Nahid Zeinali

San Diego, CA|Nahid-zeinali@uiowa.edu
GitHub | LinkedIn | Google Scholar | Portfolio | ORCID

AI Researcher and Data Scientist with deep experience in machine learning, deep learning, and LLM in biomedical AI. Experienced in building scalable ML pipelines integrating text and structured biomedical data using transformer architectures and embedding-based representation learning. Proven ability to develop end-to-end AI systems for healthcare and scientific discovery, including multimodal Retrieval-Augmented Generation (RAG) and foundation model evaluation.

PROFESSIONAL EXPERIENCE

AI Research Scientist, California Medical Innovation Institute, San Diego, CA ***Feb 2025- Present***

- ***DMP Bridge (Link):*** Led and developed an open-source multimodal GenAI pipeline that converts funder-specific Data Management Plan (DMP) PDFs into interoperable machine-actionable metadata using LLMs, document AI, and FAIR-aligned workflows.
- ***DMP Chef (Link):*** Led and implemented a Retrieval-Augmented Generation (RAG) pipeline leveraging GPT-4, Gemini, and Llama models to automatically generate NIH-compliant and FAIR-aligned Data Management Plans (DMPs).
- ***LLM Evaluation for NIH DMP Generation (Link):*** Led and evaluated GPT-4.1 and Llama 3.3 for NIH-compliant DMP generation using automated metrics, statistical analysis, and human-centered evaluation, with GPT achieving the highest overall user satisfaction and stronger SBERT and ROUGE performance.
- ***NIH DMP Generation with LLMs (Link):*** Led and designed reproducible prompt-engineering and structured generation workflows for automated NIH Data Management Plan creation using large language models.
- ***FAIR Ophthalmology Data Standardization Pipeline (Link | White Paper):*** Collaborated on FAIR-aligned data standardization pipelines for ophthalmology imaging datasets, improving machine learning readiness, interoperability, and future analytics/model development.
- ***FAIR Scientific Poster Schema & Machine-Readable Pipeline (Link |Schema |User Experience| Software Architecture***): Collaborated on schema evaluation and JSON-based FAIR workflows to transform scientific posters into machine-readable and interoperable research objects for AI-driven discovery and reuse.
- Prepare technical reports, manuscripts, and grant applications with investigators

Graduate Research & Teaching Assistant, Computer Science Department, University of Iowa ***Aug 2021- Jan 2025***

- ***Symptom-GPT (Link | Article):*** Led and developed transformer-based clinical NLP and LLM models for named entity recognition and cancer symptom detection from large-scale oncology 1M EHR clinical notes, achieving F1 scores up to 0.989 for nausea/vomiting and 0.912 for anxiety detection.
- **Symptom-BERT *(Link | Article):*** Led and developed transformer-based clinical NLP and LLM models for further pretrain and fine-tune transformer models on oncology 1M EHR data to detect 13 cancer symptom groups, achieving micro-F1 = 0.933 in internal validation and micro-F1 = 0.831 in external validation.
- **Care-BERT *(Article | Article):*** Led and built transformer-based clinical NLP models to identify care priorities and life-sustaining treatment preferences from EHR narratives, achieving F1 = 0.941 and AUC = 0.978 in internal validation.
- **Spiritual-BERT *(Link | Article):*** Led and developed a Bio Clinical BERT-based NLP system to detect under-documented spiritual care information from 3.6M+ EHR notes, achieving F1 = 0.938 in internal validation and F1 = 0.832 in external validation.
- **OASIS – Oncology Associated Symptoms & *Individualized Strategies (Link)***: Collaborated with clinicians, researchers, and technical collaborators on AI- and NLP-driven oncology symptom monitoring research using EHR data to support personalized cancer care and clinical decision support.
- **Patient-Reported Outcomes & EHR Symptom Concordance (*Article*):** Collaborated and analyzed concordance between patient-reported and provider-documented cancer symptoms, identifying significant documentation gaps across multiple symptom groups.
- **Embeddings-Augmented NLP for Symptom Detection (*Article*):** Collaborated and developed embeddings-enhanced NLP pipelines for extracting cancer symptom information from 900K+ clinical notes, achieving F1 scores up to 0.937 for symptom classification tasks.
- **Cancer Symptom Prediction Using Machine Learning (*Article*):** Collaborated and built predictive ML models using structured and unstructured EHR data to forecast cancer symptom development, with Random Forest achieving macro-AUC = 0.755 and pain prediction AUC = 0.954.
- **Systematic Review of ML for Cancer Symptom Prediction (*Article*):** Led and conducted a PRISMA-guided systematic review of 42 studies evaluating machine learning approaches for cancer symptom prediction and personalized oncology care.
- **Web Analysis (*link*):** Developed an end-to-end NLP and machine learning pipeline for large-scale web scraping, sentiment classification, clustering, and topic modeling across 3,000+ online articles.

- **Demand Forecasting for E-Commerce Sales (*link*):** Built Python-based forecasting and analytics models to predict product demand and support inventory optimization for e-commerce sales data.
- Mentored master's students, fostering leadership within data science and machine learning research collaborations.
- Taught Python and machine learning courses, guiding students to develop strong skills in programming and data science.
- Published over 10 papers across machine learning and data science topics.

NLP Data Scientist, National Cancer & Frederick National Laboratory, NCAT, NIH ***Jun 2024-Aug 2024***

- **RARe-SOURCE™ (*Link*)**: Collaborated and developed AI-powered biomedical literature intelligence workflows using LLMs, semantic search, and NLP to support rare disease knowledge discovery and scientific insight extraction.
- Worked closely with biomedical researchers to convert research questions into usable AI workflows and decision-support tools.

Software Engineer, Khorshid Hospital, Isfahan ***Feb 2019 – July 2021***

- Built and enhanced EMR, PIS, and LIS systems in collaboration with clinical and business stakeholders, improving workflow efficiency and patient-care operations by 35%.
- Developed a mobile app for heart failure symptom tracking used by 3,000+ patients in rural Isfahan, expanding remote monitoring and supporting earlier clinical intervention.

Software Engineer, Parisian Institute, Tehran ***Dec 2016 – Jan 2019***

- Designed and implemented an EHR management dashboard that reduced report turnaround time by 68% and improved access to clinical information.
- Trained and mentored 1,500+ clinicians on EHR adoption and workflow best practices, supporting large-scale digital transformation in healthcare delivery.

EDUCATION

- ***Ph.D., Informatics,*** University of Iowa, Iowa City, IA, USA
- ***M.Sc., Informatics,*** University of Iowa, Iowa City, IA, USA
- ***M.Sc., Medical Informatics,*** Tarbiat Modares University, Tehran, Iran
- ***B.S, Computer Software Engineering,*** Najafabad Azad University, Isfahan, Iran

TECHNICAL PROFICIENCIES

- ***Programming Languages*:** Python, MATLAB, C, C++, C#, ASP.NET, Android, JavaScript, HTML, XML, Jupyter Notebook
- ***Data Science & Analysis:*** Pandas, NumPy, Scikit-learn, Keras, Matplotlib, Seaborn, Data Cleaning, Data Preprocessing, Exploratory Data Analysis
- **Machine Learning & AI:** Machine Learning, Deep Learning, Classification, Regression, Feature Engineering, Model Validation, Hyperparameter Tuning, Statistical Modeling, Predictive Modeling, Recommender Systems
- ***Natural Language Processing (NLP) & Large Language Models (LLMs):*** Text Mining, Named Entity Recognition, Transformers, BERT, RoBERTa, GPT, Llama, Prompt Engineering, Retrieval-Augmented Generation (RAG), Fine-Tuning, OpenAI API, Agentic AI Frameworks, Conversational AI, Transformer-based encoders/decoders, LLMs, Reinforcement learning, Agentic workflows
- **Frameworks & Libraries:** Hugging Face, LangChain, LangGraph, LangSmith
- **Inference & Performance Optimization:** Developed and benchmarked end-to-end LLM and RAG pipelines; optimized retrieval/index configurations, chunking strategies, top-k parameters, and API-based serving to improve latency, reliability
- ***MLOps Tools:*** MLflow, DVC, DagsHub, Apache Airflow (Astro), Grafana, Workflow Automation
- ***CI/CD & DevOps:*** Git, GitHub Actions, Docker, AWS SageMaker, Astro
- ***Cloud & Distributed Computing:*** AWS, GCP, HPC, Linux
- ***Database & Data Management:*** PostgreSQL, MS SQL Server, MongoDB, FAISS, ChromaDB, Cassandra, Pinecone
- ***Analytics & Statistics:*** SPSS, SAS, STATA, Power BI, experimental evaluation, statistical testing

HONORS & AWARDS

- ***Excellent Award Research, University of Iowa*** ***Spring 2025***
- ***Ballard and Seashore Dissertation Fellowship, University of Iowa*** ***Fall 2024***
- ***Student Impact Grant, University of Iowa*** ***Spring 2024***
- ***Research and Travel GPSG, GSS Award, the University of Iowa*** ***2021-2024***
- ***Recruitment Fellowship, IGPI (Per Year), University of Iowa*** ***2021-2024***

COMMUNITY AND LEADERSHIP SERVICE

- ***Reviewer Volunteer, Informatics Summit 2025, AMIA Organization, etc.*** ***Fall 2024***

- ***P2P Mentor & Mentee Volunteer, University of Iowa*** ***Fall 2024***
- ***Student Volunteer, AMIA 2024 Annual Symposium, AMIA Organization*** ***Fall 2024***
- ***Student Volunteer, ISO Organization, University of Iowa*** ***2021-2023***

PUBLICATION

- A. AlBashayreh, **N. Zeinali,** et al. "Advance Directives and Dementia: How Illness Trajectories Influence Goals-of-Care". Under review
- A. AlBashayreh, **N. Zeinali,** et al. "Artificial Intelligence Reveals Language Disparities in Person-Centered Spiritual Care: Access and Timing Among Older Adults". Under review
- **N. Zeinali**, B. Patel, et al**. "**Evaluating the Performance of LLMs in Creating NIH Data Management Plans". Under review
- **N. Zeinali**, S. White, et al. "Using Large Language Models to Detect Anxiety and Nausea/Vomiting Documentation in Clinical Notes of Patients with Cancer", *CIN Journal (2025).*
- **N. Zeinali**, Stephanie Gilbertson-White, et al., "Machine Learning Approaches to Predict Symptoms in People with Cancer: A Systematic Review," *JMIR Cancer,2024.*
- ***N. Zeinali**, S., White, et al. "*Symptom-BERT: Enhancing Cancer Symptom Detection in EHR Clinical Notes*." Journal of pain and symptom management (2024). (*selected as "Year in Review" by the Nursing Informatics Working Group at AMIA2024.)*
- A. AlBashayreh, **N. Zeinali,** et al. "Goals-of-Care in Older Adults with Heart Failure, Cancer, and Dementia: Classifying Comfort and Life-Sustaining Preferences Using Priorities-BERT", *Innovation in Aging Journal (2025).*
- A. AlBashayreh, **N. Zeinali**, Stephanie Gilbertson-White. An Informatics Approach to Characterizing Spiritual Care Documentation in Electronic Health Records of Older Adults." *ACI journal, 2025*
- A. AlBashayreh, **N. Zeinali**, Stephanie Gilbertson-White. "Innovating the Detection of Care Priorities in Heart Failure Using Large Language Models." *Innovation in Aging 8 (Suppl 1), 1339.*
- A. AlBashayreh, A. Bandyopadhyay, **N. Zeinali**, et al. "Natural Language Processing Accurately Differentiates Cancer Symptom Information in EHR Narratives." *JCO clinical cancer Informatics,2024.*
- S.G. White, A. AlBashayreh, A. Bandyopadhyay, **N. Zeinali,** et al., "Special Section on Patient-Reported Outcomes and Informatics: Predictors of Concordance Between Patient-Reported and Provider-Documented Symptoms in the Context of Cancer and Multimorbidity." *ACI,2024*
- A. Bandyopadhyay, A. AlBashayreh, **N. Zeinali,** et al. **"**Using real-world EHR data to predict the development of 12 cancer-related symptoms in multimorbidity. Predictive." *Open JAMIA (Journal of American Medical Informatics Association), 2024.*
- Elham Nazari, **Nahid Zeinali**, et al. "Application of Big Data Analysis in Healthcare based on the six building blocks of health systems' Framework: A Survey." *Dokkyo Journal of Medical Sciences 2021.*
- **Nahid Zeinali,** Abbas Asosheh, et al. "Provide interoperability model to interact in hospital information systems." *Journal of Health and Biomedical Informatics, 2017.*
- **Nahid Zeinali**, Abbas Asosheh, et al. "The Conceptual Model to Solve the Problem of Interoperability in Health Information Systems." *2016 8th International Symposium on Telecommunications (IST).*
- Zeinab Delaram, **Nahid Zeinali,** et al. "The Common Applications of Social Networks in Healthcare." *Health Information Management 2016.*

PRESENTATION & POSTER

- **N. Zeinali (Presenter),** B. Patel, et al. "Evaluating the performance of large language models in drafting data management plans." 2026 BOSC conference
- B. Patel, **N. Zeinali,** et al**. "**Toward Envision Portal: Designing a FAIR and AI-Ready Framework for Ophthalmic Imaging sharing and Discovery**"**, ARVO 2026 Annual Meeting.
- **N. Zeinali (Presenter),** B. Patel, et al. "Evaluating the Effectiveness of an Open-Source Large Language Model in Drafting NIH Data Management Plans" International Data Week (SciDataCon 2025)
- **N. Zeinali (Presenter)**, A. AlBashayreh, et al. "Comparison of BERT Implementations for Enhanced Cancer Symptoms Extraction from Electronic Health Records." *2024 IEEE First International Conference on Artificial Intelligence for Medicine, Health, and Care (AIMHC), Laguna Hills, CA, USA, 2024, pp. 18-19, Doi: 10.1109/AIMHC59811.2024.00011.*
- **N. Zeinali (Presenter),** Stephanie Gilbertson-White, et al. "Advanced Detection of Nausea/Vomiting and Anxiety in Patients with Cancer." *AMIA 2024 Annual Symposium.*
- A. AlBashayreh, **N. Zeinali**, et al. "Leveraging Spiritual-BERT for Characterizing Spiritual Care Documentation in EHRs of Older Adults with Heart Failure Leveraging Spiritual-BERT for Characterizing Spiritual Care Documentation in EHRs of Older Adults with Heart Failure." *AMIA 2024 Annual Symposium.*
- A. AlBashayreh, **N. Zeinali**, et al. "Innovating the Detection of Care Priorities in Heart Failure Using Large Language Models." *Poster/GSA 2024*
- A. AlBashayreh, **N. Zeinali**, et al. "Disparities in Advance Directive Completion and life-sustaining Treatment Preferences in older adults." *Poster/ Annual Assembly Hospice and Palliative Care 2024*
- **N. Zeinali (Presenter),** Stephanie Gilbertson-White, et al. "Leveraging Large Language Models for Named Entity Recognition of Anxiety and Nausea and Vomiting in Patients with Cancer." *AMIA Informatics 2025*.