Data Science Community for Health and Care Newsletter February 2025

Jennifer Struthers, Amaia Imaz-Blanco

2025-02-25

Welcome to the latest newsletter from the Data Science Community for Health and Care, brought to you by the NHS England Data Science Professional Development Functional Team.

The newsletter team are always happy to receive constructive feedback, and we invite you to send us any contributions you may have.

If you cannot access something of interest to you, please reach out.

Thanks for reading! – newsletter team

Probabilistic Data Linkage in NHS England Goes Open Source!

The Data Linkage Hub has been developing a probabilistic data linkage pipeline to enhance the quality of data linkage in the NHS. We're excited to share that the first iteration is now openly available on Github!

Each stage of the pipeline has been shaped by extensive research into best practices, rigorous testing on samples of the real data, feasibility assessments of computational demands, and thorough evaluation. Our approach is incremental, with planned iterative improvements to ensure the highest standards of accuracy, efficiency, and scalability.

While this pipeline is not yet in use, it represents the foundation of what we plan to deploy within the Federated Data Platform (FDP). Our next step is to quality assure the pipeline, refining it further before implementation.

If you'd like to learn more, visit our website or get in touch!

Events

Lots of exciting things coming up! See the full calendar here, and a small selection below.

An Alan Turing Institute workshop on AI Ethics and Governance in Practice programme

Thursday 27th February, 14:00-15:00, Online

This workshop will introduce the AI Ethics and Governance in Practice programme developed by experts of The Alan Turing Institute's Ethics and Responsible Innovation theme.

This programme provides public sector organisations with the Process-Based-Governance (PBG) Framework, a structured framework to put ethical values and practical principles into practice across the AI project lifecycle. It is an integral part of the UK's National AI Strategy and expands on the UK Government's official Public Sector Guidance on AI Ethics and Safety.

NB: This event is only available to public sector employees

Statistical Methods for Health Equity Webinar: Catalina Vallejos (University of Edinburgh)

Thursday 6th March, 16:00-17:00, Online

Catalina will talk about "Using Routine Healthcare Data to Predict Future Health," exploring how electronic health records (EHR) can be leveraged to identify individuals at risk of adverse health events and improve patient outcomes. She will also discuss her work on SPARRAv4, a risk prediction tool soon to be deployed across Scotland to support primary care interventions.

Rethinking good technology: feminism, AI and ethical futures

Tuesday 11th March, 18:30-20:00, Online and In-Person (Old Theatre, Old Building, LSE, London)

What does it mean for technology to be 'good' in an age dominated by AI? Can a feminist perspective help guide us towards more ethical digital futures?

In this event Eleanor Drage talks about HEAT, a somewhat anarchic regulation tool that takes a feminist approach to helping companies meet the EU AI Act's obligations. The toolkit has been developed as part of a project led by Drage, which is committed to an in-depth response to regulation that goes beyond mere compliance by working towards a pro-justice and sustainable future with AI.

AI UK 2025

Monday 17th - Tuesday 18th March, All Day, London

Hosted by The Alan Turing Institute, AI UK is an in-depth exploration of how data science and AI can be used to solve real-world challenges. Our diverse programme was thematically structured around the latest innovations from across the AI ecosystem. With a broad range of interactive content, it covers the latest thinking on fundamental AI, digital twins, algorithmic bias, AI ethics – and much more.

NB: This is not a free event

An overview of causal machine learning methodology in Electronic Health Records (EHR)

Wednesday 19th March, 12:00-13:00, Online

Dr Maurice O'Connell will give a talk on current work with Dr Matthew Sperrin and the DynAIRx team in the area of statistics and causal inference and causal machine learning.

Maurice will talk about causal inference methodology development applied to electronic health records in the areas of deprescribing or continuing/initiating medications in individuals and populations with polypharmacy and multimorbidity and selecting individuals with polypharmacy and multimorbidity for structured medication reviews.

DS: Game On 2025

Saturday 17th May, All Day, London

Join us for our 11th festival, DSF Game On 2025! Top tech speakers, incredible partners and a thriving community, all completely free.

The ballot is open for those wanting a chance to get tickets to DSF's Game On 2025. Click through the link above to find out how to apply, and the (free) tickets to those successful will be sent out in April.

There will be a mixture of talks featured at the festival, covering all things data (science, engineering, etc) and at a variety of technical levels. You can view last years playlist on YouTube here to get an idea of the talks at the events.

Big Data LDN

Wednesday 24th - Thursday 25th September, All Day, London

Big Data LDN is the UK's leading free to attend data, analytics and AI conference & exhibition.

The two day event is a hub for the Data Community to learn and share best practice, build relationships and find the tools needed to develop an effective data-driven business.

The Call For Papers closes on 14 March and all applicants will be notified by the end of April whether their talk has been accepted. Find out more here.

See more future events on the calendar.

Know of any events we should feature next month? Let us know by clicking the "Contribute" button, or here.

Knowledge Sharing

CodeCarbon

It's well known that LLM's such as ChatGPT use a lot of energy, with estimates varying widely from 5 to 200+ times more energy than a Google search. With the popularity of LLM's for everyday use growing, the total energy consumption is rising. Being aware of the energy consumption of your code is not only more important than ever, it is also easier than ever to do so.

CodeCarbon is a open-source Python package designed to help you estimate the carbon footprint of computing tasks, particularly for machine learning and other high-performance computing workloads. It does this by estimating your hardware electricity power consumption (GPU + CPU + RAM) and applies the carbon intensity of the region where the computing is done. More is explained about this calculation in the Methodology section of the documentation.

Add this (or any other tracker you know of!) to your next pipeline and start taking steps toward more sustainable computing.

NHS Python Community - What is Python?

Curious about Python in healthcare? Or maybe you're tired of answering the same questions about it? The NHS Python Community has relaunched its newsletter on Substack, kicking things off with a post answering their most frequently asked questions:

What is Python?

Why is Python used in health and care?

Why should I learn Python?

What is open source?

Check out their first post here and subscribe to stay updated with future insights delivered straight to your inbox!

AI Journal Articles

There's a never ending list of new journal papers coming out every day - here's a few we have flagged as interesting that came out in the past month or so!

- **LLMs Can Plan Only If We Tell Them** - Bilgehan Sel, Ruoxi Jia, Ming Jin
- **Thoughts Are All Over the Place: On the Underthinking of o1-Like LLMs** - Yue Wang *et al.*
- **Hallucination Mitigation using Agentic AI Natural Language-Based Frameworks** - Diego Gosmar, Deborah A. Dahl
- **Language Models Use Trigonometry to Do Addition** - Subhash Kantamneni, Max Tegmark
- **Demystifying Long Chain-of-Thought Reasoning in LLMs** - Edward Yeo *et al.*

- **GPT-4 assistance for improvement of physician performance on patient care tasks: a randomized controlled trial** - Ethan Goh *et al.*

- **Superhuman performance of a large language model on the reasoning tasks of a physician** - Peter G. Brodeur *et al.*

- **Combining Human Expertise with Artificial Intelligence: Experimental Evidence from Radiology** - Nikhil Agarwal *et al.*

Seen a cool paper or article recently you would like to spotlight? Send us a contribution!

Need a Quick Break?

How many tries will it take you?

Subscribe to the communityContributePDF Version