Md Rizwan Parvez

Scientist
Qatar Computing Research Institute (QCRI)
Hamad Bin Khalifa University
Doha, Qatar

+974 5106 9382
rizwan@ucla.edu
https://rizwan09.github.io
linkedin.com/in/rizwanparvez

RESEARCH INTERESTS (ATTRIBUTED, VERIFIABLE, AND GROUNDED AGENTIC AI)

Language – Knowledge (Attributed): retrieval-augmented generation whose answers trace back to auditable evidence – source valuation, adaptive retrieval, context filtering, and calibrated abstention (Open-RAG, FiLCo, ERA, DelucionQA, CoT–PoT, TechniqueRAG).

Language – Code (Verifiable): foundation models and multi-agent systems whose code is verified against specification and execution, from synthesis to repository-scale software engineering (RedCoder, MapCoder, CodeSim, Xolver, xCodeEval, SWE-Xplorer, PwS).

Language – Action (Grounded): long-horizon, tool-using agents whose actions are grounded in live environments – web, GUI, maps, charts, and dashboards (MapEval, MapAgent, SpatiaLab, TimeSpot, WebOperator, GUI-Primitives, DashboardQA).

EDUCATION

University of California, Los Angeles (UCLA) **Los Angeles, CA**
09/2017–06/2022 Ph.D. in Computer Science
Advisor: *Kai-Wei Chang* (UCLA NLP Group)
Committee: Guy Van den Broeck, Quanquan Gu, Yizhou Sun, Junghoo Cho

Bangladesh University of Engineering and Technology (BUET) **Bangladesh**
05/2010–09/2015 B.Sc. in Computer Science and Engineering
GPA 3.80/4.0, Rank 15th/125, Dean's List

HONORS, AWARDS, AND SELECTED IMPACTS

2026 Dataset Adopted in Claude Mythos/Fable 5
ChartQAPro (ACL Findings 2025)

2025 🏆 ICML 2025 Spotlight
MapEval, selected among 313 of 12,107 submissions (Top 2%)

2024 🏆 Top-29 Most-Cited Paper at EMNLP 2024
A Systematic Survey and Critical Review on Evaluating Large Language Models (Critics on LLM Evaluation)

🏆 Top-37 Most-Cited Paper at ACL 2024
MapCoder: Multi-Agent Code Generation for Competitive Problem Solving

6.2 M+ Hugging Face Downloads
xCodeEval (ACL 2024), widely adopted in LLM training and evaluation

Press Coverage in MarkTechPost News
Open-RAG (EMNLP Findings 2024), reaching 1.5 M+ monthly readers

2021	Top-10 Most-Cited Software Engineering ArXiv Paper, 2021 *RedCoder: Retrieval Augmented Code Generation and Summarization (EMNLP Findings 2021)*
2021, 2018	Conference Student Travel Grants *EMNLP 2021, ACL 2018*
2017, 2018	UCLA Graduate Division Fellowships *University of California, Los Angeles*
2014–2015	Professor M. A. Nasser Memorial & Undergraduate Merit Scholarships
2013–2015	Undergraduate Thesis and Dean's List Awards

PUBLICATIONS

∗ denotes equal contribution; Google Scholar

2,700+ Google Scholar citations | h-index 18 | i10-index 25. Published at ACL, EMNLP, NAACL, COLM, TMLR, ICML, NeurIPS, and ICLR.

Citation-based annotations measured by citation counts up to Sep 2026, according to Paper Digest.

Manuscripts and Pre-Prints

P.1 Mahir Labib Dihan, Tanzima Hashem, Mohammed Eunus Ali, and **Md Rizwan Parvez**. SWE-Xplorer: Iterative Tree Search with Cross-Path Reconciliation for Autonomous Software Engineering.
- Under review ICLR, 2027

P.2 Md. Nazmus Sakib, Omar Abdur Razzaque, and **Md Rizwan Parvez**. Why Are the Computer-Using Agents Still Failing?
- Under review ICLR, 2027

P.3 Md Tanzib Hosain, Azmine Toushik Wasi, Salman Rahman, Md Mofijul Islam, Mohammad Ali Moni, Mohammed Eunus Ali, Usman Naseem, and **Md Rizwan Parvez**. BRACE: An R Benchmark for Rigorous Biomedical Data Science Code Search.
- Under review ICLR, 2027

P.4 Asif Shahriar, Md Nafiu Rahman, Sadif Ahmed, Farig Sadeque, and **Md Rizwan Parvez**. AgentTell: AgentTell: Behavioural Side-Channel Leakage in Browser-Use Agents
- Under review ICLR, 2027

P.5 Md Abrar Jahin1, and **Md Rizwan Parvez**. Same Knowledge, Different Graph: Knowledge-Graph Foundation Models Are Not Schema-Invariant.
- Under review ICLR, 2027

P.6 Azmine Toushik Wasi, Mahir Absar Khan, Abdur Rahman, Wahid Faisal, Sukanta Saha, Saimon Bhuiyan, Mahdiya Rahman Sukanya, Rahatun Nesa Priti, Md. Iqramul Hoque, Munem Shahriar, Raima Islam, Sanatan Sushil, Shahriyar Zaman Ridoy, Kazi Rajwan Sultan, MD Shafikul Islam, Md Manjurul Ahsan, and **Md Rizwan Parvez**. InvariantBench: Can Large Language Models Exhibit Inherent Reasoning Across Equivalent Transformations?
- AI4Math Worskhop at ICML 2026
- Under review ICLR, 2027

P.7 Tien Anh Nguyen, Mohammed Eunus Ali, Muhammad Aamir Cheema, Tanzima Hashem, and **Md Rizwan Parvez**. HippoSpatial: Relational Scene Graph for Multi-View Spatial Reasoning.
- Under review ICLR, 2027

P.8 Wahid Faisal, Azmine Toushik Wasi, Mohammed Eunus Ali, and **Md Rizwan Parvez**. FrontierSpatial: Holistic Evaluation of Spatial Reasoning in Multimodal Large Language Models.
- AI4Math Worskhop at ICML 2026
- Under review, ICLR 2027

P.9 Xuwang Yin, Cong Geng, Hongyi Qian, Shiying Li, **Md Rizwan Parvez**, and Xu-Cheng Yin. Relay Sampling: Cross-Label Warm Starts for Short-Run Joint Energy Models.
- Under review ICLR, 2027

P.10 Azmine Toushik Wasi, Wahid Faisal, Mohammed Eunus Ali, and **Md Rizwan Parvez**. DeepVisualMath: Comprehensive Evaluation of Visual Mathematical Reasoning in Large Multimodal Models.
- Under review ICLR, 2027

P.11 Zabir Al Nazi, Shubhashis Roy Dipta, and **Md Rizwan Parvez**. Omni-Modal Dissonance Benchmark: Systematically Breaking Modality Consensus to Probe Robustness and Calibrated Abstention.
- A2A-MML, CViW, and KnowledgeMR Workshops at CVPR 2026
- Under review, WACV 2027

P.12 Tusher Bhomik, Soumik Bhattacharjee, and **Md Rizwan Parvez**. Beyond Pass or Fail: Trajectory-Grounded Failure Attribution for LLM Agents on CTF Tasks.
- Under review, MSR 2027

P.13 Md Tanzib Hosain, Shadril Hassan Shifat, Md Mofijul Islam, Tasnim Mohiuddin, Md Kishor Morol, and **Md Rizwan Parvez**. BnIdiomEval: LLMs in the Translation of Bengali Idioms.
- Under review, NAACL 2027

P.14 Ahmed Masry, Megh Thakkar, Nafis Tahmid Chowdhury, **Md Rizwan Parvez**, and Enamul Hoque. ChartQA-Video: Long-Context Spatio-Temporal Reasoning for Data Animations.
- Under review, NAACL 2027

P.15 Sheikh Tanvir Ahmed, Md Adnan Arefeen, and **Md Rizwan Parvez**. ChartOrchestra: Automating Sub-Agents in Agentic Orchestration for Chart Reasoning.
- Under review, NAACL 2027

P.16 Md Tanzib Hosain, Shadril Hassan Shifat, Jahin Sultana, and **Md Rizwan Parvez**. SWE-Discussion: Collaborative Multi-Agent Discussion for Software Issue Resolution.
- Under review, NAACL 2027

P.17 Md. Tanzib Hosain, Salman Rahman, Md Kishor Morol, and **Md Rizwan Parvez**. Xolver: Multi-Agent Generalist Problem Solving with Holistic Experience Learning.
- AI4Math Workshop at ICML 2026
- Under review, NAACL 2027

Top-performing results on 14 reasoning, math, coding, and agentic tasks, including GSM8K, Math-500, AIME'24/'25, LiveCodeBench, GPQA (Diamond), Humanity's Last Exam (HLE), BrowseComp, GAIA, OS-World (Verified), and SWE-Bench Verified (bash-only)

P.18 Md. Ashraful Islam, Mohammed Eunus Ali, Xiaoning Du, and **Md Rizwan Parvez**. VibeSweBench: Can AI Co-Worker Agents Do Real-World Software Engineering by Vibe Coding?
- DL4C Workshop at ICML 2026
- Under review, NAACL 2027

P.19 Md Abrar Jahin, Md. Akramuddoula, **Md Rizwan Parvez**, Jay Pujara, and Craig A Knoblock. Certified Obfuscation for Privacy-Preserving Learning: Reconstruction Risk and a Keyed Rotation–Noise–Permutation Mechanism.
Under review, AAAI 2027

P.20 Md Abrar Jahin, Md. Akramuddoula, Nawaz Talukder Sanglap, **Md Rizwan Parvez**, Jay Pujara, and Craig A Knoblock Refusal Is Layer-Distributed: A Causal Score for Where Safety Lives and an Adaptive Ablation That Uses It.
Under review, AAAI 2027

P.21 Md Abrar Jahin, Tithi Kirttania, Liyan Ghosh, and **Md Rizwan Parvez**. PhyloAgent: Contrastive Failure Localization via Successful-Trace Alignment in Multi-Agent LLM Systems.
Pre-print

P.22 Md. Ashraful Islam, Satak Kumar Dey, Sukarna Barua, Mohammed Eunus Ali, and **Md Rizwan Parvez**. Can AI Coding Agents Do Real-Wrold PR Reviews?
Pre-print

P.23 Md Abrar Jahin, Md. Akramuddoula, **Md Rizwan Parvez**, Jay Pujara, and Craig A Knoblock. Aethelred: Certifying Causally Faithful Explanations for Graph Neural Networks.
Pre-print

P.24 S. M. Muhtasimul Hasan, Md Abrar Jahin, **Md Rizwan Parvez**, Md. Jakir Hossen, and Dr. M. F. Mridha. Instance-BEq: Formal Faithfulness Certification for Structure-to-Instance Theorem Autoformalization.
Pre-print

P.25 S. M. Muhtasimul Hasan, Md Abrar Jahin, **Md Rizwan Parvez**, Md. Jakir Hossen, and Dr. M. F. Mridha. Conformance-Aware Repair of HTN Planning Domains: The Precision Gap and How to Close It.
Pre-print

P.26 **Md Rizwan Parvez***, Dhiman Paul*, Nabeel Mohammed, and Shafin Rahman. VideoLights: Cross-Modal Cross-Task Transformer for Joint Video Highlight Detection and Moment Retrieval.
Preprint

P.27 Zhiruo Wang, Jun Araki, Zhengbao Jiang, **Md Rizwan Parvez**, and Graham Neubig. Learning to Filter Context for Retrieval-Augmented Generation.
Preprint

Listed in AI-Guides' (30 K+ GitHub stars) most impactful RAG papers

Journal Publications

2026 J.1 Tien Anh Nguyen, Mohammed Eunus Ali, Muhammad Aamir Cheema, Tanzima Hashem, and **Md Rizwan Parvez**. UrbanAgentBench: Evaluating Multimodal Agents for Real-World Urban Task Solving in Fragmented Spatial Environments.
In submission to TMLR

J.2 Mahir Labib Dihan, Tanzima Hashem, Mohammed Eunus Ali, and **Md Rizwan Parvez**. WebOperator: Action-Aware Tree Search for Autonomous Web Agents.
In submission to TMLR

Among the top-scoring open-source agents on WebArena

J.3 Asif Shahriar, Md Nafiu Rahman, Sadif Ahmed, Farig Sadeque, and **Md Rizwan Parvez**. A Survey on Agentic Security: Applications, Threats and Defenses.
- AIWILD Workshop @ ICLR 2026
- TMLR 2026

J.4 Amira Abdalla, **Md Rizwan Parvez**, and Tanvir Alam. A Train-Once, Evaluate-Twice Protocol for Multi-Language Depression Detection with Large Language Models.
- Nature Scientific Reports 2026

Peer-Reviewed Conference Publications

2026 P.28 Md Abrar Jahin and **Md Rizwan Parvez**. GUI-Primitives: Diagnosing Spatial Reasoning Failures in Vision-Language GUI Grounding.
- EMNLP 2026

P.29 Md Rayhanul Masud, Azmine Toushik Wasi, Salman Rahman, and **Md Rizwan Parvez**. Reward Engineering for Software Tasks: A Survey of Reinforcement Learning Approaches.
- Findings of EMNLP 2026

P.30 Khang Tran, Yazan Boshmaf, Issa Khalil, NhatHai Phan, Ting Yu, and **Md Rizwan Parvez**. Poison with Style: A Practical Poisoning Attack on Code Large Language Models.
- ICML 2026

P.31 Azmine Toushik Wasi, Shahriyar Zaman Ridoy, S. M. Muhtasimul Hasan, Koushik Ahamed Tonmoy, Kinga Tshering, Wahid Faisal, Tasnim Mohiuddin, and **Md Rizwan Parvez**. TimeSpot: Benchmarking Geo-Temporal Understanding in Vision–Language Models in Real-World Settings.
- AI4Math Workshop @ ICML 2026
- ICML 2026

P.32 Azmine Toushik Wasi, Wahid Faisal, Abdur Rahman, et al., Mohammed Eunus Ali, Majd Hawasly, Mohammad Raza, and **Md Rizwan Parvez**. SpatiaLab: Can Vision–Language Models Perform Spatial Reasoning in the Wild?
- ICLR 2026

P.33 Shamit Fatin, Md Nazmul Islam Ananto, Mohammed Eunus Ali, and **Md Rizwan Parvez**. CompassLLM: A Chain-of-Thought Prompting Approach toward Geo-Spatial Reasoning for Popular Path Discovery and Synthesis Using LLMs.
- Findings of ACL 2026

P.34 Azmine Toushik Wasi, Wahid Faisal, Mst Rafia Islam, and **Md Rizwan Parvez**. Mina: A Multilingual LLM-Powered Legal Assistant Agent for Empowering Access to Justice in Bangladesh.
- AI4Law and MusIML Workshop @ ICML 2026
- Findings of ACL 2026

P.35 Raman Saparkhan, Majd Hawasly, **Md Rizwan Parvez**, and Mohammad Raza. Self-Consistency from Only Two Samples: CoT–PoT Ensembling for Efficient LLM Reasoning.
- Findings of ACL 2026

P.36 Md Hasebul Hasan, Mahir Labib Dihan, Mohammed Eunus Ali, and **Md Rizwan Parvez**. MapAgent: A Hierarchical Agent for Geospatial Reasoning with Dynamic Map Tool Integration.
- Findings of EACL 2026

P.37 Asif Azad, Mohammad Sadat Hossain, MD Sadik Hossain Shanto, M Saifur Rahman, and **Md Rizwan Parvez**. The Art of Saying 'Maybe': A Conformal Lens for Uncertainty Benchmarking in VLMs.
Findings of EACL 2026

P.38 Aaryaman Kartha, Ahmed Masry, Mohammed Saidul Islam, et al., **Md Rizwan Parvez**, Enamul Hoque, and Shafiq Joty. DashboardQA: Benchmarking Multimodal Agents for Question Answering on Interactive Dashboards.
Findings of EACL 2026

P.39 Azmine Toushik Wasi, Md. Masudur Rahman, Abdur Rahman, Mohsin Mahmud Topu, Mohammed Eunus Ali, Riashat Islam, and **Md Rizwan Parvez**. Spatial Reasoning in Multimodal Foundation Models: A Survey of Spatial Understanding in LLMs and VLMs.
AACL 2026

P.40 Shahriyar Zaman Ridoy, Azmine Toushik Wasi, Koushik Ahamed Tonmoy, and **Md Rizwan Parvez**. Geo-spatial and Geo-temporal Reasoning in Vision-Language and Large Language Models: A Review.
AACL 2026

P.41 Azmine Toushik Wasi, Sadia Tasnim Meem, Riashat Islam, and **Md Rizwan Parvez**. Large Language and Vision-Language Models as World Models: A Survey.
Findings of AACL 2026

P.42 Azmine Toushik Wasi, Abdur Rahman, Riashat Islam, and **Md Rizwan Parvez**. Is There a Dr. House in the Machine? A Survey of Clinical and Diagnostic Reasoning in LLM-based Medical Agents.
Findings of AACL 2026

2025 P.43 Mahir Labib Dihan, MD Tanvir Hassan, Md Tanvir Parvez, Md Hasebul Hasan, Md Almash Alam, Muhammad Aamir Cheema, Mohammed Eunus Ali, and **Md Rizwan Parvez**. MapEval: A Map-Based Evaluation of Geo-Spatial Reasoning in Foundation Models.
ICML 2025 **Spotlight, Top 2%**
313 out of 12,107 submissions
Catalyst to Google Ask Maps

P.44 Salman Rahman, Liwei Jiang, James Shiffer, Genglin Liu, Sheriff Issaka, **Md Rizwan Parvez**, Hamid Palangi, Kai-Wei Chang, Yejin Choi, and Saadia Gabriel. X-Teaming: Scalable Red-Teaming and Safeguarding of Interactive Multi-Turn Conversational AI Systems.
COLM 2025

P.45 Salman Rahman, Sheriff Issaka, Ashima Suvarna, Genglin Liu, James Shiffer, Jaeyoung Lee, **Md Rizwan Parvez**, Hamid Palangi, Shi Feng, Nanyun Peng, Yejin Choi, Julian Michael, Liwei Jiang, and Saadia Gabriel. AI Debate Aids Assessment of Controversial Claims.
NeurIPS 2025

P.46 Ahmed Lekssays, Utsav Shukla, Husrev Taha Sencar, and **Md Rizwan Parvez**. TechniqueRAG: Retrieval Augmented Generation for Adversarial Technique Annotation in Cyber Threat Intelligence Text.
Findings of ACL 2025

P.47 Ahmed Masry, Mohammed Saidul Islam, Mahir Ahmed, et al., **Md Rizwan Parvez**, Enamul Hoque, and Shafiq Joty. ChartQAPro: A More Diverse and Challenging Benchmark for Chart Question Answering.

Findings of ACL 2025

Evaluation Dataset Adopted by Claude Mythos/Fable 5.

Training data and evaluation suites also adopted by ByteDance Seed 2.0, Alibaba Ovis 2.5 (evaluation) and ByteDance CodeDance (training)

P.48 Md Ashraful Islam, Mohammed Eunus Ali, and **Md Rizwan Parvez**. CodeSim: Multi-Agent Code Generation and Problem Solving through Simulation-Driven Planning and Debugging.

Findings of NAACL 2025

Second best performing coding agent on HumanEval – Feb 2025

P.49 Saleh Musleh, Hamada R. H. Al-Absi, **Md Rizwan Parvez**, Anant Pai, Ghassan Salih, and Tanvir Alam. DR-VQA: Large-Language-Model-based Vision Question Answering System on Diabetic Retinopathy.

TENCON 2025

2024 P.50 Md Tahmid Rahman Laskar, Sawsan Alqahtani, M Saiful Bari, Mizanur Rahman, Mohammad Abdullah Matin Khan, Haidar Khan, Israt Jahan, Amran Bhuiyan, Chee Wei Tan, **Md Rizwan Parvez**, Enamul Hoque, Shafiq Joty, and Jimmy Huang. A Systematic Survey and Critical Review on Evaluating Large Language Models: Challenges, Limitations, and Recommendations.

EMNLP 2024 **Top-29 Most-Cited Paper at EMNLP 2024**

P.51 Md Ashraful Islam, Mohammed Eunus Ali, and **Md Rizwan Parvez**. MapCoder: Multi-Agent Code Generation for Competitive Problem Solving.

ACL 2024 **Top-37 Most-Cited Paper at ACL 2024**

Adopted as a baseline in Meta RLEF (Llama 3.1 post-training) for code generation

The top-performing coding agents on HumanEval, MBPP, and CodeContests – Dec 2024

P.52 Mohammad Abdullah Matin Khan, M Saiful Bari, Xuan Long Do, Weishi Wang, **Md Rizwan Parvez**, and Shafiq Joty. xCodeEval: An Execution-based Large-Scale Multilingual Multitask Benchmark for Code Understanding, Generation, Translation and Retrieval.

ACL 2024

Widely adopted in LLM training and evaluation with **6.2 Million+ downloads on Hugging Face**

P.53 Shayekh Bin Islam, Md Asib Rahman, K S M Tozammel Hossain, Enamul Hoque, Shafiq Joty, and **Md Rizwan Parvez**. Open-RAG: Enhanced Retrieval Augmented Reasoning with Open-Source Large Language Models.

Findings of EMNLP 2024

Press coverage in MarkTechPost News, reaching 1.5 M+ monthly readers (Oct 2024)

P.54 Mohammed Saidul Islam, Md Tahmid Rahman Laskar, **Md Rizwan Parvez**, Enamul Hoque, and Shafiq Joty. DataNarrative: Automated Data-Driven Storytelling with Visualizations and Texts.
EMNLP 2024

P.55 Ahmed Masry, Mehrad Shahmohammadi, **Md Rizwan Parvez**, Enamul Hoque, and Shafiq Joty. ChartInstruct: Instruction Tuning for Chart Comprehension and Reasoning.
Findings of ACL 2024

2023 P.56 **Md Rizwan Parvez**, Jianfeng Chi, Wasi Uddin Ahmad, Yuan Tian, and Kai-Wei Chang. Retrieval-Enhanced Data Augmentation for Question Answering on Privacy Policies.
EACL 2023

P.57 Mobashir Sadat, Zhengyu Zhou, Lukas Lange, Jun Araki, et al., **Md Rizwan Parvez**, and Zhe Feng. DelucionQA: Detecting Hallucinations in Domain-specific Question Answering.
Findings of EMNLP 2023

2021 P.58 **Md Rizwan Parvez**, Wasi Uddin Ahmad, Saikat Chakraborty, Baishakhi Ray, and Kai-Wei Chang. Retrieval Augmented Code Generation and Summarization.
Findings of EMNLP 2021 **Top-10 Most-Cited SE ArXiv Paper, 2021**
Method adopted in Microsoft AutoGen, a widely used open-source multi-agent framework with 60K+ GitHub stars
Data and evaluation suites were used in CodeT5+, the fully open-sourced LLM on par to ChatGPT on code generation up to 2023

P.59 **Md Rizwan Parvez** and Kai-Wei Chang. Evaluating the Values of Sources in Transfer Learning.
NAACL 2021

2019 P.60 **Md Rizwan Parvez**, Tolga Bolukbasi, Kai-Wei Chang, and Venkatesh Saligrama. Robust Text Classifier on Test-Time Budgets.
EMNLP 2019

2018 P.61 **Md Rizwan Parvez**, Saikat Chakraborty, Baishakhi Ray, and Kai-Wei Chang. Building Language Models for Text with Named Entities.
ACL 2018

P.62 Sarah Masud Preum, **Md Rizwan Parvez**, Kai-Wei Chang, and John Stankovic. A Corpus of Drug Usage Guidelines Annotated with Type of Advice.
LREC 2018

2016 P.63 **Md Rizwan Parvez**, Turash Mosharraf, and Mohammed Eunus Ali. A Novel Approach to Identify Spatio-Temporal Crime Pattern in Dhaka City.
ICTD 2016

Workshops and Demo Papers

2026 W.1 Md Rayhanul Masud and **Md Rizwan Parvez**. TAME: Token Attribution and Masking for Emergent misalignment.
Uncertainty-Aware NLP Workshop @ EMNLP 2026

W.2 Niloy Kumar Mondal and **Md Rizwan Parvez**. QuantumQUBO Agent: Automating Quadratic Unconstrained Binary Optimization (QUBO) Formulation Generation from Natural Language.
AI4Research Workshop @ ICML 2026

Featured in GroundWork, and Atlabyte News

2025 W.3 **Md Rizwan Parvez**. Chain of Evidences and Evidence to Generate: Prompting for Context-Grounded and Retrieval-Augmented Reasoning.
KnowledgeNLP Workshop @ NAACL 2025

W.4 Mahir Labib Dihan, Mohammed Eunus Ali, and **Md Rizwan Parvez**. MapQaTor: A System for Efficient Annotation of Map Query Datasets.
ACL 2025 (System Demonstration)

2024 W.5 Sabri Boughorbel, **Md Rizwan Parvez**, and Majd Hawasly. Improving Language Models Trained on Translated Data with Continual Pre-Training and Dictionary Learning Analysis.
ArabNLP Workshop @ ACL 2024

GRANTS CONTRIBUTED

2020 NSF CAREER, National Science Foundation
Towards Generalizable Natural Language Processing: Cross-Lingual and Cross-Task Transfer. *Proposal writing invited by Ph.D. advisor Kai-Wei Chang, UCLA. Our work on Shapley value based data selection for language modeling [P.61], Shapley value based data attribution [P.59], and downstream tasks such as text classification [P.60] served as catalysts for the grant application.*

2025 ARC Discovery Project, Australian Research Council
Advancing Geo-spatial Agentic Intelligence for Inclusive Urban Services. *Contributed to proposal writing for Prof. Mohammed Eunus Ali (Monash University). The MapEval [P.43], MapAgent [P.36], MapQaTor [W.4], CompassLLM [P.33], and SpatiaLab [P.32] projects served as the primary catalyst for the grant application.*

PROFESSIONAL EXPERIENCES

Qatar Computing Research Institute (QCRI) — **Doha, Qatar**
07/2023–current Scientist
Verifiable, attributed, and grounded NLP: code generation, RAG, and agents.

Bosch Research — **Sunnyvale, CA**
08/2022–12/2023 Research Scientist

Research Internships
06/2021–09/2021 **Google Research**, Kirkland, WA — with *Bin Zhang*
06/2020–09/2020 **Facebook AI Research (FAIR)**, Seattle, WA — with *Scott Wen-tau Yih*
06/2019–09/2019 **Salesforce Research**, Palo Alto, CA — with *Bryan McCann*
06/2018–09/2018 **Microsoft AI & Research**, Redmond, WA — with *Paul Mineiro*

TALKS

2026 **Language Agents for Software Development**
Invited Talk, University of North Texas Data Science Seminar (2026)

2021 **Language Processing with Auxiliary Supervision and World Knowledge**
Invited Talk, Microsoft Research, MBZUAI, U. of Washington, UC Santa Barbara, and AWS AI (2021)

2019 **Fast and Robust Text Classification**
Speaker, Sixth International Conference on Networking, Systems and Security (NSysS) (2019)

TEACHING EXPERIENCES

Teaching Assistant, Associate, and Fellow

2018–2020 **CSM146 Introduction to Machine Learning**, UCLA
Teaching Fellow: Fall 2018, Fall 2019, Winter 2020, Fall 2020

2020 **CS31 Introduction to Computer Science I**, UCLA
Teaching Fellow: Spring 2020

2019–2020 **CS33 Computer Organization**, UCLA
Teaching Associate: Winter 2019, Spring 2020

2017 **CS269 Natural Language Processing**, UCLA
Teaching Assistant: Fall 2017

2017 **Natural Language Processing**, University of Virginia
Teaching Assistant: Spring 2017

2016 **Principles of Software Design**, University of Virginia
Teaching Assistant: Fall 2016

Instructor

2015–2016 **Numerical Analysis**, United International University (UIU), Bangladesh
Instructor: Fall 2015, Spring 2016

2016 **C Programming Language**, United International University (UIU), Bangladesh
Instructor: Spring 2016

2015–2016 **Digital Logic Design**, United International University (UIU), Bangladesh
Instructor: Fall 2015, Spring 2016

MENTORING EXPERIENCES

PostDocs and Ph.D. Students

08/2024–08/2025 **Ahmed Lekssays** (PostDoc, QCRI)
RAG for adversarial technique annotation in cyber threat intelligence [P.46].

01/2026–present **Md Abrar Jahin** (Ph.D. student, University of Southern California)
Agent design and their failure localization [P.28, P.21].
Certified model behavior [P.23, P.20, P.19, P.24, P.25].

08/2023–present **Ahmed Masry** (Ph.D. student, York University)
Chart question answering and instruction tuning [P.47, P.55].
Chart-video spatio-temporal reasoning [P.14].
Dashboard question answering [P.38].

11/2024–present **Md Ashraful Islam** (MSc, BUET → Ph.D. student, Monash University as a *Fulbright Fellow*)
Multi-agent code generation [P.51].
Simulation-driven code synthesis and debugging [P.48].
AI co-worker agents for software engineering [P.18].
Automated pull-request review [P.22].

02/2024–present **Salman Rahman** (Ph.D. student, UCLA)
Scalable red-teaming of multi-turn conversational AI [P.44].
AI debate for controversial claims [P.45].
Multi-agent generalist problem solving [P.17].
Reward engineering for software tasks [P.29].

01/2026–present **Tien Anh Nguyen** (Ph.D. student, Monash University)
Multimodal agents for urban task solving [J.1]
Spatial Reasoning [P.7].

01/2026–present **Zabir Al Nazi** (Ph.D. student, University of California, Riverside)
Robustness and calibrated abstention in omni-modal models [P.11].

01/2026–present **Shubhashis Roy Dipta** (Ph.D. student, University of Maryland, Baltimore County)
Robustness and calibrated abstention in omni-modal models [P.11].

01/2026–06/2026 **Md Rayhanul Masud** (Ph.D., University of California, Riverside)
Reward engineering for software tasks [P.29, W.1].

08/2023–08/2024 **Khang Tran** (Ph.D., NJIT)
Poisoning attacks on code large language models [P.30].

01/2023–08/2023 **Zhiruo Wang** (Ph.D. student, Carnegie Mellon University)
Filtering context for retrieval-augmented generation [P.27].
Recipient of Google PhD Fellowship, Machine Learning and ML Foundations 2025-2026.

Undergraduate, Master, and Pre-Doctoral Students

08/2025–present **Azmine Toushik Wasi** (Shahjalal University of Science and Technology)
Spatial and geo-temporal benchmarking of VLMs [P.32, P.31, P.8].
Reasoning invariance across equivalent transformations [P.6].
LLM surveys (world models), clinical reasoning, spatial reasoning (VLMs & LLMs), geo-temporal reasoning, visual math, reward engineering [P.41, P.42, P.39, P.40, P.10, P.29].
Legal assistant agent [P.34].

01/2024–present **Mahir Labib Dihan** (Master's student, BUET)
Geo-spatial/Map reasoning evaluation [P.43, P.36, W.4].
Autonomous web agents [J.2].
Tree search for autonomous software engineering [P.1].

08/2023–present **Md. Tanzib Hosain** (American International University → Max Planck Institute)
Multi-agent problem solving [P.17], P.16.
Low-resource machine translation of Bengali idioms [P.13].

01/2026–present **Tusher Bhomik** and **Soumik Bhattacharjee** (Undergrad, BUET)
Failure attribution for LLM agents on CTF tasks [P.12].

01/2026–present **Niloy Kumar Mondal**
LLM-based QUBO formulation from natural language [W.2].

08/2024–08/2025	**Md Hasebul Hasan** (MSc, BUET → Ph.D., University of Texas, Arlington) Geospatial reasoning agent [P.36]. Map-based evaluation [P.43].
08/2024–08/2025	**Md Nazmul Islam Ananto** (Undergrad BUET → Ph.D., Penn State U.) Geo-spatial reasoning with LLMs for popular path discovery [P.33].
08/2025–06/2026	**Asif Shahriar** (BUET) Survey on agentic security: applications, threats, and defenses [J.3].
08/2025–06/2026	**Asif Azad** (BUET) Conformal uncertainty benchmarking in VLMs [P.37].
08/2024–08/2025	**Amira Abdalla** (Master's student, HBKU) Multi-language depression detection with LLMs [J.4].

PROFESSIONAL SERVICE

Organizing Committees

2022 **CodexEval** – accepted workshop proposal, joint call of ACL/EMNLP/EACL

Paper Reviewing

Session Chair

Conf. **EACL** 2026

Senior Area Chair

Conf. **ACL** 2027 (NLP and Code Models), **ACL** 2026 (Code Models), **EMNLP** 2026 (NLP and Code), **EACL** 2026 (Language Modeling), **EMNLP** 2025 (Code Models)

Area Chair / Senior Program Committee / Action Editor

Conf. **NAACL** 2027, **AAAI** 2027, **NeurIPS** 2026, **COLM** 2026, **ACL** 2025

Reviewer / Program Committee

Conf. **ACL**, **EMNLP**, **NAACL** 2019–2024, **AAAI**, **ICLR** since 2022, **ICML** since 2026, **NeurIPS** 2024–2025

Journal **JAIR** 2025

Community Service

2017–2022 **Internal Reviewer**, UCLA NLP Group
Reviewed paper submissions for each conference cycle.

2018 **Reviewer**, Southern California (SoCal) NLP Summit
Reviewed paper submissions at the symposium.

SELECTED RESEARCH MEDIA COVERAGE

2024 Open-RAG: Enhanced Retrieval Augmented Reasoning with Open-Source Large Language Models
MarkTechPost (2024.10)

REFERENCES

Note: Confidential recommendation letters are delivered via Interfolio Dossier Delivery. You may request letters using the addresses labeled "Letter Email" below. For questions, contact letter writers directly via their official emails.

Kai-Wei Chang (Ph.D. Advisor)
Professor
Computer Science, University of California, Los Angeles (UCLA)
Official Email: kwchang@cs.ucla.edu
Letter Email: send.AI4SE.AEC25AB7AA@interfoliodossier.com

Baishakhi Ray
Associate Professor
Computer Science, Columbia University
Official Email: rayb@cs.columbia.edu
Letter Email: send.NLP4SE.B815613B22@interfoliodossier.com

Mohammed Eunus Ali
Associate Professor (US-equivalent)
Software Systems and Cybersecurity, Monash University
Official Email: eunus.Ali@monash.edu
Letter Email: send.Ali.5CC421D97F@interfoliodossier.com

Updated Sep 2026