Jacob Dineen

Contact Information

📬 jdineen[at]asu[dot]edu

🌐 AI, Reasoning and Cognition (ARC) Lab

📲 (480)603-6994

🌐 jacobdineen.com

Research Interests

I focus on reasoning and alignment in large language models (LLMs), with complementary interests in explainable AI (XAI) and game theory.

Education

Arizona State University — *2022-2027 (expected)*
Ph.D. in Artificial Intelligence (GPA: 4.00/4.00)
Advisor: Prof. Ben Zhou
Committee: Profs. Muhao Chen, Chitta Baral, Vivek Gupta

University of Virginia — *2019-2021*
M.Sc. in Computer Science (GPA: 3.96/4.00)
Advisor: Prof. Madhav Marathe

Syracuse University — *2017-2018*
M.S. in Data Science (GPA: 4.00/4.00)

Grand Canyon University — *2012-2015*
B.S. Finance and Economics (GPA: 3.65/4.00)

Research

Research Assistant @ *Arizona State University* — *Present*
- (Current) LLM Research @ AI, Reasoning & Cognition (ARC) Lab
- (Previously) Cybersecurity Research @ SEFCOM

Applied Research @ *Capital One* — *2020-2021*
- Explored aspects of organization dynamics under a reinforcement learning setting. Implemented an agent-based modeling system to study managerial incentive structures on experimental program optimization.

Research Assistant @ the *University of Virginia* — *2019-2020*
- Worked in the Biocomplexity Institute and Initiative labs with a focus on graph dynamic systems and cooperative game theory/behavior modeling, under the supervision of Professor Madhav Marathe.

Publications

Li, Z., Lu, S., Wang, F., **Dineen, J**., Ye, X., Xu, Z., Liu, S., Cho, Y. M., Li, B., Chang, D., Nguyen, K., Yang, Q., Chen, M., & Zhou, B. (2025). *Unbiased Visual Reasoning with Controlled Visual Inputs*. — Pending ICLR

Li, Z., Chickering, K. R., Li, B., **Dineen, J.**, Ye, X., Xu, Z., Lu, S., Huang, Y., Shen, M., Nguyen, B., Pavuluri, J. A., Nguyen, M. S., Chavan, S., Le, N. M. T., Chen, M., & Zhou, B. (2025). *Visual Analogies: Probing Unified Generation and Reasoning*. — Pending CVPR

Liu, Q., **Dineen, J**., Huang, Y., Zhang, S., Poon, H., Zhou, B., & Chen, M. (2025). ArenaBencher: Automatic Benchmark Evolution via Multi-Model Competitive Evaluation. arXiv preprint arXiv:2510.08569. — Pending ICLR

Dineen, J., RRV, A., Liu, Q., Xu, Z., Ye, X., Shen, M., ... & Zhou, B. (2025). QA-LIGN: Aligning LLMs through Constitutionally Decomposed QA. arXiv preprint arXiv:2506.08123. — EMNLP 2025

RRV, A., **Dineen, J.**, Handa, D., Uddin, M. N., Parmar, M., Baral, C., & Zhou, B. (2025). ThinkTuning: Instilling Cognitive Reflections without Distillation. arXiv preprint — EMNLP 2025

arXiv:2508.07616.

Ye, X., Shrivastava, S., Li, Z., **Dineen, J**., Lu, S., Ahuja, A., ... & Zhou, B. (2025). CC-LEARN: Cohort-based Consistency Learning. arXiv preprint arXiv:2506.15662.	Pending ICLR
Shen, M., Xu, Z., **Dineen, J**., Ye, X., & Zhou, B. (2025). BOW: Bottlenecked Next Word Exploration. arXiv preprint arXiv:2506.13502.	Pending ICLR
Srinivasan, A., **Dineen, J**., Afzal, M. U., Sarfraz, M. U., Riaz, I. B., & Zhou, B. (2025). RECAP: Transparent Inference-Time Emotion Alignment for Medical Dialogue Systems. arXiv preprint arXiv:2509.10746.	Pending SIGCHI
Ye, X., **Dineen, J.**, Li, Z., Xu, Z., Chen, W., Lu, S., ... & Zhou, B. (2025). Evaluating Medical LLMs by Levels of Autonomy: A Survey Moving from Benchmarks to Applications. arXiv preprint arXiv:2510.17764.	Pending EACL
Xu, Z., Shen, M., **Dineen, J**., Li, Z., Ye, X., Lu, S., ... & Zhou, B. (2024). Tow: Thoughts of words improve reasoning in large language models. arXiv preprint arXiv:2410.16235.	NAACL 2025 Main
Jacob Dineen, Donald Kridel, David Castillo, and Dan Dolk. "Unified Explanations in Machine Learning Models: A Perturbation Approach". In Proceedings of the 56th Hawaii International Conference on System Sciences.	HICSS 2023
Dineen J., Haque A.S.M.AU., Bielskas M. (2021). Formal Methods for an Iterated Volunteer's Dilemma. In: Thomson R., Hussain M.N., Dancy C., Pyke A. (eds) Social, Cultural, and Behavioral Modeling. SBP-BRiMS 2021.	SBP-BRiMS 2021
Dineen J., Haque A.S.M.AU., Bielskas M. (2021) Reinforcement Learning for Data Poisoning on Graph Neural Networks. In: Thomson R., Hussain M.N., Dancy C., Pyke A. (eds) Social, Cultural, and Behavioral Modeling. SBP-BRiMS 2021.	SBP-BRiMS 2021
Dolk, D., Kridel, D., **Dineen, J.**, & Castillo, D. (2020, January). Model Interpretation and Explainability towards Creating Transparency in Prediction Models. In Proceedings of the 53rd Hawaii International Conference on System Sciences	HICSS 2020

Professional Experience

Research Engineering Intern @ *Pareto AI* *'25-*

- Focused on developing evals for LLMs.

Machine Learning Engineer @ *Spring Oaks Capital* *'22-'25*

- Developed and deployed scalable ETL and modeling pipelines using Airflow and Kubernetes for text/call efforts and offer generation, incorporating ranking recommendations and constrained optimization solutions to scheduling problems.
- Implemented robust CI/CD processes, including tests, automated builds, and deployments utilizing AWS ECR, CodeBuild, and GitHub Actions, ensuring seamless and efficient workflow.
- Assisted wrt to cloud infrastructure for the core technology stack, from containerization to resource provisioning and development environments, optimizing performance and scalability.
- Prepared and maintained comprehensive Sigma dashboards and Streamlit apps to monitor online performance metrics, providing key stakeholders with actionable insights and facilitating data-driven decision-making.

Data Scientist @ *Capital One* *'21–'22*

- Engineered and productionalized critical updates to the core codebase, impacting 30MM+ users, through advanced feature engineering, robust data pipelines, unit tests, and custom model architectures.
- Developed sequential recommendation POCs utilizing PyTorch, Huggingface, and Nvidia's Merlin/Transformers4Rec. These innovations were showcased at the Nvidia GTC Fall Summit 2022.
- Co-led and designed a bi-weekly lecture series on Deep Learning and Neural Recommendation, fostering knowledge sharing and upskilling within the team.

Ph.D. Internships @ *Capital One (2X Data Science, 1X Applied Research)* *'20-'21*

- Researched and implemented state-of-the-art neural recommendation solutions for ad tech challenges, significantly improving ad targeting and engagement.
- Developed scalable and extensible data pipelines in PySpark, leveraging novel data sources to enhance model performance and insights.
- Provided strategic insights and recommendations for integrating neural solutions into production environments, extending the impact of summer projects.
- Conducted advanced research in agent-based modeling and reinforcement learning, contributing to the Center for Machine Learning (C4ML).

Truncated for brevity

Analyst and Business Intelligence @ *Real World Marketing* *'16-'19*

Data Scientist @ *Buffalo Check LLC* *'15-'19*

Optimization Analyst @ *Voltari* *'12-'15*

Graduate Courses

Knowledge Representation, Computer Systems Security, Software Security, Planning and Learning Methods in AI, Algorithms & Research/Dissertation Hours	*ASU*
Algorithms, Machine Learning, Computer Vision, Formal Methods, Reinforcement Learning, Graph Mining, Learning Theory (Game Theory), Cloud Computing & Research Hours	*UVa*
Data Analysis and Decision Making, Business Analytics, Financial Analytics, Marketing Analytics, Advanced Information Systems, Data Science, Data Warehousing, Text Mining, Scripting for Data Analysis, and Information Policy	*Syracuse*

Skills

OS	Linux (Ubuntu), MacOS, Windows
Language	Python, Rust, x86-64, Java, JS, R, C, C++, PRISM, Bash, Vue, React
Database	MySQL, SQLite, NoSQL, MongoDB, Snowflake, Redshift, Postgres, Redis
Markup	LaTex, HTML
ML Library	PyTorch, Keras, Tensorflow, Jax, Numpy, Pandas, Polars, Dask, NLTK, Networkx, SparkML, SnowparkML, DeepGraphLibrary, HuggingFace, Botorch, Torch Geometric, Burn, TRL, vllm, VeRL
Other	Weka, Mallet, Conda/Mamba, VSCode, Git, Databricks, Docker, Snowflake Snowpark, Airflow, AWS (ECR/S3/EKS/Codebuild), Kubernetes, Helm, Sigma, Sagemaker,

		OR-Tools, Click, Streamlit

Misc.

Expert AI Trainer	pareto.ai (‘24-present)
Tutor/TA	CGCC Calculus and Linear Algebra Tutor (‘19) ASU CSE365 (pwncollege) TA
Conference Reviewer	HICSS, SBP-BRiMS, NAACL, EMNLP, EACL
Cyber Security (Certificate)	Pwn.college green belt (user: jdin) (‘22). Reverse engineering, binary exploitation, dynamic and static analysis