Hani Alomari

hani@vt.edu hanialomari.github.io Google Scholar LinkedIn GitHub

Also published as: Hani Al-Omari, Hani AlOmari, Hani B. Alomari

Research Summary

I develop **reliable multimodal AI**: vision–language and acoustic perception systems that stay accurate when their inputs are *ambiguous* (images with multiple valid meanings), *corrupted* (noisy or misspelled queries, degraded images), *biased* (models favoring superficial cues), or *missing a modality* (robots mapping rooms from echoes where cameras fail). My work spans cross-modal retrieval, visual grounding, bias diagnosis in large vision–language models, long-video and event understanding, and camera-free spatial perception, and it builds on earlier work on subjective and adversarial language (emotion, propaganda) and machine-learning security (malware detection).

At a glance: 15 archival peer-reviewed publications (CVPR, ACL, EMNLP, ECCV, NeurIPS, IEEE, ACM), including three EMNLP 2026 main-conference papers (15.3% acceptance) • 250+ citations, h-index 8 • EMNLP 2026 SAC Highlights (one of 22 papers) • Amazon Fellow • AI prototype evaluated in a U.S. hospital • Reviewer for ICLR, CVPR, ECCV, ACL ARR • Collaborations with Columbia University, Carilion Clinic, the NSF Imageomics Institute, and Adobe.

Education

Virginia Tech, Blacksburg, VA Jan 2023 – Dec 2027 (expected)
Ph.D. Candidate, Computer Science | Advisor: Dr. Chris Thomas | GPA: 4.0/4.0 | Amazon Fellow

Jordan University of Science and Technology (JUST) Feb 2020 – Jun 2022
M.S., Data Science | Advisor: Prof. Rehab Duwairi | GPA: 4.26/4.30 | Full scholarship (Outstanding Graduate Student Award)

Jordan University of Science and Technology (JUST) Feb 2016 – Jan 2020
B.S., Computer Science | Advisor: Dr. Malak Abdullah | GPA: 4.04/4.20 | Ranked #1 of 62

Research Experience

Graduate Research Assistant, Virginia Tech Mar 2023 – Present

Robust and polysemous vision–language retrieval

- **Lenses (CVPR 2026).** Built a polysemous image–text benchmark (105K images; 732K automatically generated, model-verified captions across five interpretive lenses: literal, figurative, abstract, background, emotional) and a lens-token multi-embedding model that outperforms the evaluated baselines on literal and non-literal retrieval (figurative image-to-text R@1: 56.0 vs. 28.2 for the best evaluated zero-shot baseline). *Led problem formulation, dataset design, modeling, experiments, and writing.*
- **MaxMatch (ACL 2025).** Proposed Maximal Pair Assignment Similarity (Hungarian one-to-one set matching) with two diversity losses that prevent set collapse; state of the art on Flickr30K and COCO without external data (+27.3 RSUM on Flickr30K and +21.3 on COCO 5K over SetDiv, ResNet-152 + Bi-GRU). Used as a state-of-the-art baseline by independent researchers (CVPR 2026 Findings). *Led method design, implementation, and evaluation.*
- **Query-Channel Inference (under review).** Recast test-time adaptation as test-time inference for character-corrupted text queries: estimate the corruption channel online with EM and invert it under frozen language-model priors. On the COCO-C and Flickr-C synthetic-corruption benchmarks, recovers 92% and 94% of the retrieval accuracy lost to corruption (vs. $\leq$61% for 7–8B LLM spellers), with all weights frozen.
- **JourneyBench (NeurIPS 2024, with Columbia University).** One-stop benchmark testing vision–language models on generated images, including hallucination triggers and fine-grained retrieval with distractors. *Contributed to benchmark design discussions, annotation, and implementation and evaluation of state-of-the-art models.*

Reliability, bias, and grounding in multimodal LLMs

- **Length bias in LVLMs (EMNLP 2026, SAC Highlights).** Showed that large vision–language models systematically prefer longer answer options (accuracy drops by 2–38 percentage points when the correct answer is shorter), traced the bias to late-layer suppression of visual evidence, and showed that inference-time mitigations alone are insufficient. *Contributed to study design, experiments across model families, and analysis.*
- **Part-level visual grounding via RL (ECCV 2026; NSF HDR Imageomics Institute).** Coarse-to-fine object→part grounding with self-checking and part-aware GRPO rewards; a 4B model outperforms 7B grounding LLMs and SAM3 on PascalPart, PartImageNet, and InstructPart. *Contributed to method brainstorming, experiments, and analysis.*

Video and multimodal event understanding

- **NEST (EMNLP 2026)**: 1,005 full-length movies with ∼103 grounded narrative events each, for long-video event understanding. **M^{3}EC (EMNLP 2026)**: first benchmark for multi-document event extraction and coreference across

news text, video, and audio (+30 F1 video, +40 CoNLL F1 coreference after fine-tuning). **ENTER (MAR Workshop @ NeurIPS 2024, Spotlight, non-archival; with Columbia)**: interpretable event-graph video QA. **TimeCheck (under review)**: controlled study of what improves video temporal grounding in multimodal LLMs; balancing event lengths in training improves mIoU by more than 10 points on long events. *Contributed to task design, experiments, model evaluation, and analysis.*

Spatial perception and real-world deployment

- **EchoBelief (under review).** Camera-free active acoustic mapping: a robot with a speaker and 12-microphone ring maps rooms from echoes using a persistent two-store belief and a PPO policy that decides where to move, when to listen, and when to stop. In 355 simulated test rooms with known poses, reaches 90% wall recall in 70.1% of episodes vs. 49.4–52.0% for three tuned planners, with about half the sensing acquisitions. Built **EchoRooms** (4,026 rooms, 1.58M posed 12-channel room impulse responses). *Led the project end to end.*
- **Surgical instrument recognition (CVPR 2025 FGVC Workshop; with Carilion Clinic).** Real-time, multi-view classifier for 95 ultra-fine-grained instrument types; **prototype deployed and evaluated in Carilion Clinic's sterile processing department**, where six technicians assembling a mock tray rated it 81.7 on the System Usability Scale. *Contributed to model development, experiments, and the deployment study.*
- **Envision (ongoing).** A world-model approach to composed image retrieval that imagines the target image before searching for it, aiming for accurate and efficient search over large image collections, including multi-step searches with successive edits. *Leading the project.*

Research Assistant (2020–2022) and Undergraduate Researcher (2019), JUST 2019 – Nov 2022

- **Subjective and adversarial language.** Built deep models for emotion detection in dialogue (EmoDet, EmoDet2; most-cited work, whose transformer–BiLSTM design an independent group reused in six follow-up studies), propaganda detection in news (NLP4IF 2019; ranked 7th of 25 teams in sentence-level detection), and humor/offense rating (SemEval 2021; 0.954 F1 on humor detection).
- **Machine-learning security.** Developed a LightGBM pipeline (SMOTE, normalization, PCA) for multi-family Android malware classification on CICMalDroid 2020 (Procedia CS 2023, 50+ citations).
- **Low-resource language resources.** Built So2al-wa-Gwab, a 10K-pair natively written Arabic QA dataset; led a 3-person annotation team with 3-stage quality control (ACM TALLIP 2023).

Industry Experience

Machine Learning Engineer Intern, Adobe, San Jose, CA May – Aug 2026

- Built a large-scale multi-persona image annotation pipeline for image-quality evaluation (filtering, tagging, synthetic annotation, quality validation).
- Ran a visual-embedding similarity study and developed a DINOv3 + Transformer quantile-regression model that learns quality distributions of visual attributes from Adobe Stock images.
- Introduced distortion-aware training to predict high-quality targets for degraded images, improving Q-Align on MIT-Adobe FiveK and outperforming the evaluated baselines.

Publications

Vision–language retrieval, reliability, and grounding

- Lenses: Toward Polysemous Vision-Language Understanding. **Hani Alomari**, Ali Asgarov, Chris Thomas. *CVPR 2026 (Main; acceptance rate 25.4%).* [project]
- Investigating Length Bias and Robustness in LVLMs for Multiple-Choice Question Answering. Md Atabuzzaman, **Hani Alomari**, Chris Thomas. *EMNLP 2026 (Main; acceptance rate 15.3%).* **SAC Highlights (one of 22).**
- Reasoning-Guided Part-Level Visual Grounding via Reinforcement Learning. Kazi Sajeed Mehrab, **Hani Alomari**, Najibul Haque Sarker, Chia-Wei Tang, Zaber Ibn Abdul Hakim, Anuj Karpatne, Chris Thomas. *ECCV 2026.*
- Maximal Matching Matters: Preventing Representation Collapse for Robust Cross-Modal Retrieval. **Hani Alomari**, Anushka Sivakumar, Andrew Zhang, Chris Thomas. *ACL 2025 (Main; acceptance rate 20.3%).*
- JourneyBench: A Challenging One-Stop Vision-Language Understanding Benchmark of Generated Images. Zhecan Wang, Junzhang Liu, Chia-Wei Tang, **Hani Alomari**, et al. *NeurIPS 2024 (Datasets & Benchmarks; acceptance rate 25.3%).*

Video, events, and real-world perception

- NEST: Narrative Event Structures in Time for Long Video Understanding. Ali Asgarov, Kaushik Narasimhan, Najibul Haque Sarker, **Hani Alomari**, et al. *EMNLP 2026 (Main; acceptance rate 15.3%).*
- M^{3}EC: A Multimodal Multidocument Benchmark for Event Extraction and Event Coreference Resolution. Zaber Ibn Abdul Hakim, Najibul Haque Sarker, Aafiya Hussain, **Hani Alomari**, Alvi Md Ishmam, Ali Asgarov, Chia-Wei Tang,

Chris Thomas. *EMNLP 2026 (Main; acceptance rate 15.3%).*

- Real-Time Ultra-Fine-Grained Surgical Instrument Classification. Md Atabuzzaman, Gino DiMatteo, **Hani Alomari**, et al. *FGVC Workshop, CVPR 2025.*
- ENTER: Event-Based Interpretable Reasoning for VideoQA. Hammad Ayyubi, Junzhang Liu, Ali Asgarov, Zaber Ibn Abdul Hakim, Najibul Haque Sarker, Zhecan Wang, Chia-Wei Tang, **Hani Alomari**, et al. *MAR Workshop, NeurIPS 2024 (Spotlight; non-archival).*

Language understanding, security, and information integrity

- A Comparative Analysis of Machine Learning Algorithms for Android Malware Detection. **Hani Alomari**, Qussai M. Yaseen, Mohammed Azmi Al-Betar. *Procedia Computer Science 220 (DDSW 2023).*
- So2al-wa-Gwab: A New Arabic Question-Answering Dataset Trained on Answer Extraction Models. **Hani Alomari**, Rehab Duwairi. *ACM TALLIP 2023.*
- DLJUST at SemEval-2021 Task 7: Hahackathon: Linking Humor and Offense. **Hani Alomari**, Isra'a AbedulNabi, Rehab Duwairi. *SemEval 2021.*
- EmoDet2: Emotion Detection in English Textual Dialogue Using BERT and BiLSTM Models. **Hani Alomari**, Malak A. Abdullah, Samira Shaikh. *ICICS 2020 (IEEE).*
- JUSTDeep at NLP4IF 2019 Shared Task: Propaganda Detection Using Ensemble Deep Learning Models. **Hani Alomari**, Malak Abdullah, Ola Al-Titi, Samira Shaikh. *NLP4IF Workshop, EMNLP 2019.*
- EmoDet at SemEval-2019 Task 3: Emotion Detection in Text Using Deep Learning. **Hani Alomari**, Malak Abdullah, Nabeel Bassam. *SemEval 2019.*
- MTRecS-DLT: Multi-Modal Transport Recommender System Using Deep Learning and Tree Models. Ayat Abedalla, Ali Fadel, Ibraheem Tuffaha, **Hani Alomari**, et al. *SNAMS 2019 (IEEE).*

Under review

- EchoBelief: Evidence-Anchored Spatial Memory for Active Acoustic Mapping. **Hani Alomari**, Md Atabuzzaman, Alaa Algargoosh, Chris Thomas. *Under review, ICLR 2027.*
- Mind the Channel: Repairing Character-Corrupted Queries for Robust Cross-Modal Retrieval. **Hani Alomari**, Ali Asgarov, Chris Thomas. *Under review, ACL ARR (Oct. 2026).*
- TimeCheck: What Really Improves Video Temporal Grounding in Multimodal LLMs? Ali Asgarov, Adhitya Srinivasan, **Hani Alomari**, Zijia Lu, Chris Thomas. *Under review, ACL ARR (Oct. 2026).*

Honors, Awards, and Research Support

- **SAC Highlights**, EMNLP 2026 (one of 22 papers selected among 2,704 main-conference papers).
- **Amazon Fellowship**, Amazon–Virginia Tech Initiative for Efficient and Robust Machine Learning (2026–2027).
- CCI SWVA Cyber Innovation Scholar, Commonwealth Cyber Initiative ($1,500, 2026).
- Travel grants from the Sanghani Center, CS Department, and GPSS, Virginia Tech ($2,000, 2025).
- Research supported by the NSF HDR Imageomics Institute (OAC-2118240), the VT Academy of Data Science Discovery Fund (Award 238695), and the Carilion Clinic Research Acceleration Program.
- Outstanding Graduate Student Award (full M.S. scholarship), JUST (2020–2022).
- Outstanding Undergraduate Student Award (ranked #1 of 62), JUST (2020); multiple Honors List mentions.

Service, Teaching, and Presentations

- **Reviewer:** ICLR (2027); CVPR (2026); ECCV (2026); ACL Rolling Review (Feb/May 2025, Feb/Aug 2026); WACV (2024).
- **Graduate Teaching Assistant, Virginia Tech:** CS4804 Intro to AI (Spring 2025); CS5914 AI Tools in Software Development (Fall 2023, Spring/Fall 2024); CS1064 Intro to Programming in Python (Summer 2023/2024); CS5824/ECE5424 Advanced Machine Learning (Spring 2023).
- **Presentations:** Posters at CVPR 2026 (Denver, CO) and ACL 2025 (Vienna, Austria).

Technical Skills

Programming:	Python, C++, Java, Bash, SQL, LaTeX
ML & Data:	PyTorch, Transformers, TensorFlow, Keras, scikit-learn, NumPy, Pandas, spaCy, NLTK; multimodal LLM fine-tuning (SFT, GRPO/PPO)
Tools:	Conda, Git, Docker, Weights & Biases, Codex CLI, Claude Code, GitHub Copilot
Languages:	English, Arabic