Andrew Ellul
AIML Solutions Architect

London, United Kingdom
+44 7512 631 643
andrew.t.ellul@gmail.com
https://www.linkedin.com/in/andrew-ellul-30bbb274/

Accomplished AI/ML Cloud Architect with over 8 years of expertise in designing scalable, intelligent systems. Recognized thought leader with numerous public speaking engagements and advisory roles to AWS' top startup clients. End-to-end proficiency in Machine Learning, specializing in Natural Language Processing. Currently spearheading architectural scalability and ML innovation for UK's leading AWS AI startups.

Skills

- **Programming Languages |** • Python • Java
- **Frameworks |** • Pandas • Numpy • NLTK • NetworkX • SKLearn
- **Concepts |** • Topic Modelling • Clustering (K-Means) • Dimensionality Reduction • Hyperparameter Optimisation • Generative AI Fine Tuning • RAG optimization • MLOPs • AWS • Cloud Architectures • Patterns for Scalability, Observability, Security, SaaS Multi-Tenancy
- **Domains |** • AI • ML • NLP • Graph Science • Computational Social Science • Startups

Experience

Amazon Web Services, London (UK) - *AIML Startup Solutions Architect |* **August 2023 - Present**
- Provided technical advisory consulting on AWS, AIML, Generative AI to startup CTOs, VPs and Engineering teams
- Created, designed and currently leading the AWS UKIR Startup AI Safety / Responsible AI GTM strategy
- Developed Generative AI RAG solutions on Amazon Bedrock with Amazon OpenSearch VectorDB, LlamaIndex, LangChain, evaluation frameworks TruEra and RAGAS. Optimised with Sentence Window Retrieval and query caching
- Developed Text2SQL solution architecture on AWS using CodeLLaMa, Amazon ECS with few-shot prompting solution, later with supervised-fine-tuning against the Spider 1.0 dataset. Published as open-source on aws-samples Github [1]
- Led monthly full-day workshops on MLOps with Amazon SageMaker educating customers on data cleaning, preprocessing with Pandas, SageMaker Data Wrangler, training with in-built and customer docker containers with SageMaker Training, running HPO (hyperparameter optimization) strategies (Grid, Random, Bayesian, Hyperband) with SageMaker AMT and model deployment & inference strategies (batch, real-time, serverless & multi-model endpoints).
- Presented various works at large public events. Text2SQL solution at AWS NLP Conference 2023, Text2SQL SFT research review at AWS EMEA Tech Summit 2023, RAG Best Practises at AWS EMEA Builders Community, CausalAI with CausaLens and Bedrock at AWS London Summit 2024
- Led quarterly team retrospectives, conducted hiring interviews and weekly AIML mentorship of junior AIML SA colleagues. Awarded UKIR Startup SA Technical Excellence award (2023)

Technologies: MLOPs, Data Science, RAG, VectorDBs, Text2sql

Amazon Web Services, Munich (DE) - *AIML SMB Solutions Architect |* **April 2022 - August 2023**
- Provided technical advisory to AWS' largest SMB in DE, plus 27 other SMBs in FinTech, Manufacturing and Adtech
- Developed observability solution with Amazon CloudWatch metric filters for IoT AI customer, resulting in 15% reduction in DevOps team time spent on production scaling issues due to logging triggers and automation.
- Architected security strategy around PAM, perimeter security and SIEM integration for CISO of largest AWS SMB
- Architected solution to help data science team scale ML experiment throughput by moving from Jupyter on EC2 instances to an MLOps architecture based on SageMaker. Increased experiment output from 3 trials per week to 7
- Developed and published open-source utility for Hyperparameter Optimisation on Amazon SageMaker AMT with vega-altair, sagemaker SDK to visualise and explore hyperparameter space for optimization [2]. Tool demoed by SageMaker AMT Applied Science Manager at re:Invent 2022 and is on roadmap for integration into SageMaker Core
- Developed text matching solution for transcription and translation which overcomes subtitle-to-voice misalignment after translation. My solution resulted in customer succeeding to scale pipeline from 20 video translations per week to 90
- Ambassador for Amazon Deep Racer (Reinforcement Learning) and AIML Demos. Spoke at 3 AWS Summits
- Amazon global intern advisory panellist. Continuously mentoring (1-hr per week) 2 Junior SAs and interns in EMEA.

Technologies: Cloud, Observability, MLOps, HPO, Deep Racer & RL, NLP

Amazon Web Services, Munich (DE) - *Data Scientist (Part-Time) |* **October 2020 - April 2022**
- Adjacent to Full-time Master's studies, worked as a Data scientist in a sales BI team running experiments exploring how AWS customers adopt new services and grow their architecture organically during their cloud journey with AWS
- Built a topic modelling pipeline with LDA, NLTK to explore linguistic evolution of customer conversations as indicator of SA effectiveness in influencing technical curiosities and interests of customer during their growth in the cloud
- Built a Recommendation Engine for AWS Services Clustering on 6 Million+ customer records with K-means and PCA. Prototyped a graph visual tool for service adoption with NetworkX. Developing hypotheses for exploratory analysis, testing and validating theories. Optimized Numpy code with vectorized operations

Technologies: Python (Pandas, Numpy, NetworkX, NLTK), SageMaker, Jupyter Notebooks, K-Means, PCA, LDA

Capgemini, Berlin (DE) - *Lead Technology Consultant* | September 2016 - April 2020

- Lead Technical Product Owner on a Global e-Commerce insurance platform for VWFS. Built and deployed customizable AWS-based e-commerce platform across 7 EU countries for 4 brands (VW, Skoda, Seat, Audi)
- Planned feature rollout, MVP design, user stories and technical spec. Ran workshops with business & technical Architects. Designed system architecture, UML, Sequence Diagrams of APIs. Monthly reporting to CTO of VWFS DU
- Functional architect in a 12-person team for a complex engine emissions and efficiency calculation system at BMW
- Prototyping SDE, developed PoCs with new technologies for a new enterprise CI-CD testing pipeline for BMW

Technologies: Java 8, CI-CD (Jenkins), AWS, Systems Design

SAP (Signavio), Berlin (DE) - *Backend Software Engineer* | January 2016 - July 2016

- Lead backend engineer for Signavio's new process mining product. Implemented a Petri-net data analysis tool to investigate process data and serve to a front-end data analysis tool based on research by Wil van der Aalst

Technologies: Java 8, NoSql, MongoDb, Sql, Hibernate

Goldman Sachs, London (UK) - *Technology Analyst* | Summer 2014
J.P. Morgan Chase, Bournemouth (UK) - *Technology Analyst* | July 2013 - June 2014

Projects, Publications

- [1] - 2024 - "Building an LLM-powered Natural Language to SQL Application"
- https://github.com/aws-samples/amazon-sagemaker-generativeai/blob/main/usecases/text_to_sxql/text2sql.ipynb
- [2] - 2022 - "Optimize hyperparameters with Amazon SageMaker Automatic Model Tuning"- Open source publication of visualisation tool to inspect hyperparameter space after HPO jobs on SageMaker AMT [2022] https://github.com/aws-samples/amazon-sagemaker-amt-visualize; https://aws.amazon.com/blogs/machine-learning/optimize-hyperparameters-with-amazon-sagemaker-automatic-model-tuning/
- 2021 - Analysis of Covid hate-speech online - https://github.com/interimblue/idp/,
- 2021 - Polaraziation Quantification on Social Networks - https://github.com/interimblue/guided-research-polarisation

Education

Technical University of Munich, DE - *M.Sc Informatik (Major: ML)* | September 2019 - April 2022 (Interrupted)

Note: Due to personal circumstances (Illness in family) & lack of distance learning support I had to re-prioritise, and consequently left my studies before completion, after completing 80% of Master's degree, including thesis submission and acceptance.

- Courses: Machine Learning, Natural Language Processing, Cloud computing, Compiler construction, Guided research (NLP for online radicalisation), computational social science and graph science, virtual machines and assembly, Distributed systems
- Research: "Quantifying Polarization in Social Networks" (2021, Distinction), "Multi-Task DL for Hatespeech detection" (2020, Distinction)
- Thesis: "A survey of topic modelling methods for an automated journalist"

University of Manchester, UK - *B.Sc Computer Science with Year in Industry* | September 2011 - June 2015

- Grade: 1st Class (Highest UK Grade category)
- Thesis (Distinction): "A Java based desktop tool for graph and social network analysis" (Distinction)

Speaking, Awards, Certifications

Speaking

- "A Simple Way for Growing Businesses to Migrate Their Data to the Cloud"- Executive blogpost on cloud migration strategies [2022]
- https://aws.amazon.com/blogs/smb/a-simple-way-for-growing-businesses-to-migrate-their-data-to-the-cloud/
- "Getting RAG Right" [2024] - AWS EMEA Builder's Community - Talk and Whiteboarding session on best practises for deploying RAG applications
- "Sustainability in Cloud Computing and AI" [2024] - University of Leeds - Guest Lecture on sustainability in Cloud Computing and AI
- "State of the Art Text 2 SQL" [2023] - AWS EMEA Tech Summit - Talk and Code Demo on SOTA methods for text to SQL and supervised fine tuning
- "Reinforcement Learning with Amazon Deep Racer" [2023] - AWS Berlin Summit 2023 - AWS Deep Racer demo and talk on reinforcement learning
- "Security in the AWS Cloud" [2023] - (German) - BCT Technology Days, Karlsruhe, Germany - Talk in German at Siemens partner BCT on Security in AWS
- "Reinforcement Learning with Amazon Deep Racer" [2022] - AWS Berlin, Paris Summits 2022 - AWS Deep Racer demo, talk on reinforcement learning
- "Intelligent Storefront - Computer vision demo with Amazon Rekognition" [2022] - (German) - DigiMesse 2022 Koblenz

Awards, Certifications

"UKIR SUP SA Tech Excellence Award" - For outstanding contribution to AIML startups	**December 2023**
AWS Certified Solutions Architect – Professional;	**June 2023**
AWS Certified Machine Learning – Specialty;	**February 2023**
AWS Certified Analytics – Specialty;	**November 2022**
AWS Certified Developer – Associate;	**January 2022**
AWS Certified Solutions Architect –	**December 2021**
Associate; Amazon Web Services Cloud Practitioner	**October 2020**
Professional Scrum Product Owner I	**October 2018**
"Highly Commended" (2015) - Undergraduate Awards	**October 2015**