

Perspectives and Experiences with Large Language Models in Healthcare: Survey insights from the workforce

Jennifer Sumner, Yuchen Wang, Si Ying Tan, Emily Hwee Hoon Chew, Alexander Wenjun Yip

Submitted to: Journal of Medical Internet Research
on: October 10, 2024

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 5

Supplementary Files..... 28

 Figures 29

 Figure 1..... 30

 Figure 2..... 31

 Figure 3..... 32

 Figure 4..... 33

CONSORT (or other) checklists..... 34

 CONSORT (or other) checklist 0..... 35

Perspectives and Experiences with Large Language Models in Healthcare: Survey insights from the workforce

Jennifer Sumner^{1*} BSc, PhD, MSc; Yuchen Wang^{2*} BSc; Si Ying Tan^{1*} PhD; Emily Hwee Hoon Chew¹ BSc; Alexander Wenjun Yip^{1*} MBBS

¹Alexandra Research Centre for Healthcare in a Virtual Environment Alexandra Hospital Alexandra Hospital Singapore SG

²National University of Singapore School of Computing National University of Singapore Singapore SG

*these authors contributed equally

Corresponding Author:

Jennifer Sumner BSc, PhD, MSc

Alexandra Research Centre for Healthcare in a Virtual Environment

Alexandra Hospital

Alexandra Hospital

Alexandra Hospital-Medical Affairs

Alexandra Road, Singapore

Singapore

SG

Abstract

Background: Large Language Models (LLMs) can potentially transform how we use data, including within the healthcare sector. Understanding stakeholder perspectives on new technology is crucial for successful implementation.

Objective: 1. Investigate users' uptake, perceptions, and experiences regarding LLMs in healthcare.
2. Analyse survey responses based on demographic profiles.

Methods: We administered an electronic survey to elicit stakeholder experiences and views of LLMs. Survey domains included: Demographics, user experiences of LLMs, motivations for using LLMs, and perceived impact on functional roles. The survey targeted healthcare providers, support staff, healthcare students and academics in health-related fields. Respondents were adults (>18 years) aware of LLMs.

Results: We received n=1083 responses, of which n=845 were analysable. Users primarily adopted LLMs for speed, convenience, and productivity. Non-users were more likely to be healthcare workers ($p<0.001$), older ($p<0.001$), and female ($p<0.01$). While 79% agreed that the user experience was positive, 46% found the generated content unhelpful. Non-users were less likely to report LLMs meeting unmet needs (45% versus 65%) or improving functional roles (63% versus 75%). Free-text opinions highlighted concerns regarding autonomy, outperformance, and reduced demand. Respondents also predicted changes to human interactions, including fewer but higher quality interactions and a change in consumer needs, which would require provider adaptation.

Conclusions: Despite the reported benefits of LLMs, the non-user profile—healthcare workers, older individuals, and females—underscores the need for targeted education and support to address adoption barriers and ensure the successful integration of LLMs in healthcare. Anticipated role changes, evolving human interactions, and the risk of the digital divide further emphasise the need for careful integration and ongoing evaluation of LLMs in healthcare to ensure equity and sustainability. Clinical Trial: N/A

(JMIR Preprints 10/10/2024:67383)

DOI: <https://doi.org/10.2196/preprints.67383>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

Please make my preprint PDF available to anyone at any time (recommended).

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

✓ **No, I do not wish to publish my submitted manuscript as a preprint.**

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org/preprint/67383>



Original Manuscript

Perspectives and Experiences with Large Language Models in Healthcare: Survey insights from the workforce

Jennifer Sumner^{1,2}

SiYing Tan²

Yuchen Wang³

Emily Hwee Hoon Chew²

Alexander Wenjun Yip²

¹ Medical Affairs – Research Innovation & Enterprise, Alexandra Hospital, National University Health System, Singapore

² Alexandra Research Centre for Healthcare in a Virtual Environment (ARCHIVE), Department of Healthcare Redesign, Alexandra Hospital, National University Health System, Singapore

³ School of Computing, National University of Singapore, Singapore.

Corresponding Author: Jennifer Sumner, Jennifer_sumner@nuhs.edu.sg

Alexandra Hospital, Medical Affairs – Research Innovation & Enterprise, National University Health System, Singapore

Keywords: Digital health, artificial intelligence, large language models, survey research, healthcare

ABSTRACT

Background

Large Language Models (LLMs) are transforming how we use data, including within the healthcare sector. However, frameworks including the Unified Theory of Acceptance and Use of Technology highlight the importance of understanding the factors that influence technology use for successful implementation. We conducted a cross-sectional survey to:

1. Investigate users' uptake, perceptions, and experiences regarding LLMs in healthcare.
2. Contextualise survey responses by demographics and professional profiles.

Materials and Methods

We administered an electronic survey to elicit stakeholder perspectives of LLMs (healthcare providers and support functions), their experiences with LLMs, and its potential impact on functional roles.. Survey domains included: Demographics (6 questions), user experiences of LLMs (8 questions), motivations for using LLMs (6 questions), and perceived impact on functional roles (4 questions). We launched the survey electronically, targeting healthcare providers or support staff, healthcare students and academics in health-related fields. Respondents were adults (>18 years) aware of LLMs.

Results

We received n=1083 responses, of which n=845 were analysable. Of the 845 respondents, 221 had yet to use an LLM. Non-users were more likely to be healthcare workers ($p<0.001$), older ($p<0.001$), and female ($p<0.01$). Users primarily adopted LLMs for speed, convenience, and productivity. While 79% agreed that the user experience was positive, 46% found the generated content unhelpful. Regression analysis showed that the experience with LLMs is more likely to be positive if the user is male (Odds Ratio (OR) 1.62, Confidence Interval (CI) 1.06, 2.48), and increasing age was associated with a reduced likelihood of reporting LLM output as useful (OR 0.98, CI 0.96-0.99). Non-users compared to LLM users were less likely to report LLMs meeting unmet needs (45% versus 65%) (OR 0.48 (0.35-0.65), and males were more likely to report that LLMs do address unmet needs (OR 1.64 (1.18-2.28). Furthermore, non-users compared to LLM users were less likely to agree that LLMs will improve functional roles (63% versus 75%) (OR 0.60 (0.43-0.85). Free-text opinions highlighted concerns regarding autonomy, outperformance, and reduced demand for care. Respondents also predicted changes to human interactions, including fewer but higher quality interactions and a change in consumer needs as LLMs become more common, which would require

provider adaptation.

Conclusions

Despite the reported benefits of LLMs, non-users—primarily healthcare workers, older individuals, and females—appeared more hesitant to adopt these tools. These findings underscore the need for targeted education and support to address adoption barriers and ensure the successful integration of LLMs in healthcare. Anticipated role changes, evolving human interactions, and the risk of the digital divide further emphasise the need for careful implementation and ongoing evaluation of LLMs in healthcare to ensure equity and sustainability.

INTRODUCTION

The healthcare sector faces substantial workforce challenges driven by ageing populations, increased chronic disease prevalence, and growing healthcare demands (1, 2). Consequently, technology is increasingly adopted to improve the quality and efficiency of care while alleviating workforce burdens (3, 4). Among the many technological developments, Artificial Intelligence (AI) has garnered much interest in the healthcare sector due to its ability to analyse, interpret and generate actionable insights from large volumes of complex health data, transforming how care is delivered and managed (3). Most recently, Large Language Models (LLMs) have been changing how we use text, numeral, audio and visual data, which will have widespread implications in the healthcare sector (5-9).

Large language Models (LLMs), a specific type of conversational AI, are trained to understand and generate human-like text (7). While the foundational concepts of LLMs have been around for some time, there has been a significant leap in development in recent years. General purpose models like OpenAI's ChatGPT, Google Gemini Ultra, Meta Llama 3, Anthropic Claude 3, as well as domain-specific LLMs (BlueBERT, Copilot, Med-PaLM) (7, 9, 10) are at the forefront of LLM capabilities, and interest in how these tools can benefit healthcare is only growing (7, 9). For example, LLMs to draft responses to patient messages, create structured medical notes of physician-patient interactions, provide clinical decision support, screen and enrol research participants, and enhance learning and education are only a few areas where LLMs are being applied in healthcare (7, 9, 11-17).

While the promise of LLMs in healthcare is substantial, it is also essential to understand their utility and challenges in real-world scenarios. For instance, early LLMs tended to 'hallucinate' or provide inaccurate information, which would represent a significant risk if adopted in the healthcare context (16, 18). Newer innovations, including the creation of domain-specific LLMs, the incorporation of human feedback on responses, and the application of real-time domain-specific knowledge to enhance the performance of general LLMs (non-medically trained), are now helping to minimise this issue (7, 19, 20). Another limitation is a lack of clarity on how LLMs generate responses. The ongoing development of explainable AI models will aid in overcoming this problem by providing transparency in the decision-making process (21-23). Finally, the need for adequate safeguards and regulation are other common concerns. As regulations evolve, such as the liability rules outlined by the European Commission, ethical safeguards and clear frameworks for implementing LLMs into practice will establish (24).

Given the challenges with LLMs, the perspectives of healthcare workers, who represent both users and stakeholders who will influence and be influenced by others using these tools (e.g., patients, other support functions), becomes particularly important for successful deployment. Moreover, understanding the influence of different backgrounds is also key. For instance, it is known that age and gender can significantly influence how individuals perceive and interact with technology (25-27). Older individuals can be more sceptical and less trusting of technologies, including AI (26). Conversely, younger generations may be more accustomed to digital technologies and report more positive views of technology (26). Likewise, males are more likely to use digital applications, reporting more favourable views and being more confident technology users (27). Frameworks like the Unified Theory of Acceptance and Use of Technology (UTAUT) highlight the importance of such factors in influencing technology use (25).

By examining healthcare providers' perceptions of LLMs, we aim to identify gaps between user expectations and reality, and how this might vary by profile. This understanding is key for adoption, ensuring that these models align with users' needs and ultimately improves patient care and healthcare efficacy. The purpose of this study is to explore professional views on LLM use in healthcare to inform future deployments. Specifically, we aim to:

- 1) Investigate users' uptake, perceptions, and experiences regarding LLMs in healthcare.
- 2) Contextualise survey responses by demographics and professional profiles.
- 3)

MATERIALS AND METHODS

This study is reported according to the Consensus Based Checklist for Reporting of Survey Studies (CROSS) (28). In this cross-sectional study, we developed and administered an electronic survey to elicit stakeholder perspectives and experiences of LLMs and their potential impact on functional roles. A visual overview of the study flow is reported in Supplement 1.

Setting and study population

We adopted a convenience sampling approach. Recruitment occurred through a combination of electronic email blasts (distributed through the National University Health System email lists) and word of mouth (adverts sent to department leads and a prompt included in the survey invitation encouraging sharing among peers). The survey was aimed at healthcare professionals, healthcare professional students, or academics in -related fields, regardless of institution. The survey was first launched on 13th July 2023, and a reminder email blast was sent on 21st August 2023. The survey eventually closed on 16th October 2023. Eligibility criteria were as follows:

Inclusion criteria

- Any healthcare providers (any profession or setting), students training in any healthcare discipline, or academics from health-related fields
- Adults aged 18 years or older
- Awareness of LLMs

Exclusion criteria

- Participants from fields unrelated to healthcare
- No awareness of LLMs such as ChatGPT

We did not exclude those who had yet to use an LLM, so long as they had heard of them. For those yet to use an LLM, survey questions relating to user experience were omitted.

Survey design

We developed the survey iteratively. First, JS and SYT reviewed the literature to identify common themes and issues related to the research questions. A draft survey was then shared with the wider project team to brainstorm new topics not otherwise captured and to refine questions. Finally, we assessed the content validity of the survey by circulating the survey with colleagues not within the

project team (i.e., healthcare providers and healthcare researchers) prior to launch. Feedback was used to improve the clarity of the survey before launch.

The final survey comprised of four domains: Demographics (6 questions), user experiences of LLMs (8 questions), motivations for using LLMs (6 questions), and perceived impact on their functional role (4 questions). The questionnaire took about 10 minutes to complete, and responses were anonymous. The survey consisted of multiple-choice questions, Likert scales, binary responses (i.e. Yes/No) and open-ended questions. The survey was launched in English on Qualtrics^{XM} (<http://qualtrics.com>), an online survey platform. Although Singapore is a multi-lingual country, English is the default business language. Finally, a single question was included at the start of the survey: “Have you heard of LLMs such as ChatGPT before?” to assess eligibility. A copy of the survey is included in Supplement 2.

Data analysis

Analyses were performed in STATA v15.0 (STATA Corp, College Station, Texas, USA) and Microsoft[®] 365 EXCEL v16.78 (Washington, USA). Summary statistics are presented as means with standard deviations (SD) or proportions. Demographics for non-users and early-LLM adopters were compared using a t-test or chi² test as appropriate. Significance was set at $p < 0.05$.

For categorical questions, the proportion of those who strongly or partially agreed with a question was calculated. Proportions were calculated for the whole sample and by specific demographical traits (sex, age group (18-29 years, 30-40 years, 41-50 years, ≥ 50 years), ethnic group (Chinese, Malay, Indian, Other), residency status (Singaporean or permanent resident, employment pass or other pass holder), highest education level (0-level/diploma or higher degree) and job category (healthcare provider or student or healthcare support function)). The calculated proportions were then used to generate a heat map in Excel to visualise the data.

Ordinal regression (for Likert scale questions) or logistic regression (for binary questions) were used to explore whether participant characteristics were predictive of positive survey responses. Independent variables included: sex, age group, ethnic group, residency status, highest education level, job category, and whether they were an LLM user or not. For logistic regression, the dependent variable was binary coded (positive vs. non-positive). For ordinal regression, the dependent variable was collapsed to three categories (positive, neutral, non-positive). Odds ratios (ORs) with 95%

confidence intervals (CIs) were computed for each predictor to interpret the strength and direction of associations. Robust standard errors were used in each analyses. Models were assessed for goodness-of-fit using the Hosmer-Lemeshow test for logistic regression and pseudo-R² metrics for ordinal regression.

For the open-ended survey questions, we performed a content analysis (29). First, the free text was checked and corrected for spelling mistakes, slangs and abbreviations. Second, a subset of the data was analysed and used to generate an initial set of themes. The themes were generated by one researcher and verified by the project team. Third, the themes were applied to the full dataset, further modifications were made to the code book if required through discussion among the study team. Finally, the coded data was then checked by a researcher to verify the appropriateness and accuracy of the coding.

Ethical considerations

The study was reviewed and approved by the National University of Singapore Ethical Review Board (NUS-IRB-2023-211). The online survey was self-administered, with the Participant Information Sheet and Consent Form shown at the start of the online survey. Participants could only proceed after indicating their consent to participate in this research. Data were anonymous and analysed at the aggregate level. Participants were not individually compensated but had the chance to enter a lucky draw, with fifty S\$20 prizes awarded at random.

RESULTS

We received a total of 1083 responses. After data cleaning 940 completed responses remained. A further 95 respondents were excluded from the dataset as they had never heard of LLMs such as ChatGPT; their characteristics are reported in Supplement 3. Participant characteristics for the final analysis sample (n=845) are reported in Table 1. Of the 845 respondents, 221 had yet to use an LLM. Non-users were more likely to be older, female, employment pass holders, and healthcare providers $p<0.01$.

Table 1 Survey respondent characteristics, overall and by LLM users and non-LLM users

Characteristic	All survey respondents N=845	Non-LLM users N=221	LLM users N=624
Mean age, years (SD)	35.49 (9.43)	37.72 (10.03)	34.70 (9.09)***
Male, n (%)	240 (28)	45 (21)	195 (31)**
Ethnicity, n (%)			
Chinese	707 (84)	174 (79)	533 (85)
Malay	43 (5)	15 (6)	28 (5)
Indian	51 (6)	14 (6)	37 (6)
Other	44 (5)	18 (9)	26 (4)
Residency status			
Singaporean	783 (93)	195 (88)	588 (94)**
Employment pass	48 (6)	23 (11)	26 (4)
Other (i.e. student)	14 (1)	3 (1)	10 (2)
Education, n (%)			
O-level/N-level	8 (1)	2 (1)	6 (1)
Diploma/A-level	160 (19)	44 (20)	116 (19)
Higher degree	677 (80)	175 (79)	502 (80)
Role/Occupation			
Healthcare provider/student	552 (65)	164 (74)	388 (62)***
Healthcare administration/support functions	293 (35)	56 (26)	236 (38)

Abbreviations: Standard Deviation (SD)

Healthcare provider: Doctors, nurses, allied health, dentist, pharmacy staff; **Healthcare**

administration/support functions: Operations, IT, finance, research, corporate roles related to healthcare

Non-LLM user versus LLM user: * $p<0.05$; ** $p<0.01$; *** $p<0.001$

Survey Responses

Results for the regression analyses are reported in Table 2. The two most frequent and significantly associated factors for positive survey responses were being a non-LLM user (associated with 7 questions) and increasing age (associated with 4 questions). The subsequent results are presented in three sections according to the three survey domains: 1. User experiences of LLMs, 2. Motivations for using LLMs, and 3. Perceived impact on functional roles.

Table 2 Statistically significantly participant characteristics associated with positive survey responses, Odds Ratio's, 95% confidence intervals and p values

Survey domain	Question	Significantly associated characteristic	Odds ratio (confidence intervals and p value)
User experiences	Positive user experience	Male	1.62 (1.06-2.48, 0.02)
	LLMs are accurate	None	-
	LLMs are useful	Increasing age	0.98 (0.96-0.99, 0.03)
	Confident using LLMs	Non LLM user	0.31 (0.22-0.44, <0.001)
	Recommends LLMs	Non LLM user	0.10 (0.07-0.15, <0.001)
Motivations for use	LLMs are not overhyped	Increasing age	0.97 (0.96-0.98, 0.001)
		Non LLM user	0.67 (0.50-0.90, 0.008)
	No social expectation to use LLMs	Increasing age	0.96 (0.95-0.97, <0.001)
		Singaporean/permanent resident	1.66 (1.04-2.64, 0.03)
	LLMs address an unmet need	Male Non LLM user	1.64 (1.18-2.28, 0.003) 0.48 (0.35-0.65, <0.001)
Perceived impact on functional role	LLMs will improve functional roles	Chinese	2.52 (1.06-5.97, 0.03)
		Malay	3.17 (1.07-9.36, 0.03)
		Non LLM user	0.60 (0.43-0.85, 0.004)
	LLMs will impact human interactions	Increasing age	1.02 (1.00-1.03, 0.007)
	LLMs are not a threat to functional roles	Higher degree holder	1.84 (1.26-2.70, 0.002)
		Non LLM user	0.65 (0.46-0.91, 0.01)
	LLMs should be used in healthcare	Higher degree holder Non LLM user	1.80 (1.09-2.98, 0.02) 0.62 (0.38-0.98, 0.04)

User experiences of LLMs

Roughly half of users reported rarely using LLMs (52%), and 40% reported weekly use. More than 70% rated the overall experience, the perceived accuracy, and their confidence in using LLMs highly, and 62% would recommend LLMs to others. Conversely, less than half (46%) agreed that the content generated by the LLM was useful (Figure 1).

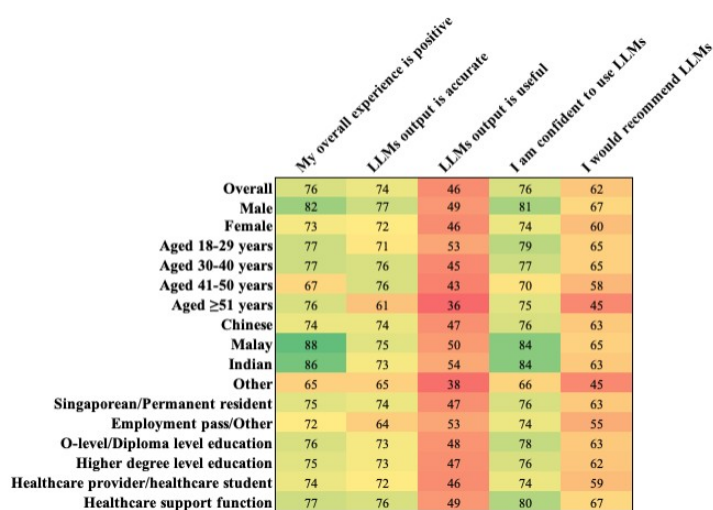


Figure 1 A heat map of question responses (% agreement) on user experience, overall and by individual demographic groups (Dark Green indicates strongest agreement and dark red indicates lowest agreement)

Regression analysis showed that the experience with LLMs is more likely to be positive if the user is male, and increasing age was associated with a reduced likelihood of reporting LLM output as useful (Table 2). Non LLM users were less likely to report feeling confident to use LLMs and less likely to recommend them to others.

Motivations for using LLMs

Approximately half of respondents used LLMs for personal and work reasons, and users (65%) were more likely to report LLMs addressed an unmet need compared to non-users (45%) (Figure 2). Around a third of users and non-users agreed that external factors motivated decisions to use LLMs (i.e., social pressure or hype).

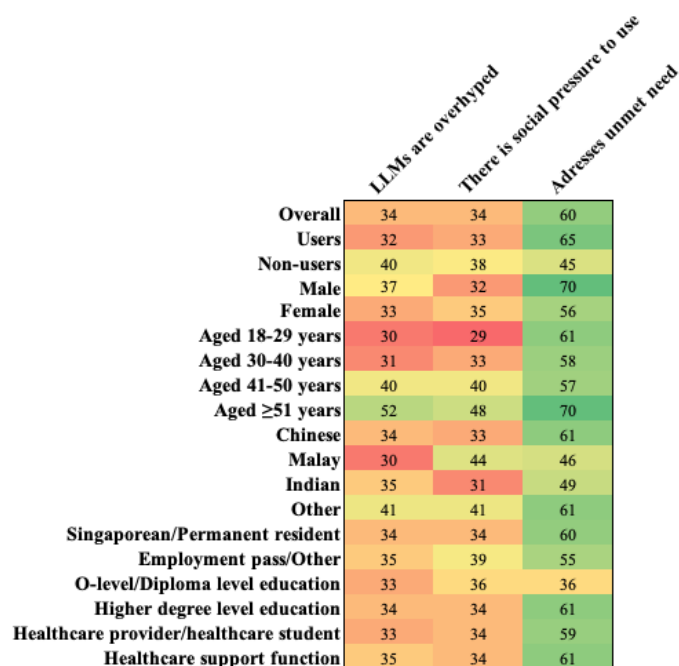


Figure 2 A heat map of question responses (% agreement) on motivations for using LLMs, overall and by individual demographic groups (Dark Green indicates strongest agreement and dark red indicates lowest agreement)

Regression analysis (Table 2) found that younger adults and LLM-users were more likely to disagree that LLMs are overhyped, and younger adults and Singaporeans/permanent residents were more likely to report there are no social expectations to use LLMs. Furthermore, males were more likely to report that LLMs address an unmet need, but non-LLM users reported the opposite.

When asked about motivations for using LLMs, the top five reasons were: speed, productivity, convenience, curiosity, and personalised responses. In terms of specific tasks, roughly two thirds of respondents reported using LLMs for ideation (63%), answering general questions (61%) and writing (57%). Other uses included entertainment (33%), answering medical questions (22%), data analysis (17%), social interaction (16%) and literature reviews (15%).

When asked why the non-users had not used LLMs yet, content analysis of free text revealed five topics. First, there was a **1) lack of trust or skepticism** among users regarding the reliability of AI-generated content, coupled with concerns about data security and the software's credibility. Additionally, some users perceived **2) no immediate need or relevance** for AI tools, relying on alternative sources for information, such as established search engines (e.g., Google). Others **3)**

lacked awareness of LLMs and their potential applications, and some did not have **4) access or an opportunity** to try LLMs, citing financial hurdles, lack of institutional investment or lack of time as reasons. Finally, some reported a lack of **5) technology literacy**, which prevented them from trying LLMs.

Perceived impact on Functional Role

Regardless of profile, agreement that LLMs should be used in healthcare was >80% (Figure 3). Regression analyses (Table 2) revealed that non-LLM users were statistically significantly less likely to agree that LLMs will improve functional roles, LLMs are not a threat to functional roles, and LLMs should be used in healthcare. Other factors significantly associated with survey responses in this domain were ethnicity, education level and age.

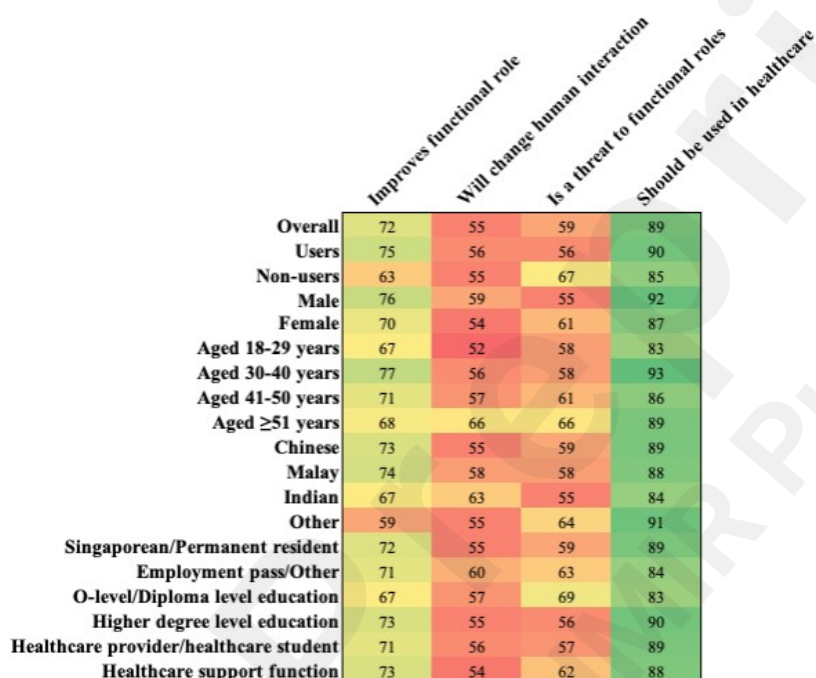


Figure 3 A heat map of question responses (% agreement) on the perceived impact of LLMs on functional roles, overall and by individual demographic groups (Dark Green indicates strongest agreement and dark red indicates lowest agreement)

When survey respondents were asked about their views on the perceived impact of LLMs on functional roles, content analysis of free-text revealed three main topics. First, there is **1) apprehension that LLMs could compete for tasks** traditionally performed by healthcare workers. For example, increasingly accurate diagnostics and scan interpretations or the automated generation of treatment plans could threaten doctors' autonomy. Second, the ability of LLMs to synthesise and interpret vast amounts of information could also lead to **2) greater accuracy and efficiency in**

practice. Consequently, LLMs could outpace doctors' ability to stay current. Third, as 3) **LLMs improve the accessibility of medical information**, there may be a reduced demand for consultations and a devaluing of domain expertise.

Despite these concerns, some healthcare workers believe that the integration of LLMs will not be immediate, given the technology's current limitations, particularly its reliability and accuracy. There is also recognition that LLMs could create new opportunities within the sector, though uncertainty remains about the extent of their impact. Importantly, participants emphasised the need for human oversight to ensure the accuracy and relevance of LLM-generated content, highlighting safety concerns such as the potential for scams, data breaches, and misinformation.

The impact of LLMs on human interactions

When asked about how LLMs might impact human interactions, three topics emerged from the analysis of free text responses: Less interactions, higher quality interactions, and a change in consumer needs (Figure 4).

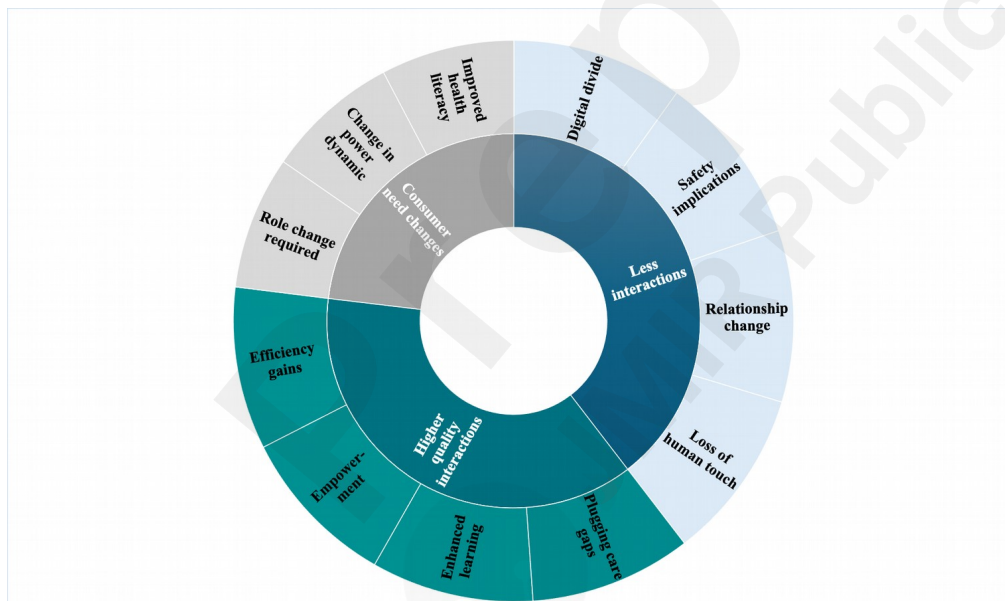


Figure 4 Qualitative themes (inner circle) and sub-themes (outer circle) on the perceived impact of LLMs on human interactions from the free text survey data

Topic 1: Less interactions

Participants reflected that introducing LLMs would likely reduce the volume of human interactions. Participants expressed concerns about a loss of human touch or empathic interactions, as LLMs may not fully replicate the nuances of human interaction. Furthermore, fewer human interactions may negatively impact the patient and provider relationship (or teacher and student), raising concerns

about the erosion of rapport and trust. The safety of reduced interactions was another concern, particularly for patients whose issues may only be fully understood with in-person discussion, creating missed opportunities to engage and address care needs. Moreover, certain aspects of human interaction were viewed as irreplaceable, highlighting the indispensable need for the human touch in healthcare settings.

Topic 2: Higher quality interactions

Respondents noted that LLMs can potentially improve the quality of human interactions. For example, integrating AI to handle simpler, less complex tasks could free up time for professionals to focus on more critical responsibilities where human interaction is essential. In turn, this was seen as a means to enhance efficiency and effectiveness. Participants expressed optimism about the empowerment derived from having greater accessibility to knowledge, leading to more informed interactions and improved learning experiences for practitioners, patients and students. Furthermore, respondents theorised that using AI could help address care gaps in situations where patients have limited access to care.

Topic 3: Change in consumer needs

Participants highlighted that LLMs would likely change consumer knowledge and attitudes (i.e., patients, students), leading to shifts in their needs and the relationship dynamics. For instance, as patients become more health-literate, their inquiries and demands will evolve. Empowered consumers, informed by LLMs, may also disrupt traditional power dynamics (i.e., patient and provider, teacher and student) where professionals are the traditional knowledge keepers. These consumer changes would likely necessitate acquiring new interpersonal skills to manage the demands of a more health-literate population, and managing misinformed consumers. Conversely, LLMs may also lead to a digital divide, with less technologically savvy individuals being left behind, exacerbating existing disparities in technology use and access to healthcare and education.

DISCUSSION

We conducted a cross-sectional survey to gather stakeholder perspectives and experiences of LLMs and their potential impact on functional roles in healthcare. We received over eight hundred analysable responses from healthcare providers, support staff, healthcare students and academics in related fields. Among the respondents, non-users, were predominantly healthcare workers, older individuals, and females. Users primarily adopted LLMs for speed, convenience, and productivity. While the overall user experience was generally positive, approximately half reported that the generated content was not useful. In contrast, nearly ninety percent of respondents felt LLMs should be used in healthcare. Non-users were less likely to recognise LLMs as addressing unmet needs or improving functional roles. Free-text opinions highlighted concerns regarding autonomy, outperformance, and reduced demand. Furthermore, respondents felt that human interactions would likely change, expecting fewer but higher quality exchanges as well as shifts in consumer attitudes and needs, which would require provider adaptation.

Most respondents rated their overall experience and confidence in using LLMs highly. LLMs have garnered significant popularity due to their user-friendly interface, accessibility and ability to generate human-like output promptly (30). Motivations for using LLMs were predominantly driven by the desire for speed, convenience, and productivity, with about 70% of respondents citing these factors. This finding aligns with current opinion on using LLMs in healthcare (5, 19, 31, 32). However, less than half found the generated output useful, indicating a gap between the positive user experience and the perceived utility of LLM-generated content. This disconnect may be explained by known issues when using LLMs, such as hallucinations, poor accuracy of responses, a lack of explainability, and suboptimal user prompting (33-35). Recent advancements in LLMs are addressing these key limitations, such as fine-tuning LLMs using domain-specific datasets to improve accuracy and reliability, new training methodologies (e.g., Reinforcement Learning with Human Feedback) to align LLM-generated outputs more closely with expert-validated data (18), real-time detection and mitigation of hallucinations, and ongoing work on explainable AI models to ensure transparency and build trust (19) (21-23, 36). Finally, as users become more familiar with LLMs, their awareness and skill at tailoring prompts to generate more accurate responses will improve.

There was a mixture of views on the perceived impact of LLMs on functional roles, but most felt LLMs should be used in healthcare. This sentiment is echoed in similar surveys, which report

favourable views on LLM use in healthcare, emphasising LLMs as co-pilots rather than role replacers (31, 32, 37). Adopting LLMs as co-pilots would mitigate the risk of using inaccurate information by maintaining human oversight. As with the perceived impact on functional roles, respondents also had mixed opinions of the impact of LLMs on professional interactions. Some foresee reduced human interactions, eroding empathy and negatively impacting the patient-provider relationship. Furthermore, LLM use in healthcare may exacerbate the digital divide, as those without access or less technologically savvy individuals are left behind (38).

These perspectives highlight the complexity and uncertainty surrounding the adoption of LLMs in healthcare settings, emphasising the importance of continued research and evaluation. Respondents reported a mix of positive and negative implications of LLMs, and these views varied by demographical profile. In particular, non-LLM users and age were associated with responses in many survey domains. Prior research has shown that age, sex and experience influence technology adoption (25). Increasing age is linked to poorer technology understanding and engagement (26, 39). Women are more hesitant to adopt technology due to an underrepresentation in AI development roles and a lower proportion of women studying STEM subjects, hindering their exposure to and interest in technology (27, 40, 41). Finally, prior experience can significantly impact future intentions to use technology (25). By taking into account the factors that influence perception and adoption, and improving education on LLM technologies, development and implementation can be enabled.

Notwithstanding the perceived benefits of LLMs in healthcare, further research is needed to evaluate the actual impact of implemented LLM technology. Evaluation is needed to establish whether promises of improved efficiency, greater accuracy and improved patient care are true or if such technology introduces new challenges. As LLM technology becomes more common place within the healthcare sector clinicians will also need training. Training will help to optimise LLM usage (i.e., prompt engineering), and how to handle patients who become active users of such technology. Finally, it is important that inequalities are not exacerbated or introduced through the introduction of LLMs into healthcare. Supporting slow adopters, including women and older adults, is also critical for successful implementation.

LIMITATIONS

There are several strengths and limitations to consider when interpreting the results. As a cross-sectional study, our study may be prone to issues with bias due to the study design. For instance, the convenience sampling approach may have led to a sample that is not fully representative of the

broader healthcare workforce. We attempted to address this by maximising our sampling approach and disseminating the survey through multiple routes. The limitation of an electronic survey format may have also resulted in a biased sample agreeing to participate, for instance, those who are more technology literate. Finally, participants may have provided socially acceptable responses, particularly as respondents had a chance to win an incentive. We attempted to mitigate this risk by making the survey anonymous. Future studies should attempt to use more diverse recruitment strategies to reach underrepresented groups.

CONCLUSIONS

LLMs can be valuable tools that can enhance and augment healthcare roles. However, healthcare inherently relies on nuanced decision-making, patient trust, physical interaction, and empathy—qualities that LLMs cannot replace. Non-users, predominantly healthcare workers, older individuals, and females in our study, remain hesitant to adopt these tools, underscoring the need for targeted education and support to overcome barriers. Anticipated role changes, evolving human interactions, and the risk of the digital divide further highlight the importance of careful integration and ongoing evaluation of LLMs to ensure equity and sustainability in healthcare.

Acknowledgements

This research is supported by Singapore Ministry of Health's National Medical Research Council RIE2025 Centre Grant Programme (Grant number: NMRC/CG3/003/2022-AH/MOH-0010130-00).

Data Availability

The data are available upon reasonable request to the corresponding author.

Competing interests

None declared.

Abbreviations

AI: Artificial Intelligence

CI: Confidence Intervals

CROSS: Consensus Based Checklist for Reporting of Survey Studies

LLM: Large Language Models

OR: Odds ratio

UTAUT: Unified Theory of Acceptance and Use of Technology

Author contributions

Conceptualisation, J.S. and SYT; methodology, J.S., S.Y.T., and Y.W.; validation, J.S., S.Y.T., Y.W., and A.Y.; formal analysis, J.S. and Y.W.; investigation, J.S., S.Y.T., and Y.W.; resources, A.Y.; data curation: J.S., S.Y.T., Y.W., and E.H.H.C; writing and editing J.S., S.Y.T., Y.W., A.Y., and E.H.H.C; supervision, J.S.,A.Y., and E.H.H.C; project administration, A.Y., and E.H.H.C; funding acquisition, A.Y., and E.H.H.C.

REFERENCES

1. World Health Organization. Ageing and health. Geneva: WHO; 2024 [Available from: <https://www.who.int/news-room/fact-sheets/detail/ageing-and-health>].
2. Azzopardi-Muscat N, Zapata T, Kluge H. Moving from health workforce crisis to health workforce success: the time to act is now. *The Lancet Regional Health – Europe*. 2023;35.
3. Thomas D, Ravi K. The potential for artificial intelligence in healthcare. *Future Healthcare Journal*. 2019;6(2):94.
4. Secinaro S, Calandra D, Secinaro A, Muthurangu V, Biancone P. The role of artificial intelligence in healthcare: a structured literature review. *BMC Medical Informatics and Decision Making*. 2021;21(1):125.
5. Clusmann J, Kolbinger FR, Muti HS, Carrero ZI, Eckardt JN, Laleh NG, et al. The future landscape of large language models in medicine. *Commun Med (Lond)*. 2023;3(1):141.
6. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med*. 2023;29(8):1930-40.
7. Cascella M, Semeraro F, Montomoli J, Bellini V, Piazza O, Bignami E. The Breakthrough of Large Language Models Release for Medical Applications: 1-Year Timeline and Perspectives. *J Med Syst*. 2024;48(1):22.
8. Tripathi S, Sukumaran R, Cook TS. Efficient healthcare with large language models: optimizing clinical workflow and enhancing patient care. *Journal of the American Medical Informatics Association*. 2024;31(6):1436-40.
9. Zhang K, Meng X, Yan X, Ji J, Liu J, Xu H, et al. Revolutionizing Health Care: The Transformative Impact of Large Language Models in Medicine. *J Med Internet Res*. 2025;27:e59069.
10. Peng L, Luo G, Zhou S, Chen J, Xu Z, Sun J, et al. An in-depth evaluation of federated learning on biomedical natural language processing for information extraction. *npj Digital Medicine*. 2024;7(1):127.
11. Garcia P, Ma SP, Shah S, Smith M, Jeong Y, Devon-Sand A, et al. Artificial Intelligence–Generated Draft Replies to Patient Inbox Messages. *JAMA Network Open*. 2024;7(3):e243201-e.
12. Kernberg A, Gold JA, Mohan V. Using ChatGPT-4 to Create Structured Medical Notes From Audio Recordings of Physician-Patient Encounters: Comparative Study. *J Med Internet Res*. 2024;26:e54419.
13. Beattie J, Neufeld S, Yang D, Chukwuma C, Gul A, Desai N, et al. Utilizing Large Language Models for Enhanced Clinical Trial Matching: A Study on Automation in Patient Screening. *Cureus*. 2024;16(5):e60044.
14. Safranek CW, Sidamon-Eristoff AE, Gilson A, Chartash D. The Role of Large Language Models in Medical Education: Applications and Implications. *JMIR Med Educ*. 2023;9:e50945.
15. Meng X, Yan X, Zhang K, Liu D, Cui X, Yang Y, et al. The application of large language models in medicine: A scoping review. *iScience*. 2024;27(5):109713.
16. Park Y-J, Pillai A, Deng J, Guo E, Gupta M, Paget M, et al. Assessing the research landscape and clinical utility of large language models: a scoping review. *BMC Medical Informatics and Decision Making*. 2024;24(1):72.
17. Yang R, Tan TF, Lu W, Thirunavukarasu AJ, Ting DSW, Liu N. Large language models in health care: Development, applications, and challenges. *Health Care Sci*. 2023;2(4):255-63.
18. Zhang J, Sun K, Jagadeesh A, Falakaflaki P, Kayayan E, Tao G, et al. The potential and pitfalls of using a large language model such as ChatGPT, GPT-4, or LLaMA as a clinical assistant. *Journal of the American Medical Informatics Association*. 2024;31(9):1884-91.
19. Tian S, Jin Q, Yeganova L, Lai PT, Zhu Q, Chen X, et al. Opportunities and challenges for ChatGPT and large language models in biomedicine and health. *Brief Bioinform*. 2023;25(1).
20. Wang D, Liang J, Ye J, Li J, Li J, Zhang Q, et al. Enhancement of the Performance of Large

Language Models in Diabetes Education through Retrieval-Augmented Generation: Comparative Study. *J Med Internet Res*. 2024;26:e58041.

21. Frasca M, La Torre D, Pravettoni G, Cutica I. Explainable and interpretable artificial intelligence in medicine: a systematic bibliometric review. *Discover Artificial Intelligence*. 2024;4(1):15.
22. Muhammad D, Bendeche M. Unveiling the black box: A systematic review of Explainable Artificial Intelligence in medical image analysis. *Comput Struct Biotechnol J*. 2024;24:542-60.
23. Bach TA, Khan A, Hallock H, Beltrão G, Sousa S. A Systematic Literature Review of User Trust in AI-Enabled Systems: An HCI Perspective. *International Journal of Human-Computer Interaction*. 2024;40(5):1251-66.
24. European Commission. Liability Rules for Artificial Intelligence Geneva: European Commission; 2022 [Available from: https://commission.europa.eu/business-economy-euro/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence_en].
25. Venkatesh V, Morris MG, Davis GB, Davis FD. User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*. 2003;27(3):425-78.
26. Czaja SJ, Charness N, Fisk AD, Hertzog C, Nair SN, Rogers WA, et al. Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). *Psychol Aging*. 2006;21(2):333-52.
27. Cai Z, Fan X, Du J. Gender and attitudes toward technology use: A meta-analysis. *Computers & Education*. 2017;105:1-13.
28. Sharma A, Minh Duc NT, Luu Lam Thang T, Nam NH, Ng SJ, Abbas KS, et al. A Consensus-Based Checklist for Reporting of Survey Studies (CROSS). *J Gen Intern Med*. 2021;36(10):3179-87.
29. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res*. 2005;15(9):1277-88.
30. Caversan F. Making sense of the chatter: The rapid growth of large language models USA: Forbes; 2023 [Available from: <https://www.forbes.com/sites/forbestechcouncil/2023/06/20/making-sense-of-the-chatter-the-rapid-growth-of-large-language-models/>].
31. Blease C, Worthen A, Torous J. Psychiatrists' experiences and opinions of generative artificial intelligence in mental healthcare: An online mixed methods survey. *Psychiatry Research*. 2024;333:115724.
32. Hosseini M, Gao CA, Liebovitz DM, Carvalho AM, Ahmad FS, Luo Y, et al. An exploratory survey about using ChatGPT in education, healthcare, and research. *PLoS One*. 2023;18(10):e0292216.
33. Andrew A. Potential applications and implications of large language models in primary care. *Fam Med Community Health*. 2024;12(Suppl 1).
34. Wang L, Chen X, Deng X, Wen H, You M, Liu W, et al. Prompt engineering in consistency and reliability with the evidence-based guideline for LLMs. *NPJ Digit Med*. 2024;7(1):41.
35. Sivarajkumar S, Kelley M, Samolyk-Mazzanti A, Visweswaran S, Wang Y. An Empirical Evaluation of Prompting Strategies for Large Language Models in Zero-Shot Clinical Natural Language Processing: Algorithm Development and Validation Study. *JMIR Med Inform*. 2024;12:e55318.
36. Farquhar S, Kossen J, Kuhn L, Gal Y. Detecting hallucinations in large language models using semantic entropy. *Nature*. 2024;630(8017):625-30.
37. Chen M, Zhang B, Cai Z, Seery S, Gonzalez MJ, Ali NM, et al. Acceptance of clinical artificial intelligence among physicians and medical students: A systematic review with cross-sectional survey. *Front Med (Lausanne)*. 2022;9:990604.
38. Reddy H, Joshi S, Joshi A, Wagh V. A Critical Review of Global Digital Divide and the Role of Technology in Healthcare. *Cureus*. 2022;14(9):e29739.
39. Stypinska J. AI ageism: a critical roadmap for studying age discrimination and exclusion in digitalized societies. *AI Soc*. 2023;38(2):665-77.

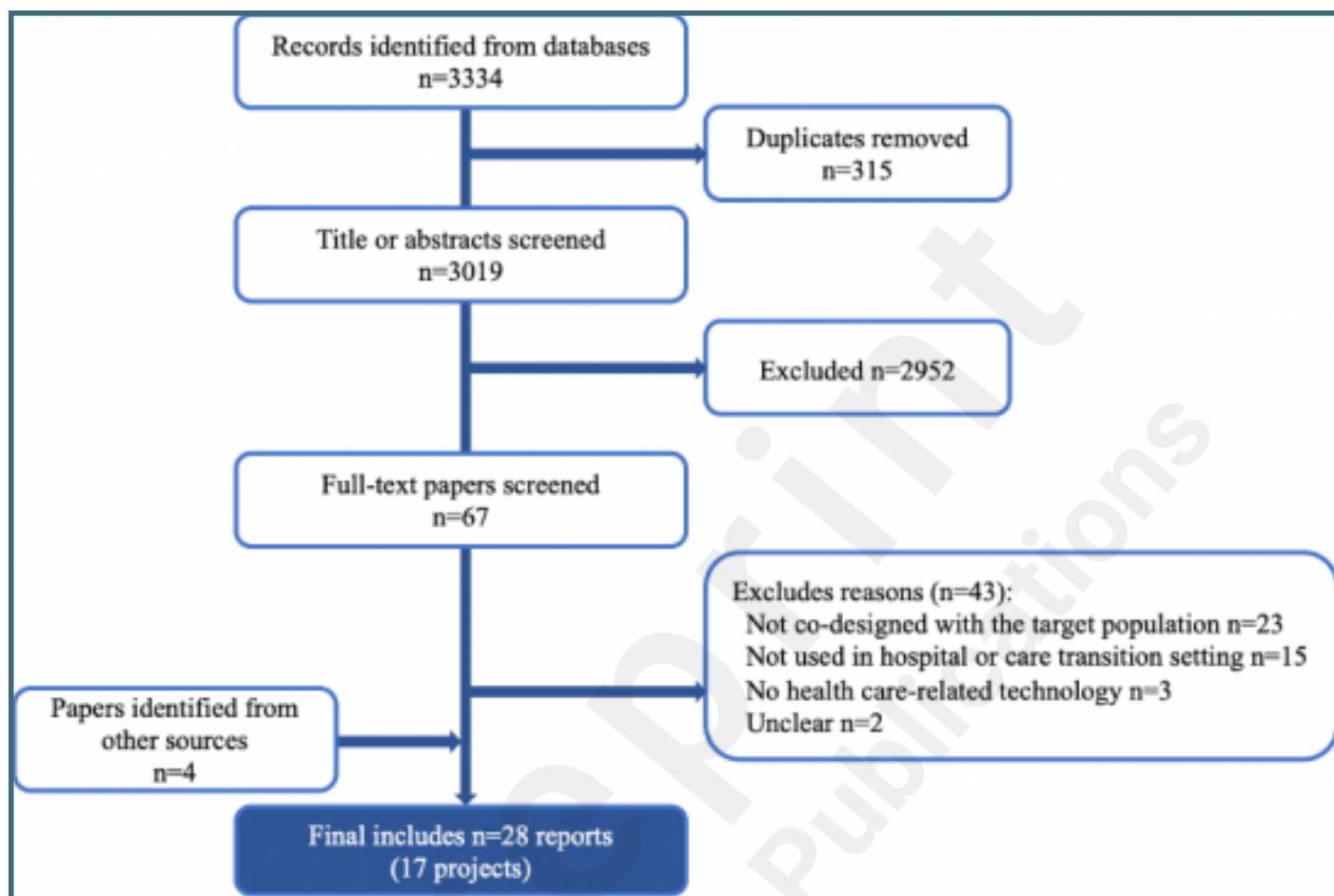
40. World Economic Forum. Why we must act now to close the gender gap in AI USA: WEF; 2022 [Available from: <https://www.weforum.org/agenda/2022/08/why-we-must-act-now-to-close-the-gender-gap-in-ai/>].
41. Buslón N, Cortés A, Catuara-Solarz S, Cirillo D, Rementeria MJ. Raising awareness of sex and gender bias in artificial intelligence and health. *Front Glob Womens Health*. 2023;4:970312.



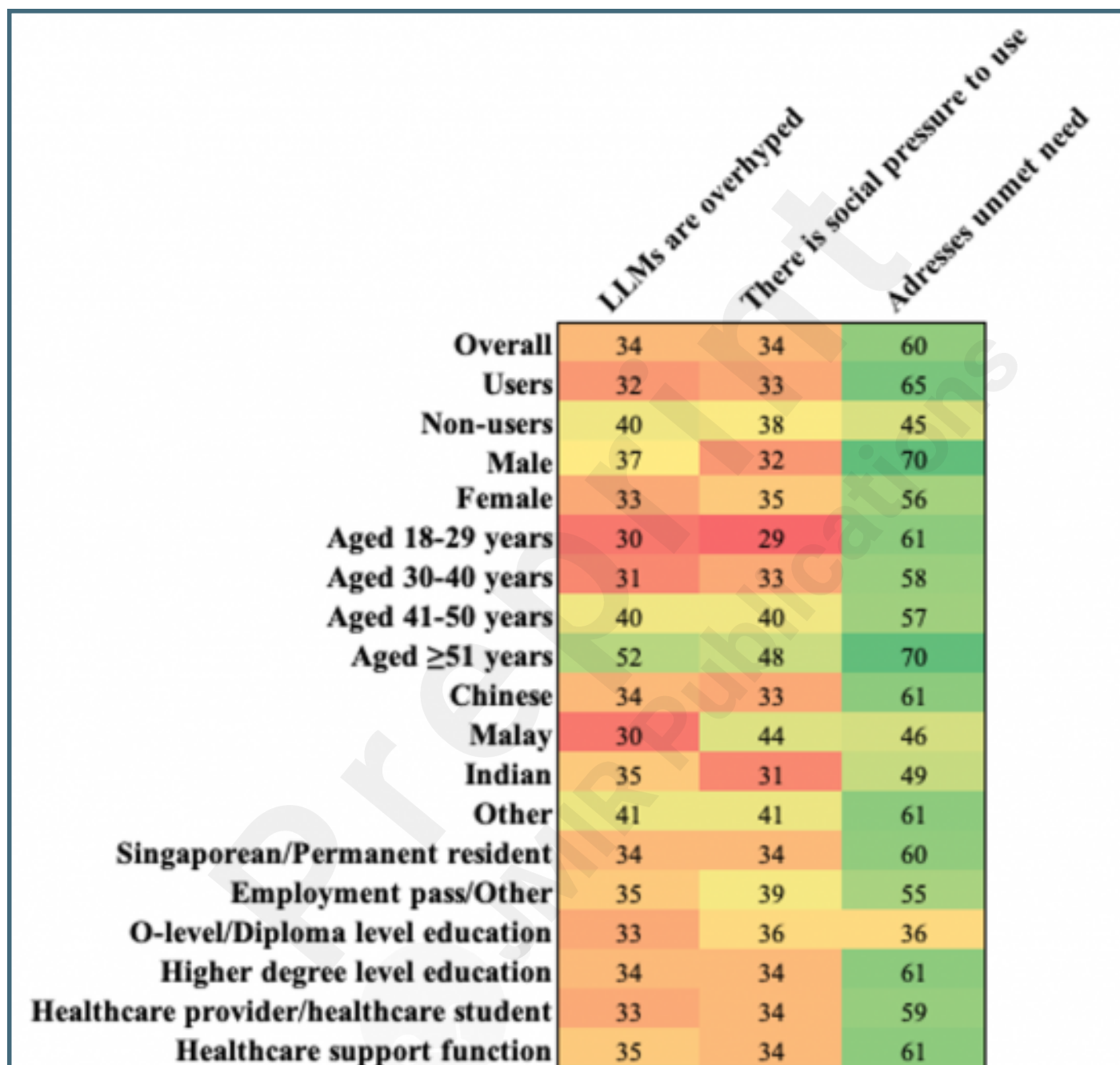
Supplementary Files

Figures

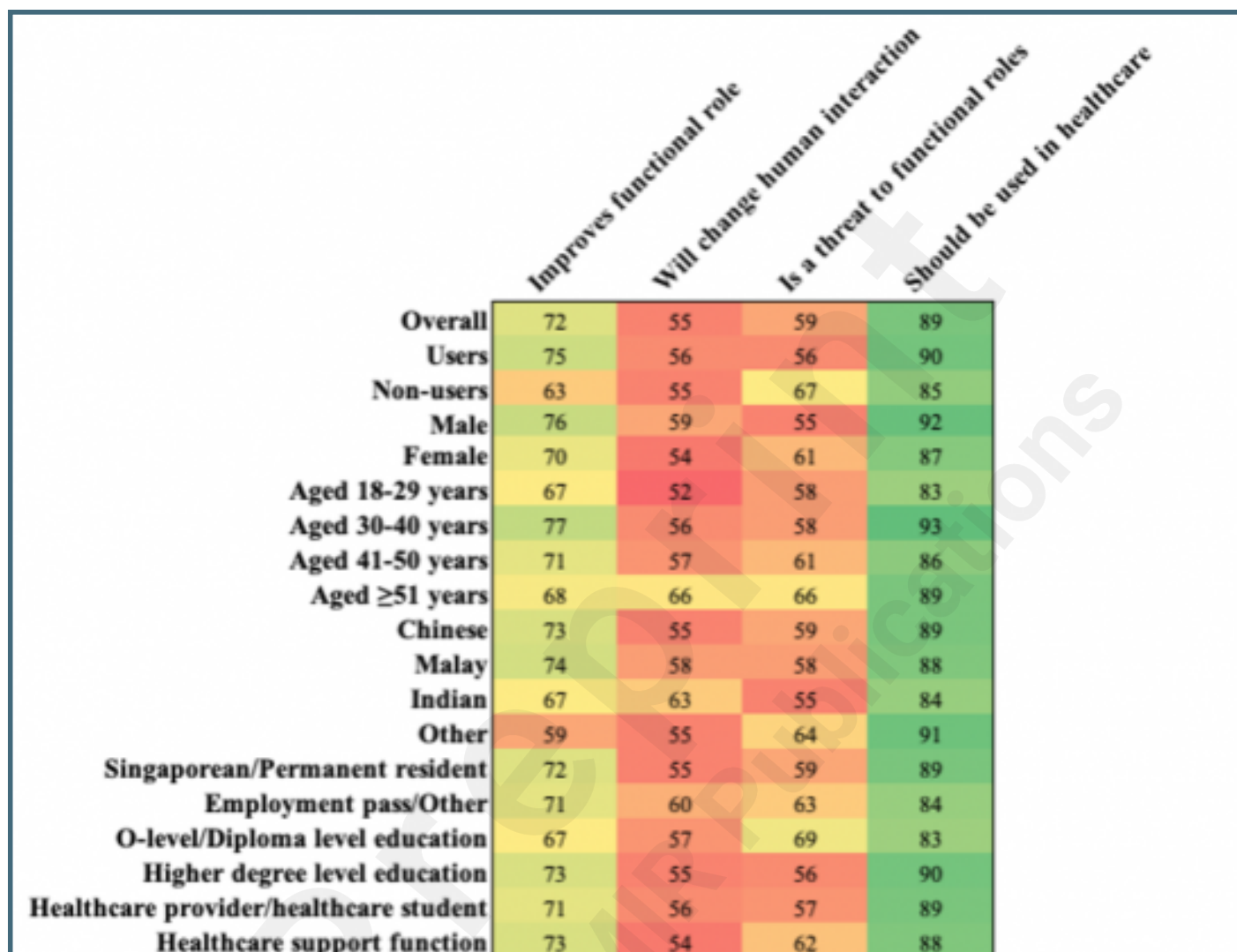
A heat map of question responses (% agreement) on user experience, overall and by individual demographic groups (Dark Green indicates strongest agreement and dark red indicates lowest agreement).



A heat map of question responses (% agreement) on motivations for using LLMs, overall and by individual demographic groups (Dark Green indicates strongest agreement and dark red indicates lowest agreement).



A heat map of question responses (% agreement) on the perceived impact of LLMs on functional roles, overall and by individual demographic groups (Dark Green indicates strongest agreement and dark red indicates lowest agreement).



Qualitative themes (inner circle) and sub-themes (outer circle) on the perceived impact of LLMs on human interactions from the free text survey data.



CONSORT (or other) checklists

CROSS checklist.

URL: <http://asset.jmir.pub/assets/31947538ef2414286b6d830f62d23d57.pdf>

