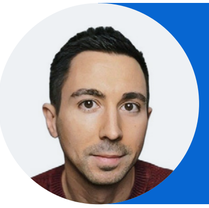
Bridging the government data divide

How could synthetic data speed up AI adoption in the public sector?

A report by SAS UK in partnership with Tussell

Contents

Introduction



By Dr Iain Brown, Head of AI & Data Science at SAS Northern Europe

Across the UK's public sector, we're seeing investment in artificial intelligence (AI) ramping up. From the launch of 'Humphrey', an AI tool designed to make the civil service more efficient, to a pledge to make the NHS the 'most AI-enabled care system in the world',[1] policymakers see the technology as a way to deliver better outcomes, more cost-effectively.

AI, deployed in the right way, isn't just a cost-saving exercise. It's an opportunity we have now to reshape public services for the better. Persistent backlogs and unnecessary bureaucracy could disappear, while innovative AI-driven services would give citizens more autonomy. These investments are only set to rise in the coming years. Political will is being matched by advances in AI – including Natural Language Processing (NLP), generative AI and agentic AI – which together pave the way for new applications.

Yet to fully realise the potential of AI, policymakers must first address what has been dubbed as the government data divide – gaps in the quality and availability of data needed to train AI models effectively and win public trust.

The purpose of this report, produced in collaboration with Tussell, a market intelligence platform for public sector procurement, is twofold.

1. Explore current and future spending across government departments and regions, both on IT projects in general and AI specifically, and what the pipeline of contracts looks like for the coming years.
2. Explain why synthetic data is key to overcoming the data challenges facing organisations – enabling them to both accelerate and gain value from AI projects.

[1] Source: UK Government

Chapter 1

AI in the public sector: Where are we now?

The UK is a leader in AI – with a market valued at £72.3bn, making it the largest in Europe and the third largest in the world.[2] At the start of 2025, the government unveiled its AI Opportunities Action Plan[3], and we're already seeing investment pour in – including a $30bn (£22bn) deal with Microsoft to develop the UK's AI infrastructure, particularly data centres.[4]

The value of AI to businesses of all sizes is becoming increasingly clear, with uses spanning forecasting and proactive decision-making; automating routine and time-consuming tasks; and new innovations that improve customer experiences and competitiveness.

It's these capabilities that will make AI a powerful force for the public sector too, at a time when many services are creaking under the weight of high and changing demand, persistent backlogs, staffing shortages and outdated processes/technologies. As many as 75% of people think that public services have worsened over the past five years and another 64% say the same about their local areas.[5]

AI could tackle many of the new and longstanding challenges that frustrate people, and lead to poorer and often inequitable outcomes. However, the road to large-scale AI adoption isn't straightforward, and like any transformation project, it has to deliver demonstrable value.

One of the biggest stumbling blocks is the government data divide.

The Public Accounts Committee has laid bare the scale of the challenge, describing it as an 'uphill struggle' – one of the key issues it identifies is that data is 'poor-quality, and often locked away in out-of-date, or 'legacy', IT systems'. Far from enhancing services, the report's authors suggest that out of date technology and poor-quality data could undermine public confidence, particularly when it comes to 'transparency in use of algorithms'.[6]

The next section highlights both the actual and projected spending on government IT and AI projects. As investment increases, so too does the need for a strong foundation of data to ensure that decisions are fair, explainable, robust, human-centric and protect fundamental rights such as privacy.

What is the government data divide?

- Data is siloed in different systems/ departments, and is out-of-date or recorded incorrectly or inconsistently (although some effort has been made to enable cross-government data sharing through initiatives such as the National Data Strategy).[7]
- Unstructured data is often difficult to manage, interpret and extract value from.
- Privacy and security limitations, especially in sectors like defence and healthcare.

[2] Source: UK Government
[3] Source: UK Government
[4] Source: BBC News
[5] Source: Ipsos
[6] Source: UK Government
[7] Source: UK Government

Chapter 2

Analysis: Government spending on IT and AI projects

The government cannot adopt AI effectively without having stable, secure IT systems and reliable data infrastructure in place. This is why investment in AI must be considered in the broader context of IT procurement.

Tussell's data uncovers an unsurprising jump in both government IT and AI contracts over the past six years – reflecting initiatives like the Government Cloud First Policy and more recent AI pilots. It also shows which parts of the public sector (both departments and regions) are leading the way with new initiatives and what's in the pipeline for the coming years.

Digital transformation is essential to reforming public services. The strong rise in IT procurement spend since 2018 shows that the government gets this and is investing accordingly. As this report demonstrates, however, a 'government data divide' is still an obstacle in the way of promoting much greater use of AI. Adoption of synthetic data would help to bridge this divide by enabling public sector teams to test, share and scale AI more safely. With clear AI governance and stronger data literacy, the UK government can seize the opportunities of AI to make public services work better for citizens.

Gus Tugendhat
Founder, Tussell

IT procurement spend

Government spending on IT procurement has risen steadily over the past six years, from just under £11 billion in 2018 to almost £19bn in 2024. The biggest increase came in 2019/20 at just over 17%, followed by 2018/19 at around 12%. New digital demands linked to Covid-19 helped drive growth in the IT market but it didn't snap back to pre-pandemic levels. Instead, spend has continued to grow from this higher baseline. Covid exposed the gaps in digital readiness, prompting government departments to start and accelerate projects.

Year	IT procurement spend
2018	£11.0bn
2019	£12.3bn
2020	£14.4bn
2021	£15.9bn
2022	£17.3bn
2023	£18.0bn
2024	£18.9bn

IT procurement spend, by buyer category in 2024

Central government leads the way in terms of IT spending with procurement spend worth just over £9bn in 2024 – more than double the second highest spender, the NHS, at £4.61bn.

Spending in the NHS increased the most out of any department, jumping by nearly 150% between 2018 and 2024 as digital transformation plans got underway. Central government spending, in contrast, grew by a more modest 53% and local government by 48%. Growth in central and local government was slower than across the total IT market – meaning that the NHS took market share from them.

Buyer	IT procurement spend
Central government	£9.1bn
NHS	£4.6bn
Local government	£2.8bn
Other	£1.3bn
Emergency services	£1.0bn

Top IT buyers (including arm's length bodies) in 2024

The Department of Health & Social Care (DHSC) is leading the way on spending at £1.45bn – well ahead of the Home Office at £1.01bn. These disparities remind us that the public sector shouldn't be considered as a single entity – each department has its own priorities, budgets and applications for technology.

Buyer	IT procurement spend
Department of Health & Social Care (DHSC)	£1.5bn
Home Office (HO)	£1.0bn
Department for Transport (DfT)	£998mn
Department for Work & Pensions (DWP)	£976mn
Department for Environment, Food & Rural Affairs (Defra)	£499mn
Metropolitan Police	£451mn
Department for Education (DfE)	£429mn
HM Treasury (HMT)	£305mn

IT procurement spend, by buyer region (excluding Central Government and NHS England)

As well as disparities between departments, the analysis also found significant regional variations in IT spending too. London leads the way in 2024, with £2.16bn, while the North-West – home to major cities like Manchester and Liverpool – was second highest after London at £987.03mn.[8]

Worryingly, there's a gap between spending in some of the most and least affluent parts of the UK. Digital technology has the potential to reduce existing economic, social and health inequalities but patchy investment could in fact exacerbate them.

Region	IT procurement spend 2024
London	£2.2bn
North West	£987mn
West Midlands	£919.4mn
South East	£914.8mn
Scotland	£636.9mn
East of England	£619.3mn
South West	£571.9mn
Yorkshire and the Humber	£544.5mn
East Midlands	£470.6mn
North East	£242.6mn
Wales	£168.2mn
Northern Ireland	£22.7mn

[8] Coverage varies across devolved nations; only English bodies are legally required to publish spend data.

AI contract value

Like general IT procurement, there's been a rise in the number of AI contracts awarded – with a 64% jump in contracts between 2023 and 2024 alone. Between 2018 and 2024, the volume of contracts jumped by a staggering 1,085%.

It's difficult to chart a straightforward trend of AI adoption because the value of contracts spikes around a major initiative, like the Met Office's £1bn supercomputing contract, awarded to Microsoft in 2021. That can mean departments working with different suppliers for different projects – which must be carefully managed to ensure data and systems are compatible, and enable data-sharing and collaboration.

Year	IT procurement spend	Contract volume
2018	£59mn	26
2019	£27mn	62
2020	£132mn	99
2021	£1.14bn	157
2022	£212mn	197
2023	£498mn	188
2024	£466mn	308

AI and data centre pipeline notices

It's not just AI award volume that's increasing – there's also a healthy pipeline of AI-related procurement activity, across a range of different buyers, between 2025 and 2027. The table below shows a total pipeline value of more than £124 million, and it's dominated by the NHS.

Organisation	Pipeline notice value
University Hospitals Birmingham NHS Foundation Trust	£91.3mn
NHS England (DHSC ALB)	£12.0mn
NHS England (DHSC ALB)	£10.5mn
Student Loans Company Ltd (DfE ALB)	£4.7mn
Crown Prosecution Services (CPS) (AGO ALB)	£3.0mn
Student Loans Company Ltd (DfE ALB)	£2.4mn
Mayor's Office for Policing & Crime (MOPAC)	£0.6mn

Taken together, insights into current and future spending highlight the government's determination to build reliable IT infrastructure even if there are still obstacles to overcome. Disparities in AI readiness between departments (e.g. central government and local government) and regions already suggest that some will be left behind by AI advances.

It's certainly welcome to see the UK positioning itself as a leader in safe and trustworthy AI with its 'Scan > Pilot > Scale' approach, detailed in the AI Opportunities Action Plan. Designed to build on successful pilot projects, and break down siloes between departments, it could drastically improve both the efficiency and outcomes of IT projects.

Still, against this backdrop of spending, the question remains: what can be done to bridge the government data divide?

Chapter 3

Unlocking the potential of synthetic data

There's no single solution to the data divide. As a minimum requirement, departments need to standardise their processes in order to capture, validate and maintain the integrity of their data. But better data management alone isn't enough, since real-world data is often limited or lacking diversity. There's also a risk that confidential information will be leaked via AI models, threatening individuals' confidentiality or national security.

Where real-world data is lacking, synthetic data can pick up the slack – enabling projects to get off the ground faster. This algorithmically generated data mimics the statistical properties and correlations of real-world data, reducing re-identification risk when properly governed and counteracting the risk of bias.

Like any form of data, it can be used to train and validate AI models, augmenting existing datasets and enabling better decision-making/scenario forecasting for 'edge cases' (such as the prevalence of diseases typically associated with old age in younger patients). As one BMJ article makes clear, it also reduces the risks around data sharing, paving the way for better practices across teams, departments and organisations.[9]

In experiments, synthetic data has shown better utility than anonymised data in some contexts.[10] It not only enables AI but accelerates it, so the value of projects could be realised faster. However, it's not without its risks and critics. Recently, the Ada Lovelace Institute warned of the potential harms associated with synthetic data – including issues like data contamination and model collapse. Its experts suggest that synthetic data could also result in hallucinations (as with generative AI) and fail to capture the nuances of real-world scenarios.[11]

However, the government's AI Opportunities Action Plan recognises the importance of exploring synthetic data and states: "We should explore use of synthetic data generation techniques to construct privacy-preserving versions of highly sensitive data sets. Government data sets are a public asset, and careful consideration should be given to their valuation." [12]

[9] Source: BMJ
[10] Source: SAS
[11] Source: Ada Lovelace Institute
[12] Source: UK Government

Overcoming the synthetic data risks

Poor data of any kind reduces the reliability of AI decision-making but synthetic data could amplify these risks – repeating ingrained biases or painting a false or misleading picture. But with good governance and continuous monitoring, you can mitigate the risks and develop a dataset that is more accurate, representative and fairer.

Challenge 1

Reinforces bias

Solution: In-built fairness and bias mitigation features are critical. For example, an exponentiated gradient reduction (EGR) algorithm considers fairness constraints during model training and adjusts the parameters to create a model that generates fair predictions and classifications.

Challenge 2

It doesn't reflect real-world complexity

Solution: Recognise that synthetic data shouldn't replace real-world data but augment it. By using diverse collection methods to collect real-world data, you can minimise biased and misleading results when creating synthetic data.

Challenge 3

Model collapse

AI models will degrade if there's no fresh data and/or an over-reliance on synthetic data, leading to potentially unfair/biased outcomes.

Solution: Ongoing capture of real-world data from a diverse range of sources to create high-quality data to train AI models. Also critical are data validation processes, including auditing and policies for labelling and identifying where synthetic data is deployed. Teams must also understand where synthetic data can and cannot be used, and its limitations.

The expert view: Dr Iain Brown

Dr Iain Brown, Head of AI & Data Science for SAS Northern Europe, explains how the public sector could use synthetic data to achieve better datasets. It could also deliver AI projects more quickly than typical pilot timelines, improve governance and reduce the time it takes for IT projects to deliver value.

"We are now starting to see signs of early interest and engagement in synthetic data in the public sector, with some departments exploring it in private previews and pilots. While the concept might not be widely known, many people in government recognise the limits in their current data. They're not able to use it to get the answers they need because it's poor-quality or siloed, even within departments.

The value of public sector AI contracts has ramped up in recent years, yet they still only account for a fraction of overall IT spending. But synthetic data could both speed up adoption, and ensure that departments get the value from current spending. Any AI investment, both current and future, won't deliver value until they can access high-quality and well-governed data.

Developing IT infrastructure is a long-term project, often spanning decades. The good thing about synthetic data is that it can be created and implemented in just a few weeks, creating a centralised, transparent and depersonalised dataset to enable decision-making. Users are able to stress test multiple scenarios to quickly and safely understand the potential outcomes, so it's particularly valuable in high-risk sectors like law enforcement and defence.

Synthetic data enables the 'test and fail' mentality needed in AI development, creating a faster and more dynamic decisioning environment. This is important as there is often concern in some parts of government about the risk of failure or delay, given the public scrutiny that comes with adopting new technology solutions.

In the case of agentic (or autonomous) AI, you also need a regular supply of data. You can't afford to wait six months for new real-world data because agentic AI models need to adapt quickly, so synthetic data can also address this issue.

All this takes us closer to a future state where government agencies are able to work collaboratively, sharing depersonalised information to develop better-informed policies and become more proactive in tackling everything from fraud to health inequalities.

You'll always need real-world data – and the better quality it is, the better your synthetic data will be. As data capture and analysis methods improve, so too will synthetic data. AI is advancing all the time too. Developments in Natural Language Processing (NLP) mean that we can extract unstructured data from any source including PDFs and image files, and synthesise it in a safe and controlled place.

It goes without saying that good data governance is needed to manage both real-world and synthetic data and AI models and decisions. That means building transparency into the whole process so that users, both technical and non-technical, can see what data went in, what actions were applied, and what factors were identified as relevant. Improving data literacy among teams will support this because it'll help them to understand the decisioning process, as well as the applications and limitations of synthetic data."

Chapter 4

AI and synthetic data use-cases in government

Analysis from Tussell highlights the AI ambitions of government departments – and the data challenges they face. As we have seen, a combination of well-governed real-world and synthetic data can be used to train AI models more effectively but what opportunities do these new capabilities bring?

tussell

Case study 1:

MOD and Project THEIA

MOD spending on IT projects has increased over the past six years from £1bn in 2018 to £1.66bn today (albeit with a drop to £943mn in 2021).

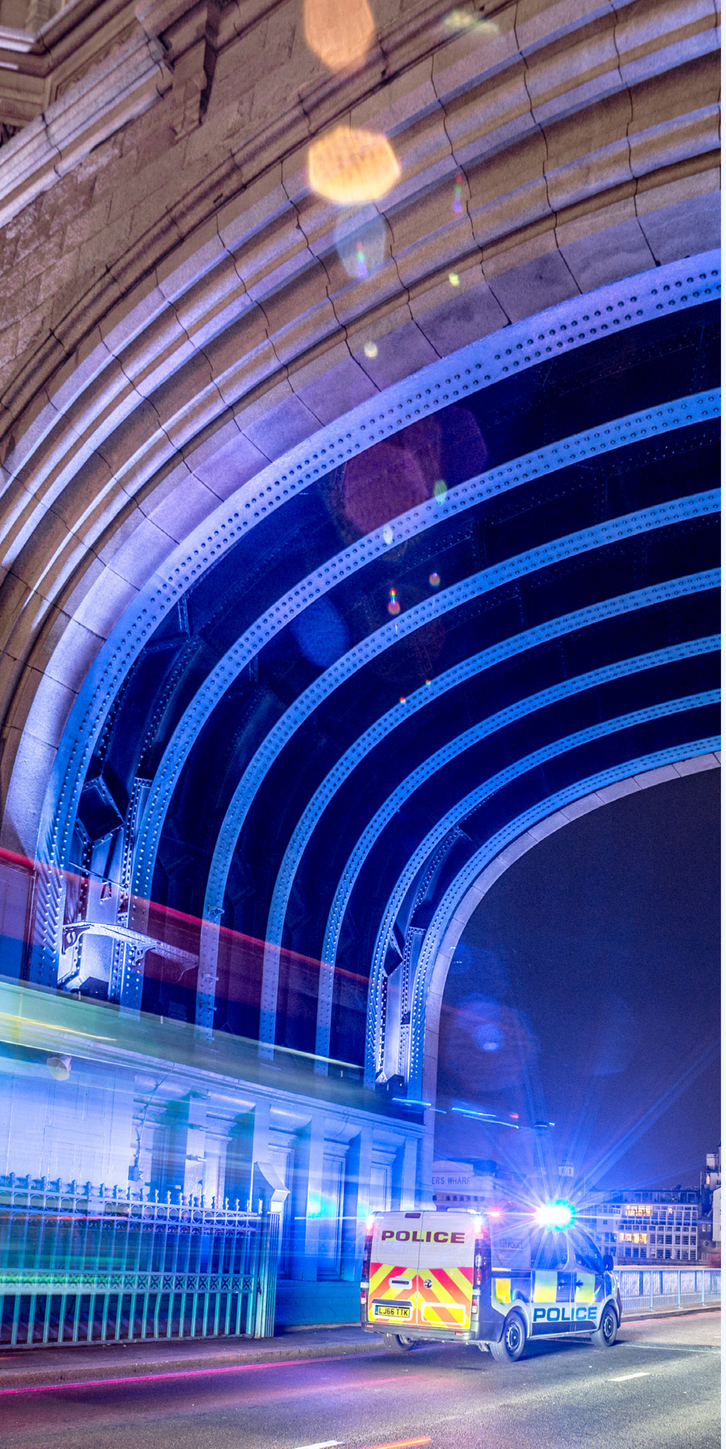
Project THEIA was an ambitious plan to digitally transform the British Army. One of its key goals was to 'integrate all information from command and control, intelligence, sensors, effectors and platforms across domains, partners and allies, and operate with maximum efficiency.'

Challenges:

Highly-classified, incomplete, inconsistently-captured and siloed data makes digital innovation difficult.

Opportunities:

- **Safe testing and validation:** Use synthetic data, which mimics real-world classified data, to train AI, reducing the security risks.
- **Scenario simulation:** Generate simulated battlefields, logistics chains or sensor data to stress-test AI models in a range of different conditions.
- **Cross-agency collaboration:** Sharing synthetic datasets with trusted academic, allied, or private sector partners without breaching confidentiality.
- **Bias mitigation:** Generating balanced datasets to ensure fairness and robustness in autonomous decision-making.



Case study 2:

Policing and Public Safety

IT spending may have been more modest compared to other government departments but it has risen year-on-year – more than doubling from just over £493mn in 2018, to £944mn in 2024. London's Metropolitan Police leads the way on spending at £451mn, nearly 12 times as much as the next highest-spending force, West Midlands Police at just over £38mn.

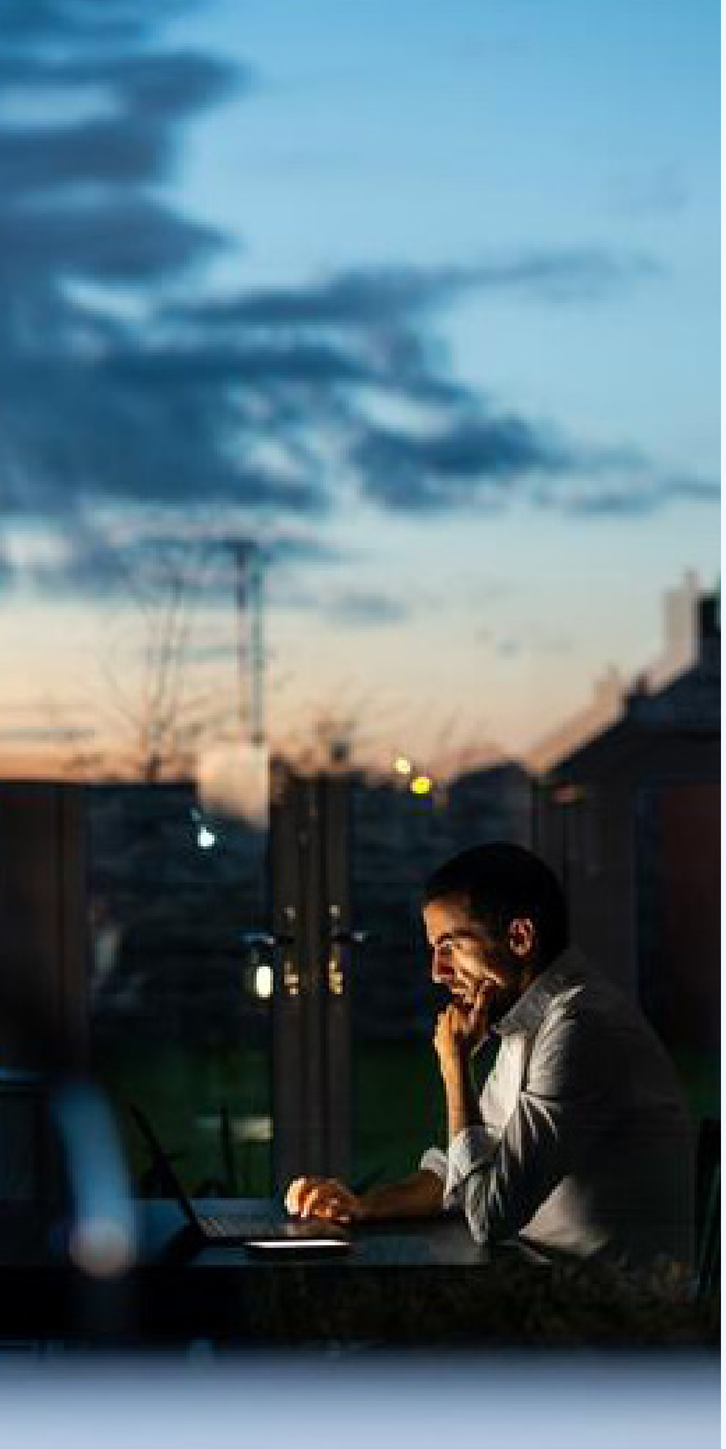
Disparities in digital capabilities could make it difficult for forces to share intelligence easily yet securely, resulting in missed opportunities to reduce crime, especially complex cases.

Challenges:

Data fragmentation and lack of integration between systems limits forces' ability to collaborate effectively. Highly-sensitive data – such as personal or criminal offence data – must also be managed in accordance with the law. Using poor-quality data in AI decision making could also lead to bias and discrimination.

Opportunities:

- **Bias mitigation:** Use synthetic data with strict controls around demographics to remove human biases in decision-making.
- **Build a picture of cases:** Train AI to detect crime hotspots using synthetic crime data.
- **Maintain privacy in predictive models:** Improve demand forecasting and resource planning models in a privacy-safe environment, while ensuring sufficient oversight of models to deliver fair outcomes.
- **Data sharing:** Share regional pilots and studies to promote best practices.



Case study 3:

Central Government Shared Services Strategy

Shared services – far from being a back office function – is a key enabler for AI in government, and its successful implementation depends on synthetic data. Shared Services contract value jumped to £1.3bn in 2024, driven by major awards with IBM, Oracle and SAP.

Challenges:

Incomplete and/or poor-quality data – often the result of outdated software – makes good decision-making and sharing best practices difficult. This can lead to inefficiencies, lack of innovation and a poorer service for citizens.

Opportunities:

- **Better decision-making:** Build a strong foundation of data to develop and scale AI applications, such as workforce planning and intelligent procurement.
- **Drive innovation:** Automate data cleansing, wrangling (transforming raw, messy, and unstructured data into a clean, structured, and usable format) and interpretation to enable innovation.
- **Collaboration:** Promote cross-departmental AI collaboration by improving data quality and governance.

Chapter 5

Next steps

As the government forges ahead with IT and AI projects, SAS recommends the following pillars are needed to help them improve and expand their datasets and bridge the current divide.

Pillar 1

Align IT and AI projects so they focus on value and fair outcomes in both the long and the short-term.

Pillar 2

Standardise data collection and governance with clear policies, communications and technologies both within and across departments to promote collaboration and efficiency.

Pillar 3

Build data literacy skills within the workforce, empowering them to understand the use-cases for data and AI, and critically assess their data's value and integrity.

Pillar 4

Explore how synthetic data could be used to augment real-world data – including unstructured and previously difficult-to-extract data.

Pillar 5

Identify the use-cases and scenarios that could be stress-tested with synthetic data in a risk-free environment.

About this data

SAS commissioned Tussell to undertake an analysis of public sector spending in the UK.

Data in this report is based on open procurement data from official sources aggregated, organised and augmented by Tussell. It includes published invoices and contracts. Some spend data may be limited due to differing transparency rules across agencies and in devolved nations, and there are security and confidentiality restrictions on MOD spend data. Data is accurate as of the date of download (11th August 2025) and there may be spending gaps in 2024 because invoices have not yet been published.

About SAS

SAS is a global leader in data and AI. With SAS software and industry-specific solutions, organisations transform data into trusted decisions. SAS gives you THE POWER TO KNOW®.

Find out about SAS' government AI and analytics solutions.

About Tussell

Tussell is the market's trusted source of insight into UK public procurement.

Our online market intelligence platform helps companies to enhance their public sector business development, and public sector bodies to get better value from their procurement. Tussell was founded in 2015 to fill a gap in the market for the provision of useful and reliable information on government contracts and spending. By consolidating trillions of pounds worth of public procurement data into a single platform, Tussell provides an unparalleled vantage point over the UK public procurement landscape.

Tussell is the media's trusted source of data on UK public procurement. With over 1,500 citations of our data across 450 different outlets since 2018, the media consistently turn to us for reliable insight and analysis into the latest developments surrounding UK public contracting and spending.

Find out more www.tussell.com.

To contact your local SAS office, please visit: **sas.com/offices**