Zain Ul Abideen

Scientific Machine Learning Researcher — BCMaterials, Basque Center for Materials, Spain

PhD candidate, University of the Basque Country (UPV/EHU) · Research software for AI-driven science

Location: Bilbao, Spain · **Email:** zainul.abideen@bcmaterials.net

zainulabideen.dev · GitHub · LinkedIn · Google Scholar · ORCID 0000-0001-9909-2061

SUMMARY

Scientific machine learning researcher building research software for AI-driven science, with over eight years of research experience in machine learning. PhD research at BCMaterials (UPV/EHU, advisor Prof. Ivan Infante) trains and applies machine learning interatomic potentials (MLIPs) for semiconductor nanocrystals and quantum dots with equivariant graph neural networks (NequIP, Allegro, MACE), combining DFT reference data with molecular dynamics. Main developer of Orchestr.AI, an open-source platform that trains, benchmarks and deploys MLIPs on European supercomputers, and of the QDSpace Desktop application. Background spans AI for drug discovery and cheminformatics, NLP, large language models, retrieval-augmented generation and agentic AI.

Publications: **16** (4 journal articles and preprint, 12 conference papers) · Citations: **583** · h-index: **15** (Google Scholar)

RESEARCH INTERESTS

Scientific Machine Learning · ML Interatomic Potentials · Atomistic Simulation · Quantum Dots and Nanomaterials · AI for Drug Discovery · Cheminformatics · Generative and Agentic AI · NLP and LLMs · Research Software and HPC

EXPERIENCE

Researcher (PhD) — BCMaterials, Leioa, Spain **Jul 2023 – Present**

Computational Materials Science Group, Prof. Ivan Infante · IKUR-HPC&AI project HPCAI11 ARK (2025–2026)

- **Orchestr.AI:** main developer of an open-source Python/PyTorch platform that trains, fine-tunes, benchmarks and runs inference with 7 MLIP architectures (SchNet, PaiNN, SO3net, FieldSchNet, NequIP, Allegro, MACE) from one YAML file, covering DFT (CP2K) data preprocessing, SLURM job dispatch and multi-GPU training on Spanish and European supercomputers.
- Built benchmarking pipelines that run parallel hyperparameter sweeps, aggregate metrics and select the best checkpoints; debugged CUDA/GPU memory and distributed-training failures across backends.
- **QDSpace:** main developer of the desktop application (Tauri/Rust, reusing the SvelteKit interface, with local MACE inference) and co-developer of the web platform, maintaining the main branch and deploying to AWS.
- Co-authored a benchmark of five GNN-based MLIPs for the dynamical stability of a 149-atom CdSe nanocrystal (force accuracy, uncertainty quantification, 1 ns MD at 300 K, active learning), trained and evaluated with Orchestr.AI. MLIPs needed 6–66 ms per MD step versus about 20 s for DFT, roughly 300–3,000 times faster.
- Co-developing a foundation-model interatomic potential for quantum dots: one transferable model across material families instead of a new force field for each system.

Computer Engineer (AI Researcher and Lab Engineer) — GIKI, Topi, Pakistan **Nov 2021 – Jun 2023**

Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, AI Research Group

- Machine learning, deep learning and NLP research in the AI Research Group.
- Taught labs for six undergraduate AI courses, including Machine Learning and Computer Vision, and developed new lab courses and teaching materials.
- Built an online cheating-detection system using ResNet-50 transfer learning for automated exam monitoring.

Research Collaborations — Jeju University (Korea) and MiLab, Universitat Rovira i Virgili (Spain) **Nov 2020 – May 2023**

- Anomaly detection in smart grids and energy-cloud environments using machine learning, DBSCAN and correlation analysis.
- Deep learning for identifying unstructured metabolomics data, including dataset collection, cleaning and processing.

Research Associate and Team Lead, NLP Lab — COMSATS University Islamabad (CUI) **Feb 2018 – Nov 2021**

- Research in NLP, text mining, machine learning and deep learning, including linguistic resources, ontologies and entity normalisation for low-resource languages.
- Lab instructor for five undergraduate computer science courses and course instructor for one.
- Led a voluntary student NLP research team of about 20 (Jan 2020 – Nov 2021): set research plans, coordinated ML and NLP platform development, and supervised student research projects.

EDUCATION

PhD in Computer Engineering — University of the Basque Country (UPV/EHU), Spain **2023 – Mar 2027 (expected)**

- Thesis: *A Generic Automated and Adaptive Machine Learning Platform for Advanced Materials Research*. Research carried out at BCMaterials, Leioa. Advisor: Prof. Ivan Infante. Funded by the IKUR Strategy (IKUR-HPC&AI).

MS in Computer Science — COMSATS University Islamabad (CUI) · CGPA 3.46 **2017 – 2020**

- Thesis: *A Machine Learning Assisted Parts of Speech Tagger for Roman Urdu*. LSTM-based POS tagger, new word and sentence corpora, a word-standardisation algorithm, a new tagset and a hybrid phonetic hashing algorithm for an unstandardised, mixed-language script.

BS in Software Engineering — COMSATS University Islamabad (CUI) **2013 – 2017**

TECHNICAL SKILLS

ML interatomic potentials	MACE, NequIP, Allegro, SchNet, PaiNN, SO3net, FieldSchNet, e3nn; equivariant GNNs, active learning, uncertainty quantification, fine-tuning, benchmarking
Atomistic simulation	ASE, DFT datasets (CP2K), OpenMM, GROMACS, MDAnalysis, DScribe, VMD, ChimeraX, py3Dmol
Cheminformatics	RDKit, Chemprop, DeepChem, PyTorch Geometric, ChemBERTa, Mordred, Mol2Vec, Open Babel
Machine learning	Python, PyTorch, PyTorch Lightning, TensorFlow, Keras, scikit-learn, XGBoost, Hugging Face Transformers, AutoML
LLMs and agents	LangChain, LangGraph, CrewAI, LoRA/PEFT, OpenAI API, Anthropic Claude API, Ollama, FAISS, ChromaDB, BM25, Sentence Transformers
NLP	spaCy, NLTK, Gensim, BERTopic, KeyBERT, YAKE
HPC, MLOps and cloud	SLURM, multi-GPU and distributed training, CUDA/GPU debugging, Linux, Conda, uv, TensorBoard, Docker, Git, GitHub Actions (CI/CD), AWS (Lambda, S3, API Gateway), PySpark
HPC systems used	Leonardo (CINECA, EuroHPC), Finis Terrae III (CESGA, Spanish Supercomputing Network), Hyperion and Atlas (DIPC)
Programming and apps	C++, C, Java, Rust, Bash, JavaScript; FastAPI, Flask, Django, Streamlit, Gradio, Svelte/SvelteKit, Tauri, PyQt5
Data and visualisation	NumPy, Pandas, Matplotlib, Seaborn, Plotly, NetworkX, Folium, Selenium, BeautifulSoup

RESEARCH SOFTWARE

Orchestr.AI — MLIP training, benchmarking and deployment platform github.com/nlesc-nano/Orchestr.AI

Trains, fine-tunes, benchmarks and deploys 7 MLIP architectures from one YAML configuration, with SLURM and multi-GPU execution on HPC. Release v0.2.0 on Zenodo: doi.org/10.5281/zenodo.22894956.

QDSpace — web platform for quantum-dot modelling quantumdotspace.org

Nanocrystal builder with facet, shape and ligand-passivation control, and a library of DFT-relaxed perovskite, II–VI and III–V structures with interactive PDOS and COOP visualisation. SvelteKit, FastAPI, MatterViz, Docker, AWS.

QDSpace Desktop — standalone desktop application (main developer)

The QDSpace quantum-dot tools as a desktop application built with Tauri and Rust, with a built-in local property calculator based on MACE.

SELECTED PROJECT EXPERIENCE

Full project portfolio with descriptions: zainulabideen.dev/projects

- **LLMs, RAG and agents:** RAG systems (LangChain, FAISS, ChromaDB, BM25 hybrid retrieval, Sentence Transformers); multi-agent systems (LangGraph, CrewAI, Gemini, Mistral-7B, Llama-3); LoRA/PEFT and transformer fine-tuning (BERT, GPT-2); LLM document analysis and resume screening (Ollama).
- **Cheminformatics and drug discovery (12 projects):** QSAR and pIC50 prediction with RDKit descriptors, Morgan fingerprints, Mol2Vec, ChemBERTa, ChemGPT, MPNN, GAT and Chemprop; multi-task learning; virtual screening, molecular docking and molecular dynamics for PRC2 inhibitors.
- **NLP:** transformer fine-tuning (BERT, RoBERTa, T5, BART, Pegasus); sentiment and emotion analysis, NER, question answering, summarisation; topic modelling (LDA, NMF, LSA, BERTopic); keyword extraction (KeyBERT, YAKE, TextRank); multilingual and Roman Urdu NLP.
- **Computer vision and bioinformatics:** CNNs and transfer learning (ResNet-50, VGG16, fastai) for medical imaging; protein, DNA and RNA sequence classification.
- **Core ML and data:** algorithms implemented from first principles (ID3, HMM, Naive Bayes, Apriori, genetic algorithms, neural networks); clustering, AutoML (PyCaret), geospatial analysis; large-scale web scraping and automation.

PUBLICATIONS

583 citations, h-index 15 on Google Scholar.

Manuscript in preparation

[M1] **Z. Ul Abideen**, M. Fernández-Pendás, M. Usman, C. Rodríguez Quintero, M. Suleymanova, I. Arganda-Carreras, I. Infante. "Orchestr.AI: A Unified Workflow for Training and Benchmarking of Machine-Learned Interatomic Potentials." In preparation, 2026.

Software

[S1] **Z. U. Abideen**, M. Fernández-Pendás, C. Rodriguez Quintero, M. Usman, I. Infante. "Orchestr.AI v0.2.0: Unified Platform for Machine-Learning Force Fields." Zenodo software release, 2026.

Journal articles and preprint (4)

[J1] M. Usman, M. Suleymanova, **Z. U. Abideen**, M. Fernández-Pendás, I. Infante. "Benchmarking Machine-Learning Interatomic Potentials for Dynamical Stability in Inorganic Semiconductor Nanocrystals: A CdSe Case Study." *arXiv:2609.15299* (preprint), 2026.

[J2] **Z. Ul-Abideen**, R. H. Ali, A. I. Zeeshan, T. A. Khan. "A Novel NLP-Based Stock Market Price Prediction and Risk Analysis Framework." *Journal of Computer Science and Technology* 24(2):e07, 2024.

[J3] I. Ul Hassan, R. H. Ali, **Z. Abideen**, A. Z. Ijaz, T. A. Khan. "Towards Effective Emotion Detection: A Comprehensive Machine Learning Approach on EEG Signals." *BioMedInformatics* 3(4):1083–1100, 2023.

[J4] I. Ul Hassan, R. H. Ali, **Z. Ul Abideen**, T. A. Khan, R. Kouatly. "Significance of Machine Learning for Detection of Malicious Websites on an Unbalanced Dataset." *Digital* 2(4):501–519, 2022.

Conference papers (12)

[C1] M. Shehzadi, R. H. Ali, **Z. Abideen**, A. Z. Ijaz, T. A. Khan. "Enhancing Flood Resilience: Streamflow Forecasting and Inundation Modeling in Pakistan." *Engineering Proceedings* 56(1):315, 2023.

[C2] T. Ahmed, A. Maaz, D. Mahmood, **Z. ul Abideen**, U. Arshad, R. H. Ali. "The YOLOv8 Edge: Harnessing Custom Datasets for Superior Real-Time Detection." *18th International Conference on Emerging Technologies (ICET 2023)*, pp. 38–43, 2023.

[C3] A. Mashhood, **Z. ul Abideen**, U. Arshad, R. H. Ali, A. A. Khan, B. Khan. "Innovative Poverty Estimation through Machine Learning Approaches." *18th International Conference on Emerging Technologies (ICET 2023)*, pp. 154–158, 2023.

[C4] N. Ali, A. Z. Ijaz, R. H. Ali, **Z. Ul Abideen**, A. Bais. "Scene Parsing Using Fully Convolutional Network for Semantic Segmentation." *IEEE Canadian Conference on Electrical and Computer Engineering (CCECE 2023)*, pp. 180–185, 2023.

[C5] U. Arshad, H. H. Ellahie, **Z. ul Abideen**, R. H. Ali. "An Enhanced Genetic Algorithm Framework for Efficient Solutions to Capacitated Vehicle Routing Problems." *18th International Conference on Emerging Technologies (ICET 2023)*, pp. 141–146, 2023.

[C6] O. K. Majeed, **Z. ul Abideen**, U. Arshad, R. H. Ali, A. Habib, R. Mustafa. "AdaptiveCloset: Reinforcement Learning in Personalized Clothing Recommendations." *18th International Conference on Emerging Technologies (ICET 2023)*, pp. 305–309, 2023.

[C7] H. Ashfaq, U. Arshad, R. H. Ali, **Z. ul Abideen**, M. H. Shah, T. A. Khan, A. Z. Ijaz, N. Ali, A. B. Siddique. "Mitigating Crop Losses: AI-enabled Disease Detection in Tomato Plants." *International Conference on Frontiers of Information Technology (FIT 2023)*, pp. 190–195, 2023.

[C8] M. Naeem, A. B. Siddique, R. H. Ali, U. Arshad, **Z. ul Abideen**, T. A. Khan, M. H. Shah, A. Z. Ijaz, N. Ali. "Performance Evaluation of Popular Deep Neural Networks for Neural Machine Translation." *International Conference on Frontiers of Information Technology (FIT 2023)*, pp. 220–225, 2023.

[C9] M. H. Ishaq, R. Mustafa, U. Arshad, **Z. ul Abideen**, R. H. Ali, A. Habib. "Deciphering Faces: Enhancing Emotion Detection with Machine Learning Techniques." *18th International Conference on Emerging Technologies (ICET 2023)*, pp. 310–314, 2023.

[C10] A. Abdullah, N. Ali, R. H. Ali, **Z. Ul Abideen**, A. Z. Ijaz, A. Bais. "American Sign Language Character Recognition using Convolutional Neural Networks." *IEEE Canadian Conference on Electrical and Computer Engineering (CCECE 2023)*, pp. 165–169, 2023.

[C11] H. Javed, U. Arshad, **Z. ul Abideen**, R. H. Ali. "A New Rhythm in A1: Convolutional Neural Networks for Music Genre Classification." *18th International Conference on Emerging Technologies (ICET 2023)*, pp. 222–227, 2023.

[C12] **Z. ul Abideen**, T. A. Khan, R. H. Ali, N. Ali, M. M. Baig, M. S. Ali. "DocOnTap: AI-based disease diagnostic system and recommendation system." *17th International Conference on Emerging Technologies (ICET 2022)*, pp. 148–153, 2022.

PRESENTATIONS

- **Poster:** Z. Ul Abideen, M. Fernández-Pendás, I. Infante. "Implementation of a Machine Learning Force Fields Platform for Quantum Dots." AI4AM 2025, Artificial Intelligence for Advanced Materials, San Sebastián, Spain, 8–10 Apr 2025.
- **Project presentation:** "HPCAI11 ARK: Advanced Machine-Learning Platform for Atomistic Simulations of Reactivity and Kinetics." IKUR-HPC&AI project session, Ikerbasque, CIC nanoGUNE, San Sebastián, Spain, 17 Jul 2026.

TEACHING, WORKSHOPS AND SERVICE

- Lab instructor for 11 undergraduate courses at GIKI and COMSATS, including Artificial Intelligence, Machine Learning, Computer Vision and Data Structures.
- Participant, BCMaterials PhD Day, Leioa, Spain (28–29 May 2026).
- Participant, Workshop: New Materials for a Better Life! 2023, BCMaterials, Leioa, Spain (26 Oct 2023).
- Instructor, Seminar on Machine Learning with Python, University of Malakand, Pakistan (2022).
- Organiser, 17th International Conference on Emerging Technologies (ICET), GIKI (2022).
- Certificate of Appreciation for excellence and consistently high performance, GIKI.
- Participant, AI and Data Science Challenge (ekipa, with the DatAd team).

LANGUAGES

English (fluent), Urdu (native).

REFERENCES

Prof. Ivan Infante (PhD advisor) — Ikerbasque Research Professor, Computational Materials Science Group, BCMaterials, Leioa, Spain.

Dr. Raja Hashim Ali (former colleague at GIKI, 2021–2023; co-author on 15 publications) — Professor, Department of Business, University of Europe for Applied Sciences, Potsdam, Germany.

Dr. Zahoor Ur Rehman (MS thesis supervisor) — Assistant Professor, Department of Computer Science, COMSATS University Islamabad, Attock Campus, Pakistan.

Contact details available on request.