2025 First Quarter Reading List

April 27, 2025

1 Reading List

Paper (Author, Year)	**Area**	**Date Read**	**Notes**
Greedy layer-wise training of deep networks, Bengio et al. [2006]	Neural Network Initialization and Deep Learning	2025 - January	
Reinforcement learning: A survey, Kaelbling et al. [1996]	Reinforcement Learning	2025 - January	
Deep Neural Network Model Forecasting for Financial and Economic Market, Chen [2022]	Time Series Forecasting and Deep Learning	2025 - January	
Organoid intelligence (OI): the new frontier in biocomputing and intelligence-in-a-dish, Smirnova et al. [2023]	Orgnoid Intelligence	2025 - January	
Organoid intelligence: integration of organoid technology and artificial intelligence in the new era of in vitro models, Shi et al. [2024]	Orgnoid Intelligence	2025 - January	
Mining the contribution of intensive care clinical course to outcome after traumatic brain injury, Bhattacharyay et al. [2023]	Deep Learning and Representation Learning	2025 - January	

Paper (Author, Year)	**Area**	**Date Read**	**Notes**
Chaotic Recurrent Neural Networks for Financial Forecast, Wang and Lee [2021]	Time Series Forecasting and Deep Learning	2025 - January	
Enabling spike-based backpropagation for training deep neural network architectures, Lee et al. [2020]	Spiked Neural Networks	2025 - January	
Reinforcement learning for portfolio management, Filos [2019]	Reinforcement Learning	2025 - January	
Using structured events to predict stock price movement: An empirical investigation, Ding et al. [2014]	Time Series Forecasting	2025 - January	
A deep reinforcement learning framework for the financial portfolio management problem, Jiang et al. [2017]	Deep Learning and Reinforecment Learning	2025 - January	
Revolutionising Financial Portfolio Management: The Non-Stationary Transformer's Fusion of Macroeconomic Indicators and Sentiment Analysis in a Deep Reinforcement Learning Framework, Liu et al. [2023]	Reinforcement Learning, Deep Learning and Sentiment Analysis	2025 - January	
Deep reinforcement learning approach for trading automation in the stock market, Kabbani and Duman [2022]	Algorithmic Trading, Deep Learning and Reinforcement Learning	2025 - January	

Paper (Author, Year)	**Area**	**Date Read**	**Notes**
Are Time Series Foundation Models Ready to Revolutionize Predictive Building Analytics?, Mulayim et al. [2024]	Time Series Foundation Models	2025 - January	
Large Concept Models: Language Modeling in a Sentence Representation Space, Barrault et al. [2024]	Large Concept Models and Representation Learning	2025 - January	
LETS-C: Leveraging Language Embedding for Time Series Classification, Kaur et al. [2024]	Time Series Classification	2025- January	
Large Language Models Are Zero-Shot Time Series Forecasters, Gruver et al. [2023]	LLM Time Series Forecasting	2025- January	
Foundation models for time series analysis: A tutorial and survey, Liang et al. [2024]	Time Series Foundation Models	2025 - January	
Evaluating Large Language Models on Time Series Feature Understanding: A Comprehensive Taxonomy and Benchmark, Fons et al. [2024]	LLM Time Series Analysis	2025 - January	
Matryoshka representation learning, Kusupati et al. [2022]	Representation Learning	2025 - January	
Signal2vec: Time series embedding representation, Nalmpantis and Vrakas [2019]	Represenation Learning	2025 - February	

Paper (Author, Year)	**Area**	**Date Read**	**Notes**
Neural basis expansion analysis with exogenous variables: Forecasting electricity prices with NBEATSx, Olivares et al. [2023]	Time Series Foundation Model	2025 - February	
Multiple-Resolution Tokenization for Time Series Forecasting with an Application to Pricing, Peršak et al. [2024]	Time Series Foundation Model	2025 - February	
Greed is all you need: An evaluation of tokenizer inference methods, Uzan et al. [2024]	Inference Methods	2025 - February	
DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning, Guo et al. [2025]	Reasoning Capability Enhancement, Reinforcement Learning in LLMs, Efficient Model Alignment.	2025 - February	This paper introduces DeepSeek-R1, a reasoning LLM trained via large-scale reinforcement learning (RL). DeepSeek-R1 combines cold start data and multi stage training to refine reasoning quality, achieving performance comparable to o1. It Leverages GRPO (Generalized Relative Policy Optimization), an RL algorithm improving upon PPO (Proximal Plociy Optimisation) by addressing reward hacking via KL divergence penalties, enhancing alignment with human reasoning patterns. However it encounters language mixing challenges.
Deepseekmath: Pushing the limits of mathematical reasoning in open language models, Shao et al. [2024]	Domain Specific LLMs and Reinforcement Learning in LLMs.	2025 - February	This paper introduces GRPO (Generalized Relative Policy Optimization), showing its performance in a math domain specific model. It matches GPT-4's performance on several academic benchmarks such as GSM8K and MATH.

Paper (Author, Year)	**Area**	**Date Read**	**Notes**
TOTEM: TOkenized Time Series EMbeddings, Talukder et al. [2024]	Time Series Foundation Model and Representation Learning	2025 - February	
Neural discrete representation learning, Van Den Oord et al. [2017]	Representation Learning	2025 - February	
Reconstructing Training Data from Real-World Models Trained with Transfer Learning, Oz et al. [2024]	Reconstruction Learning, Data Privacy in Machine Learning, and Transfer Learning Vulnerabilities.	2025 - February	This paper proposes a novel model inversion framework to reconstruct training data from models trained via transfer learning, even when only fine-tuned weights are accessible. They leverages gradient matching and feature-space priors to bypass the "domain shift" challenge in transfer learning, outperforming prior inversion methods by 2.4× on CIFAR-10 and MNIST benchmarks. They demonstrates attacks on ResNet and ViT architectures, exposing sensitive attributes, like faces and text, from models deployed in production environments .
Empowering time series analysis with large language models: A survey, Jiang et al. [2024]	LLM Time Series Analysis	2025 - February	
Scalable Numerical Embeddings for Multivariate Time Series: Enhancing Healthcare Data Representation Learning, Huang et al. [2024]	Representation Learning	2025 - February	

Paper (Author, Year)	**Area**	**Date Read**	**Notes**
Microscaling Data Formats for Deep Learning, Rouhani et al. [2023]	Deep Learning Optimization and Low-Precision Training.	2025 - February	This paper introduces microscaling, a novel data format framework designed to reduce memory and computational overhead in deep learning workflows. They combine low-precisions (4-bit) representations with dynamic scaling factors, achieving 1.8× training speedusp and 3.1× memory reductions compared to FP16/FP32 baselines while retaining model accuracy. They optimize tensor layouts for GPU memory hierarchies, improving cache utilization and reducing data movement by 40% in transformer-based models. The work advances low-precision training by balancing numerical stability and efficiency, enabling scalable training of billion parameter models on standard hardware.
Stop looking for important tokens in multimodal language models: Duplication matters more, Wen et al. [2025]	Multimodal Language Models, Optimized Token Reduction	2025 - February	Introduces the Duplication Aware Reduction of Tokens Algorithm (Dart), which removes redundant tokens based on duplication rather than perceived importance. It increases inference time by 1.99x in the tested LLMs, by selecting "Pivot" tokens and pruning tokens based on cosine similarity - leading to minimal information loss.

Paper (Author, Year)	**Area**	**Date Read**	**Notes**
ClusterKV: Manipulating LLM KV cache in semantic space for recallable compression, Liu et al. [2024]	Long Context inference, KV cache Optimization, and Memory Efficient LLM Deployment.	2025 - February	This paper addresses the dynamic importance of tokens during inference, and proposes a new clustering based selection to avoid "internal fragmentation of important tokens", whilst recognizing the slightly higher clustering overhead than page representations seen in existing methods. ClusterKV implements efficient CUDA kernels; processing the KV (key-value) heads as threads in parallel, leading to negligible accuracy losses and a 2x speed up in latency compared to a full KV cache method.
Janus: Decoupling visual encoding for unified multimodal understanding and generation, Wu et al. [2024]	Multimodal Unified Architectures, Visual Task Decoupling, and Efficient Multitask Learning.	2025 - February	This paper addresses the conflicting demands of visual encoding in multimodal models, proposing decoupled visual encoders to separate features for understanding and generation (e.g., text-to-image), akin to a lightweight mixture-of-experts (MoE) design. By decoupling visual features for distinct tasks (e.g., VQA vs. text-to-image), Janus avoids interference while retaining shared backbone efficiency, leading to SOTA performances on 12 benchmarks with 2.1× faster inference vs. dual-model approaches. Despite slight parameter overhead, its CUDA-optimized gating minimizes latency, outperforming naive MoE implementations.
Large language models understand and can be enhanced by emotional stimuli, Li et al. [2023]	Prompt Engineering	2025 - February	

Paper (Author, Year)	**Area**	**Date Read**	**Notes**
Tree of thoughts: Deliberate problem solving with large language models, Yao et al. [2023]	Reasoning in LLMs, Heuristic Search Algorithms, Prompt Engineering	2025 - February	
Chain-of-table: Evolving tables in the reasoning chain for table understanding, Wang et al. [2024]	Reasoning in LLMs, Heuristic Search Algorithms, Prompt Engineering	2025 - February	
A comprehensive survey of retrieval-augmented generation (rag): Evolution, current landscape and future directions, Gupta et al. [2024]	Retrieval-Augmented Generation	2025 - February	
Retrieval-augmented generation for large language models: A survey, Gao et al. [2023]	Retrieval-Augmented Generation	2025 - February	
A systematic survey of prompt engineering in large language models: Techniques and applications, Sahoo et al. [2024]	Prompt Engineering	2025 - March	
Deep learning for time series classification: a review, Ismail Fawaz et al. [2019]	Time Series Classification and Deep Learning	2025 - March	

Paper (Author, Year)	**Area**	**Date Read**	**Notes**
Transformers without Normalization, Zhu et al. [2025]	Activation Functions and Transformers	2025 - March	This paper introduces the Dynamic Tanh (DyT) normalization technique, providng an alternative to LayerNorm and RMSNorm. DyT eliminates the computational overhead of normalization layers, achieving 2.99× speedup in prefilling latency and 1.99× faster total inference in LLMs like LLaMA, with negligible accuracy loss, however it requires careful initialization in LLMs. The work redefines normalization's role as a learnable activation squasher rather than a statistical stabilizer, opening pathways for leaner architectures.
VOLGAN: A Generative Model for Arbitrage-Free Implied Volatility Surfaces, Vuletić and Cont [2025]	Generative Adversarial Networks	2025 - March	
Mana-net: Mitigating aggregated sentiment homogenization with news weighting for enhanced market prediction, Wang and Ma [2024]	Sentiment Weighting, Split Attention and Deep Learning	2025 - March	
Resnest: Split-attention networks, Zhang et al. [2022]	Split Attention and Deep Learning	2025 - March	
How much position information do convolutional neural networks encode?, Islam et al. [2020]	Convolutional Neural Networks and Deep Learning	2025 - March	

Paper (Author, Year)	**Area**	**Date Read**	**Notes**
An adaptive embedding procedure for time series forecasting with deep neural networks, Succetti et al. [2023]	Adaptive Embeddings and Deep Learning	2025 - March	
Train short, test long: Attention with linear biases enables input length extrapolation, Press et al. [2021]	Positional Embeddings	2025 - March	
Improving position encoding of transformers for multivariate time series classification, Foumani et al. [2024]	Positional Embeddings and Time Series Classification	2025 - March	
MTEB: Massive text embedding benchmark, Muennighoff et al. [2022]	Text Embeddings	2025 - March	
Advancing time series classification with multimodal language modeling, Cheng et al. [2024]	Text Embeddings	2025 - March	
Goat: Fine-tuned llama outperforms gpt-4 on arithmetic tasks, Liu and Low [2023]	Fine-tuned LLM and Arithmetic LLM	2025 - March	
Intriguing Properties of Positional Encoding in Time Series Forecasting, Zhang et al. [2024]	Positional Encoding and Time Series Classification	2025 - March	

Reading Statistics

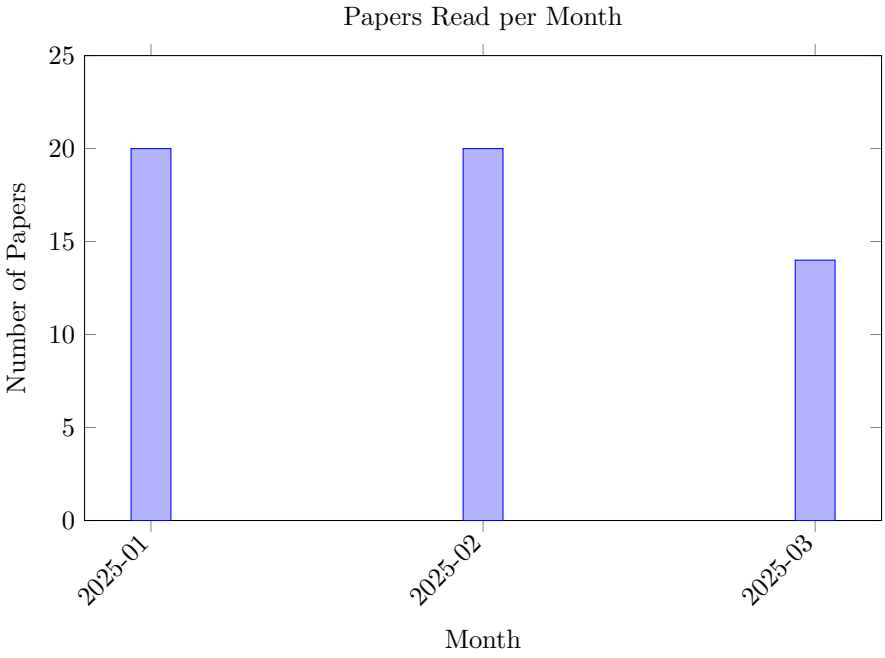

Figure 1: Monthly reading progress visualization based on completed papers

References

Loïc Barrault, Paul-Ambroise Duquenne, Maha Elbayad, Artyom Kozhevnikov, Belen Alastruey, Pierre Andrews, Mariano Coria, Guillaume Couairon, Marta R Costa-jussà, David Dale, et al. Large concept models: Language modeling in a sentence representation space. *arXiv preprint arXiv:2412.08821*, 2024.

Yoshua Bengio, Pascal Lamblin, Dan Popovici, and Hugo Larochelle. Greedy layer-wise training of deep networks. *Advances in neural information processing systems*, 19, 2006.

Shubhayu Bhattacharyay, Pier Francesco Caruso, Cecilia Åkerlund, Lindsay Wilson, Robert D Stevens, David K Menon, Ewout W Steyerberg, David W Nelson, Ari Ercole, CENTER-TBI investigators, and participants. Mining the contribution of intensive care clinical course to outcome after traumatic brain injury. *npj Digital Medicine*, 6(1):154, 2023.

Fan Chen. Deep neural network model forecasting for financial and economic market. *Journal of Mathematics*, 2022:1–10, 03 2022. doi: 10.1155/2022/8146555.

Mingyue Cheng, Yiheng Chen, Qi Liu, Zhiding Liu, and Yucong Luo. Advancing

time series classification with multimodal language modeling. *arXiv preprint arXiv:2403.12371*, 2024.

Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. Using structured events to predict stock price movement: An empirical investigation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1415–1425, 2014.

Angelos Filos. Reinforcement learning for portfolio management. *arXiv preprint arXiv:1909.09571*, 2019.

Elizabeth Fons, Rachneet Kaur, Soham Palande, Zhen Zeng, Tucker Balch, Manuela Veloso, and Svitlana Vyetrenko. Evaluating large language models on time series feature understanding: A comprehensive taxonomy and benchmark. *arXiv preprint arXiv:2404.16563*, 2024.

Navid Mohammadi Foumani, Chang Wei Tan, Geoffrey I Webb, and Mahsa Salehi. Improving position encoding of transformers for multivariate time series classification. *Data mining and knowledge discovery*, 38(1):22–48, 2024.

Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Haofen Wang, and Haofen Wang. Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*, 2, 2023.

Nate Gruver, Marc Finzi, Shikai Qiu, and Andrew G Wilson. Large language models are zero-shot time series forecasters. *Advances in Neural Information Processing Systems*, 36:19622–19635, 2023.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

Shailja Gupta, Rajesh Ranjan, and Surya Narayan Singh. A comprehensive survey of retrieval-augmented generation (rag): Evolution, current landscape and future directions. *arXiv preprint arXiv:2410.12837*, 2024.

Chun-Kai Huang, Yi-Hsien Hsieh, Ta-Jung Chien, Li-Cheng Chien, Shao-Hua Sun, Tung-Hung Su, Jia-Horng Kao, and Che Lin. Scalable numerical embeddings for multivariate time series: Enhancing healthcare data representation learning. *arXiv preprint arXiv:2405.16557*, 2024.

Md Amirul Islam, Sen Jia, and Neil DB Bruce. How much position information do convolutional neural networks encode? *arXiv preprint arXiv:2001.08248*, 2020.

Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. Deep learning for time series classification: a review. *Data mining and knowledge discovery*, 33(4):917–963, 2019.

Yushan Jiang, Zijie Pan, Xikun Zhang, Sahil Garg, Anderson Schneider, Yuriy Nevmyvaka, and Dongjin Song. Empowering time series analysis with large language models: A survey. *arXiv preprint arXiv:2402.03182*, 2024.

Zhengyao Jiang, Dixing Xu, and Jinjun Liang. A deep reinforcement learning framework for the financial portfolio management problem. *arXiv preprint arXiv:1706.10059*, 2017.

Taylan Kabbani and Ekrem Duman. Deep reinforcement learning approach for trading automation in the stock market. *IEEE Access*, 10:93564–93574, 2022.

Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore. Reinforcement learning: A survey. *Journal of artificial intelligence research*, 4:237–285, 1996.

Rachneet Kaur, Zhen Zeng, Tucker Balch, and Manuela Veloso. Lets-c: Leveraging language embedding for time series classification. *arXiv preprint arXiv:2407.06533*, 2024.

Aditya Kusupati, Gantavya Bhatt, Aniket Rege, Matthew Wallingford, Aditya Sinha, Vivek Ramanujan, William Howard-Snyder, Kaifeng Chen, Sham Kakade, Prateek Jain, et al. Matryoshka representation learning. *Advances in Neural Information Processing Systems*, 35:30233–30249, 2022.

Chankyu Lee, Syed Shakib Sarwar, Priyadarshini Panda, Gopalakrishnan Srinivasan, and Kaushik Roy. Enabling spike-based backpropagation for training deep neural network architectures. *Frontiers in neuroscience*, 14:497482, 2020.

Cheng Li, Jindong Wang, Yixuan Zhang, Kaijie Zhu, Wenxin Hou, Jianxun Lian, Fang Luo, Qiang Yang, and Xing Xie. Large language models understand and can be enhanced by emotional stimuli. *arXiv preprint arXiv:2307.11760*, 2023.

Yuxuan Liang, Haomin Wen, Yuqi Nie, Yushan Jiang, Ming Jin, Dongjin Song, Shirui Pan, and Qingsong Wen. Foundation models for time series analysis: A tutorial and survey. In *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining*, pages 6555–6565, 2024.

Guangda Liu, Chengwei Li, Jieru Zhao, Chenqi Zhang, and Minyi Guo. Clusterkv: Manipulating llm kv cache in semantic space for recallable compression. *arXiv preprint arXiv:2412.03213*, 2024.

Tiedong Liu and Bryan Kian Hsiang Low. Goat: Fine-tuned llama outperforms gpt-4 on arithmetic tasks. *arXiv preprint arXiv:2305.14201*, 2023.

Yuchen Liu, Daniil Mikriukov, Owen Christopher Tjahyadi, Gangmin Li, Terry R Payne, Yong Yue, Kamran Siddique, and Ka Lok Man. Revolutionising financial portfolio management: The non-stationary transformer's fusion of macroeconomic indicators and sentiment analysis in a deep reinforcement learning framework. *Applied Sciences*, 14(1):274, 2023.

Niklas Muennighoff, Nouamane Tazi, Loïc Magne, and Nils Reimers. Mteb: Massive text embedding benchmark. *arXiv preprint arXiv:2210.07316*, 2022.

Ozan Baris Mulayim, Pengrui Quan, Liying Han, Xiaomin Ouyang, Dezhi Hong, Mario Bergés, and Mani Srivastava. Are time series foundation models ready to revolutionize predictive building analytics? In *Proceedings of the 11th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, pages 169–173, 2024.

Christoforos Nalmpantis and Dimitris Vrakas. Signal2vec: Time series embedding representation. In *International conference on engineering applications of neural networks*, pages 80–90. Springer, 2019.

Kin G Olivares, Cristian Challu, Grzegorz Marcjasz, Rafał Weron, and Artur Dubrawski. Neural basis expansion analysis with exogenous variables: Forecasting electricity prices with nbeatsx. *International Journal of Forecasting*, 39(2):884–900, 2023.

Yakir Oz, Gilad Yehudai, Gal Vardi, Itai Antebi, Michal Irani, and Niv Haim. Reconstructing training data from real world models trained with transfer learning. *arXiv preprint arXiv:2407.15845*, 2024.

Egon Peršak, Miguel F Anjos, Sebastian Lautz, and Aleksandar Kolev. Multiple-resolution tokenization for time series forecasting with an application to pricing. *arXiv preprint arXiv:2407.03185*, 2024.

Ofir Press, Noah A Smith, and Mike Lewis. Train short, test long: Attention with linear biases enables input length extrapolation. *arXiv preprint arXiv:2108.12409*, 2021.

Bita Darvish Rouhani, Ritchie Zhao, Ankit More, Mathew Hall, Alireza Khodamoradi, Summer Deng, Dhruv Choudhary, Marius Cornea, Eric Dellinger, Kristof Denolf, et al. Microscaling data formats for deep learning. *arXiv preprint arXiv:2310.10537*, 2023.

Pranab Sahoo, Ayush Kumar Singh, Sriparna Saha, Vinija Jain, Samrat Mondal, and Aman Chadha. A systematic survey of prompt engineering in large language models: Techniques and applications. *arXiv preprint arXiv:2402.07927*, 2024.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Y Wu, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

Huaiyu Shi, Andrew Kowalczewski, Danny Vu, Xiyuan Liu, Asif Salekin, Huaxiao Yang, and Zhen Ma. Organoid intelligence: integration of organoid technology and artificial intelligence in the new era of in vitro models. *Medicine in novel technology and devices*, 21:100276, 2024.

Lena Smirnova, Brian S Caffo, David H Gracias, Qi Huang, Itzy E Morales Pantoja, Bohao Tang, Donald J Zack, Cynthia A Berlinicke, J Lomax Boyd, Timothy D Harris, et al. Organoid intelligence (oi): the new frontier in biocomputing and intelligence-in-a-dish. *Frontiers in Science*, 1:1017235, 2023.

Federico Succetti, Antonello Rosato, and Massimo Panella. An adaptive embedding procedure for time series forecasting with deep neural networks. *Neural Networks*, 167:715–729, 2023. ISSN 0893-6080. doi: https://doi.org/10.1016/j.neunet.2023.08.051. URL `https://www.sciencedirect.com/science/article/pii/S0893608023004719`.

Sabera Talukder, Yisong Yue, and Georgia Gkioxari. Totem: Tokenized time series embeddings for general time series analysis. *arXiv preprint arXiv:2402.16412*, 2024.

Omri Uzan, Craig W Schmidt, Chris Tanner, and Yuval Pinter. Greed is all you need: An evaluation of tokenizer inference methods. *arXiv preprint arXiv:2403.01289*, 2024.

Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *Advances in neural information processing systems*, 30, 2017.

Milena Vuletić and Rama Cont. Volgan:: A generative model for arbitrage-free implied volatility surfaces. *Applied Mathematical Finance*, pages 1–36, 2025.

Jeff Wang and Raymond Lee. Chaotic recurrent neural networks for financial forecast. *American Journal of Neural Networks and Applications*, 7:7, 02 2021. doi: 10.11648/j.ajnna.20210701.12.

Mengyu Wang and Tiejun Ma. Mana-net: Mitigating aggregated sentiment homogenization with news weighting for enhanced market prediction. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pages 2379–2389, 2024.

Zilong Wang, Hao Zhang, Chun-Liang Li, Julian Martin Eisenschlos, Vincent Perot, Zifeng Wang, Lesly Miculicich, Yasuhisa Fujii, Jingbo Shang, Chen-Yu Lee, et al. Chain-of-table: Evolving tables in the reasoning chain for table understanding. *arXiv preprint arXiv:2401.04398*, 2024.

Zichen Wen, Yifeng Gao, Shaobo Wang, Junyuan Zhang, Qintong Zhang, Weijia Li, Conghui He, and Linfeng Zhang. Stop looking for important tokens in multimodal language models: Duplication matters more. *arXiv preprint arXiv:2502.11494*, 2025.

Chengyue Wu, Xiaokang Chen, Zhiyu Wu, Yiyang Ma, Xingchao Liu, Zizheng Pan, Wen Liu, Zhenda Xie, Xingkai Yu, Chong Ruan, et al. Janus: Decoupling visual encoding for unified multimodal understanding and generation. *arXiv preprint arXiv:2410.13848*, 2024.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Tom Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. *Advances in neural information processing systems*, 36:11809–11822, 2023.

Hang Zhang, Chongruo Wu, Zhongyue Zhang, Yi Zhu, Haibin Lin, Zhi Zhang, Yue Sun, Tong He, Jonas Mueller, R Manmatha, et al. Resnest: Split-attention networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2736–2746, 2022.

Jianqi Zhang, Jingyao Wang, Wenwen Qiang, Fanjiang Xu, Changwen Zheng, Fuchun Sun, and Hui Xiong. Intriguing properties of positional encoding in time series forecasting. *arXiv e-prints*, pages arXiv–2404, 2024.

Jiachen Zhu, Xinlei Chen, Kaiming He, Yann LeCun, and Zhuang Liu. Transformers without normalization. *arXiv preprint arXiv:2503.10622*, 2025.