

Chaeyoon Kim

chaeyoonakim@gmail.com | +44 (0) 7856231406 | LinkedIn @chaeyoonakim | GitHub @chaeyoonakim

A certified AI Ethicist and Data Scientist with 4 years' experience within the NHS ecosystem. Specialised in large-scale analytics, machine learning engineering and healthcare workforce modelling. Skilled in developing production-ready data pipelines and machine learning solutions using Python, PySpark, and Databricks. Combines strong statistical foundations with modern AI tools, including LLM-based applications and cloud data platforms, to deliver policy-relevant analysis and scalable analytical systems — leading reproducible, secure analytics engineering across NHS workforce programmes.

EXPERIENCE

Data Scientist – NHS England (London, UK)	Apr 2024 – Present
---	--------------------

- Redesigned the CT1 training-place allocation methodology from a weighted multi-criteria model into a rule-based model using Python and Excel, presenting Tableau outputs that informed prioritisation for under-doctored deaneries during 2024 BMA industrial action negotiations.
- Working as Lead Developer for the 10-Year Workforce Plan under NHS England's Business Critical Model governance: coordinating with UDAL Data Engineers on secure Azure Data Lake environments and building SQL, Python and PySpark pipelines; co-led the migration of legacy PySpark pipelines off deprecated DBFS onto Entra-secured Data Lake containers and coordinated Windows 11 migration testing to protect model continuity.
- Led the transition of legacy Excel workforce models into reproducible Python and PySpark pipelines, addressing National Audit Office recommendations through documented assumptions, strengthened data lineage and peer-reviewed Git version control — cutting model runtime by around half.
- As Community Pharmacy modelling workstream lead, coordinated a dozen stakeholder groups — including pharmacy Deans and Chief Pharmaceutical Officers — into a single delivery scope; launched a Pythonised supply-and-demand model within six months and secured the model SRO's approval despite late changes to baseline and counterfactual assumptions.
- Designed an NLTK-based sentiment-analysis pipeline for the Reporting, Insights, Publications and Transparency programme and contributed to a paper on insights maturity for the CEO's office and Board; presented 10-Year Workforce Plan model components to the External Scrutiny Panel and the national consultation on Community Pharmacy workforce interventions.
- Providing Databricks, Python and Git training and onboarding to analysts and new data scientists, promoting coding best practices across the modelling team.

Junior Data Scientist – Health Education England (London, UK)	Jun 2022 – Mar 2024
---	---------------------

- Maintained a pythonised geo-spatial analysis and produced Tableau dashboards analysing clinical placement capacity, workforce supply trends and geographic demand across 65 medical specialties for the HEE Board's tariff-funding review, advising medical Deans on redistributing specialty training posts and informing NHS England's recommendations to the Office for Students on new dental school places.
- Supported R programming capability to predict a complex birth delivery on planning maternity labour services, with source data from Royal College of Obstetricians and Gynaecologists (RCOG) statisticians.
- Collaborated with Data Engineering and Product teams using Jira/Agile framework to manage project timelines.

Data Manager – UCL Great Ormond Street Institute of Child Health (London, UK)	Apr 2022 – Jun 2022
---	---------------------

- Collaborated with bioinformaticians and statisticians on the GSK Sotrovimab Covid-19 study, cleaning and transforming lab data.
- Designed a MySQL server for secure data transfers from entry points.

PROJECTS

Strategic planning:

- **10-Year Workforce Plan modelling:** Working with a multidisciplinary group to refresh and refine the NHS Long-Term Workforce Plan modelling (published in Jun 2022), incorporating National Audit Office feedback.

AI & Machine Learning engineering:

- **Pharmacy First agent:** Presented an LLM-powered analysis about NHS Pharmacy operation with open source at PyConUK, NHS RPySOC conference, and Health and Care Analytics (HACA) conference 2025.

- **NHS Policy Navigator:** Built an adaptive Retrieval-Augmented Generation pipeline over the published 10-Year Health Plan and public NHS news feeds, refined with a working group through a spec-driven development cycle at near-zero infrastructure cost.
- **Reducing the missed NHS appointments:** Proposed a ML solution to reduce the millions costs and increasing waitlists for a hackathon challenge set by the Number 10 Downing Street data science team.
- **AI for NHS Healthcare professionals:** Developed a proof-of-concept “NHS Career Coach” chatbot on Microsoft’s AI Studio, awarded 2nd rank in a hackathon by demonstrating how LLM-powered models can assist candidates.
- **Integrated Operational Pressures Escalation Level (OPEL) prediction:** Demonstrated ML models on NHS Federated Data Platform during a hackathon operated by Palantir Technologies and KPMG.

Teaching and outreach:

- **AI Engineering training programme:** Designed and delivered a training programme for learners from undergraduates to NHS analysts and clinicians, teaching reproducible software engineering, responsible AI and governance principles using open-source NHS datasets; accepted as a peer-reviewed oral presentation in computer science education.

Research and publication:

- Developed R and Python conversion code for a textbook, <Foundations of Programming, Statistics, and Machine Learning for Business Analytics> (published in Apr 2023), then co-authored a translated edition in Korean (published in Feb 2025).

EDUCATION

MSc Data Science (Distinction) – City St George’s, University of London (London, UK)

Sep 2020 – Oct 2021

- Awarded 1st rank for the System for the Semantic Answer Type Prediction Task; presented at the International Semantic Web Conference (ISWC) in October 2021.

PgCert Data Analytics – University of Sheffield (Sheffield, UK)

Sep 2018 – Aug 2020

- **Modules include** Computer Security and Forensics, Machine Learning and Adaptive Intelligence, Statistical Data Science in R.

BSc Electronic Engineering – Kyungpook National University (Daegu, South Korea)

Mar 2007 – Aug 2011

- **Modules include** Electronic Engineering Lab (MATLAB), Calculus, Numerical Analysis.

VOLUNTEERING

The UK-Korea Global Health Forum working group: organising a conference with senior researchers at the London School of Hygiene & Tropical Medicine (LSHTM).

Professional Mentor at the University of Greenwich, and City St George’s, University of London: supporting early career interview preparation.

Ambassador at LangChain: hosting technical meetups and organising a hackathon in central London.