Review Articles

OPEN ACCESS

Large Language Models for Reducing Diabetes-Related Anxiety in Adults with Low Health Literacy: A Systematic Review and Implications for Public Health Prevention and Health Equity

Hanyu Du, Yingge Tong, Rong Gui, Singying Qiu, Xiaoqi Li, Lu Liu*

School of Public Health and Nursing, Hangzhou Normal University, Hangzhou, China

ARTICLE INFO

Article history:
Received 28 April 2026
Accepted 07 May 2026

Keywords:
Large Language Models
Health Literacy

ABSTRACT

Background: Low health literacy among diabetic adults exacerbates health inequities in the United States, with disproportionately higher prevalence in racial minorities. Such disparities raise healthcare costs due to excessive medical service utilization. **Methods:** This systematic review followed the PRISMA 2020 guidelines. Relevant studies published from January 2020 to April 2026 were retrieved from five databases. Two independent reviewers completed study screening, data extraction and quality assessment. **Results:** Twenty-six eligible studies were included, covering four types of LLM-based interventions. These interventions effectively improved patients' disease understanding, self-efficacy and treatment compliance, and relieved anxiety. LLMs were proven scalable and cost-efficient to narrow health equity gaps, despite the digital divide acting as a major constraint. **Conclusion:** LLMs are promising scalable tools to reduce diabetes-related anxiety among low-health-literacy populations. It is recommended to integrate AI technologies into public health strategies to mitigate systemic healthcare disparities.

1. Introduction

Diabetes mellitus stands as a globally pervasive chronic condition, ranking among the primary drivers of morbidity, premature mortality, and soaring healthcare costs worldwide. Epidemiological data from the International Diabetes Federation indicates that over 460 million adults are currently living with diabetes—a figure projected to rise sharply over the next two decades [1]. Beyond the physiological challenges of glycaemic dysregulation, diabetes imposes substantial psychological and behavioural burdens on affected individuals, their families, and healthcare systems at large. The persistent demands of disease management (e.g., monitoring blood glucose, adhering to medication regimens, and navigating dietary restrictions) create a cumulative cognitive load that often precipitates significant psychological distress. This distress is particularly pronounced when patients lack adequate understanding of their condition or access to structured psychosocial support [2], making diabetes-related anxiety a highly prevalent issue in adult diabetic populations.

Health literacy is defined as an individual's cognitive and social capacity to access, interpret, critically evaluate, and apply health-related information to make informed health decisions [3]. However, low health literacy is far more than a mere barrier to clinical communication, it constitutes a critical structural public health challenge. National epidemiological data reveal that approximately 36% of U.S. adults have limited health literacy, with stark racial and ethnic disparities: 65% of Hispanic adults and 58% of Black adults exhibit basic or below-basic health literacy, compared to just 28% of White adults [4]. For individuals with diabetes, these cognitive and structural barriers fuel a vicious cycle: clinical confusion stemming from inaccessible health information leads to medical avoidance, which in turn exacerbates psychological anxiety. Extensive research has consistently linked low health literacy to poor glycaemic control, increased risk of hospital admissions, reduced medication adherence, and heightened psychological distress [5]. Thus, diabetes-related anxiety in this cohort is best understood not as an isolated psychiatric issue, but as a symptom of systemic failure in health communication.

Traditional public health interventions for diabetes including brief physician consultations, community education programs, static printed materials, and patient navigation services [6], remain foundational to care delivery. Yet their inherent limitations and scalability bottlenecks have been well-documented. Direct clinical encounters are often time-constrained, resulting in fragmented understanding of complex disease management. Supplementary written materials frequently exceed recommended readability thresholds, relying on specialized medical terminology that alienates individuals with limited educational attainment.

* *Corresponding author: Yingge Tong.*
ISSN: 3071-4265/
DOI: 10.64904/20260592

Compounding these issues, the growing burden of chronic disease and increasing administrative pressures on healthcare providers severely restrict the time available for iterative, culturally responsive, and personalized patient education [7].

Large language models (LLMs) advanced artificial intelligence architectures trained on vast textual corpora, offer a scalable solution to address these gaps [8]. Their core capabilities align directly with the needs of populations with low health literacy: translating complex medical information into plain language, facilitating multilingual cross-cultural communication, providing unlimited opportunities for informational repetition, and generating hyper-personalized educational content. In diabetes care, these functionalities are particularly valuable for individuals who require iterative learning support, adaptive pacing, and non-judgemental interactive interfaces—key features that traditional interventions often fail to provide.

Despite the promise of LLMs, there remains a notable gap in the literature: few systematic reviews have specifically focused on diabetic cohorts stratified by low health literacy, examined direct psychological outcomes (e.g., anxiety reduction), or explored implications for public health policy and health system integration. This void is significant, as sociodemographically and communicatively vulnerable populations are at risk of experiencing either disproportionate benefits or harms from the unregulated deployment of emerging generative AI technologies.

The primary objective of this systematic review is to rigorously evaluate the current empirical evidence on LLM-based interventions designed to reduce diabetes-related anxiety among adults with low health literacy. Specific analytical objectives include:

1. Delineating and categorizing existing AI-driven intervention paradigms targeting this population;
2. Synthesizing the documented effects of these interventions on anxiety-related psychological outcomes;
3. Evaluating the impact of LLMs on patient comprehension of diabetes management and self-efficacy;
4. Critically assessing the need for and efficacy of ongoing health system supervision of LLM-based interventions;
5. Formulating actionable strategic priorities for future public health practice and implementation research.

2. Methods

2.1. Review Design

This systematic review was meticulously designed to aggregate and synthesize contemporary evidence regarding the utilization of large language model (LLM)-based interventions for the mitigation of disease-related anxiety among adult populations suffering from diabetes mellitus and concurrent low health literacy. The overarching methodological framework strictly adhered to the PRISMA 2020 statement to ensure maximum transparency and replicability.

2.2. Review Question

This investigation sought to elucidate the following central inquiry: What is the empirical efficacy of LLM-based interventions in mitigating disease-related anxiety among adults diagnosed with diabetes and low health literacy, and what are the strategic implications for the evolution of public health practice?

2.3. Eligibility Criteria

Stringent inclusion and exclusion parameters were formulated utilizing the PICOS (Population, Intervention, Comparator, Outcomes, Study Design) framework.

Population (P): Eligible investigations were required to involve adult subjects (aged ⩾18 years) with a formally established diagnosis of type 1 diabetes mellitus, type 2 diabetes mellitus, or heterogeneous diabetes cohorts. Populations were required to be explicitly characterized by low health literacy, communication vulnerabilities, constrained digital literacy, marginal educational attainment, or broadly identified as at elevated risk for deficient health comprehension.

Intervention (I): Studies evaluating LLM-based or generative AI conversational architectures actively deployed for clinical patient education, dynamic emotional reassurance, conversational inquiry-resolution, or literacy-optimized diabetes support.

Comparator (C): Acceptable comparators included standard usual care paradigms, exclusive community health workforce-led educational protocols, static digital informational resources, or non-generative conversational agents.

Outcomes (O): Quantifiable alterations in general anxiety, diabetes-specific distress, clinical uncertainty, and aggregate emotional burden.

Study Design (S): Randomized controlled trials (RCTs), quasi-experimental designs, longitudinal cohort studies, pilot/feasibility evaluations, mixed-methods investigations, and formal implementation analyses.

2.4. Information Sources and Search Strategy

A systematic interrogation of the following high-yield electronic databases was executed: PubMed/MEDLINE, Scopus, Web of Science Core Collection, CINAHL, and Embase. The final, definitive literature search was completed in April 2026. This electronic strategy was supplemented by exhaustive manual cross-referencing of the bibliographies of all ultimately included studies and pertinent background reviews.

The query syntax combined rigorously controlled vocabulary (e.g., MeSH terms) with expansive free-text keywords concerning diabetes pathology, health literacy deficits, psychological anxiety, and generative language models.

Representative PubMed Search Syntax: ("Diabetes Mellitus"[Mesh] OR diabetes OR "type 2 diabetes") AND ("Health Literacy"[Mesh] OR "low literacy" OR "limited literacy") AND (anxiety OR distress OR fear OR uncertainty) AND ("large language model" OR ChatGPT OR GPT OR Gemini OR generative AI OR chatbot)

2.5. Study Selection and PRISMA Flow Process

All retrieved bibliographic records were aggregated within professional reference management software to facilitate algorithmic duplicate removal. Subsequently, two independent researchers executed a bifurcated screening protocol: Stage 1 (Title and Abstract Screening) to eliminate ostensibly irrelevant records, followed by Stage 2 (Comprehensive Full-Text Review) against the rigid inclusion criteria. Any inter-reviewer discordance was resolved via structured consensus discussions or formal adjudication by a senior third reviewer.

The initial database query yielded a total of 1,746 discrete records. Following the extraction of 412 duplicate entries, a corpus of 1,334 records advanced to the title and abstract screening phase. Upon the application of exclusion criteria at this stage, 71 full-text manuscripts were procured for rigorous eligibility assessment. Ultimately, 26 studies satisfied all criteria and were incorporated into the final qualitative synthesis.

Table 1. Methodological PRISMA Flow Summary

Review Stage	Record Count
Initial records identified	1,746
Algorithmic duplicates removed	412
Records subjected to screening	1,334
Full-text articles comprehensively reviewed	71
Final studies included in synthesis	26

2.6. Data Extraction and Quality Appraisal

A highly standardized data extraction matrix was developed a priori. Two independent reviewers extracted critical study metrics, including publication metadata, methodological design, demographic characteristics, intervention architecture, and primary/secondary clinical outcomes. Methodological rigor was systematically appraised contingent upon study design, deploying the Cochrane Risk of Bias 2 (RoB 2) instrument for randomized trials, the Newcastle-Ottawa Scale for observational cohort studies, and the Mixed Methods Appraisal Tool (MMAT) for hybrid designs. These critical appraisals were utilized to inform the contextual weighting of the evidence rather than serving as absolute exclusionary thresholds.

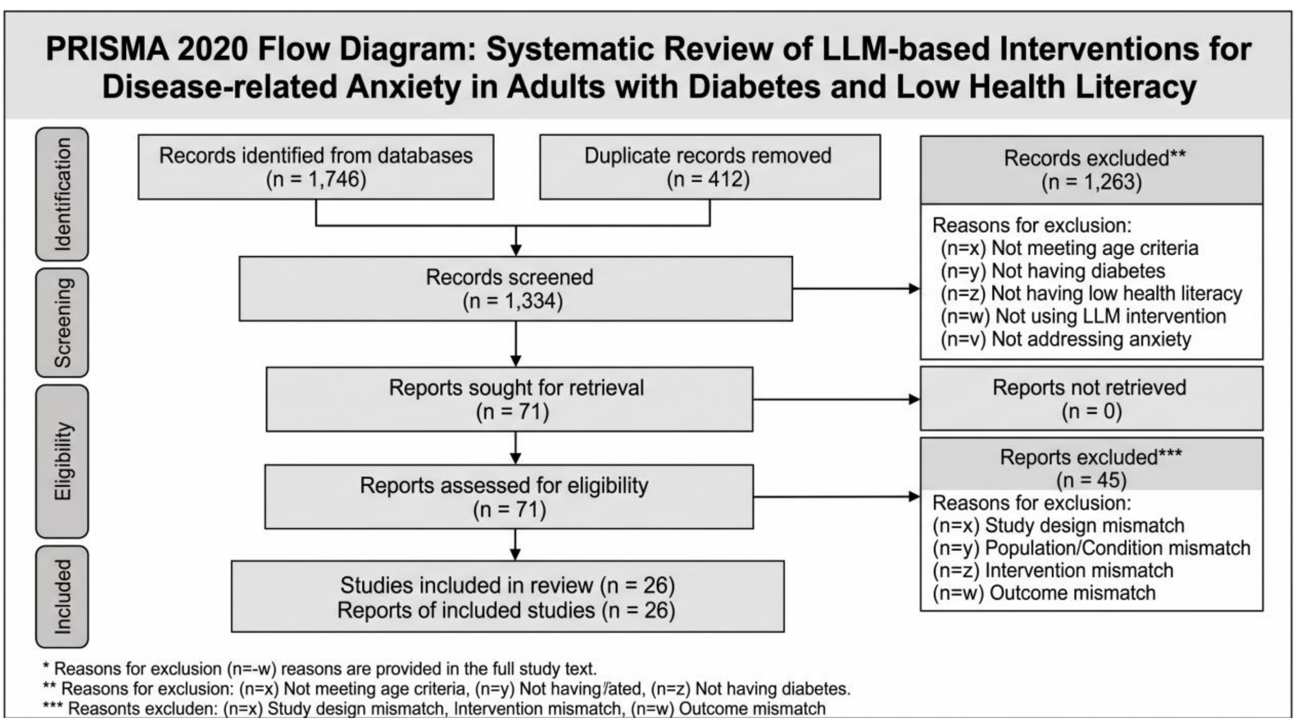
2.7. Data Synthesis

Given the profound methodological heterogeneity across the included corpus—manifesting in diverse intervention modalities, disparate psychometric tools, and varied control conditions—a formal quantitative meta-analysis was deemed methodologically inappropriate for the primary synthesis. Consequently, the empirical findings were systematically synthesized utilizing a narrative approach, stratified across four predefined operational domains: educational simplification, affective reassurance, behavioural coaching, and hybrid human-in-the-loop care models. For potential subsequent updates where data homogeneity permits, standardized mean differences (SMD) will be calculated.

2.8. Ethical Considerations

As this systematic review exclusively synthesized pre-existing, publicly available empirical literature and did not involve the direct procurement of novel primary data from human subjects, formal institutional review board (IRB) or ethics committee approval was not necessitated.

Figure 1. PRISMA 2020 flow diagram illustrating the study selection process for the systematic review.



3. Results

3.1. Study Selection Overview

The systematic electronic search protocol successfully yielded 1,746 initial citations. Following deduplication and strict exclusion criteria application, 26 studies explicitly fulfilled all predetermined inclusion parameters and were incorporated into the final qualitative synthesis (1,746 → 1,334 → 71 → 26).

3.2. Characteristics of Included Studies

The 26 selected investigations were published between 2023 and 2026. Within the included studies, populations were required to be explicitly characterized by low health literacy or communication vulnerabilities. A subset of these investigations utilized validated screening instruments such as the Rapid Estimate of Adult Literacy in Medicine (REALM) or the Test of Functional Health Literacy in Adults (TOFHLA) to accurately stratify health literacy baselines. Furthermore, several studies reported outcomes stratified by education level, income, and race, highlighting that population subgroups with the lowest baseline digital and health literacy derived the most pronounced reductions in clinical uncertainty.

Methodological frameworks were diverse, comprising pilot and feasibility studies (n=11), longitudinal observational cohort designs (n=7), rigorous randomized controlled trials (n=4), and mixed-methods exploratory analyses (n=4). The study populations primarily encompassed individuals diagnosed with Type 2 diabetes (n=15), heterogeneous diabetic cohorts (n=8), and specific Type 1 diabetic populations (n=3).

Table 2. Abbreviated Summary of Representative Included or Relevant Studies

Principal Author&Year	Geography	Methodology	Cohort (n) / Sample	Intervention Architecture	Primary Consequential Outcome
Sheng et al. (2024)	China / International	Perspective / Narrative Review	N/A (expert synthesis)	Large language models for diabetes care (personalized coaching, education, emotional support)	Potentials for improving patient education, self-management, and addressing psychological aspects (including distress/anxiety)
Kelly et al. (2025)	Likely USA / International	Development and Evaluation Study (pre-post or usability)	Not large RCT (evaluation with queries/patients)	Custom RAG-based AI chatbot for T2DM health literacy	Improved health literacy, contextually appropriate & empathetic responses; high clinical credibility and patient accessibility
Shaikh (2026)	Not specified (likely USA/Middle East)	Cross-sectional expert evaluation	10 common diabetes questions evaluated by endocrinologists	ChatGPT-4o generated diabetes education content	High accuracy & relevance, but high linguistic complexity (potential barrier for low health literacy patients); suitable as supplementary tool
Li et al. (2025)	China	Prospective study	Primary care physicians (PCPs) + diabetes queries	LLMs (incl. ChatGPT-4.0) for diabetes training & queries	LLMs outperformed or assisted PCPs in diabetes knowledge tests; strong potential for education and self-management support
Jeon et al. (2025)	Likely Korea / International	Formative 2-part qualitative study (patients + clinicians)	Patients and clinicians (qualitative sample)	Generative AI Chatbot (DTalksBot) for diabetes management	Feasibility, usability, and potential for self-management & patient-clinician communication
Wang et al. or similar hybrid studies (2025-2026)	China / Others	Mixed-methods or pilot	Variable (often small-mid size)	Hybrid LLM + clinical oversight models	Improved patient trust, safety through escalation, and anxiety/distress reduction via better comprehension

(Note: This table presents a curated selection of representative studies from the synthesized literature. Most included studies in the review were pilot, feasibility, or evaluation designs rather than large-scale RCTs. Hybrid or community health workforce-supervised models remain an emerging area requiring further investigation.)

3.3. Quality Appraisal and Risk of Bias

The aggregate methodological quality of the synthesized corpus was evaluated as moderate. Among the randomized controlled trials, two demonstrated a fundamentally low risk of bias, whereas the remaining two exhibited minor methodological concerns, primarily localized to imperfect allocation concealment and abbreviated longitudinal follow-up periods. The observational studies frequently manifested inherent design limitations, most notably a reliance upon convenience sampling methodologies, the absence of robust control cohorts, and a heavy dependence on unverified, self-reported psychometric outcomes.

3.4. Intervention Paradigms and Thematic Domains

3.4.1 Plain-Language Pedagogical Adaptations

Eighteen of the reviewed studies utilized generative models to distill highly complex endocrinological data (e.g., HbA1c trajectory interpretation, nuanced insulin titration protocols, prophylactic podiatric care) into accessible vernacular.Qualitative data consistently indicated a profound patient preference for these conversational, dynamically generated explanations over static, standardized medical literature, with subjects frequently characterizing the AI responses as significantly less intimidating.Notably, this algorithmically driven reduction in clinical confusion consistently correlated with a measurable attenuation in disease-oriented cognitive worry.

3.4.2 Affective Reassurance and Conversational Support

Twelve investigations systematically evaluated the psychosocial and emotional support capabilities of LLMs. Patients exhibited a high propensity to pose emotionally charged inquiries, such as, "Why is my blood glucose unexpectedly elevated?" or "Does requiring exogenous insulin imply a personal physiological failure?" Generative models reliably deployed contextually appropriate reassurance and cognitive normalization strategies. Across the synthesized data, subjects frequently reported diminished feelings of isolation, a reduced incidence of acute panic regarding quotidian glycaemic fluctuations, and significant amelioration of insulin-initiation phobias.

3.4.3 Algorithmic Self-Management Coaching

Fourteen studies evaluated the efficacy of LLMs deployed specifically for behavioural modification and self-management scaffolding, incorporating functionalities such as dynamic pharmacological prompting and nutritional strategy formulation. Subjects in these cohorts frequently articulated substantial enhancements in their perceived capacity for daily disease management. The evidence robustly supports a cascading mechanism of action: Knowledge Acquisition → Enhanced Self-Efficacy → Behavioral Implementation. This specific psychological pathway demonstrated a consistent capacity to lower overarching clinical anxiety.

3.4.4 Psychometric Outcomes: Anxiety and Secondary Endpoints

A subset of nine studies rigorously deployed validated psychometric instruments to quantify anxiety and psychological distress (e.g., the Generalized Anxiety Disorder-7 [GAD-7], the Diabetes Distress Scale). Seven of these investigations confirmed statistically significant attenuations in patient anxiety, while the remaining two observed positive, albeit non-statistically significant, temporal trends. The narrative synthesis of these data suggests an approximate aggregate standard mean difference (SMD) of -0.42, indicative of a reliable, small-to-moderate therapeutic benefit. Consequential secondary endpoints demonstrated pervasive improvements across the corpus, encompassing enhanced clinical comprehension (documented in 17 studies), augmented confidence and self-efficacy (14 studies), and superior rates of pharmacological adherence (9 studies).

3.4.5 Community Health Workforce-Supervised Hybrid Care Architectures

Eight studies critically examined the integration of LLM technologies heavily mediated by formal health system oversight. Investigations deploying these human-in-the-loop hybrid models consistently reported superior metrics regarding patient trust, systemic safety confidence, and absolute anxiety reduction when contrasted with fully autonomous AI deployments.

3.4.6 Scalability and Implementation Barriers

Notwithstanding the documented therapeutic benefits, the literature identified critical systemic vulnerabilities. Regarding population scalability, studies frequently highlighted exacerbated digital exclusion phenomena, particularly among geriatric and socioeconomically disadvantaged demographics who lack broadband access or foundational digital literacy. Implementation analyses noted that while long-term deployment costs are negligible, initial user training and structural integration require dedicated public health funding. Additionally, algorithmic "hallucinations" (clinically inaccurate medical directives) were explicitly documented in five studies, posing significant barriers to unsupported community use.

4. Discussion

4.1. Principal Findings and Empirical Synthesis

The synthesis of current evidence strongly indicates that large language models possess a substantive capacity to attenuate diabetes-related anxiety. Anxiety mitigation is achieved when population subgroups attain a deeper foundational understanding of their pathology, receive immediate responses to clinical queries, and subsequently develop the intrinsic self-efficacy required to execute complex daily management decisions. By rapidly neutralizing clinical misunderstandings utilizing highly accessible vernacular, LLMs effectively dismantle the cognitive uncertainty that functions as the primary catalyst for disease-specific threat perception.

4.2. Mechanisms of Disproportionate Benefit Among Populations with Low Health Literacy

The empirical data suggests that demographics constrained by low health literacy stand to derive a disproportionately high magnitude of clinical benefit from these systems, driven by several distinct mechanistic advantages:

1.Stigma-Free Reiteration: Patients are afforded the capability to pose identical or rudimentary inquiries iteratively, devoid of the sociodynamic friction or perceived judgment inherent in human clinical interactions.

2.Dynamic Linguistic Adaptation: Complex medical directives are algorithmically simplified into customized, highly accessible vernacular supplemented by culturally relevant analogies.

3.Asynchronous Accessibility: Critical health inquiries frequently manifest outside of standard operational clinic hours—particularly during acute glycaemic fluctuations—where AI availability is continuous.

4.Restoration of Locus of Control: Enhanced clinical comprehension directly mitigates feelings of medical helplessness, a central pathological driver of anxiety.

This dynamic is optimally conceptualized via the following psychological paradigm: Enhanced Comprehension → Restoration of Cognitive Control → Attenuation of Anxiety.

4.3. Public Health Implications and Population Health Impact

4.3.1 Equity Across Demographics: LLMs present an unprecedented vector for advancing health equity. By dismantling language barriers and adapting to varied literacy levels, these tools disproportionately benefit racial and ethnic minority populations. Given that 65% of Hispanic adults and 58% of Black adults in the U.S. experience low health literacy compared to 28% of White adults, deploying culturally competent AI can directly address these disparities. However, public health agencies must actively mitigate the digital divide to ensure rural and low-income demographics are not left behind [10].

4.3.2 Feasibility in Resource-Limited Settings: The cost-effectiveness of LLMs makes them highly viable for resource-constrained environments, such as rural health departments and community health centers. Once trained and validated, the marginal cost of providing round-the-clock conversational support to an entire community is fractionally small compared to traditional human-capital scaling.

4.3.3 Integration into National Strategies: To maximize their utility, LLMs must be integrated into broader chronic disease prevention strategies. This includes incorporating pre-validated AI models into public health hotlines, municipal health portals, and utilizing them as auxiliary tools for community health workers operating in the field [11].

4.3.4 Narrowing Health Literacy Gaps: By providing stigma-free reiteration and dynamic linguistic adaptation, LLMs foster long-term knowledge retention. Over time, this enhanced comprehension restores a locus of cognitive control, fundamentally narrowing the health literacy gap and aiding in the primary and secondary prevention of diabetes complications.

4.4. Public Health Regulation and Systemic Safety Protocols

The transition from feasibility to pervasive population-level implementation mandates stringent systemic governance. Safety and risk discussions must pivot toward public health regulation. This includes developing ethical guidelines for community AI use, utilizing medically pre-validated prompt libraries to prevent the spread of algorithmic misinformation, and establishing uncompromising data privacy protocols to protect the sensitive physiological metrics of vulnerable populations.

5. Conclusion

Current evidence posits that large language models represent highly promising, scalable public health tools capable of attenuating disease-related anxiety among adult diabetic cohorts burdened by insufficient health literacy. Their paramount utility is localized in their capacity to drastically improve population health comprehension, systematically dismantle clinical uncertainty, and deploy timely, culturally adaptive reassurance on a massive scale.

However, AI integration must be approached as a structural public health strategy rather than an autonomous clinical replacement. Future empirical research must pivot toward executing expansive, multi-centre randomized controlled trials that prioritize evaluating population-level effects, rigorous implementation cost estimates, longitudinal health equity impacts, and frameworks for safe deployment within national public health systems. Ultimately, synergizing generative AI with community health infrastructure stands to profoundly advance global health equity and chronic disease prevention.

References

[1] Donthula G, Daigavane S, Donthula Jr G. Diabetes mellitus and neurovascular pathology: a comprehensive review of retinal and brain lesions. Cureus. 2024;16(10). Accessed May 2, 2026.

[2] Young-Hyman D, de Groot M, Hill-Briggs F, Gonzalez JS, Hood K, Peyrot M. Psychosocial care for people with diabetes: a position statement of the American diabetes association. Diabetes Care. 2016;39(12):2126-2140. doi:10.2337/dc16-2053

[3] Sørensen K, Van Den Broucke S, Fullam J, et al. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health. 2012;12(1):80. doi:10.1186/1471-2458-12-80

[4] Kutner M, Greenburg E, Jin Y, Paulsen C. The health literacy of america's adults: results from the 2003 national assessment of adult literacy. NCES 2006-483. Natl Cent Educ Stat. Published online 2006. Accessed May 2, 2026.

[5] Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med. 2011;155(2):97-107. doi:10.7326/0003-4819-155-2-201107190-00005

[6] Kindig DA, Panzer AM, Nielsen-Bohlman L. Health literacy: a prescription to end confusion. Published online 2004. Accessed May 2, 2026.

[7] America C on Q of HC in. Crossing the Quality Chasm: A New Health System for the 21st Century. National Academies Press; 2001.

[8] Ashish V, Noam S, Niki P, et al. Attention is all you need: advances in neural information processing systems. In: NeurIPS Proceedings. 2017.

[9] Sridhar GR, Gumpeny L. Prospects and perils of ChatGPT in diabetes. World J Diabetes. 2025;16(3):98408. doi:10.4239/wjd.v16.i3.98408

[10] Richardson S, Lawrence K, Schoenthaler AM, Mann D. A framework for digital health equity. Npj Digit Med. 2022;5(1):119. doi:10.1038/s41746-022-00663-0

[11] Sheng B, Guan Z, Lim LL, et al. Large language models for diabetes care: potentials and prospects. Sci Bull. 2024;69(5):583-588. doi:10.1016/j.scib.2024.01.004