

EDUCATION

University of Arizona

Ph.D. in Systems and Industrial Engineering, MultiModal Deep Learning 2024
M.S. in Computer Science, NLP 2022
M.S. in Systems and Industrial Engineering, ML 2018

TECHNICAL STRENGTHS

AI & ML Expertise	Multimodal Learning, Reinforcement Learning, NLP, Agentic AI, Time Series
Computer Programming	Python, Java, Scala, SQL, Bash, R
Software & Tools	Docker, Git, Databricks, Kubernetes, AzureML
Deep Learning Frameworks	PyTorch, TensorFlow, HuggingFace, Keras, PyTorch Lightning

WORK EXPERIENCE

Senior AI Engineer, Armada AI

Aug 2023 - Present

- **Lead the development of a prescriptive intelligence platform** for complex industrial processes, improving overall efficiency by a projected **4.5%**. Utilize advanced ML and optimization, and leverage LLMs to translate complex data into actionable, natural language insights for operators.
- **Architect a scalable agentic AI routing system** using advanced techniques like GRPO and a novel self-correction mechanism. The resulting distilled models **outperform large-scale proprietary systems by over 10%** and are optimized for edge deployment.
- **Pioneered novel AI applications**, developing a natural language interface for robotic systems that improved task accuracy by over 20%, and leading a large-scale geospatial analysis of LEO satellite network performance to predict service degradation.
- **Mentor and train junior AI engineers and interns**, leading multiple high-impact projects and collaborating cross-functionally with product, engineering, and customer teams to deliver robust, scalable solutions.

Data Scientist, American Airlines

Aug 2022 - Aug 2023

- **Developed a deep neural network forecasting engine** that predicts flight-level traffic for the Yield Management team, reducing prediction mean square errors by 50% through the use of multimodal neural network architecture and pre-training on historical data, including time series and multi-dimensional data.
- **Led** the development of the forecasting engine and spearheaded its deployment into the airline system for **A/B testing** and potential replacement of older methods. Worked closely with the IT team on the project development using AzureML and **trained and supervised** a team of four in the use and evaluation of the approach.
- Expanded the analysis to cover the entire market of around 7,000 flights to achieve accurate predictions for future 330 days. Demonstrated the model's superior performance in capturing trends and seasonality, as well as its ability to observe sudden changes in capacity effects on demand that previous models were unable to capture.

Machine Learning Postdoctoral Researcher, Truveta

May 2022 - Aug 2022

- Contributed to the company's mission of Saving Lives with Data through the development of data-driven NLP approaches.
- Pre-trained and fine-tuned **Large Language Models (LLMs)** on clinical text data, resulting in improved overall accuracy for upstream and downstream tasks and significantly increased downstream training and inference speeds (fourfold and twofold increase, respectively).
- Achieved state-of-the-art results in the ontology alignment task, with at least a **5% improvement** in F1, Hit@1, and MRR. Developed a novel framework (Truveta Mapper) using a multi-task sequence-to-sequence LLM for unsupervised ontology alignment, which outperforms existing solutions in runtime and alignment quality.

Data Science Research Intern, RedShred

Mar 2021 - Aug 2021, Jan 2022 - May 2022

- Implemented a novel plot-processing approach utilizing object detection models and text extraction techniques, resulting in a 69.12% success rate in correctly identifying data from plots, **approaching the state-of-the-art** in the field. Deployed on the RedShred API for automatic processing of PDF documents.
- Utilized HuggingFace framework to develop and train transformer-based methods for document understanding, including segmentation and object detection. These techniques demonstrated strong performance in the field of document analysis.

Graduate Research Assistant, University of Arizona

May 2018 - Dec 2023

- **Developed and designed a novel multimodal deep learning architecture** that employs reinforcement learning to dynamically adjust network depth, achieving an approximate **10-15% reduction in mean squared error (MSE)** across diverse tasks compared to STOA models.

· **Introduced a novel Textual Question Answering architecture** that employs on-demand visual grounding to enhance answer accuracy. The 400M parameter model demonstrates **comparable performance to significantly larger models like GPT-4o**.

PUBLICATIONS

Ehsani, S., Pan, F., Hu, Q. and Liu, J., 2025. BiDepth Multimodal Neural Network: Bidirectional Depth Deep Learning Architecture for Spatial-Temporal Prediction. arXiv preprint arXiv:2501.08411.

Buynitsky, A., **Ehsani, S.**, Pallakonda, B., and Mishra, P., 2025 Camera Control at the Edge with Language Models For Scene Understanding. In *IEEE ICCAR2025 Proceedings*.

Ehsani, S., and Liu, J., 2024. From the Depths to the Surface: Navigating Spatial Temporal Data with DeepShallow Network. In *CIE51 Proceedings*.

Ehsani, S. and Liu, J., Elevating Textual Question Answering with On-Demand Visual Augmentation. *ACM Transactions on Multimedia Computing, Communications and Applications*.

Ehsani, S., Sergeeva, E., Murdy, W. and Fox, B., 2024. Predicting the Skies: A Novel Model for Flight-Level Passenger Traffic Forecasting. *arXiv preprint arXiv:2401.03397*.

Amir, M., Baruah, M., Eslamialishah, M., **Ehsani, S.**, Bahramali, A., Naddaf-Sh, S. and Zarandioon, S., 2023. Truveta Mapper: A Zero-shot Ontology Alignment Framework. *arXiv preprint arXiv:2301.09767*.

Ehsani, S., Reddy, C.K., Foreman, B., Ratcliff, J. and Subbian, V., 2021. Subspace Clustering of Physiological Data From Acute Traumatic Brain Injury Patients: Retrospective Analysis Based on the PROTECT III Trial. *JMIR Biomedical Engineering*, 6(1), p.e24698.

PATENTS

Mishra, P.K., Buynitsky, A., and **Ehsani, S.**, Armada Systems Inc, 2025. *Robotic control using natural language commands*. U.S. Patent 12,289,517.

Ehsani, S., and Mishra, P.K., Armada Systems Inc, 2024. *Natural language statistical model with alerts*. U.S. Patent 12,086,557.

Ehsani, S. and Mishra, P.K., Armada Systems Inc, 2024. *Time series data to statistical natural language interaction*. U.S. Patent 12,067,041.

Ehsani, S., and Mishra, P.K., Armada Systems Inc, 2025. *Natural language statistical model with workspaces*. U.S. Patent Pending 12,086,557.

Marcjan, C.A., Amir, M., Baruah, M., Eslamialishah, M., **Ehsani, S.**, Bahramali, A., Naddaf-Sh, S. and Zarandioon, S., Truveta Inc, 2023. *Systems and methods for ontology matching*. U.S. Patent Pending 202,400,87,687