1538.
- [688] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. "Stochastic Backpropagation and Approximate Inference in Deep Generative Models". In: *International Conference on Machine Learning*. 2014, pp. 1278–1286.
- [689] Alexandre Riazanov and Andrei Voronkov. "The Design and Implementation of VAMPIRE". In: *AI Communications*. Vol. 15. 2–3. 2002, pp. 91–110.
- [690] H. Robbins and S. Monro. "A stochastic approximation method". In: *The Annals of Mathematical Statistics* (1951), pp. 400–407.
- [691] Stephen Robertson and Hugo Zaragoza. "The Probabilistic Relevance Framework: BM25 and Beyond". In: *Foundations and Trends in Information Retrieval* 3.4 (2009), pp. 333–389.
- [692] Stephen E. Robertson and Steve Walker. "Some Simple Effective Approximations to the 2-Poisson Model for Probabilistic Weighted Retrieval". In: *Proceedings of SIGIR*. Venue/pages unspecified in this bib (see PDF source). 1994.
- [693] John Alan Robinson. "A Machine-Oriented Logic Based on the Resolution Principle". In: *Journal of the ACM* 12.1 (1965), pp. 23–41.
- [694] Robin Rombach et al. "High-resolution image synthesis with latent diffusion models". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022), pp. 10684–10695.
- [695] Bernardino Romera-Paredes et al. "Mathematical Discoveries from Program Search with Large Language Models". In: *Nature* 625 (2024), pp. 468–475.

- [696] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *Proceedings of MICCAI* (2015), pp. 234–241.
- [697] Frank Rosenblatt. “The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain”. In: *Psychological Review* 65.6 (1958), pp. 386–408.
- [698] Christopher D. Rosin and Richard K. Belew. “New Methods for Competitive Coevolution”. In: *Evolutionary Computation* 5.1 (1997), pp. 1–29.
- [699] Ludan Ruan et al. “MM-Diffusion: Learning Multi-Modal Diffusion Models for Joint Audio and Video Generation”. In: *Conference on Computer Vision and Pattern Recognition* (2023).
- [700] Donald B. Rubin. *Statistical Disclosure Limitation*. Vol. 9. 2. 1993, pp. 461–468.
- [701] Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas. “The Earth Mover’s Distance as a Metric for Image Retrieval”. In: *International Journal of Computer Vision* 40.2 (2000), pp. 99–121.
- [702] Nataniel Ruiz et al. “DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation”. In: *Conference on Computer Vision and Pattern Recognition* (2023).
- [703] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. “Learning representations by back-propagating errors”. In: *Nature* 323.6088 (1986), pp. 533–536.
- [704] Andrei A Rusu et al. “Progressive Neural Networks”. In: *arXiv preprint arXiv:1606.04671*. 2016.
- [705] Andrei A. Rusu et al. “Progressive Neural Networks”. In: *Advances in Neural Information Processing Systems*. 2016.
- [706] Alexandre Sablayrolles et al. “White-box vs Black-box: Bayes Optimal Strategies for Membership Inference”. In: *International Conference on Machine Learning*. 2019, pp. 5558–5567.
- [707] Sartaj Sahni and Teofilo Gonzalez. “P-Complete Approximation Problems”. In: *Journal of the ACM* 23.3 (1976), pp. 555–565.
- [708] Subham Sekhar Sahoo et al. “Simple and effective masked diffusion language models”. In: *Advances in Neural Information Processing Systems* 37 (2024).
- [709] Hiroaki Sakoe and Seibi Chiba. “Dynamic Programming Algorithm Optimization for Spoken Word Recognition”. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26.1 (1978), pp. 43–49. doi: [10.1109/TASSP.1978.1163055](https://doi.org/10.1109/TASSP.1978.1163055).
- [710] Tim Salimans and Jonathan Ho. “Progressive Distillation for Fast Sampling of Diffusion Models”. In: *International Conference on Learning Representations* (2022).

-
- [711] Tim Salimans et al. “Evolution Strategies as a Scalable Alternative to Reinforcement Learning”. In: *arXiv preprint arXiv:1703.03864* (2017).
- [712] Tim Salimans et al. “Improved techniques for training GANs”. In: *Advances in Neural Information Processing Systems*. 2016, pp. 2234–2242.
- [713] Tim Salimans et al. “PixelCNN++: Improving the PixelCNN with Discretized Logistic Mixture Likelihood and Other Modifications”. In: *International Conference on Learning Representations (ICLR)*. 2017.
- [714] Gerard Salton and Christopher Buckley. “Term-weighting approaches in automatic text retrieval”. In: *Information Processing & Management* 24.5 (1988), pp. 513–523. doi: [10.1016/0306-4573\(88\)90021-0](https://doi.org/10.1016/0306-4573(88)90021-0).
- [715] Arthur L. Samuel. “Some Studies in Machine Learning Using the Game of Checkers”. In: *IBM Journal of Research and Development* 3.3 (1959), pp. 210–229.
- [716] D. Samuel et al. “Diffusion Features for Personalized Segmentation and Retrieval”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. Author list abbreviated; see NeurIPS 2024 paper PDF for full metadata. 2024.
- [717] Shibani Santurkar et al. “How Does Batch Normalization Help Optimization?” In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2018.
- [718] Andrew M. Saxe, James L. McClelland, and Surya Ganguli. “Exact Solutions to the Nonlinear Dynamics of Learning in Deep Linear Neural Networks”. In: *International Conference on Learning Representations (ICLR)*. 2014.
- [719] Andrew M. Saxe et al. “On the Information Bottleneck Theory of Deep Learning”. In: *Journal of Statistical Mechanics: Theory and Experiment* (2019).
- [720] John Schulman et al. “High-Dimensional Continuous Control Using Generalized Advantage Estimation”. In: *arXiv preprint arXiv:1506.02438* (2016). Published at ICLR 2016.
- [721] John Schulman et al. “Proximal Policy Optimization Algorithms”. In: *arXiv preprint arXiv:1707.06347*. 2017.
- [722] John Schulman et al. “Proximal Policy Optimization Algorithms”. In: *arXiv preprint arXiv:1707.06347* (2017).
- [723] Stephan Schulz. “E – A Brainiac Theorem Prover”. In: *AI Communications*. Vol. 15. 2–3. 2002, pp. 111–126.
- [724] Stephan Schulz. “E – A Brainiac Theorem Prover”. In: *AI Communications*. Vol. 15. 2–3. 2002, pp. 111–126.
- [725] Jonathan Schwarz et al. “Progress & Compress: A Scalable Framework for Continual Learning”. In: *Proceedings of the International Conference on Machine Learning (ICML)* (2018).

- [726] Hans-Paul Schwefel. *Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie*. Birkhäuser, 1977.
- [727] Lee A. Segel and Irun R. Cohen. *Design Principles for the Immune System and Other Distributed Autonomous Systems*. New York: Oxford University Press, 2001.
- [728] Daniel Selsam et al. “Learning a SAT Solver from Single-Bit Supervision”. In: *arXiv preprint arXiv:1802.03685* (2019). Published at ICLR 2019.
- [729] Rico Sennrich, Barry Haddow, and Alexandra Birch. “Improving Neural Machine Translation Models with Monolingual Data”. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*. 2016, pp. 86–96.
- [730] Burr Settles. *Active Learning Literature Survey*. University of Wisconsin–Madison, 2009.
- [731] Jay Shah et al. *FlashAttention-3: Fast and Accurate Attention with Asynchrony and Low-precision*. 2024. arXiv: 2407.08691 [cs.LG].
- [732] Claude E. Shannon. “Prediction and Entropy of Printed English”. In: *Bell System Technical Journal* 30.1 (1951), pp. 50–64.
- [733] Claude Elwood Shannon. “A mathematical theory of communication”. In: *The Bell System Technical Journal* 27.3 (1948), pp. 379–423.
- [734] Zhihong Shao et al. “DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models”. In: *arXiv preprint arXiv:2402.03300* (2024).
- [735] Lloyd S. Shapley. “A Value for N-Person Games”. In: *Contributions to the Theory of Games*. Ed. by Harold W. Kuhn and Albert W. Tucker. Vol. 2. Princeton University Press, 1953, pp. 307–317.
- [736] Alok Sharma et al. “DeepInsight: A methodology to transform a non-image data to an image for convolution neural network architecture”. In: *Scientific Reports* 9 (2019), p. 11399.
- [737] Noam Shazeer. “Fast Transformer Decoding: One Write-Head is All You Need”. In: *arXiv preprint arXiv:1911.02150* (2019).
- [738] Noam Shazeer. “GLU Variants Improve Transformer”. In: *arXiv preprint arXiv:2002.05202* (2020).
- [739] Noam Shazeer et al. “Outrageously Large Neural Networks: The Sparsely-Gated Mixture-of-Experts Layer”. In: *International Conference on Learning Representations*. 2017.
- [740] Ying Sheng et al. “S-LoRA: Serving Thousands of Concurrent LoRA Adapters”. In: *arXiv preprint arXiv:2311.03285* (2024).

- [741] Abhimanyu Shetty et al. "Exposing Privacy Gaps: Membership Inference Attack on Preference Data for LLM Alignment". In: *International Conference on Learning Representations*. 2025.
- [742] Anonymous Shetty et al. "NeuroDream: A Sleep-Inspired Framework for Offline Memory Consolidation in Neural Networks". In: *arXiv preprint* (2024).
- [743] Haizhou Shi et al. "Continual Learning of Large Language Models: A Comprehensive Survey". In: *arXiv preprint arXiv:2404.16789* (2024).
- [744] Weijia Shi et al. "Detecting Pretraining Data from Large Language Models". In: *International Conference on Learning Representations*. 2024.
- [745] Hanul Shin et al. "Continual Learning with Deep Generative Replay". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2017.
- [746] Noah Shinn et al. "Reflexion: Language Agents with Verbal Reinforcement Learning". In: *Advances in Neural Information Processing Systems* 36 (2023).
- [747] Yoav Shoham. "Agent-Oriented Programming". In: *Artificial Intelligence* 60.1 (1993), pp. 51–92.
- [748] Reza Shokri et al. "Membership Inference Attacks Against Machine Learning Models". In: *IEEE Symposium on Security and Privacy (S&P)*. IEEE, 2017, pp. 3–18.
- [749] Ravid Shwartz-Ziv and Naftali Tishby. "Opening the Black Box of Deep Neural Networks via Information". In: *arXiv preprint arXiv:1703.00810* (2017).
- [750] David Silver et al. "Mastering the game of Go with deep neural networks and tree search". In: *Nature* 529 (2016), pp. 484–489.
- [751] David Silver et al. "Mastering the Game of Go without Human Knowledge". In: *Nature* 550.7676 (2017), pp. 354–359.
- [752] Karl Sims. "Evolving Virtual Creatures". In: *Proceedings of SIGGRAPH*. 1994, pp. 15–22.
- [753] Uriel Singer et al. "Make-A-Video: Text-to-Video Generation without Text-Video Data". In: *International Conference on Learning Representations* (2023).
- [754] Vaibhav Singh et al. "Beyond Cosine Decay: On the Effectiveness of Infinite Learning Rate Schedule for Continual Pre-Training". In: *arXiv preprint arXiv:2503.02844* (2025).
- [755] A. Prasad Sistla and Edmund M. Clarke. "The Complexity of Propositional Linear Temporal Logics". In: *Journal of the ACM* 32.3 (1985), pp. 733–749.
- [756] Josef Sivic and Andrew Zisserman. "Video Google: A Text Retrieval Approach to Object Matching in Videos". In: *Proceedings of ICCV*. Pages unspecified. 2003.

- [757] Malcolm Slaney and Michael Casey. “Locality-Sensitive Hashing for Finding Nearest Neighbors”. In: *IEEE Signal Processing Magazine* 25.2 (2008), pp. 128–131. DOI: [10.1109/MSP.2007.914237](https://doi.org/10.1109/MSP.2007.914237).
- [758] Leslie N. Smith. “Cyclical Learning Rates for Training Neural Networks”. In: *IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2017, pp. 464–472.
- [759] Leslie N. Smith and Nicholay Topin. *Super-Convergence: Very Fast Training of Neural Networks Using Large Learning Rates*. 2018. arXiv: [1708.07120](https://arxiv.org/abs/1708.07120) [cs.LG].
- [760] David R. So et al. “Primer: Searching for Efficient Transformers for Language Modeling”. In: *Advances in Neural Information Processing Systems*. 2021.
- [761] Jascha Sohl-Dickstein et al. “Deep Unsupervised Learning using Nonequilibrium Thermodynamics”. In: *International Conference on Machine Learning*. PMLR. 2015, pp. 2256–2264.
- [762] Kihyuk Sohn, Honglak Lee, and Xinchen Yan. “Learning Structured Output Representation using Deep Conditional Generative Models”. In: *Advances in Neural Information Processing Systems*. 2015, pp. 3483–3491.
- [763] Jiaming Song, Chenlin Meng, and Stefano Ermon. “Denoising Diffusion Implicit Models”. In: *International Conference on Learning Representations* (2021).
- [764] Jiaming Song, Chenlin Meng, and Stefano Ermon. “Denoising Diffusion Implicit Models”. In: *arXiv preprint arXiv:2010.02502* (2021).
- [765] Peiyang Song, Kaiyu Yang, and Anima Anandkumar. *Lean Copilot: Large Language Models as Copilots for Theorem Proving in Lean*. arXiv:2404.12534. 2024.
- [766] Peiyang Song, Kaiyu Yang, and Anima Anandkumar. “LeanCopilot: LLMs as Copilots for Theorem Proving in Lean”. In: *arXiv preprint arXiv:2404.12534* (2024).
- [767] Yang Song and Stefano Ermon. “Generative modeling by estimating gradients of the data distribution”. In: *Advances in Neural Information Processing Systems*. Vol. 32. 2019.
- [768] Yang Song et al. “Consistency Models”. In: *International Conference on Machine Learning* (2023).
- [769] Yang Song et al. “Score-Based Generative Modeling through Stochastic Differential Equations”. In: *International Conference on Learning Representations (ICLR)*. 2021.
- [770] Yang Song et al. “Score-Based Generative Modeling through Stochastic Differential Equations”. In: *International Conference on Learning Representations*. 2021.
- [771] Yang Song et al. “Sliced Score Matching: A Scalable Approach to Density and Score Estimation”. In: *Uncertainty in Artificial Intelligence*. PMLR. 2020, pp. 574–584.

-
- [772] Larry R. Squire. “Memory Systems of the Brain: A Brief History and Current Perspective”. In: *Neurobiology of Learning and Memory* 82.3 (2004), pp. 171–177. DOI: [10.1016/j.nlm.2004.06.005](https://doi.org/10.1016/j.nlm.2004.06.005).
- [773] Niranjana Srinivas et al. “Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design”. In: *Proceedings of ICML* (2010).
- [774] Rupesh Kumar Srivastava, Klaus Greff, and Jürgen Schmidhuber. *Highway Networks*. 2015. arXiv: [1505.00387](https://arxiv.org/abs/1505.00387) [cs.LG].
- [775] Dale O. Stahl and Paul W. Wilson. “Experimental Evidence on Players’ Models of Other Players”. In: *Journal of Economic Behavior and Organization* 25.3 (1994), pp. 309–327.
- [776] Kenneth O. Stanley and Joel Lehman. “Why Open-Endedness Matters”. In: *Artificial Life* 23.3 (2017), pp. 344–347.
- [777] Kenneth O. Stanley, Joel Lehman, and Lisa Soros. “Open-Endedness: The Last Grand Challenge You’ve Never Heard Of”. In: *O’Reilly Radar* (2019).
- [778] Kenneth O. Stanley and Risto Miikkulainen. “Evolving Neural Networks through Augmenting Topologies”. In: *Evolutionary Computation*. Vol. 10. 2. 2002, pp. 99–127.
- [779] Kenneth O. Stanley and Risto Miikkulainen. “Evolving Neural Networks through Augmenting Topologies”. In: *Evolutionary Computation* 10.2 (2002), pp. 99–127.
- [780] Nisan Stiennon et al. “Learning to Summarize with Human Feedback”. In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 3008–3021.
- [781] Joan Stiles and Terry L. Jernigan. “The Basics of Brain Development”. In: *Neuropsychology Review* 20.4 (2010), pp. 327–348.
- [782] Volker Strassen. “Gaussian Elimination is Not Optimal”. In: *Numerische Mathematik* 13 (1969), pp. 354–356.
- [783] Hongjin Su et al. “One Embedder, Any Task: Instruction-Finetuned Text Embeddings”. In: *arXiv preprint arXiv:2212.09741* (2022).
- [784] Jianlin Su et al. “RoFormer: Enhanced Transformer with Rotary Position Embedding”. In: *Neurocomputing* 568 (2024), p. 127063.
- [785] Jianlin Su et al. “RoFormer: Enhanced Transformer with Rotary Position Embedding”. In: *Neurocomputing* 568 (2024), p. 127063.
- [786] Felipe Petroski Such et al. “Deep Neuroevolution: Genetic Algorithms Are a Competitive Alternative for Training Deep Neural Networks for Reinforcement Learning”. In: *arXiv preprint arXiv:1712.06567* (2017).

- [787] Fan-Keng Sun, Cheng-Hao Ho, and Hung-Yi Lee. “LAMOL: LAnguage MOd-eling for Lifelong Language Learning”. In: *International Conference on Learning Representations (ICLR)*. 2020.
- [788] Feng Sun et al. *APOLLO: Automated LLM and Lean Collaboration for Advanced Formal Reasoning*. arXiv preprint. 2024.
- [789] Haoran Sun and Shaoning Zeng. “Hierarchical Memory for High-Efficiency Long-Term Reasoning in LLM Agents”. In: *arXiv preprint arXiv:2507.22925* (2025).
- [790] Rich Sutton. *The Bitter Lesson*. 2019. URL: <http://www.incompleteideas.net/IncIdeas/BitterLesson.html> (visited on 01/15/2024).
- [791] Latanya Sweeney. “k-Anonymity: A Model for Protecting Privacy”. In: *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10.5 (2002), pp. 557–570.
- [792] Symphony Team. “Symphony: A Decentralized Multi-Agent Platform for Cooperative Problem Solving”. In: *arXiv preprint arXiv:2407.00000* (2024). WARNING: Placeholder arXiv ID; paper may not exist as described.
- [793] Christian Szegedy. “A Promising Path Towards Autoformalization and General Artificial Intelligence”. In: *arXiv preprint arXiv:2004.12868* (2020).
- [794] Christian Szegedy et al. “Rethinking the Inception Architecture for Computer Vision”. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 2818–2826.
- [795] Esteban G. Tabak and Eric Vanden-Eijnden. “Density Estimation by Dual Ascent of the Log-Likelihood”. In: *Communications in Mathematical Sciences* 8.1 (2010), pp. 217–233.
- [796] Juntao Tan et al. “IDGenRec: LLM-RecSys Alignment with Textual ID Learning”. In: *Proceedings of SIGIR*. Also on arXiv:2403.19021. 2024.
- [797] Mingxing Tan and Quoc V. Le. “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks”. In: *Proceedings of the International Conference on Machine Learning (ICML)*. 2019.
- [798] Mingxing Tan et al. “MnasNet: Platform-Aware Neural Architecture Search for Mobile”. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 2820–2828.
- [799] Hidenori Tanaka et al. “Pruning Neural Networks without Any Data by Iteratively Conserving Synaptic Flow”. In: *Advances in NeurIPS* (2020).
- [800] R. Tanaka et al. “VDocRAG: Retrieval-Augmented Generation over Visually-Rich Documents”. In: *Proceedings of CVPR*. Author list abbreviated. 2025.

- [801] Li Tang et al. "Towards Black-Box Membership Inference Attack for Diffusion Models". In: *arXiv preprint arXiv:2405.04311* (2024).
- [802] Luming Tang et al. "Emergent Correspondence from Image Diffusion". In: *Advances in Neural Information Processing Systems (NeurIPS)*. DIFT paper. 2023.
- [803] Zhiyuan Tang et al. "AttenMIA: LLM Membership Inference Attack through Attention Signals". In: *arXiv preprint arXiv:2410.03296* (2024).
- [804] Rohan Taori et al. "Stanford Alpaca: An Instruction-Following LLaMA Model". In: *GitHub repository* (2023).
- [805] Yi Tay et al. "Transformer Memory as a Differentiable Search Index". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2022.
- [806] Guy Tevet et al. "Human Motion Diffusion Model". In: *International Conference on Learning Representations* (2023).
- [807] Nandan Thakur et al. "BEIR: A Heterogeneous Benchmark for Zero-shot Evaluation of Information Retrieval Models". In: *NeurIPS Datasets and Benchmarks Track*. Also on arXiv:2104.08663. 2021.
- [808] The Coq Development Team. *The Coq Proof Assistant*. <https://coq.inria.fr>. 2024.
- [809] The mathlib Community. "The Lean Mathematical Library". In: *CPP 2020* (2020).
- [810] The mathlib Community. *The Lean Mathematical Library*. 2020.
- [811] The Sage Developers. *SageMath, the Sage Mathematics Software System*. <https://www.sagemath.org>. 2024.
- [812] Keyu Tian et al. "Visual Autoregressive Modeling: Scalable Image Generation via Next-Scale Prediction". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2024.
- [813] Linrui Tian et al. "EMO: Emote Portrait Alive – Generating Expressive Portrait Videos with Audio2Video Diffusion Model under Weak Conditions". In: *European Conference on Computer Vision* (2024).
- [814] Zeyue Tian et al. "VidMuse: A Simple Video-to-Music Generation Framework with Long-Short-Term Modeling". In: *Conference on Computer Vision and Pattern Recognition* (2025).
- [815] Naftali Tishby, Fernando C. Pereira, and William Bialek. "The Information Bottleneck Method". In: *arXiv preprint physics/0004057* (2000).
- [816] Susumu Tonegawa et al. "Memory Engram Storage and Retrieval". In: *Current Opinion in Neurobiology* 35 (2015), pp. 101–109. doi: [10.1016/j.conb.2015.07.009](https://doi.org/10.1016/j.conb.2015.07.009).
- [817] Giulio Tononi and Chiara Cirelli. "Sleep Function and Synaptic Homeostasis". In: *Sleep Medicine Reviews* 10.1 (2006), pp. 49–62.

- [818] Hugo Touvron et al. “LLaMA: Open and Efficient Foundation Language Models”. In: *arXiv preprint arXiv:2302.13971* (2023).
- [819] Du Tran et al. “A Closer Look at Spatiotemporal Convolutions for Action Recognition”. In: *Conference on Computer Vision and Pattern Recognition* (2018).
- [820] Khanh-Tung Tran et al. “Multi-Agent Collaboration Mechanisms: A Survey of LLMs”. In: *arXiv preprint arXiv:2501.06322* (2025).
- [821] Trieu H. Trinh et al. *Towards Autonomous Mathematics Research*. arXiv:2602.10177, Google DeepMind. 2026.
- [822] Trieu H. Trinh et al. “Solving Olympiad Geometry without Human Demonstrations”. In: *Nature* 625.7995 (2024), pp. 476–482.
- [823] Trieu H. Trinh et al. “Solving olympiad geometry without human demonstrations”. In: *Nature* 625 (2024), pp. 476–482.
- [824] Grigori S. Tseitin. “On the Complexity of Derivation in Propositional Calculus”. In: *Automation of Reasoning: 2. Classical Papers on Computational Logic 1967–1970*. Originally published in 1968. Springer, 1983, pp. 466–483.
- [825] George Tucker et al. “Doubly Reparameterized Gradient Estimators for Monte Carlo Objectives”. In: *International Conference on Learning Representations (ICLR)*. 2019.
- [826] Endel Tulving. “Episodic and Semantic Memory”. In: *Organization of Memory* (1972), pp. 381–403.
- [827] Alan M. Turing. “Computing Machinery and Intelligence”. In: *Mind* 59.236 (1950), pp. 433–460.
- [828] Alan M. Turing. “On Computable Numbers, with an Application to the Entscheidungsproblem”. In: *Proceedings of the London Mathematical Society* s2-42.1 (1936), pp. 230–265.
- [829] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. *Instance Normalization: The Missing Ingredient for Fast Stylization*. 2016. arXiv: [1607.08022](https://arxiv.org/abs/1607.08022) [cs.CV].
- [830] Thomas Unterthiner et al. “Towards Accurate Generative Models of Video: A New Metric and Challenges”. In: *arXiv preprint arXiv:1812.01717* (2019).
- [831] Josef Urban, Jiří Vyskoil, and Petr Štěpánek. “MaLAREa SG1 – Machine Learner for Automated Reasoning with Semantic Guidance”. In: *International Joint Conference on Automated Reasoning (IJCAR)*. Springer, 2011.
- [832] Benigno Uribe, Iain Murray, and Hugo Larochelle. “A Deep and Tractable Density Estimator”. In: *Proceedings of the 31st International Conference on Machine Learning (ICML)*. 2014.

- [833] Benigno Uribe et al. "Neural Autoregressive Distribution Estimation". In: *Journal of Machine Learning Research* 17.205 (2016), pp. 1–37.
- [834] Leigh Van Valen. "A New Evolutionary Law". In: *Evolutionary Theory* 1 (1973), pp. 1–30.
- [835] Robert J. Vanderbei. *Linear Programming: Foundations and Extensions*. Boston, MA: Kluwer Academic Publishers, 1998.
- [836] Leonid N. Vaserstein. "Markov processes over denumerable products of spaces, describing large systems of automata". In: *Problemy Peredachi Informatsii* 5.3 (1969), pp. 64–72.
- [837] Ashish Vaswani et al. "Attention is all you need". In: *Advances in neural information processing systems*. 2017, pp. 5998–6008.
- [838] Andreas Veit, Michael Wilber, and Serge Belongie. "Residual Networks Behave Like Ensembles of Relatively Shallow Networks". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2016.
- [839] William Vickrey. "Counterspeculation, Auctions, and Competitive Sealed Tenders". In: *The Journal of Finance* 16.1 (1961), pp. 8–37.
- [840] Cédric Villani. *Optimal Transport: Old and New*. Springer-Verlag, 2008.
- [841] Pascal Vincent. "A connection between score matching and denoising autoencoders". In: *Neural Computation* 23.7 (2011), pp. 1661–1674.
- [842] Jeffrey S Vitter. "Random Sampling with a Reservoir". In: *ACM Transactions on Mathematical Software* 11.1 (1985), pp. 37–57.
- [843] Paul Voigt and Axel von dem Bussche. *The EU General Data Protection Regulation (GDPR)*. 2017.
- [844] Ellen M. Voorhees. "TREC-8 Question Answering Track Report". In: *Proceedings of TREC-8*. MRR discussed as a scoring metric. 1999.
- [845] Martin J. Wainwright and Michael I. Jordan. "Graphical Models, Exponential Families, and Variational Inference". In: *Foundations and Trends in Machine Learning* 1.1–2 (2008), pp. 1–305.
- [846] Matthew Walker. *Why We Sleep: Unlocking the Power of Sleep and Dreams*. New York: Scribner, 2017.
- [847] Bram Wallace et al. "Diffusion Model Alignment Using Direct Preference Optimization". In: *Conference on Computer Vision and Pattern Recognition* (2024).
- [848] Wan Team. "Wan: Open and Advanced Large-Scale Video Generative Models". In: *arXiv preprint arXiv:2503.20314* (2025). WARNING: Duplicate of wang2025wan.

- [849] Ang Wang et al. “Wan: Open and Advanced Large-Scale Video Generative Models”. In: *arXiv preprint arXiv:2503.20314* (2025).
- [850] Avery Wang. “An Industrial-Strength Audio Search Algorithm”. In: *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*. 2003, pp. 7–13.
- [851] Ben Wang and Aran Komatsuzaki. “GPT-J-6B: A 6 Billion Parameter Autoregressive Language Model”. In: (2021). <https://github.com/kingoflolz/mesh-transformer-jax>.
- [852] Guan Wang et al. “Hierarchical Reasoning Model”. In: *arXiv preprint arXiv:2506.21734* (2025).
- [853] Hongyu Wang et al. *DeepNet: Scaling Transformers to 1,000 Layers*. 2022. arXiv: 2203.00555 [cs.LG].
- [854] Ji Wang et al. “Symphony: A Decentralized Multi-Agent Framework for Scalable Collective Intelligence”. In: *arXiv preprint arXiv:2508.20019* (2025).
- [855] Liang Wang et al. “Improving Text Embeddings with Large Language Models”. In: *arXiv preprint arXiv:2401.00368* (2024).
- [856] Liyuan Wang et al. “A Comprehensive Survey of Continual Learning: Theory, Method and Application”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2024).
- [857] Mingqiu Wang et al. *Retrieval Augmented End-to-End Spoken Dialog Models*. arXiv:2402.01828; ICASSP 2024. 2024.
- [858] Mingzhe Wang and Jia Deng. “Learning to Prove Theorems by Learning to Generate Theorems”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2020.
- [859] Mingzhe Wang et al. “Premise Selection for Theorem Proving by Deep Graph Embedding”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2017.
- [860] Ruida Wang et al. *TheoremLlama: Transforming General-Purpose LLMs into Lean4 Experts*. arXiv preprint. 2024.
- [861] Ting-Chun Wang et al. “High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2018.
- [862] Xuezhi Wang et al. “Self-Consistency Improves Chain of Thought Reasoning in Language Models”. In: *International Conference on Learning Representations* (2023).
- [863] Fu-Yun Wang et al. “AnimateLCM: Computation-Efficient Personalized Style Video Generation without Personalized Video Data”. In: *arXiv preprint arXiv:2402.00769* (2024).

-
- [864] Zhou Wang et al. “Image Quality Assessment: From Error Visibility to Structural Similarity”. In: *IEEE Transactions on Image Processing* 13.4 (2004), pp. 600–612.
 - [865] Zhouxia Wang et al. “MotionCtrl: A Unified and Flexible Motion Controller for Video Generation”. In: *arXiv preprint arXiv:2312.03641* (2024).
 - [866] Huawei Wei, Zejun Yang, and Zhisheng Wang. “AniPortrait: Audio-Driven Synthesis of Photorealistic Portrait Animation”. In: *arXiv preprint arXiv:2403.17694* (2024).
 - [867] Jason Wei et al. “Chain-of-Thought Prompting Elicits Reasoning in Large Language Models”. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 24824–24837.
 - [868] Christoph Weidenbach et al. “SPASS Version 2.0”. In: *International Conference on Automated Deduction (CADE)*. Springer, 2001.
 - [869] Jason Weston, Sumit Chopra, and Antoine Bordes. “Memory Networks”. In: *International Conference on Learning Representations*. 2015.
 - [870] Colin White et al. “Neural Architecture Search: Insights from 1000 Papers”. In: *arXiv preprint arXiv:2301.08727* (2023).
 - [871] Alfred North Whitehead and Bertrand Russell. *Principia Mathematica*. Cambridge University Press, 1910.
 - [872] Alfred North Whitehead and Bertrand Russell. *Principia Mathematica* (2nd ed.) 1925.
 - [873] Andrew Wiles. “Modular Elliptic Curves and Fermat’s Last Theorem”. In: *Annals of Mathematics* 141.3 (1995), pp. 443–551.
 - [874] Ronald J. Williams. “Simple Statistical Gradient-Following Algorithms for Connectionist Reinforcement Learning”. In: *Machine Learning* 8.3–4 (1992), pp. 229–256.
 - [875] Ronald J. Williams. “Simple Statistical Gradient-Following Algorithms for Connectionist Reinforcement Learning”. In: *Machine Learning* 8.3–4 (1992), pp. 229–256.
 - [876] Jay G. Wilpon et al. “Automatic recognition of keywords in unconstrained speech using hidden Markov models”. In: *unspecified* (1990). Exact venue/pages unspecified in this bib; see keyword-spotting sources.
 - [877] Robin Wilson. *Four Colors Suffice: How the Map Problem Was Solved*. Revised Color Edition. Princeton University Press, 2013.
 - [878] Robin Wilson. *Four Colors Suffice: How the Map Problem Was Solved*. Includes historical survey of sixty years of work on the four-color problem. Princeton University Press, 2014.

- [879] John Winn and Christopher M. Bishop. “Variational Message Passing”. In: *Journal of Machine Learning Research* 6 (2005), pp. 661–694.
- [880] John T Wixted and Ebbe B Ebbesen. “On the Form of Forgetting”. In: *Psychological Science* 2.6 (1991), pp. 409–415.
- [881] Stephen Wolfram. *The Mathematica Book*. 5th. Wolfram Media, 2003.
- [882] David H. Wolpert and William G. Macready. “No Free Lunch Theorems for Optimization”. In: *IEEE Transactions on Evolutionary Computation* 1.1 (1997), pp. 67–82.
- [883] David H. Wolpert and William G. Macready. “No Free Lunch Theorems for Optimization”. In: *IEEE Transactions on Evolutionary Computation* 1.1 (1997), pp. 67–82.
- [884] Michael Wooldridge and Nicholas R. Jennings. “Intelligent Agents: Theory and Practice”. In: *The Knowledge Engineering Review* 10.2 (1995), pp. 115–152.
- [885] Sewall Wright. “Correlation and Causation”. In: *Journal of Agricultural Research* 20.7 (1921), pp. 557–585.
- [886] Sewall Wright. “The Roles of Mutation, Inbreeding, Crossbreeding, and Selection in Evolution”. In: *Proceedings of the Sixth International Congress of Genetics* 1 (1932), pp. 356–366.
- [887] Bichen Wu et al. “FBNet: Hardware-Aware Efficient ConvNet Design via Differentiable Neural Architecture Search”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019.
- [888] Qingyun Wu et al. “AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation”. In: *arXiv preprint arXiv:2308.08155* (2023).
- [889] Ruiqi Wu et al. “LAMP: Learn A Motion Pattern for Few-Shot Video Generation”. In: *Conference on Computer Vision and Pattern Recognition* (2024).
- [890] Tao Wu et al. “Noise Aggregation Analysis Driven by Small-Noise Injection: Efficient Membership Inference for Diffusion Models”. In: *arXiv preprint arXiv:2408.10166* (2024).
- [891] Tianxing Wu et al. “FreeInit: Bridging Initialization Gap in Video Diffusion Models”. In: *Conference on Computer Vision and Pattern Recognition* (2024).
- [892] Yuhuai Wu et al. “Autoformalization with Large Language Models”. In: *Advances in Neural Information Processing Systems (NeurIPS)* (2022).
- [893] Yuhuai Wu et al. “Memorizing Transformers”. In: *Proceedings of ICLR* (2022).
- [894] Yusong Wu et al. *Large-scale Contrastive Language-Audio Pretraining with Feature Fusion and Keyword-to-Caption Augmentation*. arXiv:2211.06687. Often cited as ICASSP 2023 in downstream references. 2022.

-
- [895] Yuxin Wu and Kaiming He. “Group Normalization”. In: *European Conference on Computer Vision (ECCV)*. 2018, pp. 3–19.
- [896] Zhan Wu et al. “Membership Inference Attacks against Large Vision-Language Models”. In: *Advances in Neural Information Processing Systems*. 2024.
- [897] Andrea Wynn, Harsh Satija, and Gillian Hadfield. “Talk Isn’t Always Cheap: Understanding Failure Modes in Multi-Agent Debate”. In: *Proceedings of the International Conference on Machine Learning (ICML)*. 2025.
- [898] Zhiheng Xi et al. “AgentGym: Evolving Large Language Model-based Agents across Diverse Environments”. In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*. 2025.
- [899] Guangxuan Xiao et al. “Efficient Streaming Language Models with Attention Sinks”. In: *International Conference on Learning Representations*. 2024.
- [900] Saining Xie et al. “Rethinking Spatiotemporal Feature Learning: Speed-Accuracy Trade-offs in Video Classification”. In: *European Conference on Computer Vision* (2018).
- [901] Sang Michael Xie et al. “An Explanation of In-Context Learning as Implicit Bayesian Inference”. In: *Proceedings of the International Conference on Learning Representations (ICLR)*. 2022.
- [902] Zhenda Xie et al. *mHC: Manifold-Constrained Hyper-Connections*. 2025. arXiv: [2512.24880 \[cs.LG\]](#).
- [903] Huajian Xin et al. “DeepSeek-Prover-V1.5: Harnessing Proof Assistant Feedback for Reinforcement Learning and Monte-Carlo Tree Search”. In: *arXiv preprint arXiv:2408.08152* (2024).
- [904] Jinbo Xing et al. “Dynamicro: Animating Open-Domain Images with Video Diffusion Priors”. In: *European Conference on Computer Vision* (2024).
- [905] Ruibin Xiong et al. “On Layer Normalization in the Transformer Architecture”. In: *International Conference on Machine Learning (ICML)*. 2020, pp. 10524–10533.
- [906] Jiazheng Xu et al. “ImageReward: Learning and Evaluating Human Preferences for Text-to-Image Generation”. In: *Advances in Neural Information Processing Systems* 36 (2023).
- [907] Jun Xu et al. “MSR-VTT: A Large Video Description Dataset for Bridging Video and Language”. In: *Proceedings of CVPR*. 2016, pp. 5288–5296.
- [908] Lei Xu et al. “Modeling Tabular Data Using Conditional GAN”. In: *Advances in Neural Information Processing Systems* 32 (2019).
- [909] Lei Xu et al. “Modeling Tabular Data using Conditional GAN”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2019.

- [910] Yuhui Xu et al. “PC-DARTS: Partial Channel Connections for Memory-Efficient Architecture Search”. In: *International Conference on Learning Representations*. 2020.
- [911] Zhongcong Xu et al. “MagicAnimate: Temporally Consistent Human Image Animation using Diffusion Model”. In: *Conference on Computer Vision and Pattern Recognition* (2024).
- [912] Linting Xue et al. “ByT5: Towards a Token-Free Future with Pre-trained Byte-to-Byte Models”. In: *Transactions of the Association for Computational Linguistics* 10 (2022), pp. 291–306.
- [913] Wilson Yan et al. “VideoGPT: Video Generation using VQ-VAE and Transformers”. In: *arXiv preprint arXiv:2104.10157* (2021).
- [914] An Yang et al. “Qwen2.5-Math Technical Report: Toward Mathematical Expert Model via Self-Improvement”. In: *arXiv preprint arXiv:2409.12122* (2024).
- [915] Antoine Yang, Pedro M. Esperança, and Fabio M. Carlucci. “NAS Evaluation is Frustratingly Hard”. In: *International Conference on Learning Representations*. 2020.
- [916] Greg Yang et al. *Tensor Programs V: Tuning Large Neural Networks via Zero-Shot Hyperparameter Transfer*. 2022. arXiv: 2203.03466 [cs.LG].
- [917] Jianchao Yang et al. “Image Super-Resolution via Sparse Representation”. In: *IEEE Transactions on Image Processing* 19.11 (2010), pp. 2861–2873.
- [918] Kaiyu Yang et al. “LeanDojo-v2”. In: *arXiv preprint* (2024).
- [919] Kaiyu Yang and Jia Deng. “Learning to Prove Theorems via Interacting with Proof Assistants”. In: *International Conference on Machine Learning (ICML)*. 2019.
- [920] Kaiyu Yang et al. “LeanDojo: Theorem Proving with Retrieval-Augmented Language Models”. In: *Advances in Neural Information Processing Systems* 36 (2023).
- [921] Mengjiao Yang et al. “UniSim: Learning Interactive Real-World Simulators”. In: *International Conference on Learning Representations* (2024).
- [922] Shuai Yang et al. “Rerender A Video: Zero-Shot Text-Guided Video-to-Video Translation”. In: *ACM SIGGRAPH Asia* (2023).
- [923] Zhilin Yang et al. “XLNet: Generalized autoregressive pretraining for language understanding”. In: *Advances in Neural Information Processing Systems* 32 (2019).
- [924] Zhuoyi Yang et al. “CogVideoX: Text-to-Video Diffusion Models with An Expert Transformer”. In: *arXiv preprint arXiv:2408.06072* (2024).
- [925] Shunyu Yao et al. “ReAct: Synergizing Reasoning and Acting in Language Models”. In: *Proceedings of the International Conference on Learning Representations (ICLR)*. 2023.

- [926] Shunyu Yao et al. “Tree of Thoughts: Deliberate Problem Solving with Large Language Models”. In: *Advances in Neural Information Processing Systems* 36 (2023).
- [927] Shunyu Yao et al. “Tree of Thoughts: Deliberate Problem Solving with Large Language Models”. In: *Advances in Neural Information Processing Systems* 36 (2023).
- [928] Jonathan S Yedidia, William T Freeman, and Yair Weiss. “Understanding Belief Propagation and Its Generalizations”. In: *Exploring Artificial Intelligence in the New Millennium* 8 (2003), pp. 236–239.
- [929] Jonathan S. Yedidia, William T. Freeman, and Yair Weiss. “Constructing Free-Energy Approximations and Generalized Belief Propagation Algorithms”. In: *IEEE Transactions on Information Theory* 51.7 (2005), pp. 2282–2312.
- [930] Samuel Yeom et al. “Privacy Risk in Machine Learning: Analyzing the Connection to Overfitting”. In: *IEEE Computer Security Foundations Symposium (CSF)*. IEEE, 2018, pp. 268–282.
- [931] Jiaqi Yi et al. “MultiFCL: Multi-Scale Federated Continual Learning”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2025.
- [932] Xin Yi, Ekta Walia, and Paul Babyn. “Generative adversarial network in medical imaging: A review”. In: *Medical Image Analysis* 58 (2019), p. 101552.
- [933] Shengming Yin et al. “DragNUWA: Fine-Grained Control in Video Generation by Integrating Text, Image, and Trajectory”. In: *arXiv preprint arXiv:2308.08089* (2023).
- [934] Shengming Yin et al. “NUWA-XL: Diffusion over Diffusion for eXtremely Long Video Generation”. In: *Association for Computational Linguistics* (2023).
- [935] Chris Ying et al. “NAS-Bench-101: Towards Reproducible Neural Architecture Search”. In: *International Conference on Machine Learning*. 2019, pp. 7105–7114.
- [936] Huaiyuan Ying et al. “InternLM-Math: Open Math Large Language Models Toward Verifiable Reasoning”. In: *arXiv preprint arXiv:2402.06332* (2024).
- [937] Da Yu et al. “Differentially Private Fine-tuning of Language Models”. In: *International Conference on Learning Representations*. 2022.
- [938] Hao Yu et al. *Evaluation of Retrieval-Augmented Generation: A Survey*. arXiv:2405.07437. 2024.
- [939] Kaicheng Yu et al. “Evaluating the Search Phase of Neural Architecture Search”. In: *Proceedings of ICLR* (2020).
- [940] Lijun Yu et al. “Language Model Beats Diffusion – Tokenizer is Key to Visual Generation”. In: *International Conference on Learning Representations* (2024).
- [941] Lijun Yu et al. “Language Model Beats Diffusion – Tokenizer is Key to Visual Generation”. In: *International Conference on Learning Representations* (2024).

- [942] Lijun Yu et al. “MAGVIT: Masked Generative Video Transformer”. In: *Conference on Computer Vision and Pattern Recognition* (2023).
- [943] Qiying Yu et al. “DAPO: An Open-Source LLM Reinforcement Learning System at Scale”. In: *arXiv preprint arXiv:2503.14476* (2025).
- [944] Weihao Yu et al. “MetaFormer Baselines for Vision”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46.2 (2024), pp. 896–912.
- [945] Yi Yuan et al. “Retrieval-Augmented Text-to-Audio Generation”. In: *Proceedings of ICASSP*. Also on arXiv:2309.08051. 2024.
- [946] Sangdoo Yun et al. “CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features”. In: *IEEE International Conference on Computer Vision (ICCV)*. 2019, pp. 6023–6032.
- [947] Manzil Zaheer et al. “BigBird: Transformers for Longer Sequences”. In: *Advances in Neural Information Processing Systems*. Vol. 33. 2020.
- [948] Neil Zeghidour et al. “SoundStream: An End-to-End Neural Audio Codec”. In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 30 (2021), pp. 495–507.
- [949] Neil Zeghidour et al. “SoundStream: An End-to-End Neural Audio Codec”. In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 30 (2022), pp. 495–507.
- [950] Eric Zelikman et al. “STaR: Bootstrapping Reasoning With Reasoning”. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 15476–15488.
- [951] Mingliang Zeng et al. “MusicBERT: Symbolic Music Understanding with Large-Scale Pre-Training”. In: *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*. 2021, pp. 791–800.
- [952] Friedemann Zenke, Ben Poole, and Surya Ganguli. “Continual Learning Through Synaptic Intelligence”. In: *Proceedings of the International Conference on Machine Learning (ICML)*. 2017.
- [953] Wenbo Zhai et al. “Noise as a Probe: Membership Inference Attacks on Diffusion Models Leveraging Initial Noise”. In: *arXiv preprint arXiv:2410.00909* (2024).
- [954] Xiaohua Zhai et al. “Sigmoid Loss for Language Image Pre-Training”. In: *Proceedings of ICCV*. 2023.
- [955] Biao Zhang and Rico Sennrich. “Root Mean Square Layer Normalization”. In: *Advances in Neural Information Processing Systems*. Vol. 32. 2019.
- [956] Hongyi Zhang, Yann N. Dauphin, and Tengyu Ma. “Fixup Initialization: Residual Learning Without Normalization”. In: *International Conference on Learning Representations (ICLR)*. 2019.

-
- [957] Hongyi Zhang et al. “mixup: Beyond Empirical Risk Minimization”. In: *International Conference on Learning Representations (ICLR)*. 2018.
- [958] Jun Zhang et al. “PrivBayes: Private Data Release via Bayesian Networks”. In: *ACM Transactions on Database Systems* 42.4 (2017), pp. 1–41.
- [959] Kai Zhang, Masahiro Mei, and Zhengli Zhu. “Sampling-based Membership Inference Attack for Autoregressive Language Models”. In: *arXiv preprint arXiv:2404.09212* (2024).
- [960] Kai Zhang et al. “Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising”. In: *IEEE Transactions on Image Processing* 26.7 (2017), pp. 3142–3155.
- [961] Linfeng Zhang et al. “Be Your Own Teacher: Improve the Performance of Convolutional Neural Networks via Self Distillation”. In: *IEEE International Conference on Computer Vision (ICCV)*. 2019, pp. 3713–3722.
- [962] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. “Adding Conditional Control to Text-to-Image Diffusion Models”. In: *International Conference on Computer Vision* (2023).
- [963] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. “Adding Conditional Control to Text-to-Image Diffusion Models”. In: *International Conference on Computer Vision* (2023).
- [964] Shiwei Zhang et al. “I2VGen-XL: High-Quality Image-to-Video Synthesis via Cascaded Diffusion Models”. In: *arXiv preprint arXiv:2311.04145* (2023).
- [965] Wentian Zhang et al. “Cross-Attention Makes Inference Cumbersome in Text-to-Image Diffusion Models”. In: *arXiv preprint arXiv:2404.02747* (2024).
- [966] Yadong Zhang et al. “LLM as a Mastermind: A Survey of Strategic Reasoning with Large Language Models”. In: *arXiv preprint arXiv:2404.01230* (2024).
- [967] Yao Zhang et al. “SwarmAgentic: Towards Fully Automated Agentic System Generation via Swarm Intelligence”. In: *arXiv preprint arXiv:2506.15672* (2025). WARNING: Duplicate of yao2025swarmagentic.
- [968] Yifan Zhang et al. “SwarmAgentic: Towards Fully Automated Agentic System Generation via Swarm Intelligence”. In: *arXiv preprint arXiv:2506.15672* (2025). WARNING: Duplicate of zhang2025swarmagentic.
- [969] Jun Zhao et al. “LongAgent: Scaling Language Models to 128k Context through Multi-Agent Collaboration”. In: *arXiv preprint arXiv:2402.11550* (2024).
- [970] Xuanlei Zhao, Zangwei Zheng, and Yang You. “PAB: Pyramid Attention Broadcast for Efficient Video Generation”. In: *arXiv preprint arXiv:2408.12588* (2024).

- [971] Zilong Zhao, Robert Birke, and Lydia Y. Chen. “TabuLa: Harnessing Language Models for Tabular Data Synthesis”. In: *NeurIPS Workshop on Table Representation Learning*. 2023.
- [972] Zilong Zhao et al. “CTAB-GAN: Effective Table Data Synthesizing”. In: *Asian Conference on Machine Learning*. 2021.
- [973] Zilong Zhao et al. “CTAB-GAN+: Enhancing Tabular Data Synthesis”. In: *Frontiers in Big Data*. 2023.
- [974] Zilong Zhao et al. “CTAB-GAN+: Enhancing Tabular Data Synthesis”. In: *Frontiers in Big Data* 6 (2023).
- [975] Zilong Zhao et al. “CTAB-GAN+: Enhancing Tabular Data Synthesis”. In: *Frontiers in Big Data* 7 (2024).
- [976] Chuanyang Zheng et al. “Thought Communication in Multiagent Collaboration”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2025.
- [977] Kunhao Zheng, Jesse Michael Han, and Stanislas Polu. “miniF2F: a cross-system benchmark for formal Olympiad-level mathematics”. In: *ICLR* (2022).
- [978] Zangwei Zheng et al. “Open-Sora: Democratizing Efficient Video Production for All”. In: *arXiv preprint arXiv:2412.00131* (2024).
- [979] Zangwei Zheng et al. “Open-Sora: Democratizing Efficient Video Production for All”. In: *arXiv preprint arXiv:2412.00131* (2024).
- [980] Zangwei Zheng et al. “Open-Sora: Democratizing Efficient Video Production for All”. In: *arXiv preprint arXiv:2412.00131* (2024).
- [981] Wanjun Zhong et al. “MemoryBank: Enhancing Large Language Models with Long-Term Memory”. In: *Proceedings of AAAI* (2024).
- [982] Defa Zhu et al. *Hyper-Connections*. 2024. arXiv: [2409.19606](https://arxiv.org/abs/2409.19606) [cs.LG].
- [983] Jun-Yan Zhu et al. “Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks”. In: *IEEE International Conference on Computer Vision*. 2017, pp. 2223–2232.
- [984] Kunlun Zhu et al. “MultiAgentBench: Evaluating the Collaboration and Competition of LLM Agents”. In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*. 2025.
- [985] Shengyao Zhuang et al. *A Setwise Approach for Effective and Highly Efficient Zero-shot Ranking with Large Language Models*. SIGIR 2024 (also arXiv:2310.09497). Conference metadata may be cited via arXiv when ACM access is restricted. 2024.
- [986] Daniel M Ziegler et al. “Fine-Tuning Language Models from Human Preferences”. In: *arXiv preprint arXiv:1909.08593* (2019).

- [987] Barret Zoph and Quoc V. Le. “Neural Architecture Search with Reinforcement Learning”. In: *Proceedings of ICLR*. 2017.
- [988] Barret Zoph and Quoc V. Le. “Neural Architecture Search with Reinforcement Learning”. In: *International Conference on Learning Representations*. 2017.
- [989] Barret Zoph et al. “Learning Transferable Architectures for Scalable Image Recognition”. In: *Proceedings of CVPR*. 2018, pp. 8697–8710.
- [990] Barret Zoph et al. “Learning Transferable Architectures for Scalable Image Recognition”. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2018, pp. 8697–8710.

