# NHS launches Foresight, a generative AI trained on 57 million patient records to predict thousands of health conditions



In a landmark initiative poised to transform healthcare, the United Kingdom's National Health Service (NHS) has developed an ambitious generative AI model named Foresight. This model is trained on the medical histories of nearly the entire population, encompassing the health records of approximately 57 million patients. By leveraging a dataset comprising over 10 billion medical events—drawn from sources such as GP visits, hospital admissions, vaccinations, and the national death registry between 2018 and 2023—the Foresight model aims to predict more than a thousand potential health conditions, from serious complications to hospital admissions.

The project, spearheaded by researchers at University College London and King’s College London in collaboration with NHS England, the British Heart Foundation Data Science Centre, and Health Data Research UK, utilises Meta’s LLaMA 2 architecture. Significantly, the training occurs within NHS England’s Secure Data Environment, utilizing cloud infrastructure provided by Amazon and Databricks. This ensures that the model is grounded in real-world clinical data, setting it apart from many existing models that often rely on curated or simulated datasets.

The scope of Foresight is not only vast but also vital for the advancement of data-driven medicine. The project signifies a unique leap towards population-scale preventative care, suggesting a future where predictive models function akin to diagnostic weather systems, identifying at-risk patients long before any symptoms arise. Dr Chris Tomlinson, lead researcher at the UCL Institute of Health Informatics, articulated the ambition succinctly: “AI models are only as good as the data on which they’re trained. Using national-scale data allows us to represent the kaleidoscopic diversity of England’s population, particularly for minority groups and rare diseases, which are often excluded from research.”

Despite its groundbreaking potential, the Foresight project walks a precarious line between innovation and ethical concern. The lack of an opt-out mechanism for individuals raises questions about patient consent and data ownership. Current governance, operating under pandemic-era provisions, allows for broader data utilisation, but this is not without controversy. Transparency in how these powerful models operate will be critical to maintaining public trust. A public contributor involved in the review process voiced this concern, stressing the importance of understanding how health data is used and ensuring that AI acts in a safe and ethical manner.

Evaluation of the model’s predictive capabilities is underway, specifically its ability to forecast outcomes retrospectively for 2023 using historical data. Success in this area could catalyse a significant shift in healthcare—from reactive treatments based on retrospective analyses to proactive interventions aimed at preventing health issues before they escalate. Such a model could assist in identifying pre-frailty conditions and subtle signs of cognitive or metabolic decline, allowing for timely interventions to prolong healthspan and functionality.

The technical challenges of combining NHS data with AI capabilities cannot be overstated. Simon Ellershaw, a PhD researcher at UCL, explained, “Combining the computing resources needed for AI with NHS data has always been challenging.” However, the collaborative efforts with various partners have facilitated this significant step. Furthermore, the researchers intend to build upon the model’s foundation by incorporating richer data sources in the future, such as clinicians' notes, blood test results, and imaging data. Such enhancements could deepen the model's capacity to analyse biological aging metrics alongside disease forecasting.

The ethical and logistical complexities surrounding the Foresight model reflect broader conversations within healthcare about the balance between individual rights and public good. Dr Vin Diwakar, National Director of Transformation at NHS England, stated that the NHS Secure Data Environment is fundamental for enabling such pioneering research, highlighting that AI has the potential to revolutionise disease prevention and treatment when applied correctly. “This will boost our ability to move quickly towards personalised, preventative care,” he asserted.

As we advance towards a deeper understanding of ageing biology and its clinical applications, the need for large-scale operationalisation of risk data is becoming paramount. Foresight not only sets a technical precedent but also presents a vision for how health forecasting might underpin a new era of preventative healthcare—a concept that could redefine the landscape of 21st-century medicine.

Indeed, while Foresight exemplifies the potential of generative AI in healthcare, it equally brings forward the pressing need for ethical frameworks that ensure security, consent, and transparency in the use of sensitive health data.

Source: [Noah Wire Services](https://www.noahwire.com)

## Bibliography

1. <https://longevity.technology/news/uk-nhs-trains-ai-on-entire-populations-health-data/> - Please view link - unable to able to access data
2. <https://www.ft.com/content/b4c57347-d64d-436a-a2f1-33b7049a74b7> - This article discusses the potential of data and technology to address challenges faced by the UK's National Health Service (NHS). It highlights how effective data governance can enhance care quality, safety, and cost-effectiveness. AI technologies, according to IBM, could reduce treatment costs by 50% and improve outcomes by 40%. The piece also emphasizes the importance of strong data infrastructure and public trust in leveraging AI for healthcare improvements.
3. <https://www.ft.com/content/97762508-4070-4e4f-9e72-5e4cc34de60c> - This article covers Microsoft's expansion into AI healthcare, focusing on a chatbot developed under Mustafa Suleyman, formerly of Google's DeepMind. It explores the potential and challenges of using AI in health systems, including trust and data privacy issues. The piece also references past controversies, such as DeepMind's data collection from NHS patients, and discusses the need for clear rules around data ownership and profit from sensitive health data.
4. <https://www.ft.com/content/149f267b-7719-47d8-92a0-d7d88eb935db> - This letter emphasizes the transformative potential of AI technologies in healthcare. It argues that AI can revolutionize patient care by efficiently analyzing vast amounts of data, surpassing human capabilities. The letter also highlights the need to remove barriers to innovation in the UK's healthcare system and the importance of medical professionals adapting to and learning to utilize AI effectively.
5. <https://arxiv.org/abs/2212.08072> - This academic paper introduces Foresight, a novel transformer-based pipeline that uses named entity recognition and linking tools to convert document text into structured, coded concepts. It provides probabilistic forecasts for future medical events such as disorders, substances, procedures, and findings. The study processed the entire free-text portion from three different hospital datasets totaling 811,336 patients, covering both physical and mental health.
6. <https://en.wikipedia.org/wiki/OpenSAFELY> - OpenSAFELY is a secure analytics platform that interfaces with NHS patient records, enabling statistical analysis by medical researchers. Initially a collaboration between the DataLab at the University of Oxford, the EHR group at the London School of Hygiene & Tropical Medicine, and electronic health record software companies, the platform is now developed by the Bennett Institute for Applied Data Science at Oxford University. It has been used to analyze risk factors associated with deaths from COVID-19 in UK hospitals.
7. <https://en.wikipedia.org/wiki/UK_Biobank> - UK Biobank is a long-term prospective biobank study in the United Kingdom that houses the de-identified biological samples and health-related data of half a million people. The volunteer participants, aged 40-69, were recruited between 2006 and 2010 from across Great Britain and consented to both share their health data and to be followed for at least 30 years thereafter. The aim is to enable scientific discoveries into the prevention, diagnosis, and treatment of disease.