MUHAMMAD HUZYAFA KHOKHAR

AI Research Engineer

Multimodal Learning · Autonomous AI Systems · Neuro-Symbolic Reasoning

mhkhokhar.com | LinkedIn | Excelleve | GitHub | mhk7960@nyu.edu

Engineering intelligent systems that autonomously acquire, represent, and reason over knowledge.

SUMMARY

AI Research Engineer specializing in multimodal learning, autonomous AI systems, and neuro-symbolic reasoning. Developed a sentence-level imagined speech decoding framework from non-invasive EEG and deterministic reasoning systems for regulated industries, while designing autonomous scientific agents that transform complex real-world data into machine-interpretable knowledge. Research centers on advancing intelligence through principled engineering, uniting representation learning, symbolic reasoning, and natural human-machine interaction.

RESEARCH EXPERIENCE

Research Engineer (Founder)

Oct 2023 – Present

excelleve.com · Neurotechnology & Brain-Computer Interfaces

- Developed a sentence-level imagined speech decoding framework from non-invasive EEG, defining a new evaluation task for a 197-sentence vocabulary where no prior published benchmark existed. Evaluated across four subjects from the ChiSCO dataset (23,000 recordings), achieving 79% Top-5 and 24.5% Top-1 accuracy.
- Designed a dual-encoder contrastive architecture combining a Multi-Scale CNN temporal encoder with a Transformer encoder aligned to frozen BERT representations through contrastive learning, enabling semantic supervision directly from neural activity while remaining language-agnostic by design.
- Developed a topology-driven EEG channel selection framework integrating Persistent Homology (Topological Data Analysis) with Phase-Locking Value analysis across 17,570 EEG trials, reducing 122 electrodes to 32 while improving signal quality by 1.92×. The learned functional topology recovered established language-processing regions without anatomical priors, providing a practical foundation for compact wearable BCIs.
- Built an EEG-conditioned language generation prototype that injects neural representations into every attention layer of a 1.5B-parameter language model, demonstrating the feasibility of open-vocabulary neural language generation with 0.738 cosine similarity between generated and intended sentences.
- Established the technical foundation for a closed-loop neural interface, where decoding serves as the first stage toward bidirectional human-machine communication by identifying the spatial and temporal organization of neural language circuits required for future stimulation-based interaction.

Founder & Research Engineer

Dec 2025 – Present

Glazion · Neuro-Symbolic AI & Trustworthy Machine Learning

github.com/HuzyafaK/RegIntel

- Designed RegIntel, an evolving neuro-symbolic reasoning architecture that separates neural perception from symbolic reasoning, exploring how deterministic AI systems can continuously acquire new knowledge without sacrificing reproducibility or auditability.
- Developed an Inductive Logic Programming (ILP) knowledge-discovery framework that learns symbolic rules from regulatory evidence, enabling deterministic reasoning systems to evolve beyond manually engineered rule sets. Applied to 1,265 real regulatory documents, the framework discovered a previously uncaptured compound GMP pattern that eliminated 18 false negatives remaining after more than 200 hours of expert rule engineering.
- Built a hybrid inference pipeline combining neural document understanding, hallucination verification, symbolic reasoning, and cryptographically signed HMAC-SHA256 audit trails to produce deterministic decisions aligned with 21 CFR Part 11 data-integrity requirements.
- Evaluated the system across 1,415 real regulatory documents spanning FDA guidance, SEC regulations, and ClinicalTrials.gov, achieving 100% decision accuracy, 100% safety recall, 0 false negatives, and an F1 score of 0.968 under the project's evaluation protocol.
- Collaborated with regulatory experts including former FDA scientists, ASTM leadership, and healthcare executives to validate system behavior and support deployments in regulatory compliance and industrial data intelligence.

- Secured early commercial traction with two active engagements: a Lighthouse Partnership with ASTM on its Data Vault Program, making industrial sensor data AI-ready and monetizable, with a working demo delivered and executive-level discussions underway; and mediCLARUS, automating lost-revenue recovery for primary care clinics and health plans with Ruben R. Cardenas (former Chief Medicare Officer, Aetna/CVS Health).

SELECTED SYSTEMS

Autonomous Manufacturing Intelligence Agent — github.com/HuzyafaK/Manufacturing-Intelligence-Agent

Autonomous Scientific Data Engineering · Agentic AI

- Designed an autonomous neuro-symbolic research agent that transforms heterogeneous additive-manufacturing sensor data into AI-ready datasets by studying unfamiliar manufacturing processes, scientific literature, and technical standards to determine preprocessing, feature engineering, labeling strategies, and analytical workflows.
- Architected an end-to-end scientific reasoning system that retrieves domain knowledge, generates and executes data-processing pipelines, evaluates intermediate results, and iteratively refines its own workflow, reducing dependence on manually engineered process-specific pipelines.
- Produced foundation-model-ready datasets in Apache Parquet together with structured JSON knowledge representations for semantic querying and agent interoperability (MCP-compatible), demonstrated on the Peregrine dataset spanning Laser Powder Bed Fusion (L-PBF), Electron Beam Powder Bed Fusion (EB-PBF), and Binder Jetting manufacturing processes.
- Technically reviewed by Mark Burhop (Siemens 2022 Inventor of the Year, ASTM F42).

WildNet — github.com/HuzyafaK/WildNet

Large-Scale Bioacoustic Learning

- Developed a multimodal deep-learning architecture combining CNN feature extraction, channel attention, and dual Transformer encoders for large-scale wildlife audio classification, achieving 88% ROC-AUC across 206 bird species under approximately 1:500 class imbalance.
- Achieved approximately 76% of the competition-winning score with only two competition submissions, demonstrating strong generalization under limited optimization.

Universal Sensor Interface for VEX Robotics

Embedded Systems · Hardware Interoperability

- Reverse-engineered VEX's undocumented analog communication protocol and designed an Arduino interoperability layer that enabled commodity off-the-shelf sensors to operate as native VEX peripherals without modifying the robot's control ecosystem.
- Reduced sensing hardware cost by approximately 90% ($100 → $10) while maintaining 100% accuracy across 1,000+ operational cycles, enabling rapid hardware iteration under competition constraints and deployment at the VEX Asia Open Championship.

RESEARCH EXPERTISE

Research Areas: Multimodal Learning · Brain-Computer Interfaces · Neuro-Symbolic AI · Autonomous AI Systems · Representation Learning · Agentic Systems

Machine Learning: PyTorch · Transformers · CNNs · Contrastive Learning · Self-Supervised Learning · Multimodal Representation Learning · Hugging Face

Reasoning & AI Systems: Knowledge Representation · Inductive Logic Programming · Symbolic Reasoning · Hallucination Verification · Tool-Using Agents · Retrieval-Augmented Systems · Evaluation Framework Design

Signal Processing: EEG Signal Processing · Persistent Homology (TDA) · Phase-Locking Value · Spectral Analysis · Neural Signal Decoding

Healthcare & Clinical Operations: Revenue Cycle Management · Clinical Workflow Automation · Imaging Center Operations · Insurance Verification Systems · Healthcare Regulatory Compliance (21 CFR Part 11)

Engineering: Python · C++ · Docker · AWS · API Design · FAISS · Git

EDUCATION

New York University Shanghai Expected May 2028

B.S. Computer Science · Minor: Neurolinguistics

- Research Focus: Brain-Computer Interfaces & AI
- Medalist — VEX Asia Open Championship
- Promising Award — VEX U SJTU World Championship China Selection