

GHAZANFAR ALI, PH.D.

APPLIED AI RESEARCH SCIENTIST | RESEARCH ENGINEER | TECHNICAL LEAD

Multimodal AI, embodied agents, human motion, computer vision, LLM/RAG systems and real-time deployment

Seoul, South Korea | ghazanfar.6785@gmail.com | [LinkedIn](#) | [ORCID](#) | [Google Scholar](#)

PROFESSIONAL SUMMARY

End-to-end AI research builder and Assistant Professor with 9+ years of experience turning ambiguous research ideas into datasets, models, benchmarks, real-time products, user studies, and peer-reviewed publications. Deep expertise in multimodal and embodied AI, co-speech motion, contrastive retrieval, LLM/RAG agents, computer vision, and Unity/web deployment. Combines research-scientist depth with research-engineering execution: owns architecture, implementation, optimization, evaluation, demonstration, and technical leadership across Python, PyTorch, Unity, Node.js, Three.js, and production-oriented server-client systems.

SELECTED IMPACT

- Built a real-time conversational digital-human stack across LLM, ASR, TTS, gestures, facial/body animation, Python services, Unity, desktop, and web; gesture generation returns each two-second segment in under 50 ms.
- Architected ConGRets for approximately 0.5 ms/query CPU retrieval and under 100 MB memory, with 99% speaker-style and 89% content-to-cluster accuracy.
- Improved noisy 2D-to-3D motion matching from 18% to 97%, scaling an animation library from 56 to 2,035 motions and a rule base from 84,000 to 210,000 mappings.
- Built grounded RAG agents over 49.6 million characters of historical records, reaching 88.7 factual accuracy, 92.2 reliability, and 91.1 reasonableness.
- Delivered a medical embodied-agent pipeline that improved answer F1 from 0.492 to 0.779 and was validated in a 113-participant user study.
- Developed a 2.4M-parameter mobile AR segmentation system with 0.971 mIoU and statistically significant HCI gains in a 24-participant study.

CORE CAPABILITIES

AI/ML: Python, PyTorch, PyTorch Lightning, transformers, BERT/Sentence-BERT, contrastive learning, retrieval, diffusion models, RAG, LLM integration, clustering, representation learning, model evaluation.

Motion and digital humans: motion capture, motion tokenization, 2D/3D pose processing, OpenPose, co-speech gesture, facial/body animation, blendshapes, animation blending, Unity3D, Blender.

Systems and web: Python backends, sockets, HTTP, streaming, queues, ASR/TTS APIs, Node.js, Three.js, JavaScript, WordPress, session management, logging, cloud-hybrid and server-client architecture.

Research delivery: dataset design and curation, benchmarking, latency/memory profiling, human studies, crowdsourced evaluation, IRB study preparation, technical writing, team leadership, stakeholder demonstrations.

EXPERIENCE

Assistant Professor

Mar 2026 - Present

Department of AI, Gachon University | South Korea

- Lead applied research and two PhD students across live conversational agents, scalable monocular-video motion ingestion, and persona modeling for LLM-based embodied agents.
- Designed and built LiveAgent end to end: custom LLM chatbot; ASR/TTS; gesture, face and body animation; Python socket/HTTP backend; Unity, desktop and Node.js/Three.js clients; session/logging services.
- Reduced interaction stalls through token streaming, parallel TTS/gesture execution, and server/client queues; fast-model configurations achieve approximately two-second first-word latency followed by continuous conversation.
- Originated MotionBridge and MotionStage, defining a consumer-GPU path from large-scale monocular video to retargetable motion and a synthetic-data renderer with exact ground truth.

InnoCORE Postdoctoral Researcher

Sep 2025 - Feb 2026

LLM Center, KAIST | South Korea

- Implemented the major components of ControllableGesture, enabling text-only authoring of persistent physical personas through explicit handedness, gesture amplitude/speed, prosody, emotion, and semantic controls.
- Built a 99,022-unit, 55-hour, 30-speaker control store and a shared latent architecture supporting both retrieval and decoded motion generation; expert HCI evaluation is being prepared.

Postdoctoral Researcher

Sep 2023 - Aug 2025

Korea Institute of Science and Technology (KIST) | South Korea

- Originated ConGRets, a scalable, text-only gesture-retrieval architecture with motion-derived zero-shot style adaptation and real-time deployment characteristics.
- Built RIDGE, combining an LLM-generated rule base with a deep-learning fallback; improved semantic alignment to GCA 0.73 versus 0.61 and 0.52 baselines.
- Co-developed a high-density historical RAG agent and integrated evidence-grounded answers, voice, and synchronized animation into an interactive Unity avatar.
- Built evaluation harnesses comparing seven recent motion-generation systems across model footprint, VRAM/RAM, and inference latency.

Research Assistant

Sep 2017 - Aug 2023

Korea Institute of Science and Technology (KIST) | South Korea

- Developed GestureCLR and a multilingual text-to-gesture stack with translation, word-level alignment, clustering, Python services, and real-time Unity rendering; validated in a 51-participant study.
- Delivered ASAP, an automated screenplay-to-storyboard/3D/VR pipeline with 93% simple and 87% complex text-to-action accuracy.
- Built medical, no-code, mobile-AR, and HoloLens agent systems spanning RAG, TTS, lip synchronization, 48 facial expressions, object recognition, spatial mapping, and cloud offloading.
- Mentored approximately seven graduate researchers and interns and routinely translated research prototypes into conference-ready live demonstrations.

SELECTED SYSTEMS PORTFOLIO

LiveAgent

Technical lead | Working real-time platform

Reusable multi-client conversational-agent infrastructure for interactive demos and HCI studies.

- Python, sockets/HTTP, streaming LLM, ASR/TTS, gesture server, face/body animation, Unity, desktop, Node.js/Three.js, logging and sessions.

ConGRets

Originator | Paper under review

Low-latency text-to-gesture retrieval with zero-shot speaker style from unlabeled motion and decoded-frame evaluation against generative baselines.

- Global motion-style encoder, BERT-mini text backbone, shared 10D latent space, curated motion library, overlap-aware segmentation, runtime blending.

MotionBridge and MotionStage

Architect and primary developer

Consumer-GPU video-to-motion ingestion plus a companion synthetic-data specification for exact motion ground truth and coherent body/hand data.

- System design follows a 368-work literature audit and prioritizes throughput, yield, usable motion-hour cost, and deployment accessibility.

GQNet and Motion EvalSuite

Lead developer | In development

Human-aligned quality estimation and a unified seven-axis benchmark for co-speech gesture and text-to-motion systems.

- Approximately 35 verified metrics, artifact sensitivity, human-score learning, GENE meta-evaluation design, and resource profiling of seven popular systems.

Embodied RAG Agents

Architecture, implementation and deployment

Historical and medical conversational agents combining dynamic retrieval, source grounding, Unity avatars, speech, lip sync, gesture, and expression.

- Historical pipeline used dynamic chunking, query regeneration, date metadata filtering, and token-budget-aware retrieval; medical pipeline paired Llama2-based RAG with expressive behavior.

Behavioral Imposter Test

Live system delivered | IRB under review

Embodied-persona fidelity study with eight matched personas and 48 controlled builds across lexical, knowledge, voice, and gesture channels.

- Own the LiveAgent delivery, study design, and controlled personas; a PhD student is adapting the platform for crowdsourced evaluation.

INDUSTRY AND WEB LEADERSHIP

Web Development Advisor and Team Lead

Jul 2015 - Present

Penumbra Digital | Remote

- Lead the web-development team and partner with designers to deliver websites and digital platforms for confidential B2B, corporate, NGO, and technology clients across Pakistan, the Middle East, and the United States.
- Own the full delivery lifecycle: technical scoping, WordPress or project-specific JavaScript implementation, deployment, reliability, and post-launch services.

Lab Engineer

Feb 2014 - Jul 2017

University of Central Punjab | Pakistan

- Planned and supervised practical computer-programming laboratories serving approximately 600 undergraduate students.

EDUCATION AND RECOGNITION

Ph.D. in AI-Robotics

2017 - 2023

KIST School, University of Science and Technology (UST), South Korea

Dissertation: Scalable Hybrid Approach of Co-speech Text-to-Gesture Generation for Interactive Digital Humans.

B.E. in Computer Engineering

2009 - 2014

National University of Sciences and Technology (NUST), Pakistan

Final-year project: driving simulator for autonomous vehicles.

Awards: Best Paper Award, CASA 2024 | Academic Excellence Award, UST-KIST School, 2023.

SELECTED PUBLICATIONS

1. 2026. Jeongha Lee, Yujin Kim, **Ghazanfar Ali**, Suhyun Kim, and Jae-In Hwang. "Through Van Gogh's Eyes: Global Style Transfer with Diffusion Model." *ECCV*.
2. 2026. Hwang Youn Kim, **Ghazanfar Ali**, Jeongha Lee, and Jae-In Hwang. "Lightweight Speech-Conditioned Upper-Face Animation for Virtual Agents via Emotion-Liveness Composition." *IEEE Access*.
3. 2025. **Ghazanfar Ali**, Hwang Youn Kim, and Jae-In Hwang. "RIDGE: Rule-Infused Deep Learning for Realistic Co-Speech Gesture Generation." *Computer Animation and Virtual Worlds*, 36(4), e70034.
4. 2025. Jeong Ha Lee, **Ghazanfar Ali**, and Jae-In Hwang. "A Retrieval-Augmented Generation System for Accurate and Contextual Historical Analysis." *Computer Animation and Virtual Worlds*, 36(4), e70048.
5. 2025. **Ghazanfar Ali**, Woojoo Kim, Muhammad Shahid Anwar, Jae-In Hwang, and Ahyoung Choi. "Expanding Multilingual Co-Speech Interaction." *IEEE Access*, 13, 145144-145157.
6. 2024. Hwang Youn Kim, **Ghazanfar Ali**, and Jae-In Hwang. "Enhancing Doctor-Patient Communication in Surgical Explanations." *Computer Animation and Virtual Worlds*, 35(3), e2236. Best Paper Award, CASA 2024.

Complete publication list: *Google Scholar and the Academic Master CV*.

RESEARCH MANUSCRIPTS

Under review: ConGRets: Contrastive Learning for Scalable, Low-latency Co-speech Gesture Generation with Zero-shot Speaker Style Adaptation. First author and research lead.

Full draft, internal review: What Does a Motion-Hour Cost? Video-to-Motion Ingestion, Motion Tokenisation, and the Unreported Economics of Motion Corpora. First author.

PROFESSIONAL RECOGNITION AND SERVICE

- Best Paper Award, CASA 2024; Academic Excellence Award, UST-KIST School, 2023.
- Reviewer for Cognitive Computation, Scientific Reports, Discover Computing, and Discover Artificial Intelligence; Program Committee Member, RoSE'26.
- Presenter, ASAP, SIGGRAPH Asia 2023 Real-Time Live!; additional research demonstrations for KOCCA Korea and institutional venues.

REFERENCES

Available upon request.