Intelligent Data Categorisation:
New Insights for Healthcare

June 2025

Intelligent Data Categorisation
Foreword

We are not the first to highlight that the NHS is under-utilising the wealth of data at its disposal. Across our work with healthcare organisations, we consistently see the same challenges: siloed, unstructured or inaccessible data that hinders decision-making and slows progress. In the current climate, marked by spiralling economic pressures, post-COVID supply chain disruption and an NHS that is being asked to improve care and deliver transformation at scale whilst simultaneously reducing costs, this is a missed opportunity the system can no longer afford.

The NHS 10 Year Plan highlighted data and AI as a key transformation required for the future health system and without the ability to reliably access and analyse data at a system-wide level, the authors' ambitions will remain out of reach.

Intelligent data categorisation offers a practical, and scalable response. Recent advances in Large Language Models have made it possible to automate and dramatically improve data quality, turning fragmented datasets into usable, trustworthy intelligence. Akeso recently attended the NHS Confed Conference, where AI was a key topic and it was highlighted that 80% of data in the NHS is currently unstructured. At a time when integrated care, neighbourhood models, and the Federated Data Platform are moving to the forefront of national policy, the capability to harness your data is no longer a 'nice to have'; it's essential infrastructure.

Investment in high-quality data has a multiplier effect. With it, organisations can unlock the insights needed to identify process inefficiencies, drive cost-saving initiatives, and design new models of care with clarity and confidence.

At Akeso, we've been at the forefront of this shift: our team has developed and deployed AI-driven categorisation tools across clinical, operational, and commercial domains beginning with a medicines optimisation tool in 2019. We've seen first-hand how intelligent categorisation can help NHS organisations and suppliers translate data into measurable savings, increased efficiency, and, ultimately, better patient outcomes.

We believe operational and non-clinical use cases represent the most effective and lowest-risk starting point for AI adoption in health and care. These domains are offer immediate value and greater potential as they are often overlooked areas of value. They also help build the internal capabilities and confidence needed for more complex clinical applications.

Nevertheless, we must confront the broader question: AI...Curse or Cure? While public debate often swings between extremes, our experience suggests that, with the right safeguards and a clear purpose, AI is neither a silver bullet nor existential threat. It's a powerful enabler of better decisions, better use of resource, and better care.

Our view is clear. With the right focus, intelligent categorisation offers a high-impact, low-risk path to AI adoption, starting where it counts, and scaling with confidence, to achieve in-year cost reduction, operational improvements and the ability to make integrated care a reality.

Dr Luke Taylor
Head of Analytics

Intelligent Data Categorisation
Contents

Intelligent Data Categorisation
Executive Summary

The NHS operates in a data-rich but insight-poor environment. Across both clinical and operational domains, vast volumes of information remain unstructured, inconsistent, or incompatible between systems, limiting the ability to make informed decisions, optimise performance, and deliver better care.

This whitepaper explores the transformative potential of intelligent data categorisation, powered by AI, to overcome this challenge. From reducing the burden of manual work to enhancing visibility over procurement and enabling a smoother transition to the Federated Data Platform (FDP), categorisation is a critical enabler of a modern, connected healthcare system.

The background to categorisation and standardisation is laid out, alongside the recent advances in AI which enable its transformation.

Drawing on real-world case studies, including NHS Trusts and national suppliers, we demonstrate how intelligent categorisation can be deployed across a range of use cases:

- Non-Pay Spend: Increasing visibility of non-pay spend and identifying procurement inefficiencies, unlocking cost-saving opportunities through product switching and procurement strategy.

- Clinical Coding & Notes: Reducing manual workload and enabling structured analysis of clinical narratives with very high accuracy.

- FDP Transitions: Accelerating the mapping of legacy data systems to standard models, reducing effort and improving data readiness at speed.

The evidence shows that categorisation, long proven in clinical settings like ICD-10 and dm+d, can now be expanded into operational domains with scalable, AI-driven tools. The benefits include improved financial oversight, enhanced interoperability, more efficient workflows, and ultimately, more effective patient care.

However, categorisation is not a plug-and-play solution. It requires adaptable tooling, clinical and operational context, and human oversight. Through our work with NHS organisations, Akeso has developed a repeatable model combining semantic AI, healthcare taxonomies, and domain expertise to deliver actionable, trusted insights.

As cost pressures increase and digital transformation accelerates, the case for intelligent categorisation is no longer theoretical. It is practical, proven, and ready to scale.
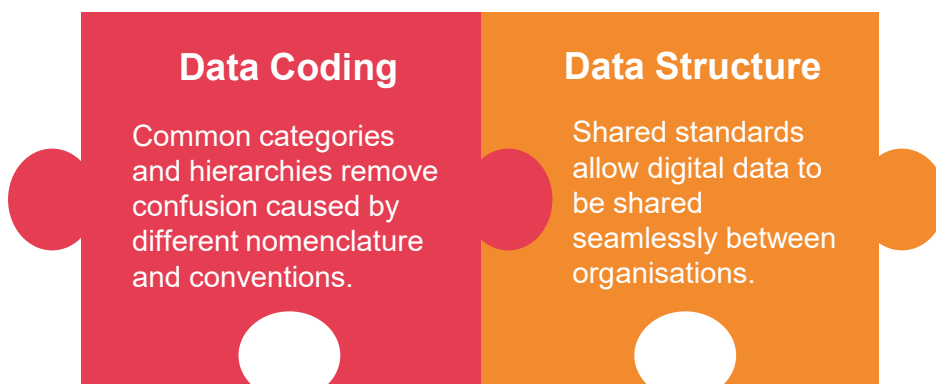
Intelligent Data Categorisation
The Need for Structured Data

The NHS is clear in its priority to improve data quality and interoperability within the service. The 'analogue to digital' mantra within the NHS 10 Year Plan [1] and the rollout of the Federated Data Platform (FDP) point to the NHS chiefs' acknowledgement that improved digital maturity leads to reduced costs, released staff time and improved patient care. The provocatively named "Data Saves Lives" policy paper from DHSC outlines the wasted time and poor decision-making that can result from bad data quality and calls for "unified architecture, open standards and better interoperability" [2].

This priority is now set against the drastic cost-cutting targets faced by all levels of the NHS. [3] Much of the attention on structuring data has, rightly, been focused on frontline services. Multiple factors within the data ecosystem of current NHS and wider healthcare system work against the goal of a more efficient service. Hospitals, community services, GPs and pharmacies all have different conventions and standards based on their local context which can cause confusion and error when it comes to sharing information [4], a problem which will be exacerbated under the recent plans to bring interdisciplinary teams together into neighbourhood health services [1]. But other, less visible, parts of the healthcare system are often missed in the discussion and these overlooked areas can present significant opportunities for savings that cannot afford to be missed.

Data Interoperability

Data Coding

Common categories and hierarchies remove confusion caused by different nomenclature and conventions.

Data Structure

Shared standards allow digital data to be shared seamlessly between organisations.

There are two types of data standardisation that must occur to enable this shift. The first, data coding, is the categorising of information into standard codes, such as classifying diseases and conditions into a medical hierarchy. The second, data structure, is the standardisation of fields and nomenclatures across different systems in order to improve interoperability. Even if data is well coded in two systems, if it is coded and stored with different conventions then they won't be able to share information effectively.

Intelligent Data Categorisation
The Need for Structured Data

Especially in clinical settings, the NHS has many examples of successful data standardisation: ICD-10 and SNOMED CT standardise disease classification and medical terminology, OPCS-4 codifies activity and FHIR ensures a common language for sharing data digitally between systems; many more coding standards are specified in the NHS Data dictionary.

This data standardisation brings many benefits such as improved interoperability, a smoother tracking of patient pathway, reduced manual effort from clinicians, reduced human error and a greater ability to get data-driven insights at all levels of the health service[5]. For example, in prescriptions, secondary care tends to use dose-based prescriptions, whereas product-based prescriptions are used in primary care. In the NHS, coordinating medicines is done using the NHS Dictionary of Medicines and Devices (dm+d) which encodes medicine information into standardised SNOMED CT codes and translates between dose-based and product-based prescriptions. All of this is shared in a standardised FHIR format between organisations.

Example: Standardising Prescriptions

At Somerset NHS Foundation Trust, they estimate that the implementation of dm+d into their prescription system resulted in an over **50%** reduction in time taken to create medicines on admission list and over a **90%** reduction in transcription during medication history taking and discharge. There were also numerous secondary benefits, such as improved medicine management, decision support, an increase in integrated care and reduction in unwarranted variation [5].
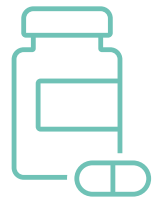
While the benefits of data standardisation are clear, there are costs. Data on the ground is often not pre-categorised and classifying each element to fit within taxonomies requires additional effort from the staff on the ground. Outside of clinical settings, data standardisation across the healthcare sector is less consistent but the benefits are no less considerable. This whitepaper presents several use cases of areas that would benefit from further intelligent categorisation.

Our hypothesis for what is holding back this work is the cost of manually processing the data. While often mandatory within a healthcare setting, the investment needed to unlock the benefits is harder to realise for back-office functions. An ability to categorise data automatically can unlock the benefits that this brings without the human effort that would otherwise be required. Recent advances in AI and Large Language models have enabled this to be attainable by healthcare organisations.

Intelligent Data Categorisation
AI: Curse or Cure?

The field of AI focused on language, Natural Language Processing (NLP), has seen significant advancements thanks to Large Language Models (LLMs). In recent years, there has been a paradigm shift in language processing, facilitated by the introduction of LLMs, that has enabled advanced and reliable semantic understanding.

LLMs are the sophisticated AI systems trained on huge amounts of textual data, capable of interpreting meaning based on the surrounding context and generating human-like text. They underpin transformative applications ranging from translation systems to intelligent chatbots, and are used across a wide range of industries, including highly regulated industries such as banking, to uncover insights [6].

A key improvement driving the breakthrough in language NLP is in *how* LLMs are able to understand the meaning of words. Transformer-based LLMs use 'word embeddings'; where words with similar meaning are represented to be mathematically similar. The embeddings use the context of the surrounding phrase to disambiguate similes, ensuring 'doctor' (a clinician) is understood as being similar to 'nurse' but 'doctor' (as in, to alter) is not. Similarly, a bandage and a dressing may be different terms, but LLMs can understand they belong to the same functional category.
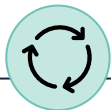
Adaptability

LLMs are pretrained on vast and diverse datasets, making them highly adaptable. Their extensive training ensures familiarity with virtually any input, enabling consistent, reliable performance across any scenario.

Accuracy

LLMs capture context and subtle nuances of language effectively. This context-sensitivity results in more accurate and meaningful analysis, closely mirroring human understanding of language.

Scalability

Information is represented mathematically in LLMs, enabling rapid and efficient processing. LLMs can handle millions of items simultaneously, allowing large scale applications that would be impractical manually.

While previous uses of NLP have failed to meet the required standard for widespread use. This capability is increasingly being recognised and regulated in healthcare. The NHS has issued detailed implementation guidance for tools like Ambient Scribing, requiring safety assessments, DPIAs, and integration with Electronic Patient Records (EPRs), illustrating a proactive approach to safe deployment of generative AI in clinical workflows [7]. These tools hold clear promise within clinical settings to improve patient care and save clinical time, and their real-world impact is already being demonstrated in other regulated industries such as banking and law [8].

Akeso

Our Experience:
Non-Pay

Poor data standardisation in non-pay spending
prevents actionable insights that can drive operational
and financial improvements

Intelligent Data Categorisation
Our Experience: Non-Pay

Non-Pay spend, a rich source of financial and operational opportunity, is often overlooked when investigating cost-saving initiatives. In this section we will investigate real case studies and explore how intelligent categorisation can be applied to Non-Pay.

Why Non-Pay?

Healthcare providers are frequently purchasing the same product from different suppliers at vastly different prices, often without realising. With no shared standard across NHS organisations, it's nearly impossible to compare supplier prices or spend profiles, making it difficult to spot cost inefficiencies or argue for better deals. Without a standardised description of what is being purchased, organisations lack visibility into what they're spending, and more importantly, where they could be saving.

Non-Pay expenditure, defined as the spend excluding salaries and wages, holds immense potential for cost-saving and productivity improvement. This data, typically captured in accounts payable (AP) and purchase order (PO) systems, records millions of transactions containing valuable details such as product price, volume, purchase date, and supplier. But despite this richness, much of the information exists in an unstructured format, especially in item descriptions held as free text. The result is a lack of insight: organisations are sitting on data that could be working for them but isn't.

Driving Spend Control and Procurement Optimisation

When non-pay expenditure is intelligently categorised, healthcare organisations gain visibility into purchasing behaviour, supplier trends, and product usage patterns. This foundational clarity enables both detailed analytics and strategic optimisation. By transforming unstructured and often overwhelming volumes of transactional data into a structured format, organisations can unlock valuable insights that would otherwise remain hidden. This allows procurement teams to:

- Track cost variation across products and suppliers
- Benchmark spend across departments or trusts
- Streamline approvals and reduce wasteful purchases
- Identify price discrepancies and procurement inefficiencies
- Monitor spend by addressable category

With this level of insight, procurement leaders are better positioned to drive efficiencies, reduce operating costs, and extract more value from every purchasing decision.

Intelligent Data Categorisation
Our Experience: Non-Pay

Applying Categorisation in Practice

Intelligent categorisation offers a practical and scalable approach to resolving some of the most persistent challenges in NHS procurement. By grouping similar or medically equivalent items, organisations can reduce variation, consolidate spending on products, and unlock significant cost savings. While a degree of manual validation is still required, particularly in clinical contexts, the process can be largely automated, improving speed, accuracy, and resource efficiency.

These categorisation methods are not only theoretically sound but are also being successfully deployed across NHS organisations. The following case studies demonstrate how we have applied intelligent data categorisation to drive measurable improvements in spend control, procurement optimisation, and commercial strategy.

Case Study: Delivering Control Through Hierarchy-Driven Spend Review

Akeso worked with a major NHS Trust to improve control over in-year non-pay expenditure without compromising patient care. Applying our intelligent data categorisation tool to the Trust's purchase order data, we enabled a semi-automated spend review process where essential purchases, like medical items, were fast-tracked for approval, while non-essential items were flagged for deferral.

Receive Purchase Order data	Categorise spend using AI models	Automatically flag categories for review	Manual review of flagged spend by the Trust

In addition, price benchmarking was used to group similar products and recommend more cost-effective alternatives. The result was a clearer oversight for the Trust, a more manageable workload for reviewers, and improved overall governance, all without disrupting frontline care.

Case Study: Delivering Reduced Costs Through Strategic Product Switching

Example Hospital Cost Savings

Average price paid

Existing £1,035k — Switched £890k (-14%)
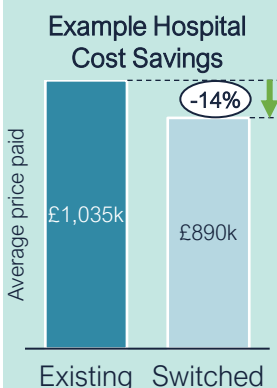
Akeso partnered with a large medical supply chain company to transform their offer to customers and deliver measurable financial outcomes. The client's strategic agenda was to highlight their own-brand products to customers as a cost-effective alternative. Using intelligent data categorisation, Akeso identified an opportunity which amounted to a 30% shift in spend to cheaper products, saving an average of 14% on product costs per item. For one hospital that equated to £145k in savings.

We also delivered an improved pricing strategy, enhanced inventory visibility, and built a stronger commercial negotiation positioning. This demonstrates the power of structured data to unlock both operational and financial value.

Use Case:
Clinical Coding Automation

Clinical coding drains expert time on repetitive tasks that could be better spent on complex coding cases and quality improvement

Intelligent Data Categorisation
Use Case: Clinical Coding Automation

Clinical coding is a high-value, high-cost task. In 2020, the NHS employed over 3,000 clinical coders at an estimated cost of approximately £100 million *, [9] . It takes roughly two years to qualify as a clinical coder, and many go on to specialise further within particular domains [10]. Despite this level of training, a large portion of their time is spent on repetitive and routine work, reading through clinical notes and manually converting them into structured codes. Each case takes an average of 10 minutes to process and often involves little to no interpretation or follow-up [11].

A 2022 study explored the feasibility of AI-led automation in clinical coding, concluding that while the potential was evident, significant technical and organisational challenges remained [11]. Since then, developments in Natural Language Processing, especially the release of tools like ChatGPT, have made rapid progress in addressing many of those early limitations. In 2023, a study using an LLM to convert free-text clinical narratives into structured FHIR format achieved an exact match accuracy of over 90% against human annotations, demonstrating a major breakthrough in automating clinical data standardisation at scale. [12]

Metric	Manual Coding	AI + Human
Time per Case	10 Min	3-5 Min
Error Risk	Medium - High	Low
Suitability for Complex Cases	✓	✓
Time Spent on Repetitive Tasks	High	Low

We envisage a semi-automated workflow, tied in with the NHS' push for AI scribes [1] , in which AI transcribes clinical notes and performs an initial coding pass, highlighting relevant text and assigning confidence scores to each recommendation. This structured output enables coders to efficiently verify, amend, or prioritise cases based on complexity and risk, embedding AI into existing review processes rather than replacing them.

With NHS spending ~£100M a year on clinical coding: AI can help save time, reduce errors, and let human experts focus where they're needed most.

* Estimated based on average clinical coding salaries plus associated costs

Use Case:
FDP Transition

Manually mapping legacy systems to FDP is a laborious process that is slowing Trust adoption and undermining the urgency of nationwide data consolidation
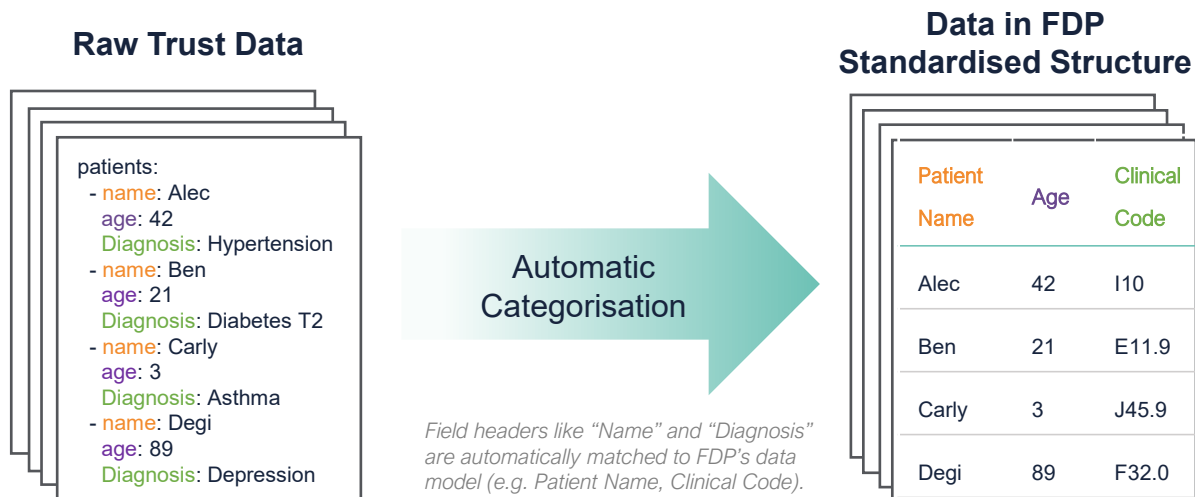
Intelligent Data Categorisation
Use Case: FDP Transition

The Federated Data Platform (FDP) is a major transformation programme that aims to unify how health data is structured and shared across England. With just 45 Trusts onboard and another 108 signing a memorandum of understanding, the majority of NHS organisations remain at the very start of their journey [13]. Yet, NHS England has written to all Trusts requesting they transition by mid-2026, placing significant urgency on this shift. [14]

The aim of the FDP is to standardise the data model that each Trust is working with. One of the first steps that every Trust will need to go through is to map their existing systems onto the FDP common data model. Mapping every field header, and the values within each field, to the new FDP ontology for every system that the Trust uses for appointments, bookings, clinics, patient databases, theatre management, workforce organisation and more is a manual and onerous task. A tool which can reduce the effort to transition to the FDP should improve the slow take-up of the FDP amongst Trusts and speed up the time to realising benefits from the FDP. The mapping process is a very similar challenge to information categorisation.

Raw Trust Data

Data in FDP Standardised Structure

patients:
- name: Alec
  age: 42
  Diagnosis: Hypertension
- name: Ben
  age: 21
  Diagnosis: Diabetes T2
- name: Carly
  age: 3
  Diagnosis: Asthma
- name: Degi
  age: 89
  Diagnosis: Depression

Automatic Categorisation

Field headers like "Name" and "Diagnosis" are automatically matched to FDP's data model (e.g. Patient Name, Clinical Code).

Patient Name	Age	Clinical Code
Alec	42	I10
Ben	21	E11.9
Carly	3	J45.9
Degi	89	F32.0

Intelligent categorisation can assist the FDP Transition by automating the restructuring of raw trust data into a standardised format

Most Trusts operate within 30-70 separate clinical and operational systems, including EPR/PAS, ePrescribing, LIMS, many of which can contain over 100 tables (not accounting for derived views). The FDP has nearly 600 required fields that will need to be mapped to the corresponding one amongst the thousands of tables. Assuming Trusts have dedicated resource to manually map datasets onto the new FDP ontology, even if only 30 minutes are spent per field and an automated tool could save 75% of that time it would save 215 months of FTE across the NHS's 170 trusts which are yet to join the FDP.

Using a tool for the process would mean a more consistent mapping between similar categories across Trusts, so that data sharing is more easily facilitated in the future.

Akeso

Data Standardisation
What's Next?

Categorisation and data standardisation are critical enablers of a modern, interoperable NHS. Its implementation, particularly in clinical domains, has already delivered substantial benefits, including reduced transcription burden, enhanced patient safety and improved decision support. These successes, and the increasing importance of data interoperability during the rollout of Neighbourhood Health services, encourage its use to be extended into non-clinical domains.

Despite its value, categorisation remains a challenging and resource-intensive task. The work is often manual, slow, and complex. Successful standardisation demands adaptable tools, deep domain knowledge, and functional understanding of the wider healthcare system.

At Akeso, we believe that AI has now reached the required competence and reliability to enable the expansion of categorisation and standardisation within the NHS.

Akeso has built a general-purpose, proprietary tool that categorise a wide variety of data into a structured format. It has already shown tangible operational and financial outcomes enabling in-year spend control for an NHS Trust and optimising the product catalogue of a national medical supplier and our team has continued to explore untapped areas for the NHS to investigate further.

This whitepaper has outlined specific areas in which to begin this adoption, focusing on areas away from direct clinical use that are safer for AI pilots but which still deliver impact:

- Non-Pay spend analysis
- Automating areas of clinical coding
- Transitioning data ontologies to the FDP

Having supported transformational change across more than 20 NHS Trusts, regional partnerships, and national healthcare programmes, Akeso has developed a deep, practical understanding of what it takes to deliver successful data categorisation at scale. The following principles have consistently proven critical to successful implementation :

- Prioritise high-volume, high-impact datasets that underpin clinical or operational decision-making. These domains deliver the fastest return on investment and the clearest pathway to organisational buy-in.

- Embed validation mechanisms into the operational pipeline from the outset. Integrating lightweight validation steps, particularly in clinically sensitive areas, not only improves data quality but also builds trust among end users.

- Foster cross-functional collaboration by bringing together clinical stakeholders, procurement leads, data architects and transformation teams. So that technical delivery is aligned with frontline priorities.

- Ensure strong alignment between technical outputs and organisational goals. Categorisation efforts are most effective when directly tied to operational pressures and create traction beyond the project team.

The value of data standardisation has already been proven. By using AI to automate and accelerate the categorisation process, the health system can begin to build an ecosystem of structured, interoperable data. Akeso is ready to support that journey. We combine deep health system expertise with powerful, adaptable technology to help organisations unlock the full value of their information.

Akeso

AUTHORS

Harry Style
Senior Data Scientist

Rio Carasco
Data Scientist

CONTACT US

Luke Taylor
Head of Analytics

T: (+44) 20 3011 1381
E: luke.taylor@akeso.co.uk

www.akeso.co.uk

REFERENCES

[1] DHSC (2025). 10 Year Health Plan for England: fit for the future *GOV.UK*

[2] DHSC (2022). Data saves lives: reshaping health and social care with data *GOV.UK*

[3] NHSE (2024). 2025/26 priorities and operational planning guidance *NHS England*

[4] NHSE (2023). Building an integrated care system intelligence function *NHS England*

[5] Health Innovation NENC (2021). Clinical coding in the NHS: An article exploring the impact and importance. *Health Innovation NENC*

[6] IBM (2025). AI in banking. *IBM Think*

[7] NHSE (2025). Ambient Scribing Guidance for AI Assistants. *NHS England*

[8] Nature Medicine (2022). Federated learning in healthcare: a primer. *Nature*

[9] NHS Digital (2021). Clinical coding workforce data pack. *NHS Digital*

[10] NTH NHS Trust (2025). A day in the life of a clinical coder. *North Tees and Hartlepool NHS Foundation Trust*

[11] Dong et al. (2022). Automated clinical coding: what, why, and where are we?. *Nature*

[12] Li et al. (2023). FHIR-GPT Enhances Health Interoperability with Large Language Models. medR*xiv*

[13] NHSE (2024). FDP uptake and benefits. *NHS England*

[14] Digital Health (2024). NHS England says Trusts must start using FDP within two years. *Digital Health*

Akeso

Thank you