

Quality data and artificial intelligence (AI)

The November 2022 launch of ChatGPT dramatically heightened interest in artificial intelligence (AI), but – like other previous digital innovations – it also raised awareness of fundamental weaknesses in the data used to populate such systems. In the built and managed environment, if we don't have assured good quality data, AI will not deliver the promised future benefits.

Beware the latest buzzwords

Over recent years, government, industry and media discussions about digital transformation have frequently been quick to seize upon the latest technological advances. New buzzwords are sometimes recklessly seized upon as offering the latest 'must-have' or universal panacea.

For example, the Internet of Things (IoT) and 'Big Data' were, according to the technology research firm Gartner, at the peak of inflated expectations in 2013.¹ Artificial intelligence (AI) was an innovation trigger in 2018, just as the hype about 'Digital Twins' peaked.² In the built and managed environment, like BIM and blockchain before them, these various technologies have all been promoted as the latest potential 'silver bullet'

The current surge in AI interest followed the November 2022 launch of the OpenAI chatbot and virtual assistant ChatGPT, which dramatically heightened (already often unrealistic) expectations. In less than three months, it became the fastest-growing consumer application in history, gaining over 100 million global users. It has spawned numerous competing products across just about every sector of industry and commerce, including the built and managed environment.

As of August 2024, the new Labour Government has also been moving to support appropriate adoption and use of AI alongside other digital initiatives. The Government Digital Service, the Central Digital and Data Office and the Incubator for AI have been consolidated in the Department for Science, Innovation and Technology, while a new AI Opportunities Action Plan is also being developed. DSIT will be working closely with the Cabinet Office and the Treasury to maximise the potential of digital, data and technology to enable greater sharing of better information.

Despite media anticipation, the July 2024 King's Speech did not include a bill to regulate AI, but government consultation about controls on those working to develop powerful AI models is expected soon. AI-related requirements will also be written into other forthcoming legislation. When

¹ Schofield, Jack (2013), "[Gartner's 2013 Emerging Technologies hype cycle focuses on humans and machines](#)", *ZDNet* (19 August 2013).

² Hickey, Alex (2018), "[Gartner serves up 2018 Hype Cycle with a heavy side of AI](#)", *CIO Dive* (20 August 2018).

an AI Bill does appear, it is likely to focus on the large language models (LLMs) that underpin AI products such as ChatGPT and Microsoft's Copilot.

Build on good data foundations

We should be using this brief hiatus to build sound data foundations for the future. AI requires good quality information. As the saying goes, if we put rubbish in, we get rubbish out. AI tools need to be populated, ideally, with assured quality data of trusted provenance. Amid growing concern about data reliability and misinformation, this has been a recurring theme in UK government thinking for many years, and also reflects advice to the new Labour administration from leading data scientists.

In 2017, for example, the UK government's *Industrial Strategy* set four Grand Challenges to put the UK at the forefront of future industries. The first Challenge focused on AI and the data economy.³ This imperative was then woven into the 2019 *National Data Strategy*,⁴ founded on four pillars enabling best use of data: data foundations (ensuring the quality of data), data skills, data availability (making it appropriately accessible, mobile and re-usable), and responsible data (ensuring its lawful, secure, fair, ethical, sustainable and accountable use). The first pillar – good quality data – is vital:

“The true value of data can only be fully realised when it is fit for purpose, recorded in standardised formats on modern, future-proof systems and held in a condition that means it is findable, accessible, interoperable and reusable. By improving the quality of the data, we can use it more effectively, and drive better insights and outcomes from its use.”⁵

The *National Data Strategy* emphasises this core consideration regarding AI:

“... poor data foundations can be a real blocker for driving the transformative power of data. For example, **when the source data needed to power AI or machine learning is not fit for purpose, it leads to poor or inaccurate results**, and to delays in realising the benefits of innovation.”⁶ [*emphasis added*]

This view has since also been repeatedly and strongly supported by data scientists. Professional journals report, for example: “Without good data, even the best machine learning algorithms cannot perform well,”⁷ and “Data governance is the foundation of trustworthy AI.”⁸ In November 2023, the Open Data Institute's annual summit stressed the need to focus on the data (“without data, there is no AI”).⁹ In May 2024, ODI co-founder Sir Nigel Shadbolt told a London audience: “...the quality and provenance of data are too often overlooked in business decisions about ... AI tools. Yet data plays a pivotal role in determining these systems' reliability, effectiveness - and value to the bottom line.”¹⁰

³ In the 2017 [Industrial Strategy](#), ‘AI and Data Economy’ was the first of four ‘Grand Challenges’. The others related to Clean Growth, Mobility (of people, goods and services) and Ageing Society (innovating to meet the needs of an ageing society).

⁴ DCMS (2019) [National Data Strategy](#).

⁵ DCMS (2019) [National Data Strategy](#), s.2.6, see also expanded discussion in s.4.

⁶ DCMS (2019) [National Data Strategy](#), s.4.

⁷ Whang, S.E., Roh, Y., Song, H. et al. (2023) “[Data collection and quality challenges in deep learning: a data-centric AI perspective](#).” *The VLDB Journal* 32, 791–813.

⁸ Marijn Janssen, Paul Brous, Elsa Estevez, Luis S. Barbosa, Tomasz Janowski (2020) “[Data governance: Organizing data for trustworthy Artificial Intelligence](#)”, *Government Information Quarterly*, Volume 37, Issue 3.

⁹ [ODI Summit 2023: Data Changes](#), 7 November 2023.

¹⁰ Sir Nigel Shadbolt (2024) [The Achilles' Heel Of AI: How A Major Tech Risk To Your Business Could Be One You Haven't Heard Of - And What You Should Do](#) (695th Lord Mayor's Lectures, 24 May 2024).

And recently, Shadbolt called on ministers to build “a trustworthy data infrastructure for AI because **the feedstock of high-quality AI is high-quality data**” [*emphasis added*].¹¹

Without (quality) data, there is no (quality) AI

But what do we mean by high-quality data? ISO 8000 defines data quality as the “degree to which a set of inherent characteristics of data ... fulfils requirements” – effectively, data’s fitness for its intended purposes. However, there is no universal consensus about what those characteristics or dimensions might be, particularly in the built and managed environment.

For example, the UK Government’s Data Quality Hub says quality can be measured using six dimensions: completeness, uniqueness, consistency, timeliness, validity and accuracy, though different data uses will need different combinations of these dimensions.¹² ISO 19650-4:2022 Clause 7 lists acceptance criteria regarding information exchanges including Common Data Environment (CDE) requirements, conformance, continuity, communication, consistency, and completeness.¹³ And 2021 guidance from the Centre for Protection of National Infrastructure (CPNI – now the National Protective Security Authority) lists 10 quality considerations: relevance, clarity, accessibility, consistency, provenance, timeliness, completeness, accuracy, validity and cost/benefit/risk.¹⁴

From nima’s perspective, excitement about the latest digital buzzwords is to be welcomed – it means there is an active interest in digital transformation – but business leaders in the built and managed environment cannot just throw AI tools at a problem and hope it will be magically resolved. Doing so without paying attention to quality data will also increase the amount of computer processing involved, adding environmental and financial costs.

AI, data and a future connected built and managed environment

Organisations, before they deploy and use AI solutions, will need to ensure that their underlying source data has the quality characteristics necessary to meet its intended purposes. Some organisations are already well advanced in this respect, with strong data governance structures, processes and standards for enterprise-wide information management. But there will be others that have yet to start creating these vital foundations, and still more who are plagued by ‘data debt’ as a result of not addressing their data foundations.

Moreover, as organisations standardise their internal approaches, they will also need to incorporate future national industry standards and rules that will enable and govern inter-organisational sharing of high-quality data. To foster greater collaboration in addressing challenges such as climate change, the Labour Party’s manifesto featured plans for a regulatory innovation office¹⁵ to set rules about cross-sectoral sharing of data. This has been welcomed by experts from the Connected Places

¹¹ “[UK government sets up AI action plan unit](#)”, *Computer Weekly*, 29 July 2024.

¹² Government Data Quality Hub (2020) [Government Data Quality Framework](#)

¹³ [ISO 19650-4:2022](#): Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) — Information management using building information modelling. Part 4: Information exchange

¹⁴ CPNI (2021): [Establishing high level information need and management requirements; Guidance document](#)

¹⁵ [Change](#), Labour Party Manifesto 2024, p.35

Catapult,¹⁶ who also applauded Labour plans for a national data library to support a data-sharing infrastructure.¹⁷

These are important proposals that will have a profound impact right across the built and managed environment. AI may be the latest shiny ‘silver bullet’, but nima believes the successful future deployment and exploitation of AI (and any other digital innovations) will depend on establishing strong foundations: **high-quality data managed through effective data governance structures, processes and standards**. This is key to the outcomes nima strives for, and, building on its current work with government and industry in refreshing the ‘Information Management Mandate’, Nima will be aiming to help define and develop these foundations.

¹⁶ Stanton, Justin (2024) “Why the new government plans must embrace data,” *BIMplus* (22 July 2024), referencing a [Digital Twin Hub blog post](#) by Sarah Hayes, Mark Enzer, Justin Anderson and Geoffrey Stevens on behalf of the Connected Places Catapult.

¹⁷ [Change](#), Labour Party Manifesto 2024, p.35