Dr. ChatGPT: Revolutionary or Evolutionary?

GSBGEN 390: Independent Study | Fall 2025

Final Report

Alexa Murray

Executive Summary

The United States enters the "ChatGPT era" of health information with a population that is chronically ill, unevenly served, and largely unable to parse the healthcare system's language and mechanics. Long waits for appointments, geographic distance from care, opaque insurance rules, and low health literacy have pushed patients online for decades, first to Google and WebMD, and now to AI chatbots. This report asks: **is consumer use of ChatGPT for health fundamentally different from the WebMD/Google era, or is it simply the latest interface for an old behavior—using whatever tools are available to patch structural gaps in care?**

Drawing on national surveys, comparative studies of diagnostic accuracy, and emerging research on LLM-based health tools, this report arrives at a nuanced answer. On the demand side, ChatGPT does not create a new impulse. People have been "pre-diagnosing" themselves online since the late 1990s, and a stable share of adults report that searching symptoms is their first response to feeling unwell. What is new is the interaction style and the depth of what users can bring into the conversation. Instead of skimming static pages, patients now paste full visit notes, imaging reports, and symptom narratives into a chat window and ask for personalized, plain-language explanations. Usage is especially concentrated among younger adults, urban residents, and people facing language or literacy barriers—the same groups who report the greatest difficulty navigating traditional care.

On the supply side, ChatGPT's strengths map almost one-to-one to structural weaknesses in U.S. healthcare. It offers 24/7 access and infinite follow-up when the average patient waits over a month for an appointment barely over 15 minutes. It can reason (imperfectly) across body systems when patients commonly must see five or more fragmented specialists to get a diagnosis. It can iteratively simplify jargon for users with limited health literacy, and act as a judgment-free space for stigmatized topics. And it can accept longitudinal symptom logs, lifestyle context, and wearable data in free text and surface patterns over time in a way that static EHR notes and WebMD pages do not.

However, the same properties that make ChatGPT attractive also introduce new risks. Large language models hallucinate, lack stable provenance, and operate outside the accountability structures that govern clinicians and regulated medical content. Across controlled studies, LLM diagnostic accuracy tends to be in the range of non-expert physicians and meaningfully below experts, with performance varying widely by use case. Static sites like WebMD may be incomplete and sometimes commercially biased, but they do not invent trial results, and they are governed by formal editorial policies. In practice, this means ChatGPT is "better" than both WebMD and the traditional system on attributes users care about—time, context, and personalization—but worse on the two attributes that matter most in high-stakes decisions: grounding in verifiable evidence and formal responsibility.

The report argues that the long-term impact of ChatGPT will depend less on model IQ and more on product architecture. To retain and deepen health usage, ChatGPT would need to evolve from a stateless chat window into a centralized repository and playground for health information. This would include a user-controlled health data playground with clear visualization, a living, editable health profile that conditions every answer on diagnosis, meds, goals, and constraints, insurance-aware guidance that surfaces realistic, cost-sensitive options, and visible grounding, citations, and guardrails around high-risk actions to build trust. In that configuration, ChatGPT would go beyond answering questions to become the center of how a patient stores, interprets, and acts on their health information.

The central thesis of this report is that ChatGPT health use is currently best understood as a continuation of existing behavior in a new wrapper, but the wrapper matters. LLMs transform self-directed health search to an ongoing, personalized dialogue that can incorporate records, context, and longitudinal data. Whoever successfully couples that new interaction pattern with trustworthy grounding and clear user context including insurance, will become a de facto consumer operating system for health decision-making, sitting in the gap between a strained medical system and a chronically ill public.

Health Baseline: A Sick, Hard-to-Navigate System

The United States is bluntly, quite sick. Recent data show that in 2023, 76.4% of US adults, about 194 million people, reported at least one of twelve common chronic conditions (examples include obesity, hypertension, high cholesterol, heart disease, diabetes), and more than half (51.4%, ~131 million) reported two or more.[1] Chronic illness is no longer a problem of the elderly: prevalence has risen across young, midlife, and older adults.[2]

Access to care is uneven. 20% of Americans do not have a primary care physician.[3] Even for those who are "in the system," timely access is difficult. New patients in major metropolitan areas wait an average of 31 days for an appointment across six common specialties, up 19% since 2022 and nearly 50% since 2004.[4] Geographic barriers compound this: rural Americans live about 10.5 miles, or 17 minutes by car, from the nearest hospital—roughly twice as far as suburban or urban residents, all while rural hospitals continue closing.[5]

Even when people do reach care, health literacy is a major limiting factor. National assessments show that only about 12% of US adults have proficient health literacy, and federal summaries conclude that nearly 90% of adults may lack the skills needed to manage their health and prevent disease.[6] These gaps are especially pronounced among older adults and marginalized groups, and they are directly linked to worse outcomes, higher hospitalization rates, and higher costs.[7]

Taken together, the picture is stark: a chronically ill population, long delays and distance to reach clinicians, and a system whose language and mechanics are unintelligible to most of the people it serves. In that environment, it is unsurprising that patients have looked elsewhere for answers, historically to Google and WebMD, and now, increasingly, to AI tools like ChatGPT. The rest of this report asks whether this new behavior is fundamentally different from the WebMD era, or simply the latest interface for an old pattern: people trying to fill structural gaps in a strained health system with whatever information source they can find.

How People are Using ChatGPT to Access and Understand Health Information

When people feel a twinge, a rash, or a new worry about their body, an increasing share are now typing into a chat window rather than a search bar. One in six U.S. adults (17%) use AI chatbots such as ChatGPT, Gemini, or Copilot at least once a month to find health information or advice.

[1] CDC - Trends in Multiple Chronic Conditions. https://www.cdc.gov/pcd
[2] NAM (National Academy of Medicine) - Health Basics: Chronic Disease. https://nam.edu/
[3] Yahoo/YouGov Poll - Primary Care Physician Access. https://www.yahoo.com/
[4] AMN Healthcare - Survey on Physician Appointment Wait Times. https://ir.amnhealthcare.com
[5] American Hospital Association - Rural Hospital Closures. https://www.aha.org/new
[6] Health.gov - National Action Plan to Improve Health Literacy. https://health.gov/
[7] CHCS (Center for Health Care Strategies) - Health Literacy. https://www.chcs.org/res

Yet only about 29% of adults say they trust these chatbots to provide reliable health information.[8] This combination of high curiosity and wariness is the starting point for understanding how people are actually using "Dr. ChatGPT."

Survey data from Australia offers a window into *what* people ask. Among these users, the most common questions were:

- learning about a specific health condition (48%)
- finding out what their symptoms might mean (37%)
- figuring out what to do next (36%)
- understanding medical terms or jargon (35%).[9]

Critically, 61% of users in that study had asked ChatGPT at least one "higher-risk" question—the kind of decision (changing a medicine, delaying care, starting a new treatment) that would usually warrant professional clinical advice. This high-risk use was especially common among people born in non-English-speaking countries and those who spoke a language other than English at home (both 95%).[10] In other words, some of the people facing the biggest structural barriers to care are those pushing ChatGPT furthest into quasi-clinical roles.

It should be noted that this behavior is not appearing in a vacuum. For decades, patients have been consulting "Dr Google" and WebMD before, after, and sometimes *instead of* seeing a clinician. More than one-third (36%) of US adults said searching their symptoms online is their very first response when feeling unwell, and one-third (33%) said they *always* search their symptoms online before consulting a professional.[11] Older surveys from Pew reached similar conclusions: around 35% of U.S. adults have gone online to try to diagnose a medical condition for themselves or someone they know, a pattern that gave rise to the "Dr Google" moniker. (CBS News) Seen against that backdrop, ChatGPT could be seen less as an entirely new behavior, and possibly the next evolution of a deeply entrenched self-triage habit.

The novelty with ChatGPT is in the *interaction style*. Rather than skimming pages of links, patients can paste in long, messy descriptions of their lives—symptom timelines, lab values, parts of visit notes, even full MRI reports—and get immediate, personalized responses back. In the widely publicized case of a U.S. child with chronic pain, his mother fed the text of his radiology report and clinical history into ChatGPT after three years and 17 different doctors without a diagnosis. The model suggested tethered cord syndrome, which was later confirmed by

[8] KFF - Health Misinformation in Artificial Intelligence. https://www.kff.org/public-opinion/kff-health-misinformation-tracking-poll-artificial-intelligence-and-health-information/
[9] RACGP (Royal Australian College of General Practitioners). https://www.racgp.org.au
[10] RACGP (Royal Australian College of General Practitioners). https://www.racgp.org.au
[11] Pacific Lutheran University - Most Googled Health Questions. https://absn.plu.edu/

a neurosurgeon and treated.[12] Similar stories have made ChatGPT feel to many like a free, always-on differential-diagnosis machine, even if that is not what it was built or validated to be.

Why are people putting this much weight on a tool they don't fully trust? The same structural forces show up again here, but now you can see how they funnel users into ChatGPT:

Cost and access barriers. In 2023, 27% of American adults skipped some form of medical treatment because they couldn't afford it.[13] ChatGPT, while imperfect, is free and always available.

Insurance confusion. 7 in 10 insured adults (72%) are not confident in navigating their health plan.[14] ChatGPT can easily process complex insurance policies and break them down into approachable language.

Health literacy and jargon. 90% of adults prefer doctors who use simpler, jargon-free language.[15] ChatGPT can easily rephrase "acute exacerbation of diastolic heart failure" as "your heart is having trouble relaxing and filling with blood, which is making you short of breath."

Desire for personalization. With traditional online health information, the patient reads the generic information and then applies it to themself. ChatGPT streamlines this process, allowing the patient to provide additional context in plain English. This gives patients far more personalized and nuanced guidance faster than traditional online search.

Time. Many physicians find the rushed, transactional nature of modern appointments draining, contributing to physician burnout. And patients feel unheard, leading to further dissatisfaction with the existing medical system and reluctance to seek out care. ChatGPT is available *at* any time *for* any time. It will answer any question as many times as requested, without judgment or the need to rush out to the next appointment.

How People Historically Used WebMD and Google

To better understand whether ChatGPT's use is fundamentally novel, one must understand the history and evolution of WebMD. While ChatGPT has made a splash, people have been seeking health information online for years. From its inception, WebMD tapped into the public's endless questions about health.

---

[12] Radiology Business - ChatGPT Diagnosis Case Study. https://radiologybusiness.com/topics/artificial-intelligence/after-seeing-17-different-doctors-boy-rare-condition-receives-diagnosis-chatgpt
[13] USA Facts - How Many People Skip Medical Treatment Due to Healthcare Costs https://usafacts.org/articles/how-many-people-skip-medical-treatment-due-to-healthcare-costs/
[14] PAN Foundation - Health Insurance Navigation Poll. https://www.panfoundation.org
[15] PMC (PubMed Central) - Medical Language Study. https://www.ncbi.nlm.nih.gov/pmc

Symptom self-diagnosis through the symptom checker is a core use case. Users also seek information on specific conditions, treatments, and medications through WebMD pages and Q&A articles. WebMD became the internet's de facto "first opinion" for millions of people experiencing a health worry or trying to learn about a diagnosis.

Research and expert investigations paint a mixed picture of WebMD's reliability and accuracy. On one hand, WebMD employs medical professionals and touts a thorough editorial review policy to vet its content for "accuracy and reliability" and to ensure it was written in "plain language."[16] However, independent analyses have found significant limitations in WebMD's diagnostic accuracy and completeness. A 2015 study evaluated 23 online symptom checkers (including WebMD's) using standardized patient vignettes. On average, the correct diagnosis was listed as the top result in only 34% of cases.[17] Harvard researcher Dr. Ateev Mehrotra, who led this study, warned that "if symptom checkers are seen as a replacement for seeing a physician, they are likely an inferior alternative", given doctors' diagnostic accuracy is roughly 85–90% in similar cases.[18] The consensus from experts is that WebMD is "an okay starting point" for laypeople to learn basics, but it is not consistently reliable as a sole resource.[19] A 2019 Outline article quipped that WebMD occupies a "lucrative, brilliant limbo" where its information is technically doctor-approved and credible, yet "we're not supposed to take it too seriously."[20]

In sum, WebMD altered patient behavior. People now often research first, then seek care (or decide not to). As of the mid-2020s, many clinicians acknowledge that tools like WebMD are "here to stay" and the best approach is to guide patients in using them wisely rather than forbid them.[21]

Where ChatGPT Outperforms the Traditional Healthcare System

Turning now to ChatGPT, we can see how it could be considered a natural evolution of the WebMD model. ChatGPT's strengths line up directly with a number of structural weaknesses in U.S. care: time, integration across specialties, support for low-literacy and marginalized groups, and longitudinal context.

Time and Access Access has traditionally been the headline problem in healthcare. Simulation work from UChicago suggests that delivering all guideline-recommended preventive and chronic

[16] The Outline - Thank You WebMD, But It's Not Cancer
https://theoutline.com/post/8256/webmd-is-a-hypochondriacs-dream
[17] The Atlantic - Online Symptom Checkers are Often Wrong.
https://www.theatlantic.com/health/archive/2015/07/online-symptom-checkers-inaccurate-webmd/398654/
[18] The Atlantic - Online Symptom Checkers are Often Wrong.
https://www.theatlantic.com/health/archive/2015/07/online-symptom-checkers-inaccurate-webmd/398654/
[19] Vox - The Truth About WebMD. https://www.vox.com/2016/4/5/11358268/webmd-accuracy-trustworthy
[20] The Outline - Thank You WebMD, But It's Not Cancer
https://theoutline.com/post/8256/webmd-is-a-hypochondriacs-dream
[21] California Healthline - 'Dr. Google' Meets Its Match: Dr. ChatGPT.
https://californiahealthline.org/news/article/chatgpt-chatbot-google-webmd-symptom-checker/

care would require 26.7 hours per day for a typical panel.[22] This is vastly more than the 18 minutes patients receive. With ChatGPT, information access is turned into a prolonged, on-demand visit, something the current system simply cannot provide at scale. Patients can ask multiple follow-ups about the same lab result, return to the same thread days later, and iteratively refine their understanding.

Cross Body System Reasoning U.S. care is fragmented by design. Doctors train around one body system and treat only that system. But many high-burden conditions, especially those most prevalent today, involve multiple body systems. Recent work finds that about 35% of Medicare beneficiaries see five or more physicians in a year, often across several practices.[23]

LLMs are not trained as "multi-specialty clinics," but because GPT-4–class models ingest material from many specialties at once, they can generate differential diagnoses that span organ systems. In controlled studies using internal medicine vignettes, GPT-4's differential diagnosis lists matched or exceeded the performance of experienced physicians on complex cases.[24]

Longitudinal Memory and Pattern Detection Neither WebMD nor Google has memory beyond browser history. The healthcare system does record longitudinal data, but often in siloed EHRs, and as a series of sporadic, single points in time. Additionally, much of the day-to-day variation in symptoms, diet, stress, and sleep never enters the record.

LLMs invite a different behavior: ongoing narrative logging. Because they accept unstructured text (and, increasingly, images and other modalities), patients can log symptoms when they occur, in natural language, append contextual details (meals, exercise, menstrual cycle day, stressors), and periodically ask the model to "look for patterns."

The model does not require consistent logging or rigid data structure to function (though both would likely result in higher quality results). ChatGPT can provide hypotheses ("these symptoms cluster after large late-night meals" or "they often precede your period by 2–3 days"), which the patient can then test on their own or bring to their clinician.

This use tilts LLMs beyond the episodic question-answering currently supported by Google/WebMD towards continuous health management. It also complements the shift toward longitudinal health data from wearables and home devices.

Health Literacy Several comparative studies now suggest that LLM-generated medical explanations are, on average, competitive in quality with top search results. Other work finds ChatGPT outperforms Google Search for general medical knowledge acquisition but performs

[22] JAMA Network - Primary Care Visit Duration Study. https://jamanetwork.com
[23] PMC (PubMed Central) - Medicare Fragmentation Study. https://pmc.ncbi.nlm.nih
[24] PMC (PubMed Central) - GPT-4 Diagnostic Performance Study. https://pmc.ncbi.nlm.nih

worse when offering specific treatment recommendations.[25] In practice, LLMs are not automatically simpler, but they are uniquely capable of becoming simpler on request, a capability that static articles and time-pressed visits rarely match consistently.

Reliability All of these strengths sit atop two massive liabilities; hallucination and lack of formal accountability.

Investigations by journalists and clinicians have documented strong performance in some cases and dangerous oversights in others: a physician reviewing 12 real ChatGPT medical conversations for *The Washington Post* gave four answers a perfect score and four a failing grade.[26] By contrast, WebMD and similar sites can be incomplete or biased, but are static and curated; they do not invent journal articles that do not exist. Additionally, clinicians misdiagnose, especially in groups historically underserved, but they operate within licensing, malpractice, and institutional frameworks.[27] The mismatch of high-risk use, medium reliability, and no accountability, is the main reason most experts argue that ChatGPT should be positioned as a supplementary tool, not a replacement.

What Would Keep People on a ChatGPT Health Product?

Right now, people use ChatGPT for health because it gives them things the traditional healthcare system under-supplies. To *keep* those users and not lose them to a future "Google Health Companion" or other competing products, ChatGPT will have to evolve beyond a smart chat box to an intelligent health data storage and engagement platform.

The "home for your health data" vision requires three big pieces:

1. a durable health data "vault" with clear visualization,
2. a continuously updating personal context graph (including insurance), and
3. a visible trust and safety layer that seems more rigorous than the average wellness app.

Evolving From Chat Window to Health Vault

Today, most people's health information is scattered across patient portals, PDFs of imaging reports, Apple Health logs, wearable dashboards, emails, and sticky notes. EHR interoperability remains aspirational; even systems using the same vendor often struggle to share data in a standardized way.[28]

[25] ResearchGate - ChatGPT Performance in Medical Treatment Recommendations. https://www.researchgate.net/publication/372853869
[26] The Washington Post - ChatGPT Medical Conversation Analysis. https://www.washingtonpost.com
[27] Yale School of Medicine - ChatGPT High-Risk Use Survey. https://medicine.yale.edu
[28] PMC - EHR Interoperability Challenges. https://pmc.ncbi.nlm.nih.gov/articles/PMC5565131

If a ChatGPT-like system wants to be the place people *stay*, it has to offer something the portals and point solutions don't: a single, user-controlled vault where they can upload and organize records, see a timeline of their health, visualize trends in health metrics, and ask questions of their health data.

Conceptually, you could imagine a left-hand navigation with: a "Timeline" view (major events, color-coded by system), a "Metrics" view (graphs for labs, vitals, sleep, etc.), a "Documents" view (structured access to notes and reports), and a "Conversations" view that links each Q&A thread to the underlying data it drew from. If ChatGPT becomes the *only* place where your scattered records and daily logs are stitched into a coherent, searchable narrative, it secures tremendous user lock-in,

A Living "Health Profile" Most current interactions with health LLMs are relatively stateless: you ask a question, get an answer, and the system knows almost nothing about *you* beyond what's in that one prompt or its limited memory. For health, that's not good enough. To feel truly valuable and safe, responses should be grounded in a persistent, user-visible profile, including: demographic basics (age, sex, pregnancy potential), key diagnoses and surgeries, medications and allergies, health goals, constraints (night shift, caregiving, financial limits), and known sensitivities (suspected lactose intolerance, iron-deficiency history, prior ACL tear).

That context has to update as the user learns and changes. If someone discovers they're sensitive to a specific food, gets a new diagnosis, starts a GLP-1, or fractures a bone, that should immediately influence future advice. You can think of this as a personal context graph: a structured layer the model consults every time it answers a health question, and that the user can see and edit. This level of personalization is not possible on static articles or even with traditional, deterministic software. High-quality, personalized health advice at scale requires advanced, persistent, contextual understanding.

Insurance-Aware Guidance Insurance is another major friction keeping people from using traditional care. About half of insured adults report difficulty with at least one aspect of their insurance, including understanding what's covered or what they'll owe.[29]

For an LLM health companion, this is an opportunity. Users could upload or select their insurance plan, see in plain language what's in-network, what's covered, where they are in their deductible and out-of-pocket maximum, track remaining benefits (1 vision exam left, 2 preventive dental visits left), and get recommendations that explicitly reflect those constraints. This is exactly the kind of friction Google is well-placed to attack. But if ChatGPT or a similar system builds it first, it creates another axis of lock-in: users will internalize that "this is the place that knows my coverage and tells me what is realistic for *me*."

[29] KFF (Kaiser Family Foundation) - Insurance Navigation Survey. https://www.kff.org/medicaid/kff-survey-shows-complexity-red-tape-denials-confusion-rivals-affordability-as-a-problem-for-insured-consumers-with-some-saying-it-caused-them-to-go-without-or-delay-care/

Trust through Grounding and Guardrails People hesitate to use LLMs for health because of a lack of trust. People's willingness to share digital health data depends heavily on specific privacy protections like transparency, oversight, and the ability to delete data.[30] Trust is created through concrete design choices like clear explanations, visible grounding, and evidence that data aren't being misused.[31]

If ChatGPT becomes the place that feels both most useful and most explicit about its limits and data practices, its early brand advantage can harden into long-term loyalty even if competitors match its raw model capabilities.

Is ChatGPT "Better," or Just Different?

Relative to WebMD, ChatGPT does not create the desire to self-diagnose; that pattern has been entrenched since the late 1990s. However, it does change the shape of that behavior. Relative to the traditional healthcare system, ChatGPT offers capabilities in exactly the places the system is structurally weak: time, integration, and accessibility for those with language and literacy barriers. In that sense, it is "better" at what patients are increasingly asking for: more interpretable information, more continuity of care, and more help engaging with their own health.

At the same time, ChatGPT is worse on the two attributes that matter most in high-stakes decisions: grounding in verifiable evidence, and formal accountability. In part, this is just a reality of large language models. However, thoughtful UI design can minimize the negative effects of these model byproducts. Given these realities, we must be intentional about where to deliberately plug its strengths into the health information ecosystem and where to keep it firmly behind human judgment.

WebMD, on the other hand, provides always accessible, verified, digestible health information. With simple input, users can get general but, for the most part, physician-reviewed content. If the only difference between WebMD and ChatGPT is information input, that's not a particularly durable moat. WebMD could easily add a chat feature and close the gap. Like with Google Search, WebMD has historical user behavior and distribution advantages that could allow it to catch up if OpenAI/ChatPGT doesn't continue to invest to solidify their lead.

Overall, ChatGPT is a significant new player in health information access and understanding. It addresses clear deficiencies in the current medical system, even with its flaws. But imperfect information is better than no information, and it is clear from the success of WebMD that users will look past accuracy for any scrap of health understanding. For ChatGPT, the user behavior is

[30] JAMA Network - Consumer Views on Sharing Health Data. https://jamanetwork.com/journals/jama-health-forum/fullarticle/2824634
[31] Nature - Trust in AI-Enabled Health Data Systems. https://www.nature.com/articles/s41746-025-01510-8

there. Right now, the gap between it and WebMD is small, with the potential to become a chasm with the right product decisions.