Microsoft AI

How people use Copilot for Health

Beatriz Costa-Gomes* Pavel Tolmachev* Eloise Taysom Viknesh Sounderajah
Hannah Richardson Philipp Schoenegger Xiaoxuan Liu Matthew M Nour
Seth Spielman Samuel F. Way Yash Shah Michael Bhaskar Harsha Nori
Christopher Kelly Peter Hames Bay Gross Mustafa Suleyman Dominic King

Microsoft AI

March 2026

Abstract

We analyze over 500,000 de-identified health-related conversations with Microsoft Copilot from January 2026 to characterize what people ask conversational AI about health. We develop a hierarchical intent taxonomy of 12 primary categories using privacy-preserving LLM-based classification validated against expert human annotation, and apply LLM-driven topic-clustering for prevalent themes within each intent. Using this taxonomy, we characterize the intents and topics behind health queries, identify who these queries are about, and analyze how usage varies by device and time of day. Five findings stand out. First, nearly one in five conversations involve personal symptom assessment or condition discussion, and even the dominant general information category (40%) is concentrated on specific treatments and conditions, suggesting that this is a lower bound on personal health intent. Second, one in seven of these personal health queries concern someone other than the user, such as a child, a parent, a partner, suggesting that conversational AI can be a caregiving tool, not just a personal one. Third, personal queries about symptoms and emotional health queries increase markedly in the evening and nighttime hours, when traditional healthcare is most limited. Fourth, usage diverges sharply by device: mobile concentrates on personal health concerns, while desktop is dominated by professional and academic work. Fifth, a substantial share of queries focuses on navigating healthcare systems such as finding providers, and understanding insurance, highlighting friction in the delivery of existing healthcare. These patterns have direct implications for platform-specific design, safety considerations, and the responsible development of health AI.

1 Introduction

Health is among the most high-stakes domains in which people interact with conversational AI.[†] For many users, conversational AI is becoming increasingly important in their medical journeys (Huo et al. 2025), ranging from a first point of contact when a symptom appears, to medication questions and understanding interactions with healthcare professionals and the health system. The quality of these AI-powered medical interactions is thus central to individual wellbeing (Goldberg et al. 2024), and understanding how these interactions currently play out is crucial for improving health-related conversational AI experiences.

Conversational AI is a stepwise change in how humans interact with information platforms and digital technology. Unlike web search, which returns ranked web pages for a single query, chat interfaces support multi-turn dialogue in which users can ask a question, add context, and correct course, producing responses tailored to their specific situation. In the early web era, Eysenbach and Köhler (2002) documented how "Dr. Google" reshaped patient-provider relationships, a trend later examined by Tan and Goonawardene (2017). Conversational AI represents a step change in that trajectory, and is likely to fundamentally reshape how people approach their health in a digital context (Topol 2019).

*Both BCG and PT contributed equally and have the right to list their name first in their CV. The co-first authorship order was determined via the best of five rounds in UNO.

[†]Throughout this paper, we use "AI" as shorthand for large language model (LLM)-powered conversational AI applications, such as Microsoft Copilot, ChatGPT, and Gemini, unless otherwise specified. We recognize that AI encompasses a much broader set of technologies with diverse applications and risk profile, but these are outside the scope of this paper.

A growing body of research has examined LLM capabilities in health contexts. Large language models perform competitively on medical licensing examinations (Kung et al. 2023; Nori, King, et al. 2023; Nori, Usuyama, et al. 2024) and clinical reasoning benchmarks (Singhal et al. 2023; Nori, Daswani, et al. 2025; Brodeur et al. 2025). However, strong benchmark performance does not always translate to real-world reliability, as chatbots have been shown fail in identifying the severity of health issues in triage settings (Ramaswamy et al. 2026), while users assisted by LLMs sometimes perform no better than controls at identifying conditions and choosing appropriate actions (Bean et al. 2026). Despite these limitations, studies looking at patient-facing applications suggest users find AI-generated health responses comparable to or preferred over physician responses in some settings (Lizée et al. 2024; Ruben, Blanch-Hartigan, and Hall 2025). Crucially, it is often actionable guidance rather than mere information provision (P. Lee, Bubeck, and Petro 2023) that has been empirically shown to drive engagement. Complementary work from Anthropic (McCain et al. 2025) and OpenAI (Chatterji et al. 2025) has documented affective and support-seeking interactions, indicating that conversational AI serves emotional as well as informational needs. Our previous work found that health-related queries on consumer Microsoft Copilot were the most prevalent topic category on mobile, a trend that remained consistent across temporal dimensions (Costa-Gomes et al. 2025).

However, despite this growing body of work, a basic question remains unanswered at scale: what, specifically, are people asking about regarding their health? We know that health is a dominant category of AI usage, but two dimensions remain uncharted: the *intents* behind health queries – the broad purpose of a conversation, such as seeking personalized coaching versus navigating the healthcare system – and the *topics* raised within each intent – the specific subjects within that purpose, such as tailored meal plans or finding a local specialist.

This matters for three reasons. First, building an AI health experience that meets users where they are (on the right device, at the right time, and for the right person) requires knowing what they need. Second, ensuring that AI responds appropriately to different types of health queries depends on understanding the nature of the questions being asked. Copilot is not intended to replace professional medical advice, and it is important to know when to provide information versus when to direct users to professional care. Third, because conversational AI is a new modality for health information, these usage patterns are also likely to evolve. A baseline characterization is therefore essential, both for improving the experience now and for tracking how needs change over time. More broadly, as AI systems increasingly serve health-related needs, systematic characterization of these needs becomes essential for responsible development.

In this paper, we analyze over 500,000 health-related conversations from Microsoft Copilot to provide that baseline. We make three primary contributions: (1) We develop a hierarchical intent taxonomy comprising 12 primary categories and fine-grained topic clusters within each, using a mixed-methods approach validated against human annotation. (2) We apply this taxonomy to characterize health queries at scale, revealing the prevalence of personal health intents — including symptom assessment, condition management, and emotional wellbeing — and showing that a substantial fraction of these serve caregiving rather than individual needs. (3) We analyze how intents vary by device and time of day, showing that context shapes the nature of health engagement with AI.

2 Data & Classification

2.1 Privacy considerations

All Copilot data used in this research was de-identified prior to analysis through a two-stage, privacy-preserving pipeline. In the first stage, raw conversation transcripts pass through an automated scrubbing process that detects and removes personally identifiable information (PII), including names, phone numbers, email and physical addresses, government-issued identifiers such as social security and passport numbers, and financial details such as credit card and bank account numbers. In the second stage, an LLM generates a short, privacy-preserving English-language summary of each conversation that captures the topic and intent without reproducing the user's original words. All subsequent analysis, including our clustering, operates on these summaries rather than on any form of the original text. No human researcher accessed raw conversation content at any point, i.e., the entire pipeline follows an "eyes-off" model in which only machine-based classifiers interact with scrubbed data.

Data elements were limited to coarse, non-identifying attributes and were used solely for aggregate analysis. No attempts were made to re-identify users, infer individual health status, or draw conclusions about specific individuals.

All data processing occurred within Microsoft-controlled systems with access controls and retention limits. Internal privacy review was conducted prior to and throughout the research. Data is retained only for the minimum period necessary to conduct the analysis and validate findings. All results are reported at an aggregate level. Copilot Health evaluation activities are not designed to generate generalizable knowledge, or evaluate clinical hypotheses. As such, these activities are not classified as human subjects research.

2.2 Methodology

This paper extends the methodology of the Copilot Usage Report 2025 (Costa-Gomes et al. 2025). We sampled conversations from January 2026, excluding all enterprise, educational and commercial accounts. The sample is global, with around 22% of conversations originating from the United States and the remainder from across the world; approximately 45% of conversations are in English.

Each conversation is assigned a general topic (what the conversation is about), a general intent (what Copilot is expected to do), and a privacy-preserving English summary by the same pipeline described in the previous paper. From these, we draw on a random sample of over 500,000 conversations that have been classified as "Health and Fitness" as a general topic.

For this subset, we apply a second classifier that assigns a health-specific intent from a 12-category taxonomy (see Table 1) and extracts structured attributes: including who the health query is about, and any symptoms or conditions explicitly mentioned by the user.

The taxonomy was developed prior to this study by in-house clinician scientists, informed by previous observational analyses of user intents. The intents of "Other Health/Fitness" and "Not Health" are used for telemetry purposes and will not be further explored in this paper. To validate the health-intent classifier, clinical scientists independently labeled a sample of conversations that were available for human review, showing agreement between the LLM classifier and human annotators.

To move from individual labels to the thematic categories reported in this paper, we employ an LLM-driven clustering method following TnT-LLM (Wan et al. 2024). For each health intent category, we provide an LLM with batches of conversation summaries and their extracted attributes, and prompt it to group conversations by the user's underlying journey, producing named clusters that describe coherent patterns (e.g., "General Wellness, Food Choice, and Product Comparison", "Strength Training and Fitness Routine Planning", "Finding a Local Healthcare Provider"). These clusters are ranked by prevalence to produce the ordered lists presented in the results. Device and temporal breakdowns are computed as in our previous work (Costa-Gomes et al. 2025), with mobile-to-desktop ratios for each health intent and hourly distributions across the day.

Table 1: Health Intent taxonomy.

Intent	**Description**
Health Information & Education	General information about health and wellness topics
Symptom Questions & Health Concerns	Questions about symptoms, health concerns, and understanding test results
Condition Information & Care Questions	Questions about conditions, medications, and ongoing care
Fitness, Lifestyle & Coaching	Nutrition, fitness, sleep, habit formation
Emotional Wellbeing	Coping strategies, stress management, behavioral health practices
Healthcare Navigation & Access to Care	Finding care, appointments, care coordination
Coverage & Benefits	Insurance, billing, cost, reimbursement

Intent	Description
Research & Academic Support	Literature review, exam preparation, curriculum development
Medical Paperwork	Medical documents, including forms, letters, or notes
Digital Tools & Fitness Apps	Device setup, troubleshooting, data interpretation
Other Health / Fitness Intent	Health-related but uncategorized
Not Health	Primary intent unrelated to health

3 Results

Figure 1 shows the distribution of conversations across health intents. The largest category, "Health Information & Education," accounts for over 40% of conversations. This category captures non-personalized queries, including how a medication works, what causes a condition, and general nutrition information. Its size is consistent with the finding that information seeking remains the dominant mode of health engagement online (Eysenbach and Köhler 2002). However, some queries framed in general terms may reflect underlying personal concerns, and the true share of personal health intents may be higher than the taxonomy suggests (see Section 4.6). We observe from topic clusters (Section 3) that many queries are about specific treatments and conditions rather than general health education, further supporting our theory that users may seek general information as a step towards personal decision-making.

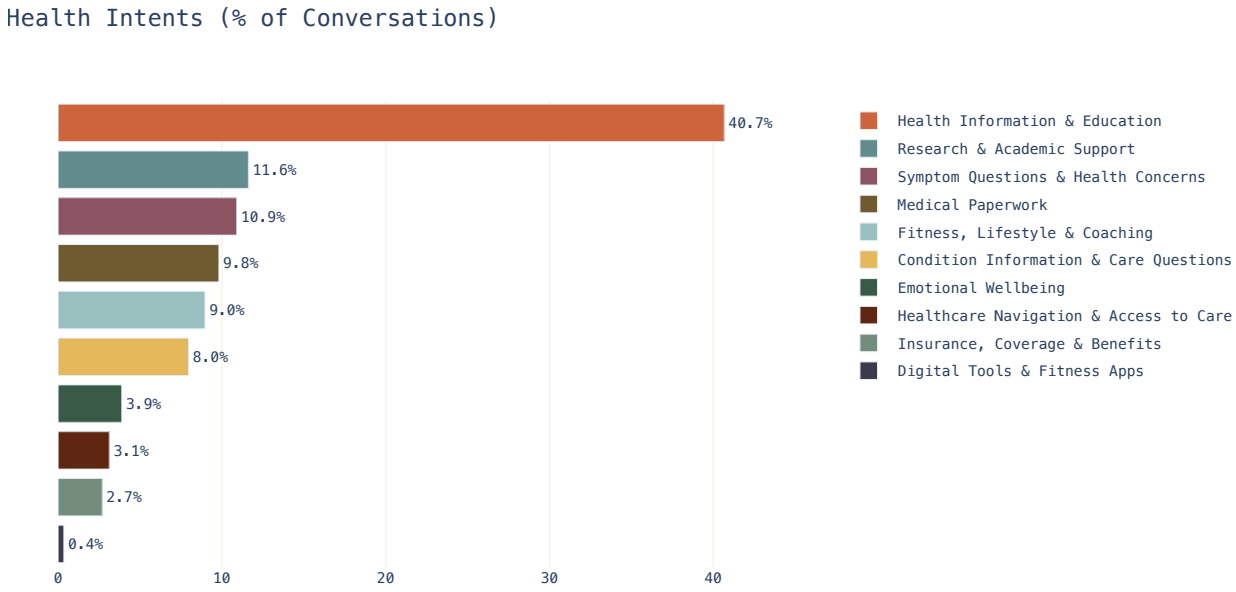

Figure 1: Distribution of health intent usage, in percentage of conversations, for the entire dataset.

Figure 2 shows how the percentage of all conversations on desktop and mobile varies throughout the day, with the former more predominant during the day and the latter at night. This pattern reflects everyday routines: during working and school hours, users have access to desktop devices and may prefer them for longer or more complex tasks, whereas in the evening and at night, when people are away from their desks, the phone becomes the primary device for health queries.

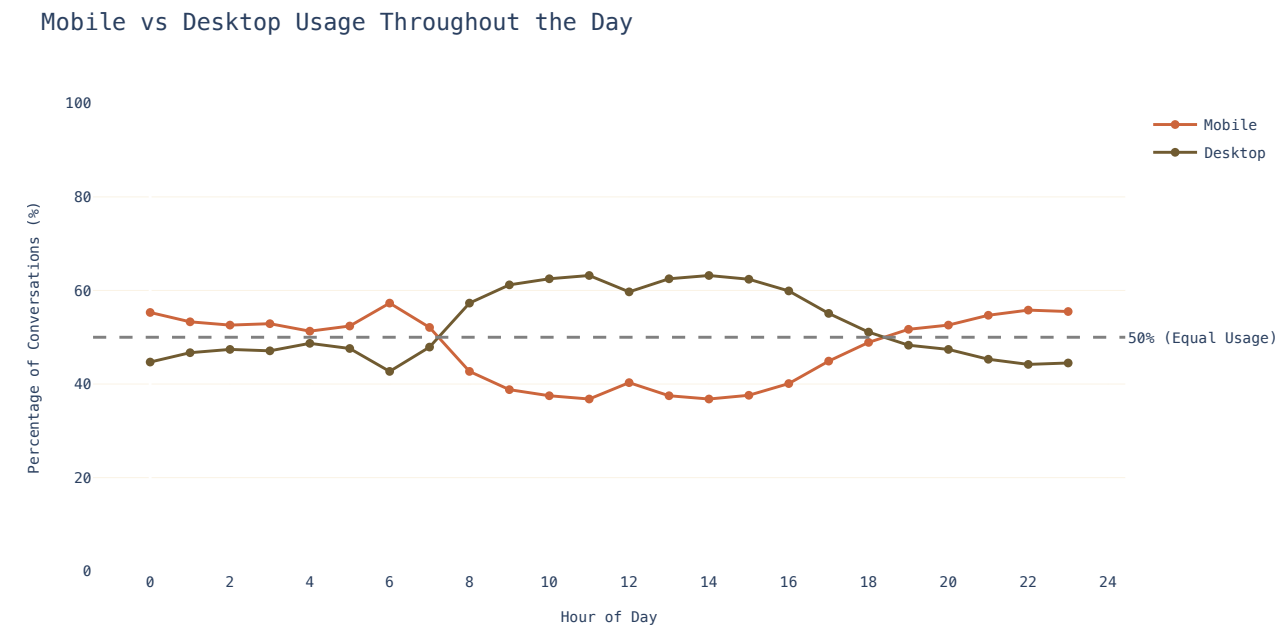

Figure 2: Average percentage of mobile vs desktop health conversations, throughout the day.

Figure 3 compares intent distributions across mobile and desktop. The most striking difference is in academic support and medical paperwork: these intents are among the most common on desktop but fall to the lower end on mobile.

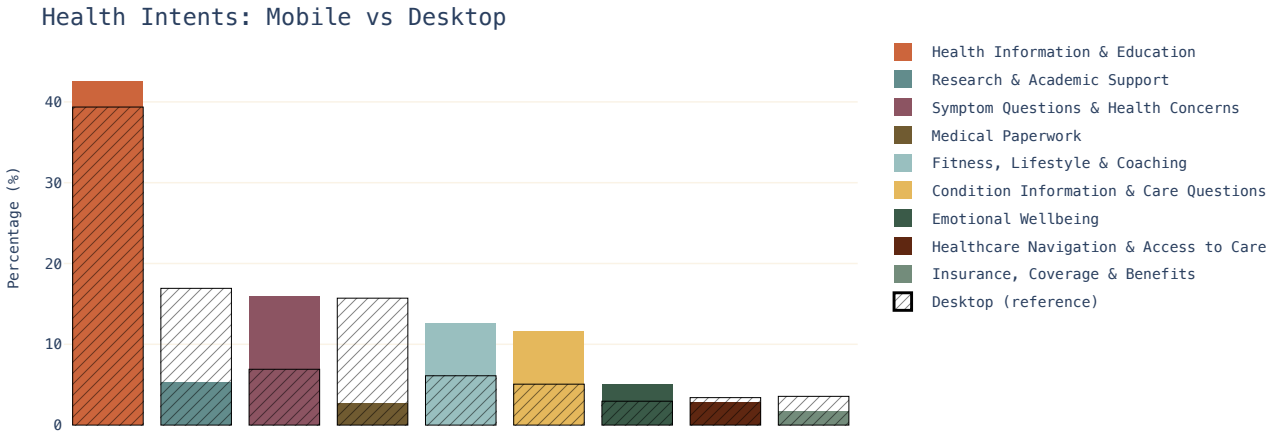

Figure 3: Average percentage of conversations per intent on mobile (block color) vs desktop (striped).

Figures 4 and 5 show the breakdown of intents per hour of the day. Although the predominance of "Health Information & Education" on both platforms is expected given Figure 3, on desktop its share decreases during working hours as "Research & Academic Support" and "Medical Paperwork" rise. This suggests that Copilot usage on desktop is often adjacent to another activity such as thesis writing, research, or processing paperwork, tasks that typically require access to other documents or files alongside the conversation. "Medical Paperwork" peaks during normal working hours, while "Research & Academic Support" rises steadily throughout the day, particularly after work and school hours when researchers and students turn to their own projects. More broadly, the desktop pattern may reflect workflows that depend on multiple windows and reference materials, which are cumbersome to manage on a mobile device.

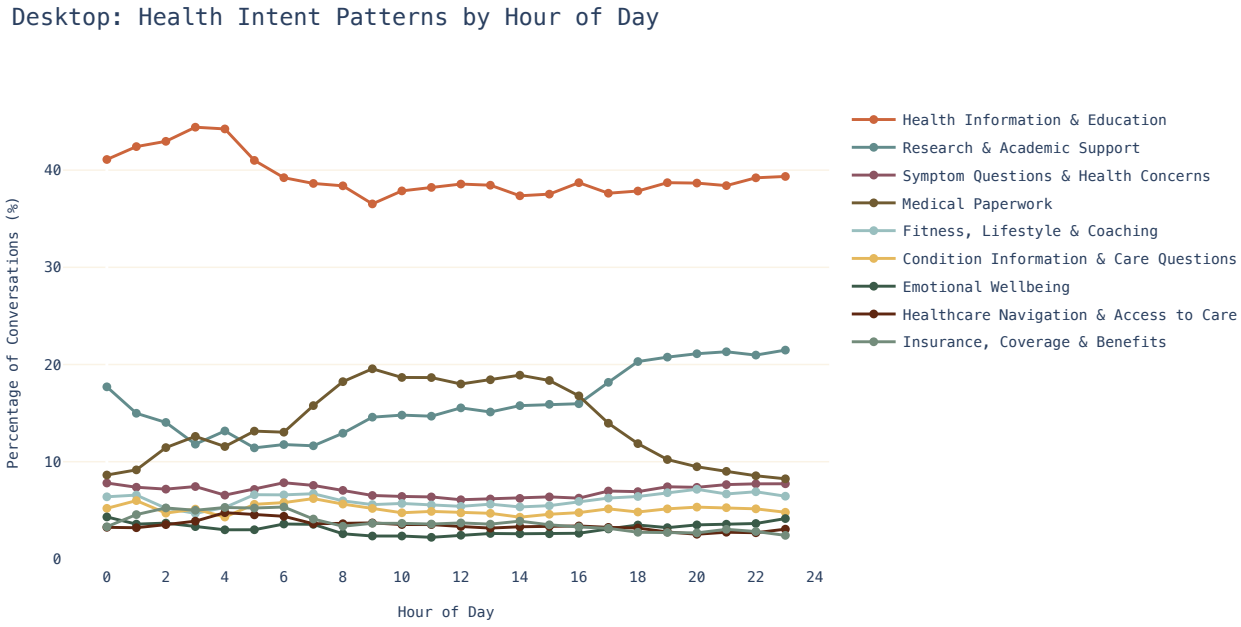

Figure 4: Health intent patterns averaged by hour of day, on desktop.

On mobile, the second most common intent is "Symptom Questions & Health Concerns", followed by queries on conditions and fitness. This is consistent with mobile devices being used primarily for personal health queries rather than work-related tasks.

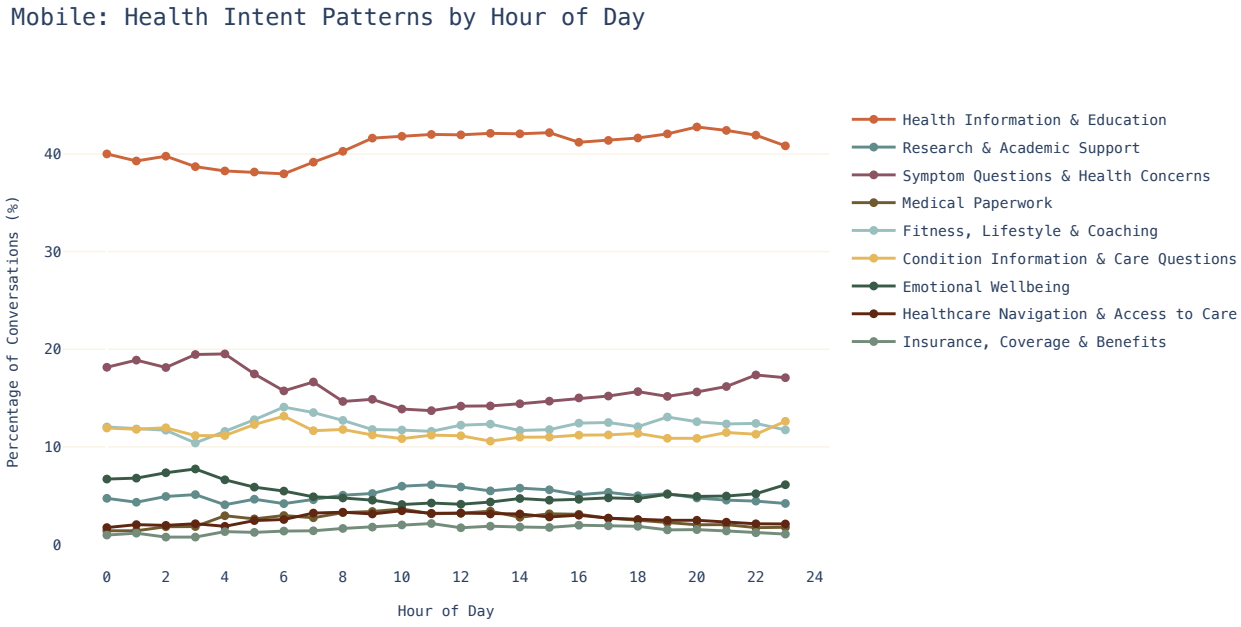

Figure 5: Health intent patterns averaged by hour of day, on mobile.

When looking at the changes throughout the day compared to morning, the distinction between types of intents becomes more evident (Figure 6), with the more personal intents (such as queries about conditions or emotional wellbeing) going up in the evening and at night and the more scholarly ones (such as research or documentation) decreasing.

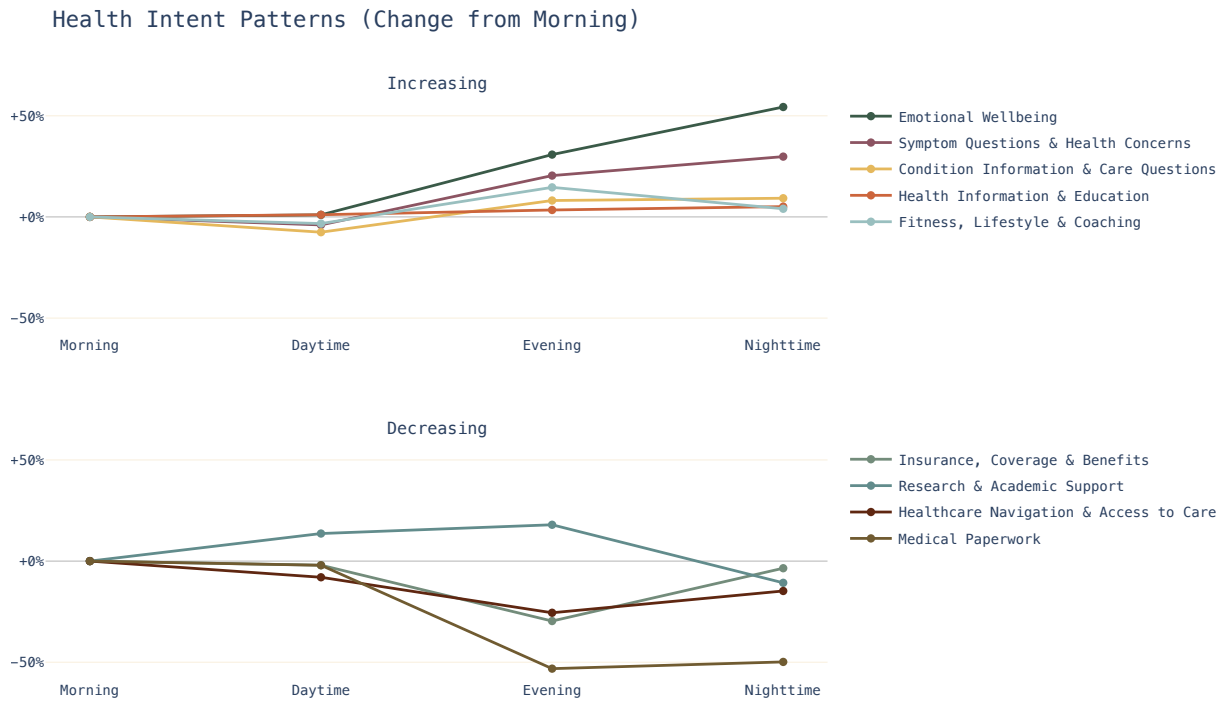

Figure 6: Temporal changes of intent usage, relative to the morning. The top graph shows the intents that increase throughout the day, and the bottom shows the ones that decrease.

We also examined who the health query is about (Figure 7). In every category, most questions are asked on behalf of the users themselves. However, across both condition information and symptom questions, one in seven conversations are on behalf of someone else, whether a child, an aging parent, or a partner.

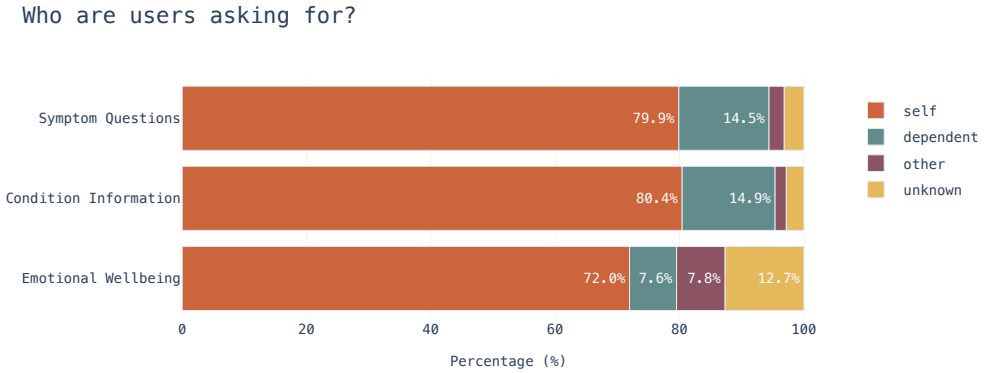

Figure 7: Percentage of conversations on three intents (symptom questions, condition information and emotional wellbeing) related to the user, a dependent, other or unknown.

Table 2 presents the five most common topic clusters for six key intents, with within-category percentages. The clusters reveal that even the broadest category, "Health Information & Education", is dominated by queries about specific treatments and conditions rather than abstract health knowledge, and that the narrower personal intents show clear concentrations around a small number of core needs.

Table 2: Top ranked topics of conversation per intent.

Intent	**Top Topics**	%
Health Information & Education	1. How specific treatments, medications, or medical procedures work	18.8
	2. Causes, symptoms, and risk factors for diseases or conditions	11.9
	3. Advice on healthy eating, supplements, and food safety	9.7
	4. Plain-language explanations of anatomy and biological concepts	8.0
	5. Everyday wellness, prevention, and self-care guidance	6.5
Symptom Questions & Health Concerns	1. Understanding new or unexpected symptoms	32.7
	2. Making sense of recurring or long-term symptoms	14.6
	3. Plain-language explanations of lab or imaging results	12.8
	4. Medication safety, side effects, and interactions	6.8
	5. Infant/child health and development	6.8
Condition Information & Care Questions	1. Questions about medications, supplements, and treatments	18.5
	2. Understanding how conditions develop and change over time	12.2
	3. Caring for minor injuries at home	12.1
	4. Everyday routines and self-care for chronic conditions	10.4
	5. Practical advice for skin and hair issues	7.3
Fitness, Lifestyle & Coaching	1. Tailored meal plans, calorie targets, and nutrition advice for specific goals	26.6
	2. How to start or progress strength and resistance training	13.4
	3. Support for setting and tracking fitness outcomes	5.2
	4. Choosing healthier foods and tracking nutrition	5.1
	5. Cardio, running, or cycling plans to build endurance	4.9
Emotional Wellbeing	1. Understanding personal emotional or behavioral health challenges	30.5
	2. Asking for practical routines to increase resilience and manage stress	13.4
	3. Support for current emotional challenges	12.2
	4. Supporting children, friends, or family members with emotional or behavioral health challenges	11.5
	5. Social, academic, or workplace-related stress	7.3
Healthcare Navigation & Access to Care	1. Finding nearby providers, clinics, or specialists	43.1
	2. Help with medical paperwork and eligibility documentation	11.5
	3. Comparing hospitals, clinics, procedures, and pricing	7.7
	4. Understanding insurance coverage and benefits	6.7
	5. Step-by-step help booking appointments	6.4

4 Discussion

The patterns that emerge from this analysis have implications for the design of health AI and for understanding health needs that existing systems may not be meeting.

4.1 The after-hours pattern

Personal health queries, particularly emotional wellbeing and symptom assessment, go up in the evening and at nighttime hours, precisely when traditional healthcare services are least accessible.

The emotional wellbeing pattern deserves particular attention. The evening increase in emotional health queries is consistent with the diurnal pattern of negative affect documented in population psychology, where negative affect rises throughout the day and peaks in the evening hours (Golder and Macy 2011). This pattern is likely multiply determined: people may also have more time for personal reflection in the evening, and the reduced availability of professional support may itself prompt queries that would otherwise be directed to a clinician. The convergence between our observed intent patterns and an independently established affective rhythm provides suggestive evidence for the construct validity of our classification approach.

4.2 The boundary between general information and personal health queries

Nearly one in five conversations involve users describing their own symptoms, interpreting their own test results, or managing their own conditions. These interactions exist on a spectrum: at one end, a user asking "what does high cholesterol mean" is seeking general education; at the other, a user describing persistent headaches alongside their medication list is seeking information specific to their own circumstances. Understanding the distribution and nature of these queries is a prerequisite for ensuring that conversational AI responses are appropriate and that users are directed to professional care when needed, which is what this taxonomy strives to enable.

4.3 Device as a signal for intent

Usage diverges sharply by device alongside previously shown lines (Costa-Gomes et al. 2025). On mobile, personal health intents are substantially more prevalent, while desktop use is dominated by research, academic support, and medical paperwork. This split suggests that device choice is not merely a matter of convenience but reflects fundamentally different modes of health engagement. This distinction has practical implications for how health AI experiences are designed across platforms, and suggests that device context could inform how responses are prioritized and presented.

4.4 Conversational AI as a health companion

One in seven queries about symptoms and conditions are asked on behalf of someone else, e.g. a child, an aging parent, a partner. This finding reframes how we should think about health AI users. The person typing is not always the person the query is about. This also has design implications: a caregiver asking about an infant's or an elderly relative's symptoms may need different information, different contextual cues, and different follow-up recommendations than someone asking about their own. It has further safety implications: when a user is asking about a dependent, it may be for reasons that affect the accuracy and completeness of the information provided.

4.5 Healthcare navigation as an unmet need

The prevalence of queries including finding providers, understanding insurance, and completing paperwork, reveals that a meaningful fraction of health AI use addresses the complexity of healthcare systems rather than health itself. Users are asking AI to help them do things that should, in principle, be straightforward: find a doctor, book an appointment, understand what their insurance covers. The fact that these queries exist at such volume provides a signal about friction in existing healthcare delivery.

4.6 Limitations

This study has several important limitations. First, our analysis was conducted exclusively on Microsoft Copilot consumer logs, representing a specific user population and platform context. While our findings directly inform design and safety considerations for this platform and similar general-purpose AI assistants, generalization to other platforms, clinical settings, or populations may be limited. Second, we observe queries but not outcomes: we cannot determine whether users subsequently sought clinical care, how they interpreted responses, or whether the information they received improved their health decisions. Third, our sample is drawn from a single month (January 2026), and seasonal effects may influence the distribution of intents. January, in particular, is associated with New Year's health resolutions, which may influence fitness and lifestyle queries relative to other months. Fourth, our taxonomy captures intent as expressed in conversation, not the underlying clinical need. The "Health Information & Education" category accounts for over 40% of conversations, and while its topic clusters suggest meaningful heterogeneity within it, its size may partly reflect the inherent difficulty of distinguishing general from personal information seeking. Conversations do not always contain sufficient context to determine whether a generally framed query (e.g., "what are the side effects of metformin") reflects casual curiosity or a user's own medication concern. This means our classifier defaults to the less-specific educational label in ambiguous cases, and the reported share of personal health intents may represent a lower bound. Future iterations of the taxonomy should explore whether this category can be further subdivided.

4.7 Future directions

This work opens several future directions. Longitudinally, tracking how intent distributions shift as conversational AI matures will reveal whether users discover new applications or converge on established patterns. Geographically, understanding how health AI usage differs across regions and healthcare systems, particularly between settings with strong primary care access and those without, will be essential for responsible global deployment. Methodologically, linking intents to response quality and downstream outcomes would move the field from characterizing what people ask to evaluating whether what they receive helps them. And from a safety perspective, the personal health intents identified here such as symptom assessment, condition management, and emotional wellbeing, arguably define some of the categories where the consequences of conversational AI responses are highest, and where investment in response quality and safety measures should be concentrated.

References

Bean, Andrew M et al. (2026). "Reliability of LLMs as medical assistants for the general public: a randomized preregistered study". In: *Nature Medicine*, pp. 1–7.

Brodeur, Peter G. et al. (2025). *Superhuman performance of a large language model on the reasoning tasks of a physician*. arXiv: 2412.10849 [cs.AI]. URL: https://arxiv.org/abs/2412.10849.

Chatterji, Aaron et al. (Sept. 2025). *How People Use ChatGPT*. Working Paper 34255. National Bureau of Economic Research. DOI: 10.3386/w34255. URL: http://www.nber.org/papers/w34255.

Costa-Gomes, Beatriz et al. (2025). *It's About Time: The Temporal and Modal Dynamics of Copilot Usage*. arXiv: 2512.11879 [cs.CY]. URL: https://arxiv.org/abs/2512.11879.

Eysenbach, Gunther and Christian Köhler (2002). "How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews". In: *BMJ* 324.7337, pp. 573–577. ISSN: 0959-8138. DOI: 10.1136/bmj.324.7337.573. eprint: https://www.bmj.com/content/324/7337/573.full.pdf. URL: https://www.bmj.com/content/324/7337/573.

Goldberg, Carey Beth et al. (2024). *To do no harm—and the most good—with AI in health care*.

Golder, Scott A. and Michael W. Macy (2011). "Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures". In: *Science* 333.6051, pp. 1878–1881. DOI: 10.1126/science.1202775. eprint: https://www.science.org/doi/pdf/10.1126/science.1202775. URL: https://www.science.org/doi/abs/10.1126/science.1202775.

Huo, Bright et al. (2025). "Large language models for chatbot health advice studies: a systematic review". In: *JAMA Network Open* 8.2, e2457879.

Kung, Tiffany H. et al. (Feb. 2023). "Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models". In: *PLOS Digital Health* 2.2, pp. 1–12. DOI: 10.1371/journal.pdig.0000198. URL: https://doi.org/10.1371/journal.pdig.0000198.

Lee, Peter, Sebastien Bubeck, and Joseph Petro (2023). "Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine". In: *New England Journal of Medicine* 388.13, pp. 1233–1239. DOI: 10.1056/NEJMsr2214184. eprint: https://www.nejm.org/doi/pdf/10.1056/NEJMsr2214184. URL: https://www.nejm.org/doi/full/10.1056/NEJMsr2214184.

Lizée, Antoine et al. (2024). "Conversational medical AI: ready for practice". In: *arXiv preprint arXiv:2411.12808*.

McCain, Miles et al. (June 26, 2025). *How People Use Claude for Support, Advice, and Companionship*. URL: https://www.anthropic.com/news/how-people-use-claude-for-support-advice-and-companionship.

Nori, Harsha, Mayank Daswani, et al. (2025). "Sequential Diagnosis with Language Models". In: arXiv: 2506.22405 [cs.CL]. URL: https://arxiv.org/abs/2506.22405.

Nori, Harsha, Nicholas King, et al. (2023). "Capabilities of GPT-4 on Medical Challenge Problems". In: arXiv: 2303.13375 [cs.CL]. URL: https://arxiv.org/abs/2303.13375.

Nori, Harsha, Naoto Usuyama, et al. (2024). "From Medprompt to o1: Exploration of Run-Time Strategies for Medical Challenge Problems and Beyond". In: arXiv: 2411.03590 [cs.CL]. URL: https://arxiv.org/abs/2411.03590.

Ramaswamy, Ashwin et al. (2026). "ChatGPT Health performance in a structured test of triage recommendations". In: *Nature Medicine*.

Ruben, Mollie A, Danielle Blanch-Hartigan, and Judith A Hall (2025). "What is Artificial Intelligence (AI)"Empathy"? A Study Comparing ChatGPT and Physician Responses on an Online Forum: Ruben et al." In: *Journal of General Internal Medicine*, pp. 1–8.

Singhal, Karan et al. (Aug. 2023). "Large language models encode clinical knowledge". In: *Nature* 620.7972, pp. 172–180. ISSN: 1476-4687. DOI: 10.1038/s41586-023-06291-2. URL: https://doi.org/10.1038/s41586-023-06291-2.

Tan, Sharon Swee-Lin and Nadee Goonawardene (Jan. 2017). "Internet Health Information Seeking and the Patient-Physician Relationship: A Systematic Review". In: *J Med Internet Res* 19.1, e9. ISSN: 1438-8871. DOI: 10.2196/jmir.5729. URL: http://www.ncbi.nlm.nih.gov/pubmed/28104579.

Topol, Eric J. (Jan. 2019). "High-performance medicine: the convergence of human and artificial intelligence". In: *Nature Medicine* 25.1, pp. 44–56. ISSN: 1546-170X. DOI: 10.1038/s41591-018-0300-7. URL: https://doi.org/10.1038/s41591-018-0300-7.

Wan, Mengting et al. (2024). "Tnt-llm: Text mining at scale with large language models". In: *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining*, pp. 5836–5847.