

Chris Ballard

Lead Machine Learning Engineer

E-mail hello@justified.ai

WWW justified.ai

WWW [linkedin.com/in/chrisballarduk](https://www.linkedin.com/in/chrisballarduk)

WWW [github.com/chrisaballard](https://www.github.com/chrisaballard)



Experienced **LEAD MACHINE LEARNING ENGINEER** with 25 years proven success delivering ML/NLP solutions for global firms and startups. Proficient in Python, NLP, LLMs and AWS, with expertise in AI project strategy, management, and technical development. Skilled in leading strategic AI programmes, prioritising business outcomes. Seeking opportunities to apply AI for positive impact.



Expertise

Expertise in **Natural Language Processing** and **Large Language Models** for data extraction, classification, NER, generative AI and search; Transfer Learning foundation models (8 years)

NLP: Transformers, LLMs, RNN, LSTM, GRU, CNN, PyTorch, Keras, Hugging Face, Spacy, Sentence Transformers, SetFit, FastText, BertTopic, Transfer Learning (8 years)

Machine Learning: supervised/unsupervised learning, zero/few-shot learning, multi-task, multi-label, neural networks, tree algorithms (Random Forest, XGBoost, CatBoost), feature engineering (13 years)

AWS SageMaker model training and deployment; **MLOps**; Deployment using Docker and ECS

Databases: SQL, ElasticSearch, MongoDB, Spark, Athena; Data pipelines (Luigi, Airflow, Dagster)

Python: Numpy, Pandas, Scipy, Scikit-Learn, Jupyter, Matplotlib, FastAPI, Flask, git, poetry, Streamlit (13 years)

Articulate and able communicate clearly with business and technical stakeholders

Experienced in defining **business and technical strategy** for new and established AI products and projects

Proficient in managing AI projects to meet rigorous deadlines using agile methodologies (Kanban/Scrum)



Achievements

Developed and deployed a multi-label email topic classifier using Transfer Learning to fine tune Sentence Transformer models using SetFit in just 8 weeks. **Reduced spam by ~20% and enhanced sales demonstrations to prospects.**

Trained a Transformer machine learning classifier to detect senior executive impersonations. Balancing costs and accuracy to deploy cost-efficient inference data pipeline handling 50k requests per minute. **Addressed a common client concern, with 9 in 10 quarantined emails detected being malicious.**

Spearheaded the development of a machine learning model and spark data preprocessing pipeline for malicious email detection. Extensive work with iterative feature engineering resulted in a **20% increase in detections and contributed significantly to improved sales PoV win rates**

Developed MVP for a climate policy [semantic search tool](#) for a climate-tech non-profit using Machine Learning to extract structured data from PDF documents. **Completed in 8 weeks and demonstrated at COP26 Climate Conference, securing valuable partnerships with funders.**

Delivered extreme Multi-label/Multi-class machine learning models, data preprocessing and inference pipeline to extract structured characteristics from eCommerce product descriptions, automating data entry process for a global market research company. Generated **\$2M in annual operational cost savings.**



Career History

Mar 2022 –
Apr 2024

Lead Machine Learning Engineer and Technical Lead

Tessian, Remote

- Malicious email detection using Machine Learning in email security scale-up
- Responsible for **complex Python data and machine learning pipeline which processes millions of emails per day across the UK, EU and US.**
- Reviewed existing system, produced recommendations for strategic improvements
- Prioritised use cases to improve detection and developed algorithm improvement roadmap
- Planning and scoping AI projects in quarterly planning cycles; presentation of roadmap and key metrics to senior leadership
- Extensive use of SageMaker for model training, optimisation, research and deployment
- Extensive use of Spark and SQL to query and process data from data lake on AWS S3
- Developed CatBoost ML model and Spark data preprocessing pipeline to classify malicious emails from email metadata. Improved detection by 20%.
- Running pilots and producing metrics to identify whether systems meet success criteria
- Produced business case for investment in LLMs which was launched as project
- Trained and deployed an LLM using few-shot learning and Transformers for email topic and intent classification and malicious conversation starter detection
- Conducting shadow deployment testing to evaluate new model performance and costs against live baselines and decide on release readiness
- Presented at [Data Science Festival Oktoberfest 2023](#) on ML with biased imbalanced data

Jul 2023 –
Dec 2023

Data Science Consultant

Dotti, Norwich, UK - Contract

- Created **AI and Data Strategy** for startup developing AI Assistant to provide women advice and support on perimenopause.
- Advised on the collection of health data from smart devices and using machine learning to predict perimenopause from symptoms.

May 2021 –
Mar 2022

Chief Technology Officer, Lead Data Scientist

Climate Policy Radar, London, UK

- Development of a new product using **AI to structure and analyse climate policy and legislation** in PDF format.
- Worked with internal climate policy domain experts and external stakeholders to understand user needs
- Presented technical strategy to company Board, Advisory Board and key partners such as the World Bank
- Managed AI projects, establishing AI roadmap and agile project management methodology, working closely with CEO, board and clients
- Evaluated methods to improve PDF text extraction, including LayoutLM and document layout analysis using Mask R-CNN.
- [Climate policy semantic search](#) system using Sentence Transformers and Elasticsearch combining Approximate Nearest Neighbour dense vector similarity with full text document scoring.
- **Deployed prototype with NextJS front end and Python backend in 8 weeks** to promote Climate Policy Radar to potential funders at COP 26 climate conference.

Jul 2017 –
May 2021

Lead Applied Data Scientist

Nielsen IQ, Oxford, UK

- **Led R&D and development of strategic machine learning projects for major global market research organisation.** Focus on automation of Operations processes to deliver productivity savings.
- Developed strategic R&D roadmap, aligning project goals to company strategy.
- Collaborated with key stakeholders and business leaders in Operations and Technology; presented project proposals and progress reports to senior leadership.
- Worked with Operations teams to identify use cases for AI; organised workshops and regular project meetings.
- Established analytic methods to measure RoI, aligning metrics to business success criteria.

- Developed Logistic Regression model to match retailer and eCommerce descriptions. **Developed in 4 weeks to solve barrier preventing new product going to market.**
- Extreme multi-label/multi-class machine learning pipeline to classify retailer product descriptions to thousands of product attributes using FastText. **Delivered \$2M annual operational cost savings.**
- **Delivered system to classify eCommerce consumer panel product reviews** to a large structured market research taxonomy. Developed multi-task classification model using a 1D Convolutional Network encoder using **PyTorch**.
- Led development of Neural Network LSTM models using PyTorch for hierarchical brand classification of eCommerce product descriptions. Improved Machine Learning training pipeline to replace multiple Jupyter notebooks with reproducible Python codebase.

Apr 2016 –
Jun 2017

Data Scientist

Nielsen IQ Brandbank, Norwich, UK

- Extraction of structured ingredients from product images. Process to resolve OCR errors in ingredients lists extracted using Faster R-CNN model. Used conditional random field to correct ingredient list segmentation errors and a language model to correct errors.
- Anomaly detection using PCA to identify errors in product nutritional content which had problem with data entry errors.

Jan 2003 –
Apr 2016

Data Scientist, Product and Engineering Manager

Tribal, Norwich, UK

Data Scientist (Jan 2014 - Apr 2016)

- Developed new analytics product for the higher education sector.
- Product roadmap and company data science strategy
- Machine learning to predict student success from multiple sources of student engagement data
- Managed remote engineering and data science team
- Led delivery to a large education provider to provide student analytics across 50+ locations; responsible for architecture, scaling the data pipeline, data analysis and model design.

Innovation Consultant (Jan 2011 - Dec 2013)

- Research project with UK university to evaluate use of machine learning to predict student outcomes

Business Intelligence Manager and Consultant (Oct 2008 - Dec 2010)

Product Manager (Oct 2006 - Oct 2008)

Engineering Manager (Sep 2003 - Oct 2006)



Education

Sep 2012 –
Jun 2015

Master of Science: Knowledge Discovery and Data Mining

UNIVERSITY OF EAST ANGLIA - Norwich, UK

- Grade: Distinction
- Obtained 90% for research dissertation: "Using diversity for dynamic optimisation of data fusion ensembles". **Published as research paper.**

Sep 1995 –
Jul 1998

Bachelor of Science: Physics

UNIVERSITY OF BRISTOL - Bristol, UK



Publications

Dynamic ensemble selection methods for heterogeneous data mining, Chris Ballard, Wenjia Wang, 2016 12th World Congress on Intelligent Control and Automation (WCICA), 1021-1026, IEEE



Interests

Drone building | Home Brewing | Running | Cycling | Reading Historical Fiction