British Journal of Healthcare Management

Leveraging artificial intelligence to support the application of quality improvement in healthcare

Clarissa Sorlie[1]
Chloe Preston[1]
Marco Aurelio[1]
Edmund Glynn[1]
Lamar Miller[2]
Akkash Purani[1]
Auzewell Chitewe[3]
Sarah McAllister[1]
Jo Moore[1]
Amar Shah[1,4]

Author details can be found at the end of this article

Correspondence to:
Marco Aurelio;
m.aurelio@nhs.net

Abstract

The use of artificial intelligence in healthcare has grown significantly, with generative AI (GenAI) offering the potential to support quality improvement processes in NHS services. The authors conducted a case study at East London NHS Foundation Trust, where quality improvement department staff tested GenAI tools over three phases. Participants used these tools to address quality improvement tasks, with feedback gathered through surveys and a standardised data collection tool. A total of 49 tests were conducted, with GenAI being particularly useful for tasks such as identifying problems, developing strategies and supporting creativity. Tests showed high levels of satisfaction, with 85% of tasks being rated helpful by users, especially in improving efficiency and generating ideas. GenAI tools also reduced time spent on quality improvement tasks, facilitated idea generation and supported data handling. However, these tools appeared to be most effective as a starting tool, requiring further human input for refinement. The authors believe that GenAI can enhance quality improvement processes in healthcare, boosting efficiency and supporting creativity, although further research is needed to refine its integration into broader quality improvement workflows.

Key words: Artificial intelligence; Efficiency; Generative AI; Quality improvement

Submitted: 9 Mar 2025; accepted following double-blind peer review: 14 May 2025

Introduction

In the last decade, the use of artificial intelligence (AI) in healthcare has grown substantially (Jimma, 2023). AI can be defined as the use of machines to simulate how humans learn and think, allowing computers to perform tasks that normally require human intelligence. AI technologies include machine learning, deep learning and natural language processing (von Gerich et al, 2022). Generative AI (GenAI), a subset of AI, can create new text, images or music, with popular tools including Microsoft Copilot and ChatGPT.

One survey identified that 29% of doctors in the UK had used AI in their practice (Hashem et al, 2024), with another reporting that 76% of respondents supported the use of AI in clinical care (Thornton et al, 2024). Uses of GenAI in healthcare include undertaking administrative duties, answering clinical queries, supporting clinical decision making and facilitating diagnostic image analysis (Bajwa et al, 2021). However, GenAI's widespread application in healthcare also raises challenges, including data quality and availability (Chomutare et al, 2022), data privacy and security (Elliott and Soifer, 2022), and concerns regarding environmental sustainability (Strubell et al, 2020).

Although existing literature has focused largely on AI's technical capabilities (Smith et al, 2021), a key challenge lies in translating AI into tools that can enhance quality improvement (Kelly et al, 2019). Quality improvement can be defined as the application of a systematic method to work through complex problems, involving the people closest to the issue in discovering new solutions, testing these in a structured way and using data to learn about the impact (Shah, 2020). Attempts to explore how AI and quality improvement can work together include the use of plan, do, study, act (PDSA) cycles to test AI-generated change ideas (Smith et al, 2021) and statistical process control charts to monitor GenAI-based clinical decision support systems (Feng et al, 2022). However, there has been little attention paid to how GenAI could support the use of a quality improvement method to help solve complex problems within healthcare; a literature search conducted by the authors in April 2025, using both PubMed and Google Scholar, yielded fewer than five relevant studies.

How to cite this article:
Sorlie C, Preston C, Aurelio M et al. Leveraging artificial intelligence to support the application of quality improvement in healthcare. British Journal of Healthcare Management. 2025. https://doi.org/10.12968/bjhc.2025.0027

What, then, is the potential scope for using GenAI to support the application of quality improvement in healthcare? This article describes a case study that explored the application of GenAI to support the use of quality improvement in an NHS trust in England, focusing on two key questions:

1. What are some of the use cases for GenAI in supporting quality improvement in healthcare?
2. How helpful can GenAI be in supporting quality improvement in healthcare, and what are some of the benefits and drawbacks?

Case study context

The case study focused on East London NHS Foundation Trust, which provides primary care, mental health and community health services to a population of 1.8 million people across east London, Bedfordshire and Luton. The trust has been using the Model for Improvement (Langley et al, 2009) as its method for quality improvement since 2012. The Model for Improvement asks three questions to guide improvement efforts:

- What are we trying to accomplish?
- How will we know that change is an improvement?
- What changes can be made that will result in an improvement?

PDSA cycles are then used to test ideas (Langley et al, 2009). All improvement work at the trust also follows a standard five-step sequence of improvement, which includes identifying the quality issue, understanding the problem using tools, developing a strategy and change ideas, testing changes and implementation. A range of infrastructure exists to support this, including a central quality improvement department that is responsible for delivering the organisation's annual quality improvement plan (Institute for Healthcare Improvement, 2024).

This case study took place within East London NHS Foundation Trust's quality improvement department and was undertaken in three phases, modelled on the Institute for Healthcare Improvement's 90-day innovation cycle, which attempts to bring rapid learning to a problem (Martin and Mate, 2018). These phases included a set-up phase in December 2023–February 2024, a testing phase in March–August 2024 and a consolidation phase in August–September 2024.

All 16 members of the quality improvement department tested the GenAI tools. The team included 13 people in roles that support the application of quality improvement methods across the trust and three people in roles that support the operational functioning of the team. Seven team members had clinical backgrounds (four registered mental health nurses, two allied health professionals and one clinical psychologist), with the remaining nine staff being non-clinical. Fourteen team members had been trained in the use of quality improvement methods, either via the Institute for Healthcare Improvement's improvement advisor training or a 6-month continuing professional development-accredited improvement coach course (Frasquilho et al, 2023).

Platforms tested

The primary platform tested was Microsoft Copilot, integrated within Office365, as this was aligned with the organisation's information governance requirements. Team members had the option to test other platforms, such as ChatGPT and Gemini, provided no identifiable information was used. To support safe experimentation, the team developed a set of internal guidelines, outlining principles for testing, ethical considerations and data protection measures. Staff were asked to ensure that their mandated data awareness training was up to date and that they were aware of any other relevant trust guidelines and policies. They were also asked not to input commercially sensitive or person-identifiable information into the AI platforms used.

Team members interacted with the GenAI tools individually as part of their day-to-day work, using them when they felt it might be beneficial to undertake a test. A test in this case refers to a member of the department accessing a GenAI platform and using a prompt to answer a question or provide support around a quality improvement project. For example, a test could include asking a GenAI tool to provide a list of evidenced-based interventions to reduce the number of missed appointments in an outpatient clinic.

Measuring effectiveness

Following the completion of a test, team members self-documented each instance of use on a structured feedback form to develop an understanding of how GenAI could be used to support quality improvement and its perceived helpfulness. This tool was accessible via a Microsoft Teams channel and included four questions:

1. Which GenAI platform did you use?
2. What was the reason you used GenAI?
3. Were there any helpful prompts you used?
4. How helpful or not was the use of GenAI in this instance? (Please write either: very unhelpful, unhelpful, helpful, or very helpful).

The outcome measure was the perceived helpfulness of the GenAI tool in each instance, measured using the 4-point Likert scale described in the fourth question. When completing the feedback form, team members needed to consider whether the output was relevant to the task, accurate, understandable, time-saving and whether it contributed new insights or increased their confidence in completing the task. Scores were collated on a Microsoft Excel document, with responses thematically grouped by three members of the project team. This phase of the project ran from March to August 2024.

After the testing period, an anonymous survey was undertaken to understand the team's experiences of using GenAI to support their quality improvement work. This survey was administered in August 2024 using Microsoft Forms, with questions designed to elicit advantages and disadvantages experienced. Questions included:

- Choosing from the options below, how has the use of GenAI impacted your quality improvement work? (Options: improved time management; improved productivity; reduced tasks of low value; enhanced decision making; no significant impact)
- Please provide a description of your experience of using GenAI in your quality improvement work and how it has helped (free text)
- Do you have any negative experiences of using GenAI to support your quality improvement work? (Options: yes, no, maybe)
- Please describe any negative experiences of using GenAI in your work (free text).

Responses were initially thematically analysed using Copilot and then manually reviewed by two team members for accuracy.

Use of generative AI tools in quality improvement

Overall use

A total of 49 discrete tests were carried out between March 2024 and August 2024, with a median of three tests carried out every 2 weeks. Of the 49 tests, 16 were carried out using ChatGPT, 30 using Copilot, two with both ChatGPT and Copilot and one with Gemini. Tests were grouped into five categories: identifying and understanding the problem; developing a strategy and testing; data and measurement; supporting administrative tasks; and enhancing creativity. Figure 1 shows the number of tests carried out in each category as a Pareto chart (Langley et al, 2009).

Helpfulness ratings of AI tools

A total of 13 tests were conducted to identify and understand the problem being explored in quality improvement work, with 10 (76.9%) of these being deemed helpful or very helpful (Table 1). This included the use of Copilot to summarise articles, conduct literature reviews and develop quality improvement tools such as process maps or cause and effect diagrams. However, team members noted that the visual tools (such as driver diagrams or fishbone diagrams) created often 'did not reflect their thinking' or were 'not in the correct format'. Additionally, when Copilot was prompted to interpret a large Pareto chart, the platform was unable to read the chart axis and the test was rated unhelpful.

Eleven tests were conducted to develop a strategy for change, all of which were rated as helpful or very helpful. Examples included using GenAI tools to develop aim statements, driver diagrams and change ideas. One team member used a GenAI tool to support the planning of a PDSA cycle, which was rated as helpful. However, one team member noted limitations to creating a driver diagram using GenAI, as although Copilot could reorganise

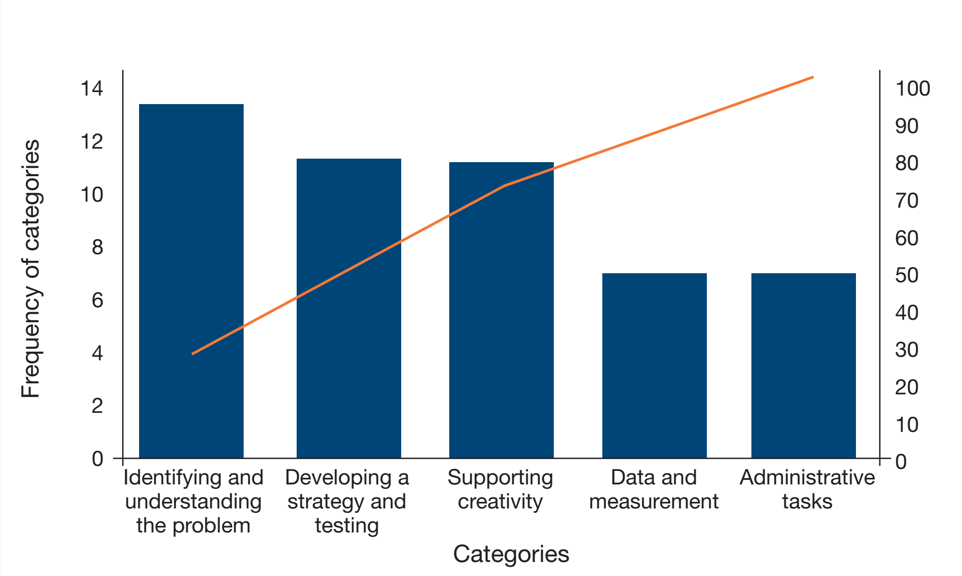

Figure 1. Pareto chart of categories of generative artificial intelligence tests carried out (*n*=49).

Table 1. Summary of tests carried out with generative artificial intelligence tools

Domain	Number of tests	Helpfulness rating, *n* (%)	
		Helpful or very helpful	Unhelpful or very unhelpful
Identifying and understanding the problem	13	10 (76.9%)	3 (23.1)
Developing a strategy and testing	11	11 (100.0)	0 (0.0)
Information and data handling/analysis	7	6 (85.7)	1 (14.3)
Enhancing creativity	11	9 (81.8)	2 (18.2)
Operational efficiency (administrative tasks)	7	5 (71.4)	2 (28.6)

their notes into bullet points that listed primary and secondary drivers, it did not produce the actual diagram.

Of the seven tests to support information and data handling/analysis, six (85.7%) were rated as helpful or very helpful. This included developing measurement plans, creating fictitious data for educational purposes, picking the correct control chart for a set of data and theming qualitative data. Although Copilot was unable to create a pareto chart, ChatGPT was able to do this. The unhelpful test occurred when a team member attempted to calculate the number of days between two events from a set of data, which was unsuccessful because of an inaccurate calculation by the tool.

Of the 11 tests intended to enhance creativity, nine (81.8%) were rated as helpful or very helpful. These activities included support with planning activities and icebreakers and agendas for team building. The two unhelpful tests occurred when trying to create an image to illustrate a concept for a training session and to generate an icebreaker exercise. In the latter instance, the ideas produced were rated unhelpful because they had been previously used, rather than being unsuitable.

Seven tests were carried out to support administrative tasks, with five (71.4%) being deemed helpful or very helpful. Tasks included removing word repetition from reports, summarising long topics and developing a title for a questionnaire. GenAI was also considered helpful for summarising a list of articles from a website. However, these tools were deemed unhelpful for summarising a set of notes, with the team member stating that the notes generated were too short, with 'too much of less relevant information'. GenAI was also unable to summarise a table in a book and, when inserting the picture into ChatGPT, it was unable to recognise it.

Team member feedback

Of the 16 team members, 14 responded to the anonymous survey about their experiences. Twelve team members reported that GenAI had helped them to improve efficiency or productivity in relation to their quality improvement work, while four reported that it improved time management.

In response to the question asking whether they had any negative experiences of using GenAI during the project, four team members said no, eight said yes and two said maybe. Three themes emerged from the qualitative feedback (given in the free-text responses) regarding team members' positive and negative experiences of using GenAI to support quality improvement work: operational efficiency; enhanced creativity; and information retrieval and data handling (Table 2).

Operational efficiency

Using GenAI was described as having a considerable impact on time efficiency and productivity across a range of tasks. Team members noted how GenAI tools had helped to speed up administrative tasks, such as typing up notes and creating documents, with one writing:

> **'I can craft an excellent document… in a fraction of time that it would take me otherwise without AI.'**

Another team member mentioned that these tools helped to sense check work, to see if they had missed anything in their root cause analysis or driver diagram. Others commented on the use of GenAI tools to make written documents more concise and reduce the time taken to edit, which was seen as leading to higher quality outputs.

However, one respondent reported that some tasks took longer to complete with GenAI:

> **'[It takes] more time to think of the right prompt… And the time would have been better used thinking myself than trying to come up with a prompt that didn't help me anyway.'**

Enhanced creativity

GenAI tools were seen as a catalyst for creative thinking, helping the team to generate ideas and reducing thinking time:

> **'It's faster to get ideas, I can then go and research them.'**

Table 2. Summary of themes emerging from the anonymous survey

Theme	Positive experiences	Negative experiences
Operational efficiency	Reducing time to type notes and create documents Ability to sense check work to see if there are gaps Making written documents more concise	Crafting the right prompt can take longer
Enhanced creativity	Quickly generating aim statements, driver diagrams and measurement plans Creating icebreakers and designs for facilitated sessions Generating ideas not previously thought of	None reported
Information retrieval and data handling	Acting as a starting point for thinking Ability to quickly recall and retrieve facts and figures Ability to theme large datasets Retrieving Microsoft Excel formulas	Can hallucinate sources Retrieves sources that are not relevant Cannot visualise data in appropriate graphs

This included crafting aim statements, change ideas, driver diagrams, measurement plans and PDSA plans. One team member summarised their experience, saying:

'AI can create driver diagrams… that need further editing but will often have a good starting structure.'

Information retrieval and data handling

Team members felt that GenAI was useful for summaring information, with one stating:

'It helped me get hold of key information… such as names, references, or quotes.'

Several team members commented on how GenAI tools could speed up thinking and information retrieval, with one noting that this was a faster way of generating ideas, which they could then research themselves, 'rather than trawling the internet'. It was also used to request recommended reading related to the topic of how to write articles about improvement work. That said, some team members highlighted that the GenAI tools also created fictious sources or 'hallucinated' answers, and that this was not immediately clear, leading to potential for inaccurate outputs. It could also provide outputs that were not what team members had requested.

Team members described finding GenAI helpful for aspects of data analysis, such as 'improved time taken for theming large qualitative data sets' and 'supporting me to do complex Excel formulas'. However, others shared examples of inaccurate information being generated, such as the tool selecting an inappropriate statistical process control chart for the data type provided. Two respondents noted that the tool used was unable to visualise data very well, with one commenting that it was unhelpful for data synthesis, analysis and visualisation.

Potential benefits of using artificial intelligence in quality improvement work

Team member notes and feedback regarding the 49 tests conducted during this project can provide insights into the potential uses of GenAI tools to support quality improvement in healthcare. From the work carried out, a driver diagram (or theory of change; Langley et al, 2009) was developed to demonstrate these potential uses (**Figure 2**). Based on both the quantitative data on the types of tests conducted and the qualitative data on team member experiences, the authors believe that GenAI tools could support four areas related to quality improvement:

- Use of quality improvement tools or methods
- Information and data handling/analysis
- Enhancing creativity
- Operational efficiency.

Although quality improvement projects rely on a diverse range of stakeholders to provide different perspectives (Silver et al, 2016), the learning from this project suggests that GenAI could supplement this by generating ideas that were previously unconsidered and supporting the rapid retrieval of evidence to inform effective development of change ideas for testing. The latter element has previously been noted as a barrier to effective quality improvement work (Solomons and Spross, 2011).

Change is a socio-technical endeavour, and GenAI may be able to support the technical aspects of quality improvement work, while also offering a starting point for certain steps in an improvement project. For example, several team members reported using GenAI to support the initial development of artefacts to understand a problem, such as fishbone diagrams. They also used these tools to develop strategies through driver diagrams and build initial measurement plans. The view that AI tools can offer a starting point, rather than end point, for tasks has been suggested previously (Thomson Reuters, 2024). This is important to consider in the context of its use in quality improvement, which typically involves people closest to the problem developing solutions. Additionally, the authors observed that GenAI could help to stimulate creative thinking by retrieving ideas that

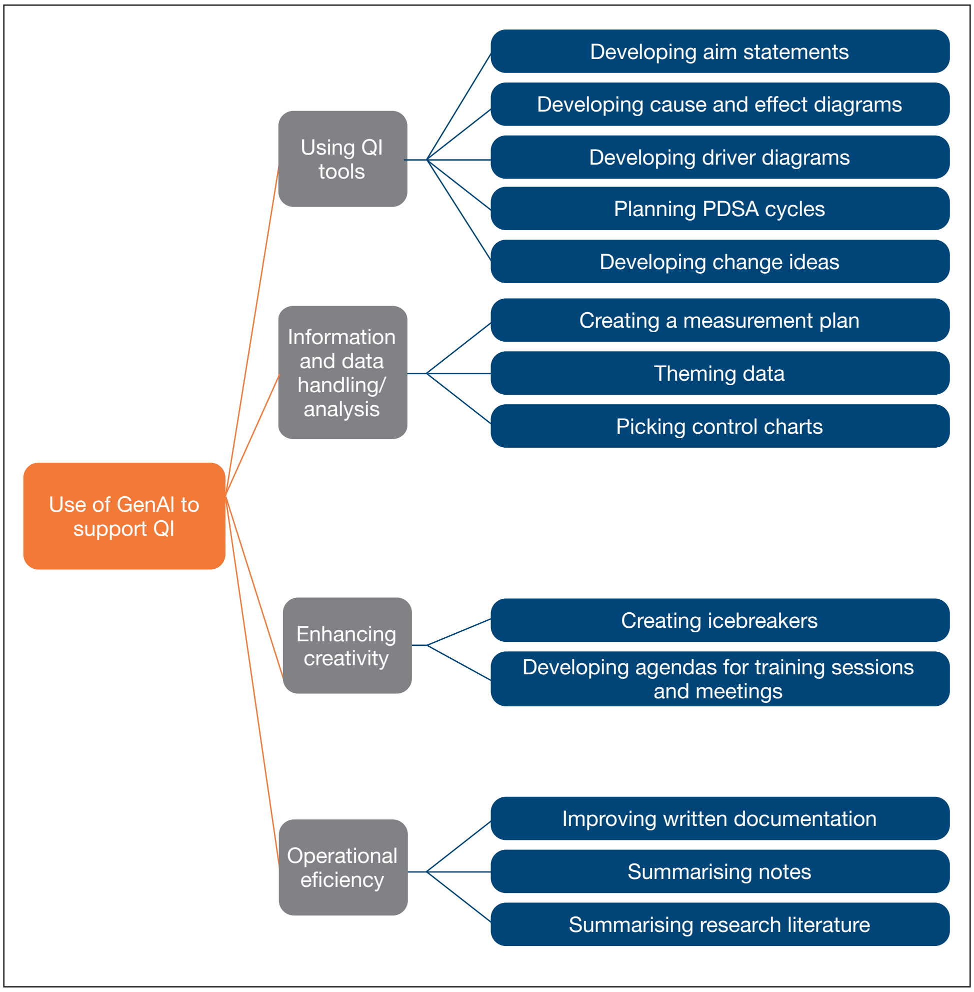

Figure 2. Driver diagram demonstrating uses of generative artificial intelligence (GenAI) tools to support quality improvement (QI) work. PDSA=plan, do, study, act.

others may not have thought of, which has also been noted in the literature, particularly in the field of education (Koivisto and Grassini, 2023; Liu et al, 2023).

Several team members reported that using GenAI helped to speed up their work. This is important, as a commonly cited barrier to effective quality improvement is the time available for frontline staff to undertake it meaningfully (Alexander et al, 2022). Team members in this case study found that GenAI could reduce the time burden of quality improvement-related tasks, particularly in areas such as information retrieval, summarising and analysing data and generating ideas. These tasks are common when understanding a problem or developing a strategy for quality improvement work, and are often time consuming. Similar efficiency gains from the use of AI have been reported in other fields, such as education (Noy and Zhang, 2023) and manufacturing (Okpala et al, 2023).

Taking a structured approach to experimenting with the potential uses of GenAI to support quality improvement could benefit other organisations looking to undertake similar work. Adopting a structured and phased approach could help healthcare leaders to understand when and where GenAI could add value in a controlled way. The importance of small-scale iterative testing for AI has been emphasised in the literature (Sundberg and Holmström, 2024), with quality improvement frameworks such as PDSA cycles potentially offering a practical and safe way to do this.

While this study was limited to a single team in one NHS trust, the use of widely recognised improvement methodologies and generic digital tools such as Office365 suggests that this approach, and the subsequent learning, could be broadly transferable to other, similar organisations. However, generalisability may depend on factors such as local context, digital infrastructure and local quality improvement capability. Future work should thus explore similar approaches across a broader range of settings and evaluate additional outcomes, such as time saved by using GenAI tools and, crucially, the accuracy of outputs.

Challenges and drawbacks of artificial intelligence

Despite several benefits being observed, team members also encountered limitations in using GenAI to support quality improvement in practice. For example, the phenomenon of hallucinations, where the GenAI platform produces incorrect, misleading, or nonsensical information (Nah et al, 2023), was an important challenge noted by users. This emphasises the importance of manually checking sources provided by these tools to ensure that they are accurate.

The authors also observed that GenAI was limited in its ability to create specific visuals, with literal representations of the words inputted during the prompt being provided, rather than a quality improvement tool output. For example, GenAI was unable to create a visual representation of a fishbone diagram, but was able to produce an organised list that could then be translated into a visual. While providing a good starting point, users require some quality improvement knowledge to accurately assess these lists. This emphasises the need for human oversight of outputs, as well as the development of a large language model that is trained to answer domain-specific problems in quality improvement (Noy and Zhang, 2023).

Team members who participated in this project had varying levels of AI literacy, and it was realised early on that supporting people to develop well-constructed prompts was vital. In such situations, developing a continuously updated prompt library can help to support learning (Bansal, 2024). A summary of some prompts developed as part of this work are shown in Table 3.

These challenges were more pronounced in real-time applications, where GenAI outputs required human validation, slowing decision making. Switching between platforms also disrupted workflow continuity, and varying levels of AI literacy and trust in outputs made seamless integration into live meetings difficult. For GenAI to be used effectively in real-time settings, it would need to be embedded into existing workflows, with efforts made to build user confidence.

Further limitations and ethical considerations

This work was undertaken as an explorative study to support the practical application of GenAI to support quality improvement work; as such, several limitations exist. First, the sample size of 49 tests alongside qualitative feedback from 14 people is relatively small and it is thus not possible to draw significant generalisable conclusions about the usefulness of GenAI in supporting quality improvement. The authors would encourage others to conduct large, longitudinal research studies to understand this issue more fully.

Second, the authors were unable to quantify any efficiency or productivity gains made from this work. Although team members noted this as a benefit in the qualitative feedback, future work should incorporate measurement of time saved to further asses gains associated with using GenAI.

Third, all tests were conducted by individuals who were experienced in the use of quality improvement methodology. The outcomes may not reflect the experiences of those with limited quality improvement knowledge, who may require more guidance in interpreting or applying GenAI outputs. The authors would encourage others to test the use of GenAI in supporting quality improvement work with a wide range of different individuals.

Fourth, the tests were conducted over a 6-month period. As both GenAI and user knowledge evolves, results captured early in the study may differ from those produced later in the process. It is possible that a test initially rated as unhelpful might be considered more helpful if repeated after the user had gained more experience or the GenAI model had improved.

Fifth, although the authors' primary motive was not to understand which GenAI platform works best in supporting quality improvement work, the fact that most tests were carried out with either Copilot or ChatGPT limits the ability to understand if different platforms would produce different results. Future work should consider intentionally experimenting with different platforms or considering a design that allows the assessment of how different platforms perform.

Table 3. Summary of prompts developed as part of this work		
Domain	**Topic**	**Prompt**
Using quality improvement tools	Developing a cause and effect diagram	A quality improvement project would like to understand the main causes for missed appointments in older adults memory clinics. Can you create an Ishikawa analysis, based on the clinical literature, around the main causes for missed appointments for older adults memory clinics in the following domains? Please add at least three subdomains, with four causes in each subdomain – people, process, resources, environment, culture
	Developing a driver diagram	Imagine you are an experienced improvement advisor trained with the Institute for Healthcare Improvement. Develop a detailed driver diagram for an equity project to improve access to community podiatry services for homeless people
	Develop a plan-do-study-act (PDSA) cycle	Imagine you are a team of specialist diabetes nurses and health psychologists. Please design a PDSA test for the change idea 'providing clear pre-appointment communication, including potential benefits'. The purpose of this test is to learn what should be included in the letter
Information and data handling	Creating a measurement plan	A memory clinic is starting a quality improvement project with the following aim: 'By 31 December 2024, reduce the rate of missed appointments in memory clinics by 50%'. Their primary drivers are: Primary driver 1: improve patient communication, primary driver 2: enhance accessibility, primary driver 3: streamline appointment scheduling Can I have a measurement plan on a table, which includes the following columns: a) Type of measure (it can be outcome, process or balancing measure), b) name of the measure, c) operational definition of the measure, d) summary of data collection plan
Enhancing creativity	Sourcing ideas for the causes behind a problem	Can I have a 500-word literature review around the causes of missed appointment in in the UK? Please ensure there are in-line citations with number in brackets '()', as a full list of references in the end (Vancouver style)
	Developing an icebreaker	Please suggest some ideas for an icebreaker activity. The activity will be delivered virtually and will need to last 10 minutes

In addition to these practical limitations, there are important ethical considerations related to this work. The use of AI to support decision making, particularly in safety critical settings, is prone to scrutiny because of potential impact on human lives (Wang and Chung, 2022). In a study of public perception of AI by the UK government, healthcare was rated as the industry that should be subject to most oversight when it comes to AI use (Department for Science, Innovation and Technology, 2024). The 'black box' nature of AI can lead to concerns around the trustworthiness of results generated (Cutillo et al, 2020), with potential for bias driven by these tools being trained on broad datasets that reflect societal systemic biases (Gallegos et al, 2024). According to a global study published by the University of Melbourne, many users do not check the outputs of AI (Gillespie et al, 2025), meaning there is a risk that users who are unfamiliar with such limitations will accept the information provided without scrutiny.

Although data privacy was carefully managed in this project by ensuring that no identifiable information was inputted into the GenAI models, this remains a critical concern, particularly

Key points

- The project described in this article suggests that use of generative artificial intelligence (GenAI) tools can support the application of quality improvement methodology in a healthcare organisation.
- GenAI can be a useful starting point to support thinking or complete routine tasks, but should not replace human expertise and judgement.
- Structured iterative testing can be useful in supporting the safe use of GenAI in quality improvement work.

in healthcare settings (Bajwa et al, 2021). This work reinforces the notion that GenAI should support, not replace, human judgement in quality improvement, with human oversight remaining essential. Future research should explore how to build trust in GenAI tools over time, how to mitigate bias in quality improvement-specific use cases and how to establish governance frameworks for the safe and ethical use of these tools in healthcare improvement work.

Conclusions

By following a structured and deliberate testing approach, this work was able to identify ways in which GenAI could be used to support quality improvement work in healthcare. The feedback from team members indicates that GenAI can be a valuable tool, supporting key activities in quality improvement such as understanding a problem, enhancing creativity and developing strategies. Despite some limitations, the ability of these tools to retrieve and summarise information, generate ideas and save time underscores its practical uses in healthcare settings.

That said, several key challenges remain with AI-generated outputs, including hallucinations that produce incorrect information and difficulties in creating accurate quality improvement-specific visuals. The learning from this work highlights the importance of improvement skills and experience in iteratively developing prompts when using GenAI platforms to support this work. Developing a large language model that is specifically trained on quality improvement tools could potentially help to address this challenge and improve the usefulness of GenAI in this context.

In future, research should prioritise longitudinal assessments of GenAI's effectiveness and explore additional ways of measuring its impact beyond those explored in this project, such as time saved, accuracy and trust. Specific areas for future work could include enhancing user confidence, improving integration into existing workflows and developing training programmes to build AI literacy among healthcare professionals. Addressing these areas will be crucial for scaling the safe and effective use of GenAI in healthcare quality improvement.

Author details

[1]East London NHS Foundation Trust, London, UK

[2]London Interdisciplinary School, London, UK

[3]National Institute for Health and Care Excellence, UK

[4]NHS England, UK

Acknowledgement

An AI tool (Microsoft Copilot) was used to thematically analyse survey responses as part of this work. Generated themes were then manually checked by two authors.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

Alexander C, Tschannen D, Argetsinger D, Hakim H, Milner KA. A qualitative study on barriers and facilitators of quality improvement engagement by frontline nurses and leaders. J Nurs Manag. 2022;30 (3):694–701. https://doi.org/10.1111/jonm.13537

Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. Future Healthc J. 2021;8(2):e188–e194. https://doi.org/10.7861/FHJ.2021-0095

Bansal P. Prompt engineering importance and applicability with generative AI. J Comput Commun. 2024;12(10):14–23. https://doi.org/10.4236/jcc.2024.1210002

Chomutare T, Tejedor M, Olsen Svenning T et al. Artificial intelligence implementation in healthcare: a theory-based scoping review of barriers and facilitators. Int J Environ Res Public Health. 2022;19(23):16359. https://doi.org/10.3390/ijerph192316359

Cutillo CM, Sharma KR, Foschini L et al. Machine intelligence in healthcare—perspectives on trustworthiness, explainability, usability, and transparency. NPJ Digit Med. 2020;3(1):47 https://doi.org/10.1038/S41746-020-0254-2

Department for Science, Innovation and Technology. Public attitudes to data and AI tracker survey: wave 4 report. 2024. https://www.gov.uk/government/publications/public-attitudes-to-data-and-ai-tracker-survey-wave-4/public-attitudes-to-data-and-ai-tracker-survey-wave-4-report (accessed 30 October 2025)

Elliott D, Soifer E. AI technologies, privacy, and security. Front Artif Intell. 2022;5:826737. https://doi.org/10.3389/frai.2022.826737

Feng J, Phillips RV, Malenica I et al. Clinical artificial intelligence quality improvement: towards continual monitoring and updating of AI algorithms in healthcare. NPJ Digit Med. 2022;5(1):66. https://doi.org/10.1038/s41746-022-00611-y

Frasquilho F, Brittin K, Chitewe A, Aurelio M, Ballingall N, Lloyd R, Shah A. Building capacity and capability for quality improvement: developing an organisational approach. British Journal of Healthcare Management. 2023;29(6):1–14. https://doi.org/10.12968/bjhc.2023.0003

Gallegos IO, Rossi RA, Barrow J et al. Bias and fairness in large language models: a survey. Computational Linguistics. 2024;50(3):1097–1179. https://doi.org/10.1162/coli_a_00524

Gillespie N, Lockey S, Ward T, Macdade A, Hassed G. Trust, attitudes and use of artificial intelligence: a global study 2025. 2025. https://figshare.unimelb.edu.au/articles/report/Trust_attitudes_and_use_of_artificial_intelligence_A_global_study_2025/28822919?file=54013232 (accessed 30 October 2025)

Hashem Y, Esnaashari S, Morgan D et al. One in four UK doctors are using artificial intelligence: exploring doctors' perspectives on AI after the emergence of large language models. 2024. https://www.turing.ac.uk/news/publications/one-four-uk-doctors-are-using-artificial-intelligence (accessed 30 October 2025)

Institute for Healthcare Improvement. Fostering an improvement culture: learning from East London NHS Foundation Trust's improvement journey over 10 years. 2024. https://www.ihi.org/library/publications/fostering-improvement-culture-learning-east-london-nhs-foundation-trusts (accessed 30 October 2025)

Jimma BL. Artificial intelligence in healthcare: a bibliometric analysis. Telematics Inform Rep. 2023;9:100041. https://doi.org/10.1016/j.teler.2023.100041

Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019;17(1):195. https://doi.org/10.1186/s12916-019-1426-2

Koivisto M, Grassini S. Best humans still outperform artificial intelligence in a creative divergent thinking task. Sci Rep. 2023;13(1):1–10. https://doi.org/10.1038/s41598-023-40858-3

Langley GL, Moen R, Nolan KM, Nolan TW, Norman CL, Provost LP. The improvement guide: a practical approach to enhancing organizational performance (2nd edition). San Francisco (CA): Jossey-Bass Publishers; 2009

Liu ZY, Vobolevich A, Oparin A. The influence of AI ChatGPT on improving teachers' creative thinking. Int J Learn Teach Educ Res. 2023;22(12):124–139. https://doi.org/10.26803/ijlter.22.12.7

Martin L, Mate K. IHI innovation system. 2018. https://www.ihi.org/library/white-papers/ihi-innovation-system (accessed 30 October 2025)

Nah FF-H, Zheng R, Cai J, Siau K, Chen L. Generative AI and ChatGPT: applications, challenges, and AI-human collaboration. J Inf Technol Case Appl Res. 2023;25(3):277–304. https://doi.org/10.1080/15228053.2023.2233814

Noy S, Zhang W. Experimental evidence on the productivity effects of generative artificial intelligence. Science. 2023;381(6654):187–192. https://doi.org/10.1126/science.adh2586

Okpala C, Igbokwe N, Nwankwo CO. Revolutionizing manufacturing: harnessing the power of artificial intelligence for enhanced efficiency and innovation. Int J Eng Res Dev. 2023;19(6):18–25

Shah A. How to move beyond quality improvement projects. BMJ. 2020;370:m2319. https://doi.org/10.1136/bmj.m2319

Silver S, Harel Z, McQuillan R et al. How to begin a quality improvement project. Clin J Am Soc Nephrol. 2016;11(5):893–900. https://doi.org/10.2215/CJN.11491015

Smith M, Sattler A, Hong G, Lin S. From code to bedside: implementing artificial intelligence using quality improvement methods. J Gen Intern Med. 2021;36(4):1061–1066. https://doi.org/10.1007/s11606-020-06394-w

Solomons NM, Spross JA. Evidence-based practice barriers and facilitators from a continuous quality improvement perspective: an integrative review. J Nurs Manag. 2011;19(1):109–120. https://doi.org/10.1111/j.1365-2834.2010.01144.x

Strubell E, Ganesh A, McCallum A. Energy and policy considerations for modern deep learning research. *Proceedings of the AAAI Conference on Artificial Intelligence.* 2020;34(9):13693–13696. https://doi.org/10.1609/aaai.v34i09.7123

Sundberg L, Holmström J. Innovating by prompting: how to facilitate innovation in the age of generative AI. Business Horizons. 2024;67(5):561–570. https://doi.org/10.1016/j.bushor.2024.04.014

Thomson Reuters. Future of professionals report. AI-powered technology & the forces shaping professional work. 2024. https://www.thomsonreuters.com/content/dam/ewp-m/documents/thomsonreuters/en/pdf/reports/future-of-professionals-report-2024.pdf (accessed 30 October 2025)

Thornton N, Binesmael A, Horton T, Hardie T. AI in health care: what do the public and NHS staff think? 2024. https://www.health.org.uk/reports-and-analysis/analysis/ai-in-health-care-what-do-the-public-and-nhs-staff-think (accessed 30 October 2025)

von Gerich H, Moen H, Block LJ et al. Artificial intelligence-based technologies in nursing: a scoping literature review of the evidence. Int J Nurs Stud. 2022;127:104153. https://doi.org/10.1016/j.ijnurstu.2021.104153

Wang Y, Chung SH. Artificial intelligence in safety-critical systems: a systematic review. Ind Manage Data Syst. 2022;122(2):442–470. https://doi.org/10.1108/IMDS-07-2021-0419