

Curiosity, reassurance and convenience: Understanding public motivations for using Generative AI in symptom assessment

Austen El-Osta

a.el-osta@imperial.ac.uk

Imperial College London <https://orcid.org/0000-0002-8772-4938>

Marvin Kopka

Imperial College London <https://orcid.org/0000-0003-3848-1471>

Sami Altalib

Imperial College London <https://orcid.org/0000-0001-7404-8486>

Ana Luisa Neves

Imperial College London <https://orcid.org/0000-0002-7107-7211>

Aldo Faisal

Imperial College London <https://orcid.org/0000-0003-0813-7207>

Henry Potts

University College London <https://orcid.org/0000-0002-6200-8804>

Gabriella Spinelli

Brusel University London <https://orcid.org/0000-0003-1717-7868>

Research Article

Keywords: Generative artificial intelligence, large language models, symptom assessment, self-care, digital health, public trust, eHealth literacy, NHS integration, transparency, health technology adoption

Posted Date: March 8th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9004371/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare potential competing interests as follows: HWWP provides or has provided consultancy for Flo Health Inc., Prova Health, Adora Digital Health and Thrive Therapeutic Software Ltd. He has PhD students employed by or previously employed by and with fees

paid by or previously paid by AstraZeneca, Patients Know Best and BetterPoints Ltd. The other authors declare no conflict of interest.

Abstract

Introduction

Generative artificial intelligence (GenAI) tools such as large language models are increasingly used for health information and symptom appraisal. While diagnostic benchmarking studies are expanding, little population-level evidence explains why individuals use GenAI for symptom assessment or how motivations relate to trust, perceived accuracy and preferences for integration within health systems.

Methods

We analysed data from the 2025 RADIANT VOICES cross-sectional survey of UK adults (n=1449). Participants reported GenAI use, motivations for symptom-related use, trust across use cases and comparative accuracy expectations. Internal consistency supported aggregation of trust ($\alpha=0.90$) and perceived accuracy ($\alpha=0.87$) scales. To compare symptom-assessment users with non-users, we applied 1:1 nearest-neighbour propensity score matching (caliper 0.2) across sociodemographic and digital-literacy covariates, achieving standardised mean differences <0.10 . Weighted regression models with Benjamini–Hochberg correction were used for outcome comparisons.

Results

Overall, 62.8% reported using GenAI; 49.7% of these had used it to assess symptoms. The most common motivations were curiosity (72.3%), reassurance following a diagnosis (60.7%) and convenience (32.1%). Trust was highest for low-risk informational uses (self-care advice 75.3%) and lowest for urgent-care guidance (38.0%). Participants generally rated GenAI as less accurate than general practitioners (55.8% perceived lower accuracy) but more accurate than traditional non-AI symptom checkers (35.8% perceived higher accuracy). After matching, symptom-assessment users demonstrated higher overall trust (mean difference 0.60, $p<0.001$), higher perceived accuracy (mean difference 0.30, $p<0.001$) and greater support for NHS integration (OR=4.06, 95% CI 2.85–5.79). Support for transparency was near-universal (97.0%).

Conclusion

Public use of GenAI for symptom assessment is widespread and primarily motivated by curiosity, reassurance and convenience. Trust is risk-dependent, favouring informational tasks over urgent decisions. Integration strategies should prioritise scope clarity, human oversight and transparency.

What is already known on this topic

- Online symptom checkers show variable diagnostic and triage performance, often with risk-averse outputs.
- Large language models (LLMs) can generate conversational health responses but remain vulnerable to hallucination and reliability failures.
- Public support for AI in healthcare is conditional on transparency, human oversight and regulation.

- Little population-level evidence explains why people use generative AI (GenAI) for symptom assessment or how motivations relate to trust and integration preferences.

What this study adds

- This study provides one of the first population-level examinations of why people use GenAI for symptom assessment.
- It shows that use is not primarily diagnostic but motivationally structured around three uncertainty domains: curiosity (epistemic), reassurance (affective) and convenience (structural).
- Trust follows a clear risk gradient. It is high for low-stakes informational tasks and low for urgent care. Users display markedly higher trust and fourfold greater support for NHS integration.
- Transparency expectations approach consensus (97%), establishing disclosure as a normative baseline.

How this study might affect research, policy or practice

- The findings reposition GenAI from a diagnostic competitor to a behavioural self-care instrument. Innovation strategies should prioritise scoped informational use, embedded escalation routes and explicit transparency.
- Regulatory science should incorporate behavioural surveillance, monitoring how GenAI shapes help-seeking, anxiety and service utilisation rather than focusing solely on accuracy metrics. The study findings reinforce transparency and human oversight as public expectations, not optional safeguards.

Introduction

Generative AI, a type of artificial intelligence that creates new, original content and video by learning patterns from massive datasets has entered the everyday life of the public, especially for health use cases and assessing symptoms. Before the rise of large language models (LLMs), most digital symptom assessment for laypeople relied on rule-based or probabilistic “symptom checker” applications [1-5]. Independent benchmarks consistently found wide variability in both diagnostic suggestion and triage, with a tendency toward risk-averse advice and imperfect safety, highlighting the need for caution in unsupervised use [3, 6, 7].

LLMs offer a qualitatively different interaction by parsing open-ended narratives and providing conversational, context-sensitive guidance aligned with how people describe symptoms. Early studies show they can generate highly rated responses to patient questions and are sometimes perceived as more empathetic than clinicians in vignette settings, though they remain non-expert tools [8]. At the same time, frontier medical models such as Med-PaLM have achieved passing or expert-level performance on medical QA benchmarks, yet do not match real-world clinical practice [9-12]. At the

same time, hallucinations and other reliability failures remain well documented, with medical applications posing domain-specific risks that require guardrails and oversight [13, 14]

Regulatory and normative guidance has moved in response. The World Health Organization's (WHO) 2025 guidance on ethics and governance of large multi-modal models in health emphasises safety, transparency and human oversight for generative systems used in care and public health [15]. In Europe, the AI Act establishes risk-based obligations, requiring transparency from general-purpose AI and stringent controls for high-risk health applications, with staged implementation already underway [16]. These frameworks reflect broad public expectations since in recent surveys many people express conditional support for AI in health if appropriate safeguards and disclosures are in place [17, 18].

Against this dynamic backdrop, the public's routine use of AI for health information has grown but remains uneven. In the United States, a 2024 KFF poll reported that roughly one in six adults used AI chatbots at least monthly for health information, while overall confidence in chatbot accuracy lagged behind awareness [19]. In Great Britain, official statistics show rising familiarity with AI technologies and mixed sentiment toward their benefits and risks [20].

A rich behavioural literature helps explain why people seek health information outside clinical encounters. The Common Sense Model of self-regulation posits that individuals construct illness representations and choose coping actions based on symptom perceptions, prior beliefs and affect; informational acts like searching and consulting tools are core to that process [21, 22]. Digital contexts add new constraints and affordances. The stage-model of technology-assisted self-triage posits that users consult an AI only if they cannot make sense of new symptoms based on their prior experiences and knowledge [23]. Systematic reviews identify sociodemographic and literacy gradients in online health information-seeking, with implications for equity and self-management behaviours [24, 25]. Yet information seeking is not uniformly beneficial. Across settings, greater online search is positively correlated with health anxiety and, in some individuals, with "cyberchondria," marked by reassurance seeking- loops that can amplify distress [21, 26]

These behavioural determinants intersect with system design. Traditional online symptom checkers provided structured pathways that were often perceived as limited in nuance [27]. Qualitative work suggests that people used them for sensemaking, second opinions and action guidance, but were ambivalent about diagnostic reliability and valued them more for self triage-triage than for diagnosis [28]. Newer studies of symptom checker app users report motivations that include curiosity, reassurance, preparation for consultations and convenience, whereas concerns include generic outputs, over-caution, and uncertainty about provenance [29]. LLM-based interfaces may alter that calculus since they can accept free text-text narratives, summarise options in everyday language and sustain follow up, which could increase perceived usefulness for lay reasoning even if factual accuracy remains imperfect [28].

The UK context provides additional relevance. The NHS 111 service embeds algorithmic triage into the urgent care "front door," with evaluations showing limited displacement of call volumes but improved

accessibility for some users [30]. As LLMs become widely accessible outside the NHS, members of the public are newly able to conduct symptom appraisal in general purpose chatbots that are not explicitly -health regulated-. This raises governance questions about appropriateness, triage safety, privacy and the need for clear disclosure when AI is involved in care, reflected in both policy guidance and patient surveys [15-17].

Understanding why people use generative AI for symptom assessment is consequential for clinical safety, service design and regulation. Motivations likely span instrumental aims (deciding whether to self-care or seek care, identifying appropriate services) and epistemic or affective aims (curiosity, reassurance after diagnosis, second opinions without appointment friction). Pragmatic factors such as speed, 24-hour access and anonymity—particularly for sensitive concerns—may also drive use [31]. Conversely, concerns about provenance, inaccurate or alarmist outputs, limited eHealth literacy or prior negative experiences may deter engagement. The impact on help-seeking and system workload is therefore ambiguous: GenAI may support appropriate self-care through clear explanations or increase utilisation by prompting precautionary care. Evidence from algorithmic symptom checkers suggests both effects are plausible, depending on advice quality and users’ cognitive–affective context [6-8].

Pertinently, Public comfort is higher when AI is transparent, human-supervised and accountable, with clear escalation pathways to clinicians [17, 18]. Equity considerations are therefore central. Equity is central, as differences in access, literacy and trust may shape who benefits or is exposed to harm. Despite rapid technical and regulatory developments, empirical evidence linking motivations for GenAI symptom use with trust, perceived accuracy and preferences for integration or disclosure remains limited.

The aim of this study was to investigate *why* members of the public used Generative AI (GenAI) tools to assess their symptoms and to understand their perspective on them. Specifically, the study sought to (i) quantify the prevalence and reasons for using GenAI for symptom assessment among UK adults, (ii) assess their trust levels in various symptom-assessment use cases and their expected accuracy in comparison to other decision support, (iii) compare trust, accuracy expectations, and support for NHS integration between individuals who use GenAI for symptom assessment and those who do not, and (iv) identify implications for responsible design and communication of GenAI-enabled symptom-assessment tools in healthcare.

Methods

Survey

This study used data from the RADIANT VOICES survey (**bmjinnov-2025-001497**), which was developed and refined by the lead author and refined by the wider RADIANT CERSI consortium including input from lay contributors (see PPI and acknowledgement section). The survey asked participants a series of questions about their attitudes and experiences with GenAI for health use cases. The online survey,

which was hosted on Qualtrics XLM, was pre-tested for usability and technical functionality by all authors and was designed as an open survey.

Overall, the survey included 58 questions across 9 pages with no randomization of items or questionnaires (**Supplementary File 1**). Adaptive questioning was used in cases of questions that required specific experience (e.g., only those who used GenAI to assess symptoms were asked why they did so). Completeness checks were included on each page by requiring participants to answer every question or select a “Not sure” option. The survey included a back button so that participants could review their answers.

Reporting of this study followed the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). The survey design, recruitment procedures, consent process, data protection measures, prevention of duplicate entries, response and completion rates, and analytic methods are described in accordance with CHERRIES guidance (**Supplementary File 2**).

Participants

Overall, we aimed to recruit a nationally representative sample of UK adults based on feasibility between 3 October 2025 and 7 November 2025. Participants were required to be at least 18 years old, UK residents and to speak English. The survey was shared on the Prolific Academic Panel and initial contact with potential participants was online. Participation was voluntary, and participants from Prolific were reimbursed.

Data Analysis

To prepare the data, we first conducted IP deduplication by assessing whether participants participated more than once from the same IP within 24 hours. If this was the case, only the first submission was retained. We did not use cookies for user identification nor any log file analysis. Only completed questionnaires were included in the analysis, and no participants were excluded due to their duration.

We analysed multinomial data by assessing the proportions of participants who selected each option, as well as raw numbers and a 95% confidence interval. We then visualised the proportions and confidence interval in a bar plot. To assess trust and performance expectations (which were on Likert scales), we used top box scoring to assess positive and negative responses and report the percentages with 95% confidence intervals and raw numbers. To visualise these data, we used divergent bar plots with separate neutral responses to better show differences between high and low trust and performance expectations.

Next, we were interested in comparing respondents who used GenAI for symptom assessment with those who did not with respect to trust in GenAI, expected accuracy, support for integrating GenAI into NHS symptom checkers, and support for transparency about AI involvement in care. To do this, we

calculated internal consistency of trust and expected accuracy items to determine whether items are similar enough to build mean scores. Because internal consistency was acceptable ($\alpha=0.9$ for questions on trust in GenAI for symptom assessment, and $\alpha=0.87$ for expected accuracy), we aggregated them into two scales for further analysis. Subsequently, we used propensity score matching to compare respondents on these outcomes as this method reduces confounding by creating treatment and control groups that are similar on baseline covariates and approximate balance typically resulting from randomization. This approach makes the treatment effects less sensitive to selection bias and confounding [32]. We calculated covariate-balancing propensity scores for all respondents in R using the MatchIt package and a multivariate logistic regression model that included participants' age, disability status, educational levels, eHealth literacy (on standardised scales), employment type, ethnicity, location, chronic condition, occupational field, internet access, residence, and sex. This was done using 1:1 nearest-neighbour matching without replacement and a caliper width of 0.2 standard deviations of the logit of the propensity score. Before matching, 687 non-symptom-assessment users and 554 symptom-assessment users were included in the dataset. After matching, 503 and 503 participants were included respectively. We assessed standardized differences before and after matching and considered differences of $<.10$ acceptable for indicating balance [33]. All variables reached this cutoff after matching, see **Supplementary Table 1**. The outcomes were then assessed using a weighted logistic regression for binary outcomes, and a weighted linear regression for metric outcomes with cluster-robust standard errors. To control the false discovery rate because of multiple testing, we employed the Benjamini-Hochberg procedure to adjust p-values and only report adjusted p-values.

Ethics, consent and data protection

The study received a favourable opinion and was approved by the Imperial College Research Ethics Committee (ICREC #7748683) and complied with UK GDPR. Participants viewed an information sheet and provided electronic informed consent before answering any items; consent pages described purpose, voluntary nature, data handling and contact details. No direct personal identifiers were collected. IP addresses and coarse location, captured by the platform, were accessible only to the core research team for fraud and duplication checks and were not used to reidentify participants. Data were stored on secure institutional servers and analysed in deidentified form.

Patient and Public Involvement and Engagement (PPIE)

Patients were not involved as research participants in the design phase. The survey instrument was developed by the research team and subsequently reviewed by Adi Behura, Ola Fagbohun, Reena Malhan, Sandra Jatacodi and Jonny Spicer, who provided patient and public involvement and engagement (PPIE) input to support refinement of wording, clarity and content. The revised instrument was then further iterated in consultation with RADIANT CERSI partners to finalise the survey prior to launch.

Results

Participant Characteristics

Overall, 1518 people visited the first page of the survey, and all of them agreed to participate, resulting in a participation rate of 100%. Out of those, 1,468 participants finished the survey, resulting in a completion rate of 97%, see **Supplementary Figure 1**. During ID deduplication, 19 were removed, resulting in a final sample of 1449 participants. Their characteristics are shown in **Table 1**.

Table 1: Participant characteristics.

Characteristic	Value
Age (years), M	44.14 (13.35)
eHealth literacy, M	2.98 (0.63)
Sex, n(%)	
Female	721 (49.8%)
Male	717 (49.5%)
Non-binary	7 (<0.1%)
Prefer not to say	4 (<0.1%)
Ethnicity, n(%)	
White	1,186 (81.8%)
Asian	83 (5.7%)
Black	77 (5.3%)
Mixed	62 (4.3%)
Other	29 (2.0%)
Prefer not to say	12 (0.8%)
Residence, n(%)	
Urban	558 (38.5%)
Suburban	609 (42.0%)
Rural	272 (18.8%)
Not sure	10 (0.7%)
Education, n(%)	
No qualifications	18 (1.2%)
GCSEs/A-levels	462 (31.9%)
University degree	645 (44.5%)
Postgraduate	302 (20.8%)
Other	22 (1.5%)
Employment, n(%)	
Employed full-time	747 (51.6%)
Employed part-time	207 (14.3%)

Self-employed	147 (10.1%)
Student	38 (2.6%)
Unemployed / looking for work	77 (5.3%)
Unable to work due to illness or disability	40 (2.8%)
Retired	150 (10.4%)
Other	27 (1.9%)
Healthcare worker, n(%)	
No	1,344 (92.8%)
Yes	101 (7.0%)
Missing	4 (0.3%)
Chronic condition, n(%)	
No	918 (63.4%)
Yes	489 (33.7%)
Prefer not to say	42 (2.9%)
Lives in London, n(%)	
No	1,251 (86.3%)
Yes	198 (13.7%)
Frequency of using symptom-assessment applications, n(%)	
Often (1-3 times per week)	24 (1.7%)
Sometime (1-3 times per month)	235 (16.2%)
Rarely (less than once a month)	626 (43.2%)
Never	562 (38.8%)
Missing	1
Frequency of using GenAI, n(%)	
Very often (daily or almost daily)	380 (26.2%)
Often (1-3 times a week)	435 (30.0%)
Sometime (1-3 times a month)	287 (19.8%)
Sometime (1-3 times a month) Rarely (less than once a month)	209 (14.4%)
Missing	3 (<0.1%)

Broader attitudes towards governance, accountability and regulation of generative AI

The main survey findings are shown in **Supplementary Table 2**. Broader attitudes towards governance, accountability and regulation of generative AI in healthcare are presented in **Supplementary Table 3**.

GenAI Use for Symptom Assessment

Among all participants, 61.2% (886/1448, 95% CI [58.6–63.7%]) had used traditional SAAs, and 62.8% (909/1447, 95% CI [60.3–65.3%]) used GenAI. Among GenAI users, 49.7% (452/909, 95% CI [46.4–53.0%]) used it to assess symptoms. About half of SAA users used GenAI to assess symptoms as well (51.8% (459/886, 95% CI [48.5–55.1%])), whereas most respondents who used GenAI to assess symptoms used an SAA before (367/452, 95% CI [77.3–84.7%])).

The most common reasons for using GenAI are shown in **Figure 1**. Most used GenAI to assess symptoms was out of curiosity (72.30% [425/588, 95% CI 68.50–75.90%]), followed by reassuring themselves of a diagnosis they received (60.70% [357/588, 95% CI 56.60–64.70%]), making a self-triage decision (35.70% [210/588, 95% CI 31.80–39.70%]), getting a second opinion (35.70% [210/588, 95% CI 31.80–39.70%]), because it was quicker than booking a GP appointment (32.10% [189/588, 95% CI 28.40–36.10%]), and to avoid bothering a healthcare professional (21.90% [129/588, 95% CI 18.70–25.50%]). Only a few respondents reported using it because they could maintain anonymity (15.00% [88/588, 95% CI 12.20–18.10%]), because they could not access another source of care (9.50% [56/588, 95% CI 7.30–12.20%]), or for other reasons (3.40% [20/588, 95% CI 2.10–5.20%]), see **Figure 2**. Other reasons that were mentioned included gaining an understanding of symptoms and diseases, preparing for appointments, and getting information for caring for others.

Figure 1: Reasons for using GenAI to assess symptoms. The bars indicate ean responses, and error bars indicate 95% confidence intervals

Trust in Use Cases

Most respondents trusted GenAI to get self-care advice for minor issues (75.3% (1072/1423, 95% CI [73.0–77.6%])), followed by promoting health literacy (74.5% (1056/1418, 95% CI [72.1–76.7%])), identifying possible causes (58.2% (827/1421, 95% CI [55.6–60.8%])), following up on a received diagnosis and monitoring a condition (43.5% (608/1399, 95% CI [40.8–46.1%])), and determining symptom seriousness (41.0% (582/1418, 95% CI [38.5–43.7%])). Only in one use case, advising on urgent care, more respondents indicated distrust rather than trust in GenAI (38.0% (538/1417, 95% CI [35.4–40.6%])), see **Figure 2**.

Figure 2: Trust in Generative AI across symptom-assessment use cases.

Performance Expectations

About a third of participants expected GenAI to be more accurate than traditional non-AI-based symptom checkers (35.8% (451/1261, 95% CI [33.1–38.5%])). This was the only source of information where more respondents expected higher accuracy from GenAI than lower accuracy. Expectations of lower accuracy for GenAI were highest when compared with general practitioners (55.8% (770/1381, 95% CI [53.1–58.4%])), followed by pharmacists (45.9% (622/1356, 95% CI [43.2–48.6%])), NHS 111 (37.8% (504/1332, 95% CI [35.2–40.5%])), and NHS.UK (38.1% (511/1340, 95% CI [35.5–40.8%])), see **Figure 3**.

Figure 3: Expected accuracy of GenAI for symptom-assessment compared to other sources.

Differences Between Users and Non-Users for Symptom Assessment

Respondents who use GenAI to assess their symptoms were OR=4.06 (95% CI [2.85–5.79]) times more likely to support integrating GenAI in the NHS' symptom checkers ($z=7.77$, $p<.001$), but they did not show any difference in thinking that users should be told if AI is involved in their healthcare (OR=0.99, 95% CI [0.28–3.47], $z=-0.02$, $p=.987$).

Respondents placed higher trust in GenAI (0.60 points on a 5-point Likert scale, $d=0.76$, 95% CI [0.88–0.63], $t=12.47$, $p<.001$) and also judged the accuracy of GenAI as higher than those who do not use GenAI to assess their symptoms (0.30 points on a 5-point Likert scale, $d=0.31$, 95% CI [0.44–0.19], $t=4.92$, $p<.001$), see **Figure 4**.

Figure 4: Differences between users who used GenAI for symptom assessment and those who did not. Groups were selected using propensity score matching to adjust for several confounders.

Discussion

Summary of principal findings

This study provides one of the first population-level insights into why members of the public use generative artificial intelligence (GenAI) tools for symptom assessment, showing that curiosity, reassurance and convenience are the main motivations.

The findings suggest that GenAI fills a gap within the healthcare system that allows patients to improve their own self-care. The conversational, context-sensitive style of LLMs appears to make information-seeking easier in situations where participants have an initial idea (like a received diagnosis) or face high informational needs (like asking many questions they do not want to bother a professional with). This aligns with the *Common-Sense Model* of illness representation, where individuals act to make sense of

symptoms through iterative information gathering and at an unaddressed need for personalised high-quality information [21].

Next to curiosity, reassurance was the second most common reason for GenAI use, which aligns with prior findings on symptom-assessment applications that were often used to validate or challenge previous diagnoses [28]. This behaviour may not only reflect a cognitive process, but also a need for emotional regulation since reassurance-seeking was found to reduce anxiety, although it may also perpetuate “cyberchondria” when users repeatedly check symptoms online [26]. The high proportion of respondents seeking reassurance indicates that GenAI may be used for self-regulation and for reducing health anxiety in the short term. Future qualitative research could examine whether conversational GenAI tools mitigate or amplify these tendencies.

Convenience-related use cases such as speed, accessibility and avoidance of “bothering” clinicians, highlight the structural pressures on primary care [34]. Given the limited appointment availability and rising digital literacy, GenAI tools offer perceived immediacy and autonomy. These findings echo prior research showing that online symptom checkers and NHS 111 services serve as substitutes when users face logistical barriers to care [30]. The finding that only a minority used GenAI because they lacked access to care suggests complementarity rather than substitution since people most likely use GenAI to *augment* decision-making, as opposed to replace formal healthcare.

These motivational patterns are broadly consistent with the stage-model of technology-assisted self-triage, which proposes that individuals consult digital tools when they are unable to make sense of symptoms using prior knowledge and experience [23]. However, our findings suggest that GenAI engagement reflects a broader conception of uncertainty than the original stage formulation implies. Curiosity appears to address epistemic uncertainty (a desire to know more), reassurance targets affective uncertainty (reducing anxiety or consolidating confidence) whereas convenience resolves practical uncertainty around timing, access and next steps. In this sense, GenAI is not used solely at the point of failed sense-making, but also after provisional understanding has already been achieved.

Notably, reassurance following a diagnosis and curiosity-driven exploration occurred even where participants had already received professional input, indicating that conversational GenAI may support additional self-care functions beyond triage. These include confidence reinforcement, iterative learning and enhanced autonomy within the illness representation process described by the Common-Sense Model [21]. Rather than displacing the stage-model, our findings extend it and highlight that while the model captures an important pathway of uncertainty reduction, it does not fully account for the expanded self-care roles enabled by conversational, explainable LLM interfaces that allow ongoing dialogue, clarification and personalised elaboration [28].

Trust, accuracy and expectations

Trust was highest when GenAI was used for low-risk, informational tasks such as self-care advice or health literacy promotion, and lowest for assisting in urgent-care decisions. This gradient matches established risk-trust trade-offs observed in algorithmic decision-making [35]. Users' trust appears to be influenced by the perceived stakes, a finding that has been observed across several domains [36]. The positive correlation between having used GenAI for symptom assessment and greater trust in its accuracy may be the result of bidirectional causality suggesting that use may lead users to perceive GenAI as trustworthy, but also that individuals with a higher tendency to trust may be more likely to use GenAI for symptom assessment. Longitudinal designs will be required to clarify these pathways.

The finding that participants rated GenAI as less accurate than clinicians but more accurate than traditional SAAs indicates a new middle ground in public perception. This gradient mirrors objective benchmarking results as other studies showed that LLMs such as ChatGPT and MedPaLM achieve near-clinician-level accuracy on structured diagnostic tasks yet remain prone to hallucination and overconfidence [14]. On the other hand, the average GenAI tool is better than the average SAA, although there exists variability among both [23]. Public perceptions thus appear roughly calibrated to the actual technical performance.

Implications for self-care and system integration

From a behavioural perspective, GenAI's capacity to engage curiosity and reassurance may extend the domain of self-care. By providing immediate, personalised explanations, LLMs could increase individuals' confidence to self-manage minor conditions, a critical determinant of health system sustainability [37]. Yet the same mechanism could increase healthcare utilisation if outputs suggest rare diagnoses or overload users with information, echoing prior concerns about probabilistic symptom checkers [6, 7] and broader Internet use before that [38]. However, whether GenAI ultimately lightens or adds to system burden depends on its interaction design, transparency and user literacy.

Design and governance considerations

Our results carry implications for responsible GenAI deployment in healthcare. First, interfaces intended for public use should incorporate explicit disclaimers distinguishing informational and probabilistic content from rule-based diagnostic content. Second, user interfaces could provide *structured escalation pathways* (e.g. linking to NHS 111 or verified triage tools when serious symptoms are mentioned) to ensure safety. Third, monitoring frameworks must move beyond accuracy metrics to include *behavioural outcomes*, such as help-seeking patterns and emotional responses. Integrating post-market surveillance, similar to pharmacovigilance, could enable detection of emergent harms [16]. Clinicians will also require clarity on how to interpret patients' interactions with GenAI. The boundary between self-care and professional care is blurring since patients who arrive at consultations increasingly arrive with AI-generated hypotheses, which influences differential diagnoses and shared decision-making [14]. Professional education should therefore include competencies in *AI literacy* and conversational risk

management as these skills are needed to contextualise, not dismiss, AI-informed patients. It is conceivable that, over time, domain-specific LLMs trained using retrieval-augmented generation (RAG) with only approved clinical sources will differentiate themselves in the market by signalling a higher level of trustworthiness. However, these additional measures cannot prevent unauthorised extrapolation beyond the training data, nor can they guarantee unbiased outputs. Even health-specific LLMs will therefore remain subject to ongoing monitoring, human oversight and robust governance, making user education about the capabilities and limitations of such tools pivotal.

Limitations and future directions

The principal limitation of this study is concerned with its cross-sectional design which prevents causal inference between trust, use and perceived accuracy. We acknowledge that self-selection bias is also possible because participants were digitally connected and motivated to engage in online surveys. Nevertheless, the large, demographically diverse UK sample provides a baseline for public attitudes. Future research should employ longitudinal and experimental methods to assess how repeated exposure to GenAI affects behavioural outcomes such as healthcare utilisation, anxiety trajectories, and compliance with self-care advice. Ongoing qualitative studies by the co-authors could further examine reasons for curiosity and reassurance in AI-generated health advice. We also note that those responsible for most healthcare utilisation are generally much older and this group may have differing attitudes towards AI.

Conclusion

Generative AI has entered the vernacular of everyday symptom appraisal. This study shows that people use GenAI not merely as a diagnostic shortcut but as a tool for sensemaking, reassurance and convenience. Trust is highly use case-conditional, depends on the risk involved, and is influenced by previous use. For policymakers and designers, these insights point to a dual agenda: maximise informational and self-care benefits while minimising psychological and safety risks. As generative systems evolve, their role in public health will hinge less on technical prowess than on aligning design with human decision-making, sensemaking and trust.

Declarations

Conflict of interest

HWWP provides or has provided consultancy for Flo Health Inc., Prova Health, Adora Digital Health and Thrive Therapeutic Software Ltd. He has PhD students employed by or previously employed by and with fees paid by or previously paid by AstraZeneca, Patients Know Best and BetterPoints Ltd.

The other authors declare no conflict of interest.

Funding

This research received no funding. This work was supported by the RADIANT Centre for Excellence in Regulatory Science and Innovation (CERSI), funded by the Medical Research Council (MRC), Innovate UK, the Medicines and Healthcare products Regulatory Agency (MHRA), and the Office for Life Sciences (OLS). Austen El-Osta and Ana Luis Neves are grateful for support from the National Institute for Health and Care Research (NIHR) Applied Research Collaboration NorthWest London. Marvin Kopka is grateful for support from the German Federal Ministry of Education and Research (BMBF) and the State of Berlin as part of the Excellence Strategy of the German federal and state governments through the Berlin University Alliance. The views expressed in this article are those of the authors and not necessarily those of the NIHR, the Department of Health and Social Care, or the Berlin University Alliance.

Author Contributors

All authors provided substantial contributions to the conception (AEO), design (AEO, SA, ALN, HP GS,), acquisition (AEO, AA, SA, MA, AB) and interpretation (AEO, MK, SA, AF) of study data and approved the final version of the paper. AEO took the lead in planning the study with support from co-authors. MK carried out the data analysis with support from AEO. AEO is the guarantor.

Consent for publication

Not applicable.

Twitter: @austenelosta @kopka_sci @HWWPotts @ImperialSCARU

References

1. Pairon, A., H. Philips, and V. Verhoeven, *A scoping review on the use and usefulness of online symptom checkers and triage systems: How to proceed?* Frontiers in Medicine, 2023. **Volume 9 - 2022**.
2. Kopka, M., et al., *Characteristics of Users and Nonusers of Symptom Checkers in Germany: Cross-Sectional Survey Study*. J Med Internet Res, 2023. **25**: p. e46231.
3. El-Osta, A., et al., *Factors Influencing the Use of Online Symptom Checkers in the United Kingdom: Cross-Sectional Study*. JMIR Formative Research, 2025. **9**.
4. Wetzel, A.-J., et al., *Unveiling Usage Patterns and Explaining Usage of Symptom Checker Apps: Explorative Longitudinal Mixed Methods Study*. J Med Internet Res, 2024. **26**: p. e55161.
5. Kopka, M., L. He, and M.A. Feufel, *ChatGPT Is Still Not Good Enough at Giving Care-Seeking Advice, or Is It?* medRxiv, 2025: p. 2025.05.13.25327519.
6. Semigran, H.L., et al., *Evaluation of symptom checkers for self diagnosis and triage: audit study*. BMJ, 2015. **351**: p. h3480.

7. Wallace, B. and C.P. Martin, *Embodying the Glitch: Perspectives on Generative AI in Dance Practice*. arXiv preprint arXiv:2210.09291, 2022.
8. Ayers, J.W., et al., *Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum*. JAMA Intern Med, 2023. **183**(6): p. 589-596.
9. Singhal, K., et al., *Large language models encode clinical knowledge*. Nature, 2023. **620**(7972): p. 172-180.
10. Singhal, K., et al., *Toward expert-level medical question answering with large language models*. Nat Med, 2025. **31**(3): p. 943-950.
11. Kopka, M., N. von Kalckreuth, and M.A. Feufel, *Accuracy of online symptom assessment applications, large language models, and laypeople for self-triage decisions*. npj Digital Medicine, 2025. **8**(1): p. 178.
12. Kopka, M., et al., *The RepVig framework for designing use-case specific representative vignettes and evaluating triage accuracy of laypeople and symptom assessment applications*. Scientific Reports, 2024. **14**(1): p. 30614.
13. Huang, L., et al., *A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions*. ACM Trans. Inf. Syst., 2025. **43**(2): p. Article 42.
14. Bélisle-Pipon, J.-C., *Why we need to be careful with LLMs in medicine*. Frontiers in Medicine, 2024. **Volume 11 - 2024**.
15. Organization, W.H., *Ethics and governance of artificial intelligence for health: large multi-modal models. WHO guidance*. 2024: World Health Organization.
16. Pehlivan, C.N., *Report: The EU Artificial Intelligence (AI) Act: An Introduction*. Global Privacy Law Review, 2024. **5**(1).
17. Platt, J., et al., *Public Attitudes Toward Notification of Use of Artificial Intelligence in Health Care*. JAMA Network Open, 2024. **7**(12): p. e2450102-e2450102.
18. Pew Research Center, *What the data says about Americans' views of artificial intelligence*. 2023.
19. KFF, *Poll: Most Who Use Artificial Intelligence Doubt AI Chatbots Provide Accurate Health Information*. 2024.
20. ONS, *Management practices and the adoption of technology and artificial intelligence in UK firms: 2023*. 2025.
21. Leventhal, H., L.A. Phillips, and E. Burns, *The Common-Sense Model of Self-Regulation (CSM): a dynamic framework for understanding illness self-management*. J Behav Med, 2016. **39**(6): p. 935-946.
22. Hagger, M.S., et al., *Perceived behavioral control moderating effects in the theory of planned behavior: A meta-analysis*. Health Psychology, 2022. **41**(2): p. 155.
23. Kopka, M., et al., *Technology-supported self-triage decision making*. npj Health Systems, 2025. **2**(1): p. 3.

24. Zhao, Z., et al., *Challenges and opportunities of AI-enabled monitoring, diagnosis & prognosis: A review*. Chinese Journal of Mechanical Engineering, 2021. **34**(1): p. 56.
25. Zhang, C. and Y. Lu, *Study on artificial intelligence: The state of the art and future prospects*. Journal of Industrial Information Integration, 2021. **23**: p. 100224.
26. McMullan, R.D., et al., *The relationships between health anxiety, online health information seeking, and cyberchondria: Systematic review and meta-analysis*. Journal of Affective Disorders, 2019. **245**: p. 270-278.
27. Ceney, A., et al., *Accuracy of online symptom checkers and the potential impact on service utilisation*. PLoS One, 2021. **16**(7): p. e0254088.
28. Aboueid, S., et al., *Young Adults' Perspectives on the Use of Symptom Checkers for Self-Triage and Self-Diagnosis: Qualitative Study*. JMIR Public Health Surveill, 2021. **7**(1): p. e22637.
29. Müller, B., et al., *Headache impact and socioeconomic status: findings from a study of the German Migraine and Headache Society (DMKG)*. J Headache Pain, 2023. **24**(1): p. 37.
30. Turner, J., et al., *A qualitative interview study of NHS 111 Online user experience, in Impact of NHS 111 Online on the NHS 111 telephone service and urgent care system: a mixed-methods study*. 2021, NIHR Journals Library.
31. Dellavalle, N.S., et al., *What patients want from healthcare chatbots: insights from a mixed-methods study*. Journal of the American Medical Informatics Association, 2025. **32**(11): p. 1735-1745.
32. Caliendo, M. and S. Kopeinig, *SOME PRACTICAL GUIDANCE FOR THE IMPLEMENTATION OF PROPENSITY SCORE MATCHING*. Journal of Economic Surveys, 2008. **22**(1): p. 31-72.
33. Chen, Z., et al., *Analysis of the association between testosterone and cardiovascular disease potential risk factor apolipoprotein B in adult males without cancer: national health and nutrition examination survey 2011-2016*. Front Endocrinol (Lausanne), 2024. **15**: p. 1304344.
34. Kingsfund, *Understanding pressures in general practice*. 2016.
35. Hatherley, J.J., *Limits of trust in medical AI*. J Med Ethics, 2020. **46**(7): p. 478-481.
36. Siegrist, M., *Trust and Risk Perception: A Critical Review of the Literature*. Risk Analysis, 2021. **41**(3): p. 480-490.
37. El-Osta, A., et al., *Tools for measuring individual self-care capability: a scoping review*. BMC Public Health, 2023. **23**(1): p. 1312.
38. Potts, H.W. and J.C. Wyatt, *Survey of doctors' experience of patients using the Internet*. J Med Internet Res, 2002. **4**(1): p. e5.

Figures

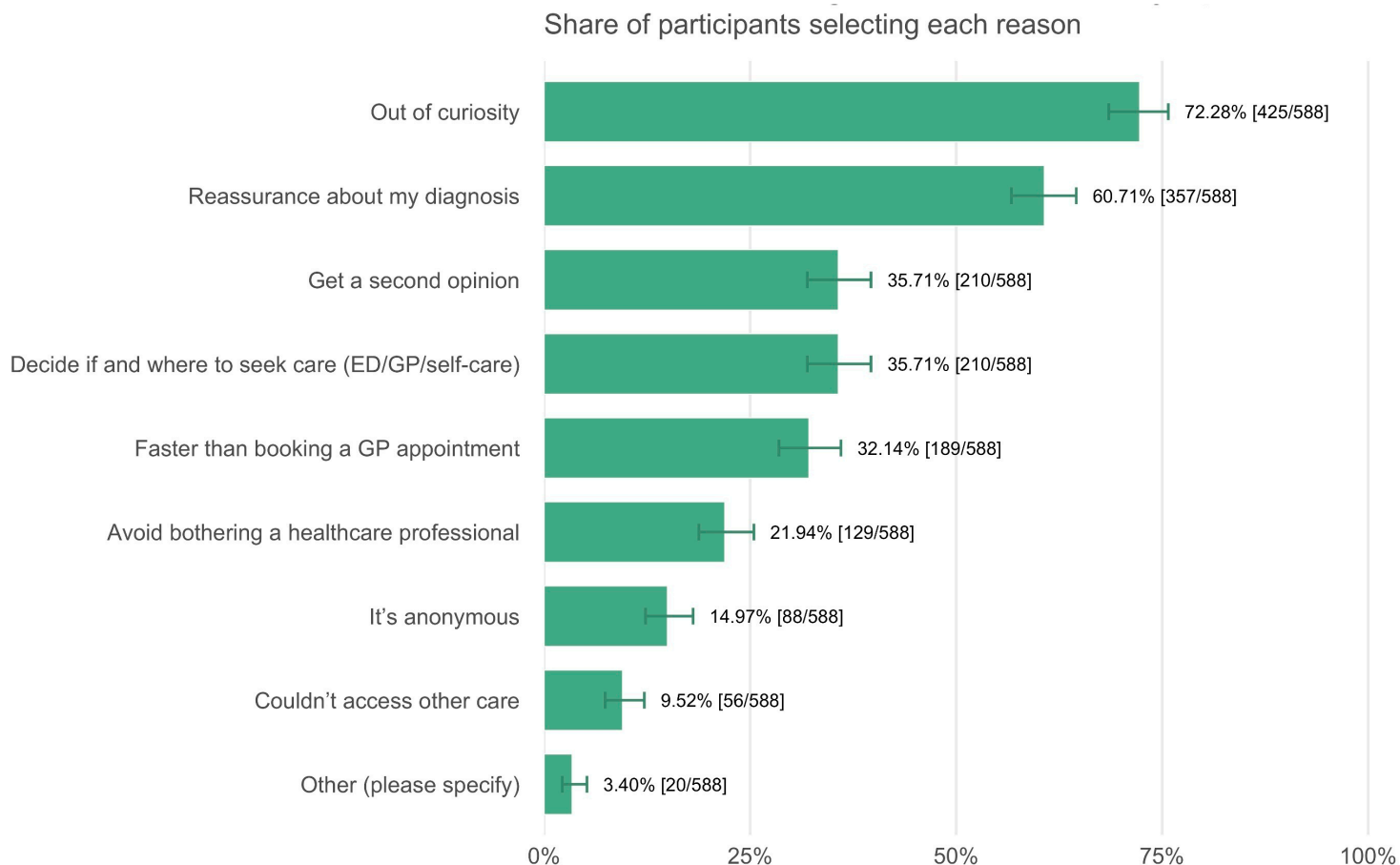


Figure 1

Reasons for using GenAI to assess symptoms. The bars indicate mean responses, and error bars indicate 95% confidence intervals

How much do you trust generative AI for...?

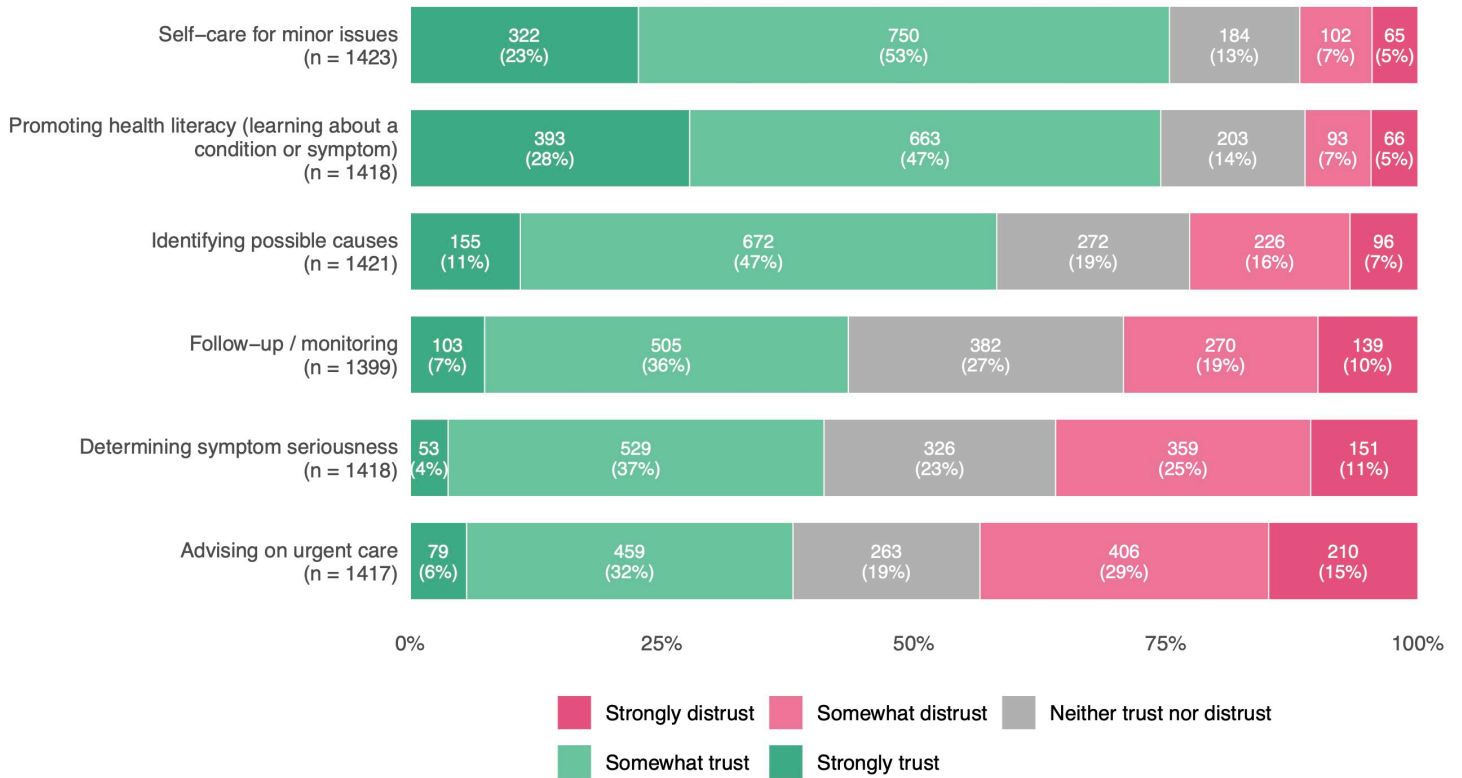


Figure 2

Trust in Generative AI across symptom-assessment use cases.

How accurate are generative AI symptom checkers compared to...?

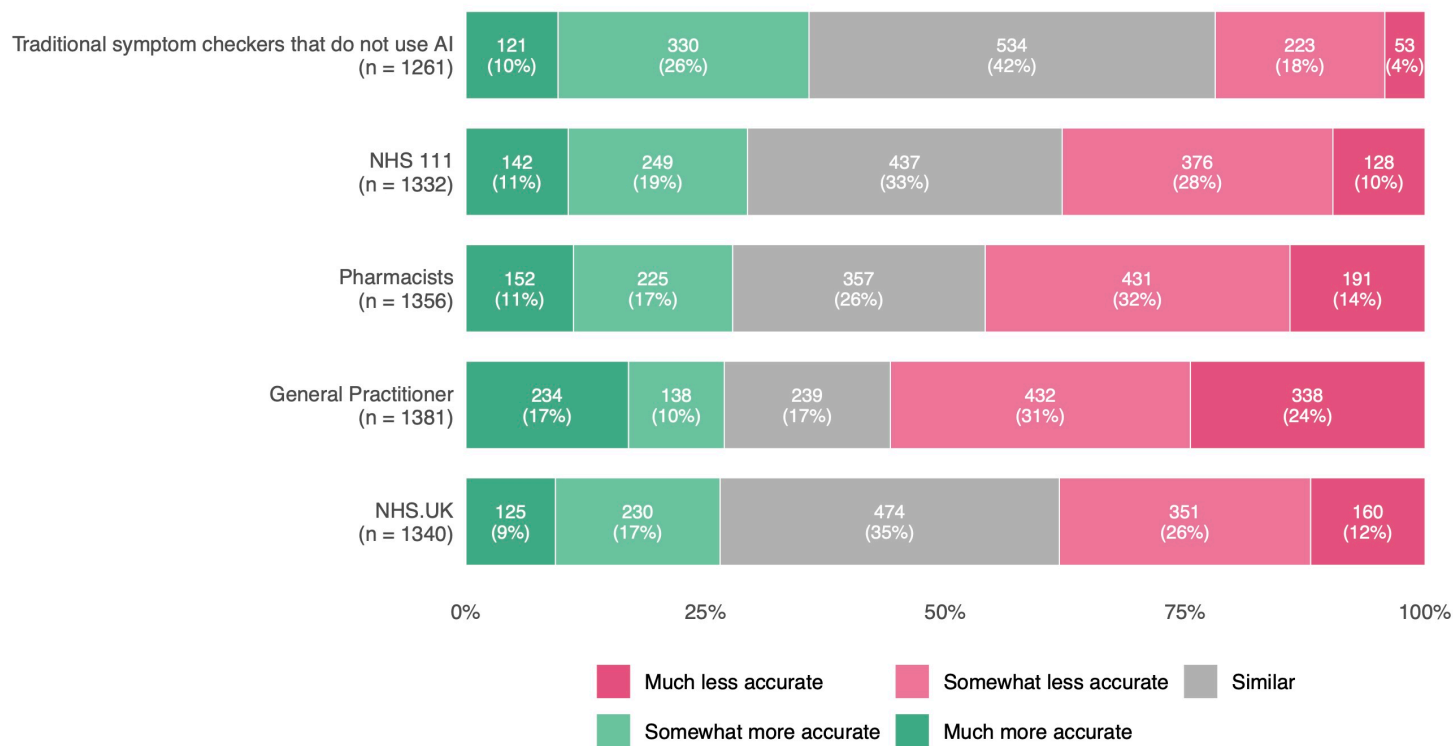
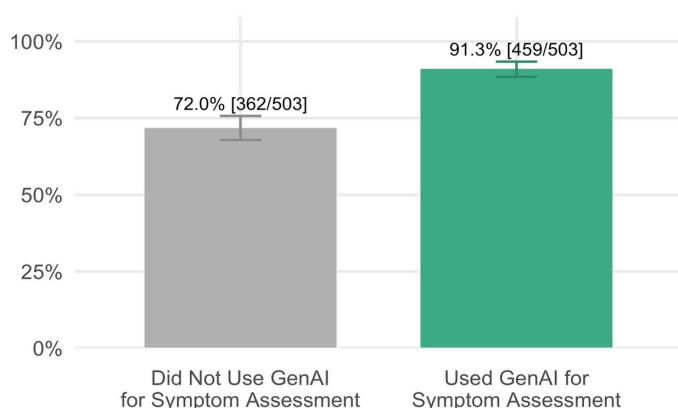


Figure 3

Expected accuracy of GenAI for symptom-assessment compared to other sources.

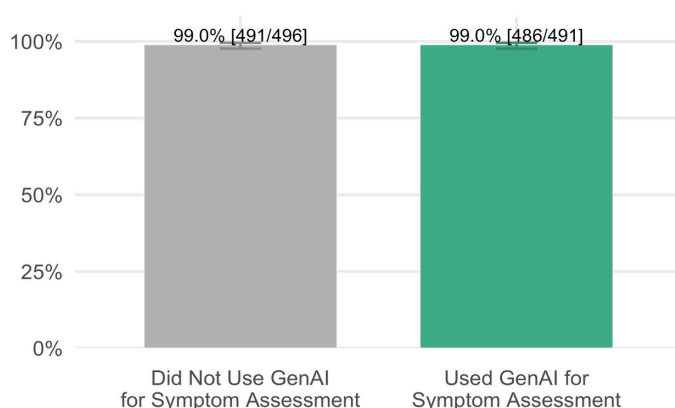
Support for integrating GenAI into NHS symptom checkers

Share supporting the implementation



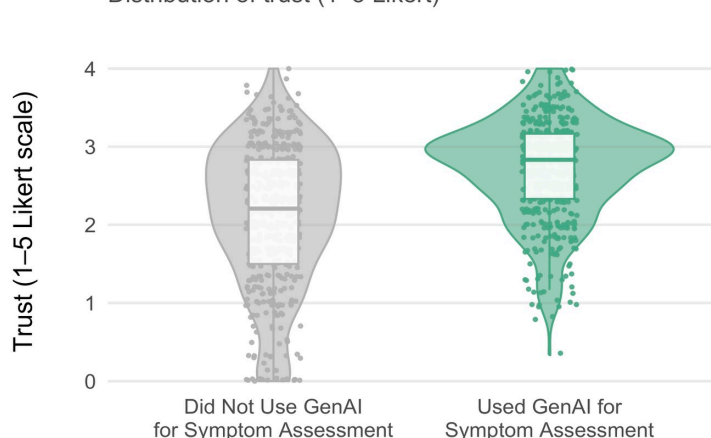
Support for Transparency About GenAI Use

Share thinking that users should be told when AI is involved in their health & care



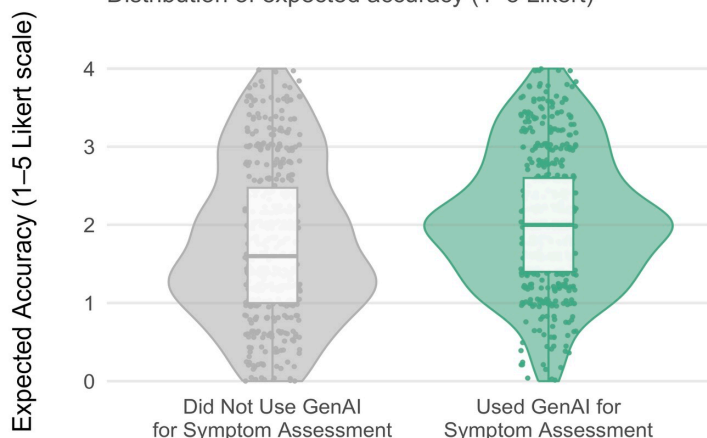
Trust in GenAI by Previous Use for Symptom Assessment

Distribution of trust (1–5 Likert)



Expected Accuracy by Previous Use for Symptom Assessment

Distribution of expected accuracy (1–5 Likert)



Note: For better visibility, the points are jittered. The bars indicate mean responses, and error bars indicate 95% confidence intervals

Figure 4

Differences between users who used GenAI for symptom assessment and those who did not. Groups were selected using propensity score matching to adjust for several confounders.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SFigure1.jpg](#)
- [SFile1SurveyExport.docx](#)
- [SFile2CHERRIESChecklist.docx](#)
- [STable1.docx](#)

- [STable2.docx](#)
- [STable3.docx](#)