

Use of large language models in German general practice: A cross-sectional survey

Hannah Maria Ott

Otth@students.uni-marburg.de

Philipps-Universität Marburg Institute of Family Medicine

Stefan Bösner

Philipps-Universität Marburg Institute of Family Medicine

Research Article

Keywords: Large language models, artificial Intelligence, general practice, primary care, family medicine

Posted Date: October 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10995850/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Large language models (LLMs) are increasingly being used in healthcare, but quantitative evidence from German general practice remains scarce. This study examined the current use of LLMs among general practitioners working in Germany, barriers to their use, their perceived impact on the physician-patient relationship and expectations regarding their future role in general practice.

Methods

A cross-sectional online survey was conducted from 5 February to 5 June 2026. Complete questionnaires were analyzed using descriptive statistics. Associations were explored using Spearman's rank correlation and the Kruskal-Wallis test.

Results

A total of 206 completed questionnaires were analyzed. Approximately 80% reported professional LLM use, with ChatGPT being the most frequently used tool (65%). LLMs were used mainly for knowledge- and decision-support tasks, whereas administrative applications were less common. More frequent use was associated with younger age ($\rho = -0.302$), fewer years of professional experience ($\rho = -0.336$) and higher self-assessed knowledge ($\rho = 0.350$), while use frequency did not differ by gender (Kruskal-Wallis $\chi^2 = 0.0117$). The main barriers included fear of inaccurate information, concerns about data protection and uncertainty regarding the legal framework and liability. 71% had never attended training on LLMs, although 64% expressed willingness to do so in the future. LLMs were used during patient encounters by 41% of respondents, including 26% not disclosing their use to patients. Most respondents expected LLMs to become an integral part of primary care and anticipated increased future use. There was strong agreement that LLMs could not replace GPs.

Conclusions

LLMs were already used among the surveyed GPs, although their use remained selective and focused mainly on knowledge- and decision-support tasks. The expected expansion of LLM use, the identified barriers and limited participation in training, highlight the need for secure applications, clear legal and liability-related guidelines, as well as practice-orientated education to support a safe integration of LLMs into general practice.

Background

Since the release of ChatGPT in November 2022, generative artificial intelligence (AI) has spread at an unprecedented rate. In only two months ChatGPT already reached over 100 million users [1]. Generative AI describes techniques that can create new content, including texts or images [2]. Large language models (LLMs) are a type of generative AI trained on large amounts of text data to generate natural language [3]. This enables LLMs to respond to text-based input without being specifically trained for each individual task [4]. Popular examples include models from the generative pre-trained transformer (GPT) family, which are used in ChatGPT [2].

Beyond the widespread public use of applications such as ChatGPT [1], LLMs are increasingly being applied in healthcare [5]. Potential applications include diagnostic support, assistance in clinical decision making and treatment related tasks [6]. LLMs may also reduce the burden of documentation by producing more standardized reports and converting notes into structured formats [7]. These functions could contribute to more efficient workflows [6]. However, a major limitation of LLMs is their tendency to generate seemingly convincing yet factually incorrect information, a phenomenon known as hallucination [8]. Reliance on LLMs may also lead to automation bias, whereby users place excessive trust in AI generated outputs compared to other sources [9]. Entering sensitive health information into LLMs further raises privacy and ethical concerns, particularly given the limited transparency surrounding how commercial providers use and store submitted data [10]. Thus, LLMs offer considerable potential for healthcare while simultaneously creating new challenges for a safe and responsible integration.

These opportunities and challenges are especially relevant in primary care. General practitioners (GPs) often serve as the first point of medical contact and therefore assess patients early in the course of illness, when symptoms are often non-specific and associated with several diseases [11]. Consequently, initial decisions have to be made based on limited information [11]. In addition, care in general practice often extends the immediate reason for consultation, requiring the simultaneous management of multiple problems. These problems are considered not only from a physical perspective but also in relation to psychological, social and other contextual dimensions [11].

Given the broad spectrum of general practice, LLMs could be relevant across a wide range of tasks. International research has therefore begun to examine whether and how these tools are already being used in routine primary care.

Trends in the use of generative AI have been documented in a series of surveys among GPs in the United Kingdom (UK). The proportion of GPs reporting professional use increased from 20% in 2024 to 25% in 2025 and 52% in 2026 [12–14]. Common applications included post-appointment documentation, generating differential diagnosis and treatment options [12, 13]. Among users, approximately 70% reported an overall reduction in workload [14]. Despite these perceived benefits, a separate UK study identified concerns among GPs regarding medico-legal issues, regulatory oversight, safety and data protection [15]. Generative AI may also affect the physician-patient relationship by transforming the traditional dyadic interaction into a triadic relationship in which LLMs used by patients are incorporated as a third party [16]. Previous research also addressed how GPs learn about AI and whether they receive

relevant training. In a survey conducted in the United States (US), about 70% of the GPs reported acquiring their knowledge about AI passively [17]. Furthermore, only 5% of the UK GPs surveyed had received training on the use of generative AI tools in the past year [18]. Furthermore, Blease et al. [14] stated in their 2026 survey, that 46% of GPs viewed generative AI as a collaborative technology to be used alongside physicians, while 69% believed it to play at least a moderate role in routine care within five years.

Evidence from Germany remains limited. Qualitative and mixed-methods studies, some focusing exclusively on GPs and others also including physicians from different specialties, have identified increased efficiency, improved diagnostic quality and support in treatment decisions as potential benefits of AI. Concerns included the fear of being replaced, liability issues, lack of transparency, data protection and changes of the physician-patient relationship [19–21]. The only cross-sectional survey focusing solely on German GPs was conducted by Werner et al. [22]. Among 154 GPs in Saxony, 46% reported using AI professionally, particularly for therapy and speech recognition or transcription, such as dictation aids and translation programs. Lack of knowledge about suitable applications was the main barrier.

However, the described studies all addressed AI broadly rather than specifically focusing on generative AI or LLMs.

Consequently, there is very limited data available about the use of LLMs by GPs in Germany. This study aims to address that gap by examining how frequently GPs use LLMs in routine practice and which models they apply. It further assesses the areas of application and the perceived helpfulness within these areas. In particular, barriers to use will be identified to develop strategies for better integration of LLMs. The study will also examine GPs' level of knowledge and training regarding LLMs, as well as the effects on the physician-patient relationship and the future role of LLMs in general practice.

Methods

Study design and setting

We conducted a quantitative cross-sectional online survey to examine the current use and perceptions of LLM-based tools in German general practice. Data were collected from 5 February to 5 June 2026 via LimeSurvey (Version 6.6.6 + 230904; LimeSurvey GmbH). The questionnaire was implemented as an open web-based survey accessible via a web link or QR code. Participation was voluntary and could be discontinued at any time. No incentives were offered. The survey was anonymous, since no identifying information was collected. Consequently, responses could not be linked to individual participants or practices.

The questionnaire encompassed 23 questions organized into several thematic areas. It was designed to be completed in about 10 minutes. It predominantly consisted of closed-ended single-response, multiple-response and matrix questions. Additional free-text fields allowed the participants to mention

tools and paid subscriptions not covered by the predefined response options. All questions were configured as mandatory in LimeSurvey. No adaptive or conditional questioning was used. Where appropriate, response options such as “no LLM use” were provided, allowing physicians who did not use LLMs to complete the questionnaire.

At the beginning of the survey, participants were provided with a brief definition of LLMs to establish a common understanding of the term. For simplicity, “LLMs” was used throughout the questionnaire as an umbrella term encompassing both underlying language models and user-facing LLM-based application.

The reporting of the study was guided by STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement for cross-sectional studies and the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) guideline for web-based surveys [23, 24].

Questionnaire development

A structured online questionnaire was developed specifically for this study. The selection of thematic domains was informed by German studies [19–22] and selected international studies [12, 17, 18].

A preliminary version of the questionnaire was presented and discussed at the monthly scientific colloquium of the Department of Family Medicine Marburg. Based on the feedback received, questions and response options were revised. Before the final launch on LimeSurvey, the questionnaire was pretested for usability and technical functionality.

Study population and recruitment

Eligible participants were physicians working in general practice, including specialists in general practice, specialists in internal medicine working in general practice and general practice trainees. These groups are collectively referred to as general practitioners (GPs) in this study. Physicians from other specialties and clinical physicians without primary care function were excluded. The eligible types of practices were solo practices, group practices and medical care centers (MVZs).

Primary recruitment was conducted by email using a mailing list maintained by the Department of Family Medicine at Philipps University Marburg. The list included general practitioners and general practice trainees in the German federal state of Hesse. 434 solo practices and 434 group practices were contacted, comprising 1678 physicians in total. The invitation provided information about the study aim, eligibility criteria, estimated completion time and the voluntary and anonymous nature of the participation. It further contained a link to the questionnaire and detailed participant information.

Additional recruitment took place during advanced training courses for teaching physicians of the Department of Family Medicine in Marburg, which were attended by approximately 170 individuals. The questionnaire link or QR code was also distributed through training courses of the Kompetenzzentrum Weiterbildung Hessen with around 180 general practice trainees attending. The study was further advertised via the podcast “Allgemeinmedizin konkret- spannende Fälle aus der Sprechstunde”. On 30

March 2026, an invitation to participate was also published in the newsletter of the association of young family doctors Germany (Junge Allgemeinmedizin Deutschland). As the podcast and the newsletter were accessible beyond Hesse, physicians from other German regions were able to participate.

Statistical analysis

Given the exploratory design of the study, no power-based sample size calculation was performed, since no primary hypothesis test was planned. To guide recruitment, a precision-based sample size estimate was calculated for the proportion of GPs using LLMs in routine practice. Assuming a conservative proportion of 50%, a 95% confidence level and a margin of error of $\pm 7\%$, a sample size of approximately 200 participants was required.

Survey data were exported from LimeSurvey to Microsoft Excel (Version 16.110.1; Microsoft Corp) and reviewed for completeness. Both descriptive and exploratory statistical analyses were performed using jamovi (Version 2.7.31.0; The jamovi project). Spearman's rho was used to examine correlations between ordinal variables and between ordinal and continuous variables. Differences in ordinal outcomes across independent groups were assessed using Kruskal-Wallis test. Statistical significance was defined as $p < 0,05$.

Use of AI-assisted tools

ChatGPT and DeepL were used to improve the language, clarity and concision of the manuscript. The authors carefully reviewed and revised all AI-assisted output and assume full responsibility for the final content of the publication.

Results

Study population

A total of 287 GPs participated in the study. Of these, 206 (72%) completed the questionnaire and were included in the final analysis. The remaining 81 incomplete questionnaires were excluded. The included participants had a mean age of 45.2 years (median 42; SD 11.4) and a mean professional experience in general practice of 10.6 years (median 6; SD 10.4). Further participant and practice characteristics are summarized in Table 1.

Table 1
Participant and practice characteristics (N = 206)

Gender	n	%
Female	118	57
Male	88	43
Professional status		
Specialist in general practice	113	55
GP trainee	66	32
Specialist in internal medicine	27	13
Practice type		
Group practice	127	62
Solo practice	53	26
Medical care center	26	13
Practice location (population)		
< 2000	16	8
2000–5000	29	14
5 000–20 000	58	28
20 000–100 000	47	23
> 100 000	56	27
Patients seen during a 5-hour consultation period		
< 15	9	4
15–19	34	17
20–24	80	39
25–29	41	20
≥ 30	42	20

Frequency of LLM use

Overall, 80% (165/200) reported using LLMs professionally, whereas 20% (41/206) indicated not using them in this context. The most common response regarding the frequency of use in routine practice was “several times a month” (24%; 49/206), followed by “several times a week” (21%; 43/206). The frequency

of LLM use showed statistically significant negative correlations with both age (Spearman's $\rho = -0.302$; $p < 0.001$) and the duration of professional experience in primary care (Spearman's $\rho = -0.336$; $p < 0.001$), with older and more experienced respondents using LLMs less frequently. No significant difference in the frequency of use was observed between gender groups (Kruskal-Wallis test: $\chi^2 = 0.0117$; $p = 0.914$). Furthermore, a weak association was found between the population size of the practice location and the frequency of LLM use (Spearman's $\rho = 0.230$; $p < 0.001$).

LLMs used

ChatGPT was the most frequently used LLM-based tool (65%; 133/206), whereas Claude was used the least (2%; 5/206). In addition, 32% (66/206) of participants named other tools in the free-text field. Within this subgroup, OpenEvidence was mentioned most frequently (67%; 44/66), followed by Perplexity (23%; 15/66), Amboss AI (17%; 11/66) and Deximed AI (9%; 6/66). Most respondents (80%; 164/206) stated that they did not have paid access to an LLM to use in routine practice. Of the paid accounts mentioned, ChatGPT was the most common one (12%; 25/206).

Areas of application and perceived helpfulness

Participants reported how frequently they used LLMs across different tasks in general practice (Fig. 1). "Never" was the most frequently selected response for most applications, particularly for drafting referrals (90%; 186/206), documenting after patient encounters (76%; 157/206), photo analysis (75%; 155/206) and summarizing existing patient data (72%; 149/206). In contrast, non-use was lowest for seeking information on unfamiliar conditions (26%; 53/206), deepening medical knowledge (31%; 63/206), support with differential diagnoses (32%; 66/206) and generating treatment recommendations (33%; 67/206).

Participants rated how frequently they used LLMs for each task on a five-point scale ranging from "never" to "very often".

Participants were also asked to rate the helpfulness of LLMs in each application areas. Despite their generally limited use, "very helpful" was the most common selected response for support with differential diagnoses (36%; 75/206) and seeking information on unfamiliar conditions (36%; 75/206).

Overall, most respondents felt that LLMs made their daily work easier, with 66% (135/206) selecting "somewhat" or "very much".

Usage in patient contact

When asked about the use of LLMs during patient encounters, 30% (62/206) of participants reported never using LLMs during contact with patients. A further 29% (59/206) used LLMs only outside of the consultation period. 41% used LLMs in presence of patients, either without mentioning their use (26%; 53/206) or with their use openly disclosed to the patients (16%; 32/206).

Physician-patient relationship

Most respondents disagreed that patients would be less likely to consult their GP after already receiving help from LLMs. Around half expected LLM-generated preliminary information to induce more frequent discussions and believed that patients would respond positively to openly communicated LLM use. Respondents generally believed physicians' expertise, as well as empathy to remain central (Fig. 2).

Responses were recorded on a five-point scale from "strongly disagree" to "strongly agree".

Barriers

Fear of inaccurate information was the most prominent barrier. Concerns about data protection (42%; 86/206) and uncertainty regarding legal framework and liability (41%; 85/206) also received comparatively high ratings. By contrast, impairment of trust between physician and patient, impact on physicians' professional identity, lack of personal confidence in using LLMs, ethical concerns and inadequate technical equipment were perceived as less important obstacles (Fig. 3).

Participants rated the strength of the inhibiting effect of each factor on LLM use in their practice using a five-point scale ranging from "not strong at all" to "very strong".

Training and knowledge

The majority of participants had not received any training on LLMs (71%; 146/206). Only 18% (38/206) reported having attended a training session once. Nevertheless, 64% expressed an interest in future training focused on the practical application of LLMs into routine practice, with 48% (99/206) somewhat agreeing and 16% (33/206) strongly agreeing that they would participate.

Information about the potential applications of LLMs was primarily acquired passively, for example through conversations or news reports, rather than through active information-seeking.

Overall, 45% (93/206) described their knowledge of LLMs in medicine as somewhat or very low, compared to 35% (72/206) who rated it as somewhat or very good. Self-assessed knowledge was positively correlated with the frequency of LLM use (Spearman's $\rho = 0.350$; $p < 0.001$), indicating that greater self-assessed knowledge was associated with more frequent use in routine practice.

Verification practices

32% (65/206) reported verifying AI-generated responses "most of the time" and 26% (53/206) "always", whereas 2% (4/206) stated that they "never" verified them.

Future perspectives

The majority expected LLMs to become an integral part of primary care in the coming years and anticipated using them more frequently in their own practice in the future. The current frequency of use

was positively correlated with both the expectation of an increasing future use of LLMs (Spearman's $Rho = 0.572$, $p < 0.001$) and the assumption that they will become an integral part of general practice (Spearman's $Rho = 0.538$, $p < 0.001$).

Most participants rejected the idea that LLMs would diminish the importance of GP expertise. There was also strong agreement that LLMs could not replace general practitioners.

Discussion

Principal findings

The study found that approximately 80% of surveyed GPs used LLMs professionally. This proportion was higher than in repeated cross-sectional surveys of UK GPs, in which professional use of generative AI increased from 20% in 2024 to 25% in 2025 and to 52% in 2026 [12–14]. While the UK surveys were based on independent samples, their use of comparable core questions suggests that professional use has increased as generative AI became more widely known and established. The later timing of the present survey, conducted between February and June 2026, may thus partly explain its higher rate, although its proximity to the January 2026 survey by Blease et al. also points to methodological differences. In particular, the present study classified any response other than “never” on a five-point ordinal scale as use, thereby including very infrequent use, whereas Blease et al. applied a binary response format. Werner et al. [22] likewise reported a lower rate of AI use among GPs in Saxony (46%) between 2024 and 2025. However, comparisons are limited by the earlier data collection, broader definition of AI and different response scale.

Among the LLM-based tools included in the survey, ChatGPT was by far the most common reported one, consistent with findings among UK GPs [13].

More frequent LLM use was associated with younger age, fewer years of professional experience and higher self-rated knowledge, whereas use frequency did not differ by gender. Practice location showed a weak association, with somewhat greater use in more populous areas. The cross-sectional design precludes determining whether greater knowledge promotes LLM use or frequent use leads physicians to assess their own knowledge more favorably. Blease et al. [13] similarly found an association with younger age but not gender, while Werner et al. [22] reported lower AI use in rural practices. Although methodological differences limit direct comparison of effect sizes, the findings point in similar directions.

LLM use was concentrated primarily on knowledge- and decision-support tasks, including seeking information on unfamiliar conditions, deepening medical knowledge, support with differential diagnoses and generating treatment recommendations. By contrast, previous UK surveys found a stronger focus on administrative and documentation-related tasks [13, 14]. These different patterns should be interpreted cautiously because the application areas assessed were not identical and Blease et al. reported percentages among users, whereas those in the present study were based on the total sample.

Perceived helpfulness mirrored reported use, with knowledge- and decision-support tasks being both used more frequently and rated as most helpful.

Despite selective use in many areas, the majority of respondents perceived LLMs as making their daily work easier. This is consistent with Blease et al. [14], who found that 70% of generative AI users reported a reduction in workload.

Beyond overall patterns of use, the integration of LLMs into physician-patient interactions warrants particular attention. Most participants did not use LLMs in the presence of patients. Among those who did, more reported an undisclosed use (26%) than a disclosed use (16%). This pattern cannot solely be explained by anticipated negative reactions, as half of the participants expected patients to respond positively to openly communicated use of LLMs. These findings therefore point to a discrepancy between expected patient acceptance and current disclosure practice. Patients themselves may place considerable importance on transparency. In a US survey, 63% wanted to be informed about the use of AI in their health care [25]. Despite its broader definition of AI, the survey underlines the relevance of transparency from the patient perspective.

Patient use of LLMs may also alter the content and dynamics of consultations. Around half of the participants agreed that LLM-generated information brought to consultations led to more frequent discussions about diagnoses and treatment recommendations. Rather than necessarily being problematic, such discussions may provide opportunities for shared decision-making, while requiring GPs to evaluate and contextualize information patients obtain through LLMs. This aligns with the shift described by Xie et al. [16] from a traditional dyadic physician-patient interaction to a triadic model involving the patient, physician and LLMs used by patients. The more frequent discussions could thus reflect this additional mediation between three rather than two positions.

Most respondents did not expect patient use of LLMs to reduce GP visit. US patient data are consistent with this finding. 80% of the adults surveyed reported seeking health information from healthcare professionals in 2025, while only 32% used AI tools for this purpose. Among those who had consulted AI for physical symptoms, 58% still went to see a physician [26]. AI may therefore serve as an initial consultation tool for physical concerns, supplementing rather than replacing professional care.

The expectation that LLM use would not reduce GP visits may reflect the perceived importance of medical expertise, as most respondents assumed patients would trust physicians more than LLM-generated information. Likewise, an experimental study with 1,183 participants from German-speaking countries found that trust in treatment was highest for a human doctor and lowest for an AI system acting independently [27].

Furthermore, the majority believed that LLMs could not replace empathy in medical care, suggesting that empathy continued to be viewed as a human component.

Perceived barriers centered on reliability and safety, particularly the fear of inaccurate information, data protection and uncertainty regarding the legal framework and liability. Impairment of trust between physician and patient was the least frequently reported barrier. A survey of 2,108 UK GPs found a similar pattern, with medico-legal, regulatory, safety and data protection concerns predominating, while effects on the doctor-patient relationship ranked second to last among 12 concerns examined [15].

Participants more often rated their own knowledge of LLMs as somewhat or very low (45%) than as somewhat or very good (35%). Training participation was also limited, since 71% had never attended training, although 64% expressed willingness to do so in the future. Low participation to date may therefore not simply reflect lack of interest. Werner et al. similarly found that 56% of respondents reported a need for further training on AI use in general practice [22]. These findings highlight the relevance of developing accessible, practice-oriented training for GPs.

Beyond training, the way physicians obtain information may also influence their knowledge of LLMs. Participants reported acquiring information predominantly passively, for example through conversations or news reports. This is consistent with a US survey of 47 primary care physicians, in which 70% reported learning about AI in medicine passively, whereas 26% actively pursued education for example through lectures and professional literature [17].

Finally, most respondents expected LLMs to become an integral part of primary care in the coming years and anticipated using them more frequently in their own practices. This aligns with the increase in professional use among UK GPs observed by Blease et al. [12–14] between 2024 and 2026, as well as with the finding that 69% expected generative AI to play at least a moderate role in routine care within the next five years [14]. These findings therefore underline that the role of generative AI in primary care may continue to expand.

Associations between current use and expectations of greater future use and integration may indicate that practical experience with LLMs shapes perceptions of their future role. Conversely, physicians who anticipate a growing role for LLMs could be more inclined to use them. However, the cross-sectional design does not allow the direction of these associations to be determined.

Despite anticipating broader use, most participants rejected the idea that LLMs would diminish the importance of GP expertise or replace GPs. Similarly, Blease et al. [14] found that 46% viewed generative AI as a collaborative technology used alongside physicians, whereas only 10% regarded it as potential competition. 59% expected AI to reshape the GP role while leaving GPs central to care [14]. Overall, these findings suggest that GPs expect a transformation of their role through AI rather than their displacement.

Strengths and limitations

This study provides empirical insights into the use of LLMs in general practice in Germany, a topic on which only limited data are available. The planned sample size was achieved with 206 completed questionnaires. Providing a definition of LLMs at the beginning of the questionnaire helped to ensure a

consistent understanding of the terminology, thereby reducing the risk of misinterpretation. The broad range of aspects covered in the study, further enabled a nuanced examination of the use of LLMs in general practice. Anonymous participation may also have encouraged more candid responses, particularly regarding LLM use in patient interactions.

Several limitations should be considered. Given the rapid evolution of LLMs, the findings should be viewed as a snapshot and may soon no longer reflect current usage patterns. Voluntary participation and the exclusively online survey may have introduced selection bias, as physicians with a particular interest in LLMs, stronger views or greater digital affinity may have been more likely to participate. Moreover, 81 of the 287 individuals did not complete the survey and were excluded, creating a potential risk of non-response bias, as it is unknown whether these 81 individuals differ systematically from the cases included in the analysis. The incomplete questionnaires may be attributed to several factors, including the voluntary nature of participation and the overall length of the questionnaire. Since the federal state in which participants practiced was not recorded., the regional composition of the sample and its representativeness for Germany could not be assessed.

Furthermore, usage and knowledge were self-reported and therefore may be affected by recall and social desirability bias.

Finally, the cross-sectional study design precludes causal conclusions. The exploratory analyses examined associations separately without simultaneous adjustment for potential confounders. It therefore remains unclear whether the associations would persist after accounting for interrelated characteristics such as age, professional experience and perceived knowledge. The findings should thus be regarded as hypothesis-generating evidence and examined further in follow-up studies using longitudinal designs and multivariable analyses.

Conclusions

This study provides timely evidence that LLMs are already being used in general practice, although their integration remains selective. They were used most frequently and perceived as most helpful for knowledge- and decision-support tasks, including seeking information on unfamiliar conditions, deepening medical knowledge, support with differential diagnoses and generating treatment recommendations. Key barriers centered on reliability and safety, including the fear of inaccurate information, concerns about data protection and uncertainty regarding the legal framework and liability. Addressing these concerns will require secure applications, as well as clear legal and liability-related guidelines for medical use. Although few participants had previously received formal training, many expressed willingness to participate in future programs, indicating the need for practice-oriented education. Most participants expected LLMs to become an integral part of general practice and anticipated increasing their own use. Given these expectations, the barriers and training needs identified in this study are particularly relevant to the future implementations of LLMs and help clarify the conditions needed for their safe and meaningful integration into general practice.

Abbreviations

AI
Artificial Intelligence
CHERRIES
Checklist for Reporting Results of Internet E–Surveys
GP
General practitioner
LLM
Large Language Model
SD
Standard deviation
STROBE
Strengthening the Reporting of Observational Studies in Epidemiology
UK
United Kingdom
US
United States

Declarations

Ethics approval and consent to participate

The study was approved by the Ethics Committee of the Faculty of Medicine, Philipps University Marburg, which identified no concerns regarding professional ethics or professional law (reference number 25/316 ANZ). All methods were performed in accordance with the relevant guidelines and regulations, including the Declaration of Helsinki.

Prior to participation, participants were informed about the purpose of the study, the anonymous nature of the survey and that participation was voluntary. Informed consent was implied by voluntarily proceeding with and completing the questionnaire.

Consent for publication

Not applicable

Availability of data and materials

The dataset used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no funding. Open access funding was provided by the Open Access Publishing Fund of Philipps-Universität Marburg.

Authors ' contribution

HO and SB contributed to the conception and design of the study, as well as the development of the questionnaire. HO set up the questionnaire in LimeSurvey, prepared the recruitment materials, monitored data collection, curated and analyzed the data, prepared figures and tables and drafted the manuscript. SB contributed to participant recruitment, supervised the study and critically revised the manuscript. Both authors read and approved the final manuscript.

Acknowledgements

We would like to thank all GPs who participated in the survey for their time and valuable contribution to this study. We also thank Konrad Hierasimowicz for his technical guidance and support with the online survey.

References

1. Mesko B. The ChatGPT (Generative Artificial Intelligence) Revolution Has Made Artificial Intelligence Approachable for Medical Professionals. *J Med Internet Res*. 2023;25:e48392. <https://doi.org/10.2196/48392>.
2. Feuerriegel S, Hartmann J, Janiesch C, Zschech P, Generative AI. *Bus Inf Syst Eng*. 2024;66:111–26. <https://doi.org/10.1007/s12599-023-00834-7>.
3. Lu Z, Peng Y, Cohen T, Ghassemi M, Weng C, Tian S. Large language models in biomedicine and health: current research landscape and future directions. *J Am Med Inf Assoc*. 2024;31:1801–11. <https://doi.org/10.1093/jamia/ocae202>.
4. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med*. 2023;29:1930–40. <https://doi.org/10.1038/s41591-023-02448-8>.
5. Fareed M, Fatima M, Uddin J, Ahmed A, Sattar MA. A systematic review of ethical considerations of large language models in healthcare and medicine. *Front Digit Health*. 2025;7:1653631. <https://doi.org/10.3389/fdgth.2025.1653631>.
6. Ghnemat R, Saleh A. Large language models for clinical artificial intelligence in healthcare a systematic review. *Discov Artif Intell*. 2026;6:156. <https://doi.org/10.1007/s44163-025-00784-x>.
7. Clusmann J, Kolbinger FR, Muti HS, Carrero ZI, Eckardt J-N, Laleh NG, et al. The future landscape of large language models in medicine. *Commun Med*. 2023;3:141. <https://doi.org/10.1038/s43856-023-00370-1>.

8. Xu Z, Jain S, Kankanhalli M. Hallucination is Inevitable: An Innate Limitation of Large Language Models. 2025. <https://doi.org/10.48550/arXiv.2401.11817>
9. Natali C, Marconi L, Dias Duran LD, Cabitza F. AI-induced Deskillling in Medicine: A Mixed-Method Review and Research Agenda for Healthcare and Beyond. *Artif Intell Rev.* 2025;58:356. <https://doi.org/10.1007/s10462-025-11352-1>.
10. Li H, Moon JT, Purkayastha S, Celi LA, Trivedi H, Gichoya JW. Ethics of large language models in medicine and medical research. *Lancet Digit Health.* 2023;5:e333–5. [https://doi.org/10.1016/S2589-7500\(23\)00083-3](https://doi.org/10.1016/S2589-7500(23)00083-3).
11. WONCA Europe. The European Definition of General Practice / Family Medicine: WONCA Europe 2023 Edition. Brussels: WONCA Europe; 2023.
12. Blease CR, Locher C, Gaab J, Hägglund M, Mandl KD. Generative artificial intelligence in primary care: an online survey of UK general practitioners. *BMJ Health Care Inf.* 2024;31:e101102. <https://doi.org/10.1136/bmjhci-2024-101102>.
13. Blease C, Hagström J, Sanchez CG, Kharko A, McMillan B, Gaab J, et al. General practitioners' adoption of generative artificial intelligence in clinical practice in the UK: An updated online survey. *Digit Health.* 2025;11:20552076251394287. <https://doi.org/10.1177/20552076251394287>.
14. Blease C, Sanchez CG, Gaab J, Alderman JE, Allen LN, Kumpunen S. Generative Artificial Intelligence Adoption and the Future of the Primary Care Profession: An Online Survey of UK General Practitioners' Experiences and Opinions. 2026. <https://doi.org/10.21203/rs.3.rs-9979779/v1>
15. Kumpunen S, Lobont C, Garrard L, Fisher D, Lau R, Fallica G, et al. How are GPs using AI? Insights from the front line. London: Nuffield Trust; 2025.
16. Xie W, Liu J, Liu S. Personal Health Large Language Models and the Negotiation of Medical Authority in Clinical Care: Opportunities, Risks, and Governance. *J Med Internet Res.* 2026;28:e91727. <https://doi.org/10.2196/91727>.
17. Allen MR, Webb S, Mandvi A, Frieden M, Tai-Seale M, Kallenberg G. Navigating the doctor-patient-AI relationship - a mixed-methods study of physician attitudes toward artificial intelligence in primary care. *BMC Prim Care.* 2024;25:42. <https://doi.org/10.1186/s12875-024-02282-y>.
18. Kharko A, Garcia Sanchez C, Hagström J, Gaab J, Locher C, McMillan B, et al. General practitioners' opinions of generative artificial intelligence in the UK: An online survey. *Digit Health.* 2025;11:20552076251360863. <https://doi.org/10.1177/20552076251360863>.
19. Buck C, Doctor E, Hennrich J, Jöhnk J, Eymann T. General Practitioners' Attitudes Toward Artificial Intelligence-Enabled Systems: Interview Study. *J Med Internet Res.* 2022;24:e28916. <https://doi.org/10.2196/28916>.
20. Negash S, Gundlack J, Buch C, Apfelbacher T, Schildmann J, Frese T, et al. Physicians' attitudes and acceptance towards artificial intelligence in medical care: a qualitative study in Germany. *Front Digit Health.* 2025;7:1616827. <https://doi.org/10.3389/fdgth.2025.1616827>.
21. Heinrichs H, Kies A, Nagel SK, Kiessling F. Physicians' Attitudes Toward Artificial Intelligence in Medicine: Mixed Methods Survey and Interview Study. *J Med Internet Res.* 2025;27:e74187.

<https://doi.org/10.2196/74187>.

22. Werner A, Fischer S, Bleckwenn M, Schrimpf A. Artificial intelligence in general practice in Germany: an online survey of current use, perceived benefits, barriers, and future needs. *BJGP Open*. 2025;9. <https://doi.org/10.3399/BJGPO.2025.0166>.
23. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. *PLoS Med*. 2007;4:e296. <https://doi.org/10.1371/journal.pmed.0040296>.
24. Eysenbach G. Improving the Quality of Web Surveys: The Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *J Med Internet Res*. 2004;6:e132. <https://doi.org/10.2196/jmir.6.3.e34>.
25. Platt J, Nong P, Carmona G, Kardia S. Public Attitudes Toward Notification of Use of Artificial Intelligence in Health Care. *JAMA Netw Open*. 2024;7:e2450102. <https://doi.org/10.1001/jamanetworkopen.2024.50102>.
26. Montero A, Montalvo IIIJ, Kearney A, Valdes I, Kirzinger A, Hamel L. KFF Tracking Poll on Health Information and Trust: Use of AI For Health Information and Advice. KFF. 2026. <https://www.kff.org/public-opinion/kff-tracking-poll-on-health-information-and-trust-use-of-ai-for-health-information-and-advice/>. Accessed 10 July 2026.
27. Riedl R, Hogeterp SA, Reuter M. Do patients prefer a human doctor, artificial intelligence, or a blend, and is this preference dependent on medical discipline? Empirical evidence and implications for medical practice. *Front Psychol*. 2024;15. <https://doi.org/10.3389/fpsyg.2024.1422177>.

Figures

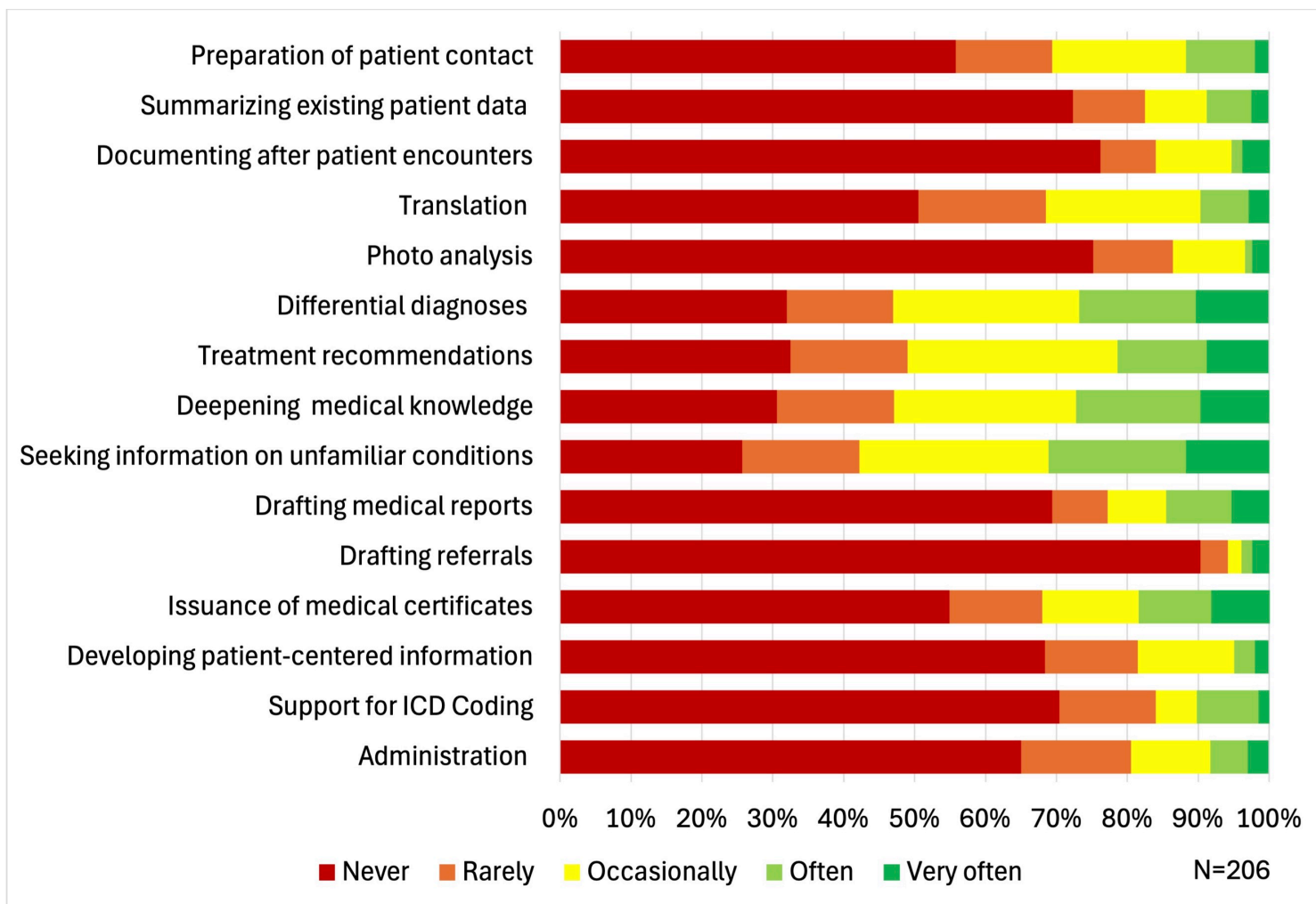


Figure 1

Frequency of LLM use across different tasks in general practice

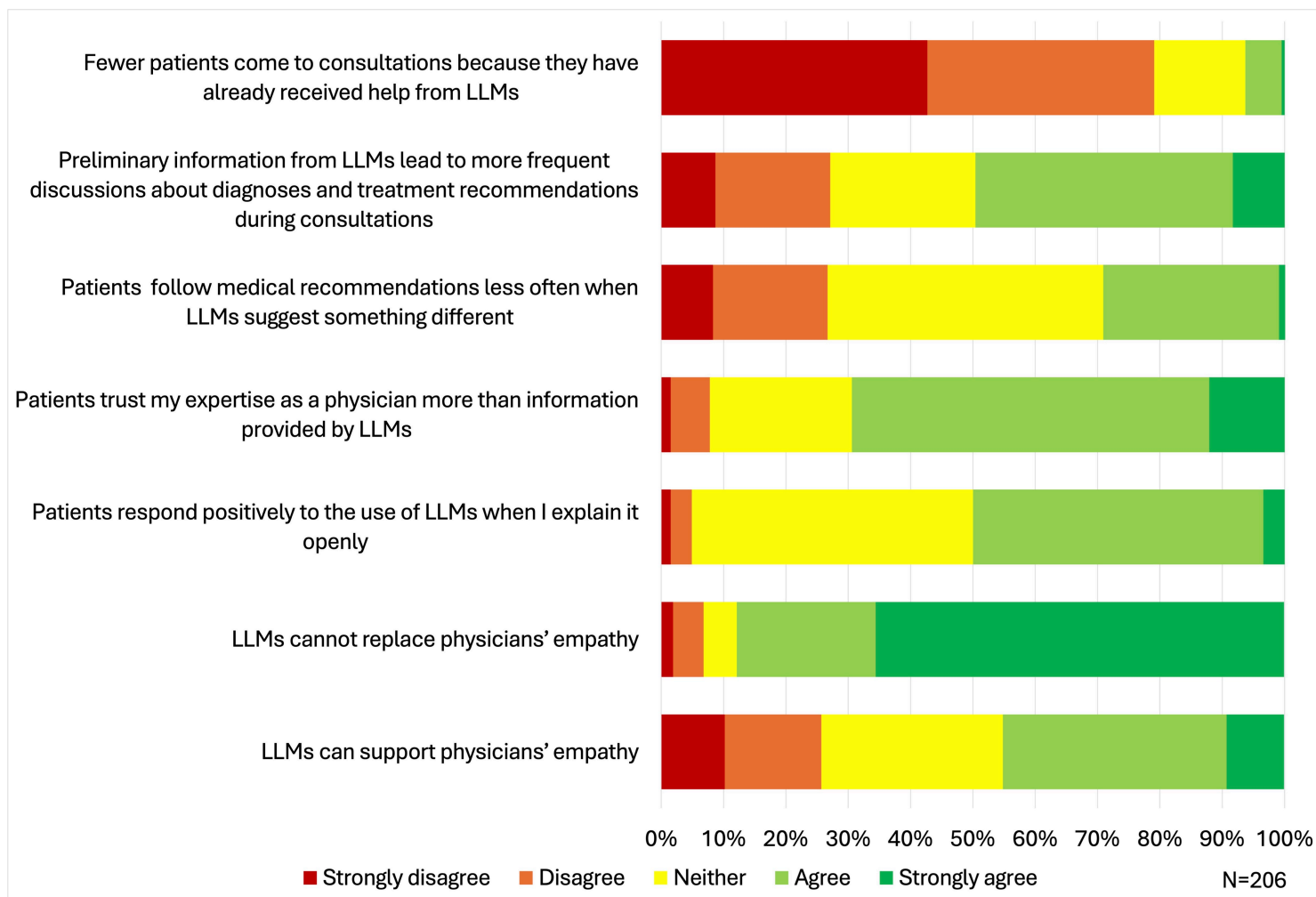


Figure 2

Perceived effects of LLMs on the physician-patient relationship

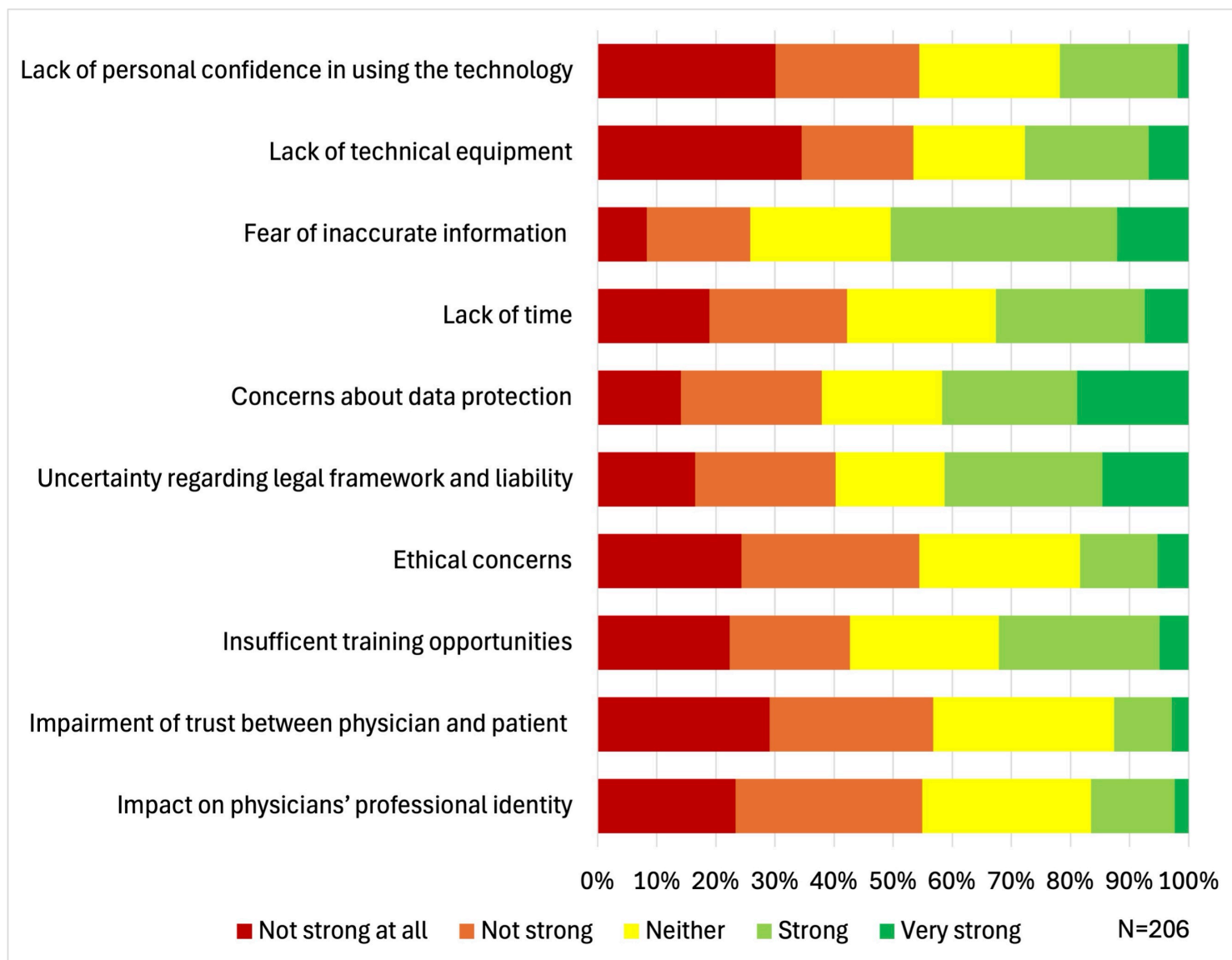


Figure 3

Perceived barriers to LLM use in general practice

Supplementary Files

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

- [OriginalGermanquestionnaire.docx](#)
- [Englishtranslationofquestionnaire.docx](#)