Umer Gupta

Flat 1, 21-23 Westbourne Grove, W2 4UA London | +44 7443063561 | umer.gupta152@gmail.com

PROFILE

Machine Learning engineer and applied scientist specialising in geospatial/computer vision applications and foundation modelling, with a strong mathematical background and diverse research experience in ML domains including Graph NNs, probabilistic deep learning, recommender systems and NLP. Apart from fundamentals, I have delivered end-to-end pipelines from data engineering and large-scale pretraining through to deployment on cloud infrastructure. My work on peatland and woodland monitoring is now an early-stage venture addressing data gaps in natural capital markets.

WORK EXPERIENCE

Founding Engineer, Calterra   **Aug 2025 - Present**

Building the AI stack for nature restoration projects, taking the platform from MVP to full production.

- Designed and implemented the core woodland mapping models using a DINOv3 backbone with Mask2Former-style heads for tree counting, canopy segmentation and species classification under a UKSA grant.
- Standardised a multi-dataset corpus of ~76k tiles and 1M+ tree annotations across diverse ecosystems, unifying heterogeneous labels and resolutions.
- Led Business Development: secured ~£50k in revenue within a month of launching; panellist at Nature Tech Collective Unconference

Lead ML Engineer, New Gradient   **Oct 2023 - Present**

Successfully delivered Peatland monitoring and restoration planning technology under an Innovate UK grant.

- Built a UK-scale peatland foundation model by pretraining a Masked Autoencoder (MAE) encoder on ~2T peatland image tiles (RGB + DSM).
- Achieved domain generalisation across multisource datasets with engineered augmentation pipelines (colour jitter, D4 transforms, dehazing, fourier domain adaptation) and active learning based sample selection.
- Implemented post-processing pipelines, including delaunay triangulation based skeletonisation for centerline delineation, morphological operations to join/smooth features. Achieved 10x mIoU improvement on incumbent DEFRA-backed project.

Future Explorer Challenge for Near Mine Exploration: Solutions for precision mining exploration using AI.

- Converted raw borehole logs (geochemistry + geophysics) into 3D voxel grids and developed a Voxel-MAE with a Single-Stride Sparse Transformer (SST) encoder over non-empty voxels (regional self-attention) for prediction on masked voxels (akin to semantic scene completion). Awarded $4,000 for the work.

Origin Peptides: Improved experimental workflows for a start-up developing sustainable technologies for protein synthesis.

- Predicted indicators of self-assembly in proteins using common protein features and latent structural representations. Led to a place in Innovate UK's £6.4M project - Sustainable Medicines Manufacturing Innovation Programme (SMMIP).

TECHNICAL PROJECTS

Research Collaborator, ETH Zurich   **Sep 2025 - Present**

Developing generative models for biological trees and dendritic arbours.

- Extended recent work on iterative graph expansion to geometric trees. Reformulated the task as an autoregressive leaf expansion + conformation prediction (single conditional for a joint diffusion process).
- Designed an SO(2)-equivariant EGNN variant using relative frames to encode geometry, with additional mechanisms such as leaf sibling matching loss to encourage realistic, diverse growth patterns.

Industry-Based Dissertation, Amazon   **2023**

- Extended the SASRec transformer-based sequential recommender from next-item prediction to 14-day future window prediction.
- Demonstrated that window-aware models improve Recall@10 by up to ~17.4% and NDCG@10 by ~25% over vanilla next-item SASRec for 14-day prediction.

Code-Mixed Sentiment Analysis   **2022**

- Fine-tuned XLM-T and other multilingual transformers on code-mixed Hinglish tweets for sentiment analysis.
- Applied latent mix-up (interpolating hidden representations), exploring within-class vs across-class mixing.

Cormeum Technologies   **2021**

- Supported batchmates from Cornell to create compact in-flight health monitoring devices for the Indian Ministry of Defence.
- Developed adaptive algorithms on ECG and PPG signals to extract indicators like heart rate, breathing rate, etc. in noisy fighter jets.

EDUCATION

2022 - 2023	**University of Edinburgh**	**Edinburgh, UK**
	MSc in Data Science	Distinction	
	Relevant Courses: Machine Learning and Pattern Recognition, Machine Learning Theory, Reinforcement Learning.	
2017 - 2020	**Sri Venkateswara College, Delhi University**	**New Delhi, India**
	Bachelor of Science (Hons) in Mathematics	GPA: 8.62 out of 10	
2016 - 2017	**College of Arts and Sciences, Cornell University**	**Ithaca, USA**
	1st Semester, Undecided	GPA: 3.67 out of 4.3	