JEM ANDREW

andrewjem8@gmail.com

**LinkedIn :** linkedin.com/in/jamesandrew‑439771333 |

**Website:** <https://jemandrewwebsite-production.up.railway.app/>

**Professional summary**

Machine Learning and Software Engineer with production experience building AI systems for the legal and construction industries. Currently developing RAG architectures and document matching pipelines with the utilisation of Claude api to process 20,000+ documents at Velitor law firm. Reduced brief preparation time by 70% through intelligent document retrieval and achieved 98% precision in duplicate detection across 8,000+ PDFs. Experience in full-stack development and software applications through previous experience at Buildchorus. Strong foundation in Python, Java, Django, PyTorch, and cloud deployment (GCP, Docker).

**Skills**

* ML/AI: RAG architecture, Prompt engineering, Vector databases (ChromaDB), LLM APIs (Anthropic Claude), Legal-BERT, BM25, PyTorch, TensorFlow, Model Fine Tuning
* Languages: Python, Java, JavaScript, SQL, R, HTML/CSS
* Backend Development: Flask, Django, Node.js, REST APIs
* Frontend Development: React, HTML, CSS
* Databases & Data Tools: MongoDB, PostgreSQL, MySQL, R Studio, Data visualisation (Matplotlib, ggplot2)
* Cloud & DevOps: Netlify, Heroku, AWS (basic), Service Workers (PWA), CI/CD pipelines, Docker, Git/Github/Gitlab, Jira, Railway
* Data Science: pandas, NumPy, R Studio, data visualisation, statistical analysis, Matplotlib, ggplot2

**Work history**

Machine learning engineer | 09/2025 - current

**Velitor Law- London**

* Developed document matching pipeline for sophisticated civil litigation case analysing 8,000+ PDFs to detect duplicates; extracted metadata to Excel, applied fuzzy date and description matching, filtered to top 3 candidates, verified with Claude Vision API
* Designed Excel-based analytics dashboard with automated confidence scoring using nested IF/VLOOKUP formulas, conditional formatting for visual triage (green for 90%+ matches, yellow for 70-89%, red for manual review), pivot tables analysing match distribution by document type, and intelligent filtering reducing 10+ candidates per document to top 3 highest-scoring matches
* Identified 20 duplicate documents with 98%+ precision despite technical challenges where matching documents had different formats (scanned vs native), inconsistent watermarks, and varying file sizes, using Python, OCR, and multi-page sampling strategies
* Built a conversational RAG system for sophisticated civil litigation, allowing lawyers to query 20,000+ case documents naturally; combined ChromaDB vector database, Legal-BERT semantic embeddings, BM25 keyword search, and intelligent query routing to specialised forensic analysis prompts
* Implemented cumulative knowledge graph storing findings across sessions (smoking guns, evidence chains, contradictions, legal arguments); system learned from each query, building progressively deeper case understanding
* Enabled queries like "find smoking guns" or "build timeline of concealment"; the system first established foundational case knowledge from pleadings, then answered increasingly sophisticated questions by building upon prior discoveries
* Produced tribunal-ready documentation (memoranda, skeleton arguments, cross-examination scripts) by synthesising cumulative knowledge; generated novel arguments, discovered new document connections, reduced brief preparation time by 70%

Software Engineer | 09/2025 - current

**BuildChorus - London, City of London**

* Developed full-stack features using Django backend and PostgreSQL database
* Built user configurability features for material procurement strategies
* Deployed applications using Docker containers on Google Cloud Platform
* Collaborated with the development team using GitLab and Jira workflows
* Worked with large PostgreSQL test databases to validate software functionality

**Projects**

**Django CV Portfolio**

*Django, Python, HTML*

* Built a full-stack Django application with a responsive design and modern UI/UX
* Deployed on Railway with Gunicorn, automated security hardening, and environment-based configuration, HTTPS redirects, HSTS headers, and WhiteNoise for efficient static file serving without external storage
* Custom eye-tracking feature on the Notion avatar's eyes that follow your cursor using real-time coordinate calculations and smooth CSS transforms.
* Integrated contact form with AJAX submission and Django admin interface

GitHub: github.com/JemAndrew/cv-website

**Holiday Cleudo Web App**

*JavaScript, HTML5, CSS3*

* Built browser-based multiplayer game supporting 20+ concurrent users
* Implemented Fisher-Yates shuffle algorithm with collision detection
* Developed offline-first architecture using service workers and local storage
* Deployed production-ready application via GitHub Pages

GitHub: github.com/JemAndrew/holiday-cluedo

Live Demo: <https://familycluedo.netlify.app/>

**Cryptocurrency Exchange Platform** *Django, PostgreSQL, REST APIs, Order Matching*

* Built a custom order matching engine with a price-time priority algorithm
* Designed database schema with relationships between orders, trades, and wallets
* Implemented REST API endpoints for mobile integration
* Developed edge case handling for partial fills and identical price orders
* Built service-oriented architecture following Django best practices
* Currently processing test orders with stable performance

**Medical AI Diagnostic System** *Pytorch, Python, CNNs, HuggingFace, Transfer Learning*

* Pioneered FDA-aligned evaluation framework comparing 4 SOTA architectures (ResNet50, InceptionV3, DenseNet121, EfficientNet-B0) achieving >90% melanoma sensitivity with McNemar's testing and stratified bootstrap validation (n=1000)
* Engineered production pipeline processing 10K+ expert-annotated dermoscopic images with intelligent class imbalance handling (6.7:1 ratio), stratified sampling, and automated medical metadata validation
* Architected modular evaluation infrastructure with custom PyTorch datasets, clinical metrics engine (sensitivity/specificity/PPV/NPV), confidence-based triage, and reproducible experiment tracking aligned with ISO 14971 standards

**Education**

**Newcastle University - Newcastle upon Tyne | MSc in Computer Science**

Computer Science, 09/2024 - 09/2025 - Predicted Distinction

Dissertation: "Comparison of Novel vs Standard CNN architectures for automated skin cancer detection"

* Developed an FDA-compliant evaluation framework comparing state-of-the-art deep learning architectures (ResNet, VGG, EfficientNet) for automated dermatological diagnosis. Conducted systematic performance analysis on clinical validation datasets, focusing on diagnostic accuracy and patient safety protocols.

**Newcastle University - Newcastle upon Tyne | BSc in Biology**

Biology, 09/2021 - 05/2024 - 2:1

Research Dissertation: "A Systematic Review of CRISPR Gene Editing in Enhancing Fusarium Head Blight Resistance in Wheat and Barley" (83%)

#HRJ#50f6f73a-7ad7-4022-89e9-2ba27065c454#