

Chaeyoon Kim

chaeyoonyunakim@gmail.com | +44 (0) 7856231406 | LinkedIn [@chaeyoonyunakim](#) | GitHub [@chaeyoonyunakim](#)

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.

EXPERIENCE

-
- | | |
|--|----------------------------|
| Data Scientist – NHS England (London, UK) | Apr 2024 – Present |
| <ul style="list-style-type: none">- Working as a Lead Developer in the community pharmacy modelling for the 10-Year Workforce Plan programme, working with NHSE and DHSC to implement the model scenario and assumptions in the core PySpark model.- Led the Excel model transition to automated pipelines, working on Databricks with Python.- Refactoring Python codebases to improve computational performance and to improve analytical capabilities for cross-functional teams.- Providing Databricks, Python and Git training to analysts and promoting coding best practices across the data analysis skilled teams. | |
| Junior Data Scientist – Health Education England (London, UK) | Jun 2022 – Mar 2024 |
| <ul style="list-style-type: none">- Maintained a pythonised geo-spatial analysis and produced data visualisation on Tableau for medical Deans and workforce planning Leads on how to redistribute the medical specialty training posts to address health inequalities and to support policy decisions.- 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 |
| <ul style="list-style-type: none">- 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.
- **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.

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 |
| <ul style="list-style-type: none">- 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 |
| <ul style="list-style-type: none">- 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 |
| <ul style="list-style-type: none">- 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.