

Dipak Meher

Fairfax, VA 22030 | +1 703-712-9477 | ✉ dipakmeher51@gmail.com | 📁 Portfolio | 🔗 LinkedIn | 📄 Google Scholar

EDUCATION

PhD in Computer Science <i>George Mason University, Fairfax, VA</i>	Aug 2024 – Present
MS in Computer Science <i>George Mason University, Fairfax, VA GPA: 3.96/4.0 Distinguished Academic Award</i>	Aug 2022 – May 2024
BTech in Computer Science and Engineering <i>SGGS, Nanded, India GPA: 8.56/10.0</i>	Aug 2016 – Nov 2020

SUMMARY

PhD student with research spanning machine learning, multilingual natural language processing, foundation models, AI safety, knowledge graphs, recommender systems, and graph-based retrieval-augmented generation (GraphRAG). Experienced in building end-to-end ML pipelines with PyTorch, TensorFlow, and Hugging Face, with two years of prior industry experience. Published in KDD, ICDM, ICKG, SDM, and IJCNN, with research spanning multilingual LLM evaluation, AI safety, knowledge graph construction, and cross-domain recommender systems. **Seeking AI/ML Internship Opportunities.**

RESEARCH EXPERIENCE

Graduate Research Assistant, George Mason University 🌐	Fairfax, VA
<i>Advisor: Prof. Carlotta Domeniconi</i>	<i>Aug 2024 – August 2025</i>

- **LINK-KG: Knowledge graph construction framework for short and long documents**
 - * Designed LINK-KG, an LLM-based pipeline for knowledge graph construction from complex legal narratives, addressing long-range coreference, entity disambiguation, and reference normalization.
 - * Developed a three-stage coreference resolution system using a type-specific prompt cache to resolve plural mentions, role shifts, and alias ambiguity while mitigating context window limitations and loss-in-the-middle issues in LLMs.
 - * Achieved a 45.21% reduction in node duplication and 32.22% decrease in legal noise over GraphRAG and CORE-KG, enabling more coherent and interpretable graphs for criminal network analysis.
- **CORE-KG: LLM-Based Knowledge Graph Construction for Human Smuggling Network Analysis**
 - * Developed CORE-KG, a modular LLM-driven framework integrating type-aware coreference resolution and domain specific prompting to construct coherent knowledge graphs from human smuggling legal cases.
 - * Introduced sequential entity extraction and legal-specific filtering to reduce attention drift, suppress procedural noise, and improve entity type accuracy in graph construction.
 - * Achieved 33.28% reduction in node duplication and 38.37% reduction in legal noise over GraphRAG, enabling cleaner and more reliable criminal network analysis.

Graduate Research Assistant, George Mason University 🌐	Fairfax, VA
<i>Advisor: Prof. David Rosenblum</i>	<i>May 2023 – Aug 2024</i>

- **Quantifying Cross-Domain User Behavior Consistency with LLMs**
 - * Proposed the first LLM-based framework to quantify user behavior consistency across domains via item feature extraction, sentiment analysis, and behavior alignment.
 - * Conducted an empirical study across 12 domain pairs and 4 state-of-the-art cross-domain recommendation (CDR) models, showing that current methods underutilize behavior consistency for better recommendations.
 - * Revealed strong user behavior consistency in similar domains (Books–Movies: 0.87) and significantly lower consistency in dissimilar domains (Electronics–Food: 0.23), highlighting the need for behavior-aware cross-domain recommendation.
- **LLMs for Transfer Learning in Recommender Systems**
 - * Co-authored research introducing an LLM-driven framework for cross-domain recommendation, improving personalization in sparse-data cold-start and warm-start scenarios.
 - * Benchmarked six baseline models (four from the SDM 2024 paper, plus DisenCDR and UniCDR), and evaluated prompt strategies with and without target-domain examples, showing that LLMs can match or outperform SOTA CDR models.
- **Vietoris-Rips Complex for Cross-Domain Cold-Start Recommendation**

- * Co-authored SDM 2024 paper, implementing and benchmarking baseline models (TGT, CMF, EMCDR, PTUPCDR) for cross-domain recommendation.
- * Developed VRCDR using Vietoris-Rips Complex and Area-based Triangulated Embedding (ATE) to model users' niche source-domain preferences as geometric structures.
- * Achieved up to 20% improvement over SOTA methods across four cross-domain tasks under extreme cold-start conditions.

Research Intern, Indian Institute of Technology, Kharagpur

Kharagpur, India

Advisor: Prof. Pawan Goyal

May 2019 – Jun 2019

- Developed an apparel classifier using deep learning on the 800K-image Deep Fashion dataset.
- Improved model robustness via data augmentation and hyperparameter tuning, achieving 89% accuracy.

WORK EXPERIENCE

PowerApps Developer, Tata Consultancy Services, Mumbai

Nov 2020 – Jul 2022

- Built 6 PCF components in TypeScript and automated workflows with Power Automate, reducing process time by 5%.
- Led a 4-member team to design optimized UI screens and create custom connectors, improving app performance by 10%.
- Delivered 10+ user stories per sprint within the SAgile framework and mentored 20 developers.

PUBLICATIONS

- **Dipak Meher**, C. Domeniconi, G. Correa-Cabrera. “*LINK-KG: LLM-Driven Coreference-Resolved Knowledge Graphs for Human Smuggling Networks*.” ICKG 2025 (IEEE), Limmasol, Cyprus. Nov 13-14, 2025. [PDF]
- **Dipak Meher**, C. Domeniconi “*Inside CORE-KG: Evaluating Structured Prompting and Coreference Resolution for Knowledge Graphs*” ICDM 2025 (IEEE), PhD Forum proceedings, Washington DC, USA. Nov 12-15, 2025. [PDF]
- **Dipak Meher**, C. Domeniconi, G. Correa-Cabrera. “*CORE-KG: An LLM-Driven Knowledge Graph Construction Framework for Human Smuggling Networks*.” KDD 2025 (ACM), SKNow-LLM Workshop, Toronto, Canada. [PDF]
- **Dipak Meher**, A. Krishna Vajjala, and D. Rosenblum. “*Understanding User Behavior in Cross-Domain Recommendation: An LLM-Based Approach*.” IJCNN 2025 (IEEE), Rome, Italy. [PDF]
- A. Krishna Vajjala, **Dipak Meher**, and D. Rosenblum. “*Cross-Domain Recommendation Meets Large Language Models*.” [In-Review] [PDF]
- A. Krishna Vajjala, **Dipak Meher**, S. Pothagoni, Z. Zhu, and D. Rosenblum. “*Vietoris-Rips Complex: A New Direction for Cross-Domain Cold-Start Recommendation*.” SDM 2024 (SIAM), Houston, USA. [PDF]

INVITED TALKS

- **Mitigating Temporal Degradation in Entity Linking** — Data Mining Lab, GMU, 2024
- **GraphRAG for Query-Focused Summarization** — Data Mining Seminar, GMU, 2024
- **Understanding User Behavior in Cross-Domain Recommendation** — IJCNN 2025, Rome, Italy
- **Knowledge Graph (KG) Construction with LLMs** — KDD 2025 Workshop (SKnow-LLM), Toronto, Canada
- **LINK-KG: Coreference-Resolved KGs for Human Smuggling Networks** — ICKG 2025, Limassol, Cyprus

ACADEMIC PROJECTS

Adversarial Evaluation of Code LLM Safety | NLP + Security

Jan 2026 – Present

- Analyzed the MALICE benchmark containing ~150K multilingual prompts to study safety robustness of Code LLMs.
- Investigated code-mixed and transliterated attacks that bypass safety guardrails while preserving programming keywords.
- Observed severe safety degradation: detection accuracy dropped from 0.93 to 0.21 with up to 3× increase in malicious code generation.

LLM Post-Training for Multilingual Reasoning | NLP

Aug 2025 – Dec 2025

- Studied multilingual reasoning failures caused by implicit translation drift across five languages on the BBQ benchmark.
- Built a translation-aided post-training framework using SFT and GRPO on synthetic translated data without labels.
- Improved compute-normalized multilingual accuracy and demonstrated strong dependence on translation fidelity using COMET.

Multimodal Sarcasm Detection using CLIP | NLP + Vision

Feb 2024 – Apr 2024

- Built a CLIP-based multimodal model for sarcasm detection from image-text pairs, achieving 0.82 F1 on benchmarks.
- Evaluated robustness on multilingual and noisy inputs, observing drop to 0.50 F1 in cross-lingual settings.
- Built a synthetic multilingual dataset using similarity-based filtering to preserve semantic intent across translations.

TECHNICAL SKILLS

Languages: Python, Java, JavaScript, C++, C, SQL, Go, React, Node.js, Angular, Spring Boot, Linux, Unix

Tools: PyTorch, TensorFlow, Scikit-learn, Pandas, Numpy, LangChain, Spacy, Ollama, Seaborn, Matplotlib