Original Paper

Racial and Ethnic Differences in Trust in AI Relative to Human Professionals for Health Care: Nationally Representative Cross-Sectional Study

Cindy H Liu[1,2], PhD; Amanda Koire[3], MD, PhD; Tiffany Yip[4], PhD

[1]Pediatrics, Brigham and Women's Hospital, Boston, MA, United States

[2]Harvard University, Boston, MA, United States

[3]Psychiatry, Brigham and Women's Hospital, Boston, MA, United States

[4]Department of Psychology, Fordham University, New York City, NY, United States

Corresponding Author:
Cindy H Liu, PhD
Harvard University
221 Longwood Ave
BL341
Boston, MA, 02115
United States
Phone: 1 5104156932
Email: chliu@bwh.harvard.edu

Abstract

Background: AI is increasingly applied across health care domains, yet public trust in AI across racial and ethnic groups and across health domains is poorly understood.

Objective: This study examined trust in AI relative to trust in human professionals and identified sociodemographic characteristics associated with this comparative trust.

Methods: A cross-sectional survey was conducted among 10,506 US adults from the Understanding America Study (UAS), a probability-based internet panel, with poststratification survey weights applied to descriptive estimates to characterize the US adult population. Data were analyzed using multinomial logistic regression, adjusted for the sociodemographic characteristics used in poststratification survey weights. Trust in AI relative to human professionals was assessed for mental health counseling (the primary outcome), health diagnoses, and diet and exercise planning. Responses were categorized as “less trust than in a human professional,” “the same level of trust as in a human professional” (parity trust), “more trust than in a human professional” (superiority trust), or “not sure.”

Results: Among the 10,506 participants (n=6385, 51.3% [percentage weighted using poststratification survey weights] female; mean age 49.3, SD 17.0 years), most trusted AI less than they trusted human professionals for health diagnosis and mental health counseling. Trust was lowest for mental health counseling (1136/10,469, 11.9% reported parity or superiority trust) and highest for diet and exercise planning (4773/10,460, 43.2%). For mental health counseling, non-Hispanic (NH) Black, Hispanic, and NH Asian adults were more likely than NH White adults to endorse parity trust, and these groups and NH multiracial adults were more likely to endorse superiority trust. NH Asian adults were 2.5 times more likely to endorse parity trust (relative risk ratio [RRR] 2.58, 95% CI 2.04-3.25; *P*<.001) and nearly 6 times more likely to endorse superiority trust (RRR 5.95, 95% CI 3.79-9.35; *P*<.001) for mental health counseling. NH Black (RRR 3.32, 95% CI 2.19-5.05; *P*<.001), Hispanic (RRR 2.41, 95% CI 1.60-3.65; *P*<.001), and NH multiracial (RRR 3.57, 95% CI 2.06-6.19; *P*<.001) adults also showed elevated superiority trust. For mental health counseling, all racial and ethnic minority groups were more likely than NH White adults to select “not sure.” Generative AI use for general purposes was associated with higher trust, whereas higher education levels were associated with lower trust.

Conclusions: NH Black, Hispanic, NH Asian, and NH multiracial adults were more likely than NH White adults to trust AI more than human professionals for mental health counseling. These data cannot show whether higher trust leads to greater use, and higher trust should not be interpreted as a preference for AI over human-delivered care. Nonetheless, this pattern informs how mental health AI tools can be designed, evaluated, and governed, as individuals with the greatest trust in these tools may be most affected if the tools perform poorly.

XSL•FO
RenderX

(J Med Internet Res 2026;28:e98253) doi: 10.2196/98253

KEYWORDS

artificial intelligence; AI; trust; mental health; racial and ethnic differences; health equity; nationally representative survey

Introduction

AI is increasingly applied across health care domains, from diagnostic imaging and clinical decision support to personalized wellness recommendations and mental health interventions [1-3]. As AI tools become more capable and accessible, understanding public trust in these technologies is essential. Prior studies have linked trust to the uptake and implementation of AI-based tools in health care [4-6]. Yet relatively little is known about how trust in AI varies across different health domains and populations.

Public trust in AI for health-related tasks is generally low but may vary by domain [7]. Tasks perceived as requiring greater interpersonal skill, emotional sensitivity, or clinical judgment, such as mental health counseling, may elicit less trust in AI than more technical or information-based tasks, such as designing a diet or exercise plan. Castelo et al [8] found that people trusted algorithms less for tasks perceived as subjective and requiring emotional sensitivity than for tasks perceived as more objective and data-driven. Similarly, Longoni et al [9] found that consumers were broadly resistant to medical AI due to "uniqueness neglect," reflecting concerns that AI cannot adequately account for people's unique characteristics. Individuals were more accepting of "noninvasive" applications such as heart monitoring and more hesitant about mental health monitoring [10]. Although more studies are evaluating trust across AI applications in health [10,11], most studies have focused on a single domain or have used nonrepresentative samples.

Sociodemographic factors may also shape trust in AI. Younger adults and those with prior AI experience tend to report more favorable attitudes toward AI [12-14]. Racial and ethnic minority groups, who face well-documented disparities in access to and satisfaction with current health care models [15], may view AI differently from non-Hispanic (NH; note that the abbreviation is used for brevity and not to reduce ethnic identity) White adults, although the direction of this association remains unclear. On one hand, experiences of discrimination or cultural barriers could lead to greater trust in AI relative to human professionals; on the other hand, historical distrust of medical institutions could extend to AI systems [16]. Understanding these patterns is particularly important in mental health, where racial and ethnic disparities in access to treatment and to human professionals are pronounced [17,18].

Most prior studies of trust in AI for health have asked whether respondents trust AI more or less than human professionals. However, this approach obscures distinctions such as whether individuals trust AI *as much as* human professionals (parity trust) or *more than* human professionals (superiority trust). Parity trust indicates that a respondent rates AI as being as trustworthy as a human professional for a task, whereas superiority trust indicates that a respondent rates AI as more trustworthy than a human professional. We make this distinction because trusting AI more than a human professional is most relevant to potential negative consequences associated with overtrust in AI as these systems continue to be developed [19,20]. Respondents who report being "not sure" whether they trust AI may represent a qualitatively distinct group from those who actively distrust it (ie, those who trust AI *less than* a human professional) and those who demonstrate parity or superiority trust.

The present study uses cross-sectional, nationally representative survey data from the Understanding America Study (UAS) on 10,506 participants to examine public trust in AI across 3 health domains: health diagnosis, diet and exercise planning, and mental health counseling. We extend prior work by leveraging a probability-based panel of US adults with poststratification survey weights. We address two research questions: (1) How does trust in AI relative to human professionals vary across health domains? (2) What sociodemographic characteristics are associated with trusting AI at least as much as a human professional across these domains? Mental health counseling was designated the primary outcome because it is a domain in which AI-based tools are proliferating rapidly [3] and in which racial and ethnic disparities in access to and quality of human-delivered care are especially pronounced [17,18]. The aim of this study is thus to characterize how public trust in AI relative to human professionals varies across health domains and sociodemographic groups, with particular attention to mental health counseling.

Methods

Study Design

This study is a secondary analysis of deidentified, publicly available data from the UAS, a probability-based internet panel maintained by the Center for Economic and Social Research (CESR) at the University of Southern California. A cross-sectional survey design was used. Poststratification survey weights were applied to descriptive estimates to characterize the US adult population.

Participants

The sample comprised 10,506 US adults aged ≥18 years from the UAS panel. Respondents with a survey weight of zero (n=227; nonprobability recruits) were excluded from the analytic sample.

Outcomes

Comparative trust in AI, that is, trust in AI relative to a human professional, was assessed across 3 health domains. Respondents were asked: "How much would you trust AI in the following daily life activities?" followed by 6 specific daily life activities presented together on a single screen. Three of the 6 activities concerned health and were analyzed here; the remaining 3 concerned financial advice, travel reservations, and driving a

XSL•FO
RenderX

taxicab. Trust in mental health counseling was considered the primary outcome: "providing counseling to address mental health issues." Two other items served as outcomes for comparative purposes: "giving a health diagnosis" and "designing a diet and exercise program." Response options were: (1) "less than a human professional or expert," (2) "as much as a human professional or expert," (3) "more than a human professional or expert," and (4) "not sure." In multinomial analyses, "less than a human professional" served as the reference category. This approach examined the predictors of parity trust ("as much as"), superiority trust ("more than"), and uncertainty ("not sure") relative to active distrust ("less than"). Respondents were asked about trust in AI in a general sense; the item did not describe a specific AI system or specify its accuracy, privacy protections, degree of human oversight, or clinical setting. Trust here therefore reflects a general disposition toward AI within each domain rather than an appraisal of a particular tool or use case.

Covariates

Age was measured in years (range: 18-104). The relative risk ratio (RRR) is reported per 10-year increment for interpretability.

For sex, UAS options were 0="female" and 1="male." Sexual orientation was not assessed.

Hispanic or Latino ethnicity was assessed (0="no" and 1="yes"). Race was assessed independently with response options of White, Black, American Indian or Alaska Native, Asian, Hawaiian or Pacific Islander, or mixed (indicating identification with 2 or more racial groups), yielding mutually exclusive race or ethnicity categories: NH White, NH Black, Hispanic (any race), NH Asian, NH multiracial, and NH other (comprising NH American Indian or Alaska Native and NH Hawaiian or Pacific Islander respondents). NH White served as the reference group.

The highest level of education was assessed, with 16 response categories ranging from "less than 1st grade" to "doctorate." Education was dichotomized as a bachelor's degree or higher vs less than a bachelor's degree. This threshold was selected both because public attitudes toward AI in health care vary across demographic groups [21] and to preserve stable cell sizes across all 6 race and ethnicity groups in the multinomial models.

A single demographic item indicated whether the respondent had a disability (0="no" and 1="yes").

Urbanicity was derived from respondents' census tract of residence using the 2010 Rural-Urban Commuting Area (RUCA) codes. The UAS collapses the 10 primary RUCA codes into 4 levels: metropolitan (RUCA codes 1-3), micropolitan (RUCA codes 4-6), small town or rural (RUCA codes 7-10), and unknown. For the present analyses, urbanicity was dichotomized as rural (small town or rural) vs nonrural (metropolitan or micropolitan).

Generative AI awareness and use for general purposes were assessed by asking, "Have you heard about or used AI applications that create humanlike text or code, such as ChatGPT, Llama, Gemini, or Copilot?" Response options were: (1) "I have never heard about them," (2) "I have heard about them, but I have never used one," (3) "I have used them," and (4) "I don't know." Respondents were classified as generative AI users if they selected option 3 and nonusers otherwise. This variable was included as a covariate to account for familiarity with generative AI, which is associated with more favorable attitudes toward AI [12-14] and could therefore account for differences in comparative trust. The item does not capture experience with health-specific AI.

Statistical Analysis

Poststratification survey weights were applied to all descriptive estimates to characterize the US adult population. Weights were computed using a 2-step process: base weights correcting for unequal probabilities of sampling UAS members, followed by poststratification weights aligning the sample with the US adult population (aged ≥18 years) on sex, race or ethnicity, age, education, and census region, benchmarked to the Current Population Survey.

Regression models were estimated without survey weights, controlling instead for the variables used to construct the weights, an approach that avoids the loss of efficiency that unnecessary weighting entails [22]. We estimated multinomial logistic regression models for each health domain, with the 4-category trust outcome (less than, as much as, or more than a human professional, and not sure) as the dependent variable and "less than a human professional" as the reference category. We report RRRs and 95% CIs. Models were estimated on complete cases: respondents missing the domain-specific trust outcome or any model covariate were excluded. Missingness was minimal (<1% for every variable). All analyses were conducted in Python (version 3.13.9; Python Software Foundation) using the *statsmodels* (version 0.14.6) library. Analytic code was developed using a human-supervised, multiphase workflow. In the first phase, an AI-assisted programming tool, Claude Code (Claude Opus 4.6; Anthropic PBC), generated draft Python scripts to implement the prespecified analytic models. This process occurred without access to or execution on the study dataset. In a subsequent phase, a separate large language model, ChatGPT (GPT-5; OpenAI), was used to support code inspection by identifying potential logical inconsistencies or deviations from the analytic plan, again without exposure to study data. All code was then reviewed line by line by the authors, revised as needed, and executed on the analytic dataset by the research team. Decisions regarding model specification, variable coding, model diagnostics, and interpretation of results were made exclusively by the authors, who assume full responsibility for the analyses.

Ethical Considerations

This study was deemed exempt from review by the Mass General Brigham Institutional Review Board because it involves secondary analysis of deidentified, publicly available data. The UAS data collection was approved by the Biomedical Research Alliance of New York (BRANY) Institutional Review Board (22-030-1044); all UAS participants provided informed consent.

XSL•FO
RenderX

Results

Sample Characteristics

Sample characteristics and the distribution of trust in AI across 3 health domains are presented in Table 1. Approximately one-third of respondents (3697/10,506, 33.2% [percentage weighted using poststratification survey weights]) reported having used generative AI applications for general purposes.

XSL•FO
RenderX

Table 1. Sample characteristics (N=10,506).

Characteristics	Sample[a]
Age (years), mean (SD)	49.3 (17.0)
Age group (years), n (%)	
18-29	874 (12.1)
30-44	2906 (33.9)
45-59	2839 (22.6)
≥60	3871 (31.3)
Sex, n (%)	
Female	6385 (51.3)
Male	4118 (48.6)
Race or ethnicity, n (%)	
Hispanic	1572 (17.3)
NH[b] Asian	780 (5.8)
NH Black	1204 (12.1)
NH multiracial	457 (3.5)
NH other (American Indian or Alaska Native and Native Hawaiian or Pacific Islander)	97 (0.7)
NH White	6386 (60.4)
Education, n (%)	
Less than high school	469 (7.9)
High school graduate or General Educational Development	1668 (29.8)
Some college or associate degree	3490 (25.9)
Bachelor’s degree	2759 (20.8)
Graduate degree	2118 (15.7)
Disability: present, n (%)	865 (9.3)
Urbanicity, n (%)	
Metropolitan	8639 (79.8)
Micropolitan	925 (10.8)
Small town or rural	941 (9.4)
Generative AI awareness and use, n (%)	
Never heard of them	1600 (17.5)
Heard of them but never used them	4832 (44.4)
Have used them	3697 (33.2)
Do not know	358 (4.6)
Trust in AI, n (%)	
Health diagnosis	
Less than a human professional	6625 (60.5)
As much as a human professional	1351 (13.1)
More than a human professional	324 (3)
Not sure	2166 (23.4)
Diet and exercise planning	
Less than a human professional	3973 (37.3)
As much as a human professional	4122 (37.2)

XSL•FO
RenderX

Characteristics	Sample[a]
More than a human professional	651 (6)
Not sure	1714 (19.5)
Mental health counseling	
Less than a human professional	7500 (67.9)
As much as a human professional	935 (9.8)
More than a human professional	201 (2.1)
Not sure	1833 (20.2)

[a]All percentages are weighted using poststratification survey weights; n refers to unweighted counts. The analytic sample excludes 227 respondents with zero weights (nonprobability recruits).

[b]NH: non-Hispanic.

Trust in AI Across Health Domains

Most respondents trusted AI less than they trusted a human professional for health diagnosis and mental health counseling; however, for diet and exercise planning, responses were more evenly divided. Trust was lowest for mental health counseling (1136/10,469, 11.9% reported trusting AI as much as or more than a human professional) and highest for diet and exercise planning (4773/10,460, 43.2%), with health diagnosis falling in between (1675/10,466, 16.1%). A substantial proportion of respondents selected “not sure” across all domains (19.5%-23.4%).

Sociodemographic Correlates of Trust

Table 2 presents the multinomial logistic regression results. Racial and ethnic minority groups were generally more likely than NH White adults to trust AI as much as or more than a human professional. These effects were most pronounced for mental health counseling, for which NH Asian adults were approximately 2.5 times more likely to endorse parity trust and nearly 6 times more likely to endorse superiority trust. NH Black, Hispanic, and NH multiracial adults also showed elevated relative risks of superiority trust for mental health counseling. Additionally, all minority groups were more likely to select “not sure” for mental health counseling, a departure from the reference category of trusting AI less than a human professional.

XSL•FO
RenderX

Table 2. Multinomial logistic regression: relative risk ratios (RRRs) for trust in AI across health domains[a,b].

Predictors	As much as a human professional, RRR (95% CI)	*P* value[c]	More than a human professional, RRR (95% CI)	*P* value	Not sure, RRR (95% CI)	*P* value
Panel 1: mental health counseling (n=10,384)						
Age (per 10 years)[d]	0.90 (0.85-0.94)	<.001	0.84 (0.76-0.93)	<.001	0.98 (0.95-1.02)	.31
Male (reference: female)	1.21 (1.05-1.39)	.009	1.24 (0.93-1.67)	.15	1.02 (0.91-1.14)	.74
NH[e] Black[f]	1.79 (1.45-2.22)	<.001	3.32 (2.19-5.05)	<.001	2.02 (1.73-2.37)	<.001
Hispanic	1.37 (1.12-1.68)	.002	2.41 (1.60-3.65)	<.001	1.67 (1.43-1.95)	<.001
NH Asian	2.58 (2.04-3.25)	<.001	5.95 (3.79-9.35)	<.001	3.06 (2.51-3.75)	<.001
NH multiracial	1.21 (0.86-1.71)	.28	3.57 (2.06-6.19)	<.001	1.31 (1.01-1.71)	.04
NH other (AI/AN[g] and NH/PI[h])	0.67 (0.24-1.87)	.45	3.12 (0.94-10.32)	.06	2.63 (1.69-4.08)	<.001
Bachelor's degree or higher	0.60 (0.52-0.71)	<.001	0.44 (0.32-0.62)	<.001	0.50 (0.44-0.56)	<.001
Disability	1.10 (0.84-1.44)	.50	1.63 (1.02-2.60)	.04	1.31 (1.10-1.56)	.002
Rural (reference: non-rural)	0.79 (0.59-1.05)	.10	0.78 (0.42-1.46)	.44	1.19 (1.00-1.42)	.06
Generative AI user (reference: nonuser)	1.74 (1.49-2.04)	<.001	1.47 (1.06-2.02)	.02	0.62 (0.54-0.71)	<.001
Panel 2: health diagnosis (n=10,381)						
Age (per 10 years)	1.02 (0.98-1.06)	.33	1.03 (0.95-1.11)	.46	1.01 (0.97-1.04)	.62
Male (reference: female)	1.46 (1.29-1.65)	<.001	1.62 (1.29-2.04)	<.001	0.98 (0.88-1.09)	.76
NH Black[f]	1.20 (0.98-1.47)	.08	1.43 (0.98-2.09)	.06	1.88 (1.62-2.18)	<.001
Hispanic	1.26 (1.06-1.51)	.01	1.58 (1.13-2.21)	.008	1.59 (1.37-1.84)	<.001
NH Asian	1.98 (1.61-2.43)	<.001	3.12 (2.18-4.45)	<.001	2.01 (1.63-2.47)	<.001
NH multiracial	1.11 (0.83-1.49)	.49	1.43 (0.84-2.43)	.18	1.11 (0.86-1.43)	.44
NH other (AI/AN and NH/PI)	0.78 (0.35-1.72)	.53	1.92 (0.68-5.40)	.22	1.84 (1.17-2.88)	.008
Bachelor's degree or higher	0.90 (0.79-1.02)	.10	0.82 (0.64-1.06)	.13	0.51 (0.45-0.57)	<.001
Disability	0.91 (0.71-1.16)	.44	1.59 (1.09-2.32)	.02	1.27 (1.07-1.50)	.005
Rural (reference: non-rural)	0.76 (0.60-0.96)	.02	0.71 (0.45-1.14)	.16	1.06 (0.90-1.26)	.48
Generative AI user (reference: nonuser)	1.35 (1.18-1.54)	<.001	1.02 (0.79-1.32)	.87	0.57 (0.50-0.64)	<.001
Panel 3: diet and exercise planning (n=10,374)						
Age (per 10 years)	1.09 (1.06-1.12)	<.001	1.00 (0.95-1.06)	.88	1.02 (0.98-1.06)	.44
Male (reference: female)	0.99 (0.90-1.08)	.82	1.17 (0.99-1.39)	.07	0.93 (0.82-1.05)	.23
NH Black[f]	1.01 (0.86-1.17)	.93	1.46 (1.11-1.93)	.007	1.77 (1.49-2.10)	<.001
Hispanic	0.97 (0.84-1.11)	.63	1.62 (1.28-2.06)	<.001	1.35 (1.14-1.61)	<.001
NH Asian	1.06 (0.89-1.27)	.51	2.18 (1.65-2.88)	<.001	1.77 (1.39-2.26)	<.001
NH multiracial	0.95 (0.76-1.18)	.64	1.11 (0.72-1.69)	.64	1.06 (0.79-1.42)	.69
NH other (AI/AN and NH/PI)	0.78 (0.47-1.32)	.36	0.96 (0.34-2.75)	.94	2.02 (1.23-3.30)	.005
Bachelor's degree or higher	1.11 (1.01-1.22)	.03	0.95 (0.79-1.15)	.62	0.53 (0.46-0.60)	<.001

XSL•FO
RenderX

Predictors	As much as a human professional, RRR (95% CI)	*P* value[c]	More than a human professional, RRR (95% CI)	*P* value	Not sure, RRR (95% CI)	*P* value
Disability	0.87 (0.73-1.03)	.12	1.18 (0.85-1.64)	.31	1.32 (1.09-1.59)	.004
Rural (reference: non-rural)	0.93 (0.79-1.08)	.34	0.59 (0.40-0.86)	.007	1.04 (0.86-1.27)	.66
Generative AI user (reference: nonuser)	1.78 (1.61-1.97)	<.001	2.19 (1.82-2.65)	<.001	0.61 (0.52-0.71)	<.001

[a]Models are estimated without survey weights.

[b]Reference outcome category: "Less than a human professional."

[c]Exact *P* values are reported for each estimate.

[d]Age is scaled per 10-year increment.

[e]NH: non-Hispanic.

[f]Reference group for race or ethnicity: NH White.

[g]AI/AN: American Indian or Alaska Native.

[h]NH/PI: Native Hawaiian or Pacific Islander.

Generative AI users reported higher trust across domains and were less likely to express uncertainty. Having a bachelor's degree or higher was associated with lower uncertainty across all 3 domains and with lower trust in AI relative to a human professional for mental health counseling. Male respondents were more likely to trust AI than female respondents were for health diagnosis and mental health counseling but not for diet and exercise planning. Disability was associated with superiority trust for health diagnosis and mental health counseling but not with parity trust. Rural residence showed inconsistent associations across domains.

Discussion

Principal Findings

Using nationally representative data from more than 10,000 US adults, this study examined trust in AI across 3 health domains and found the lowest levels of trust in AI for mental health counseling. The most pronounced racial and ethnic differences in trust were observed for mental health counseling. NH Black, Hispanic, and NH Asian adults were more likely than NH White adults to endorse parity trust, and these groups, along with NH multiracial adults, were more likely to endorse superiority trust. These differences were concentrated in superiority trust, with NH Asian adults being nearly 6 times more likely than NH White adults to trust AI more than a human counselor.

Patterns Across Health Domains

Public attitudes toward AI in health care are not monolithic. Diet and exercise planning, which may be perceived as information-based rather than relational, elicited higher trust than health diagnosis and mental health counseling. Prior work shows the same pattern: in 2022, approximately 60% of US adults reported discomfort with their own health care professional relying on AI [23], with higher acceptance for noninvasive uses such as nutritional monitoring and lower acceptance for mental health monitoring [10]. Perceived needs for empathy, clinical judgment, and therapeutic alliance may shape these views; individuals can be more averse to algorithms for subjective tasks than for objective, data-driven ones [8]; furthermore, "uniqueness neglect" may also contribute to these views [9].

Mental health counseling relies largely on the relationship between a clinician and a patient; we observed that it is a domain in which most adults trust a human professional more than AI. Utilization data are consistent with this pattern; in 2024, 5% of US college students had used chatbots for mental health support [24]. In a survey of 926 undergraduates, 31% trusted generative AI for mental health information and 13% for mental health decisions [25], and rates of mental health AI use differ by race and ethnicity in college samples [26]. In a 2023 survey of adults in India, general mental health tips were the most commonly endorsed purpose for using AI-driven mental health interventions (51.7%) and treatment the least (5.2%) [27]. Concerns about AI platforms for mental health support stem from model designs that validate and assist users regardless of the content of their interaction. The facilitation of suicidal and disordered eating behaviors and the reinforcement of delusional thoughts, described as "AI-induced psychosis," have received substantial media attention [28]. Greater uptake of AI and rapid improvement in AI applications may further increase public acceptance.

Racial and Ethnic Patterns in AI Trust

Prior survey research suggests that racial and ethnic minority groups hold more favorable views toward AI in health care. Asian respondents show higher odds of trusting health systems to use AI responsibly [16], Black and Hispanic respondents hold higher expectations for AI in improving health care access and affordability [29], and Black adults are more likely to perceive benefits of AI for mental health [30].

We consider several potential explanations for this pattern. Because the UAS in this wave did not measure access barriers, discrimination, stigma, satisfaction with care, technology attitudes, expectations about AI, or socioeconomic and generational circumstances, the explanations below should be read as hypotheses to be tested in future work rather than as mechanisms established by the present data. Possible explanations include documented barriers to mental health care, including cost, shortages of human professionals, a lack of

XSL•FO
RenderX

culturally trained human professionals, and stigma [31,32]; the capacity of digital tools to bypass such barriers through personalized, anonymous, and immediate support [33], which may hold promise for reducing disparities among minoritized groups [34]; discrimination or cultural mismatch in health care settings, which have been associated with lower trust in health systems' AI use [16]; and cultural norms that emphasize emotional restraint and stigmatize help-seeking, which may make a private, nonjudgmental option appealing [35-37].

Although AI models are not without concerns about racism and bias [38], and research has sought to address these issues [39], biases may be reduced compared to those in human responses. For instance, AI did not demonstrate preferential recommendations for opioid treatment based on race or ethnicity and sex [40]. Our findings are consistent with prior survey work reporting more favorable views of AI in health care among racial and ethnic minority groups [16,29,30] and extend that work to a nationally representative adult population.

The inclusion of the "not sure" option in the multinomial model is a notable strength. Racial and ethnic minority groups were more likely than NH White adults both to trust AI and to express uncertainty. Relative to NH White adults, these groups were less likely to default to trusting AI less than a human professional. Our approach and findings are aligned with research where modeling an "ambivalent" middle category alongside acceptance or refusal reveals racial and ethnic patterns [41,42]. The "not sure" response is unlikely to be psychologically uniform; it may reflect limited familiarity with AI, ambivalence, low health literacy, or uncertainty about the question itself. Clarifying these sources of uncertainty and their clinical implications requires further study.

The distinction between parity and superiority trust provides nuance. For mental health counseling, the racial and ethnic effects were stronger for superiority trust than for parity trust. NH Asian adults were approximately 2.5 times more likely than NH White adults to trust AI as much as a human counselor but nearly 6 times more likely to trust AI more than a human counselor. The pattern is notable given the current debate about the substitution of human clinicians with AI [43]. Superiority trust should not be equated with a preference for AI over human care; it may instead reflect curiosity or an overestimation of what AI can currently do. Furthermore, adults have been shown to overtrust AI-generated medical advice even when its accuracy is low [20]. It is also possible that the observed pattern reflects a higher prevalence of negative experiences with human clinicians in these populations. Clarifying whether the observed pattern is better explained by a positive appraisal of AI or a negative appraisal of human clinicians is a question for studies that measure those patient experiences directly.

Generative AI Use, Education, Disability, and Sex Patterns

Generative AI users were less likely to report uncertainty, which may indicate that familiarity with general-purpose AI tools makes it easier to form a judgment about AI for health tasks; because the design is cross-sectional, however, the association is equally consistent with the reverse relationship: adults who are more trusting of AI may be more inclined to use it. Education was associated with lower trust for mental health counseling, with adults holding a bachelor's degree or higher being less likely to endorse both parity and superiority trust. This association did not extend to the other 2 domains, and for diet and exercise planning it was reversed. Higher education has been associated with greater scrutiny of AI in health care [11,21], although it is also associated with higher perceived value in improving health [21]. Education may confer greater literacy, awareness of AI limitations, or stronger relationships with existing human professionals [44,45]. Disability status was not associated with parity trust for mental health counseling, yet it was associated with superiority trust. Given the data, however, we could not determine the reasons why disability was associated with superiority trust and not with parity trust. Sex differences were domain-specific: male respondents were more likely than female respondents to trust AI for health diagnosis at both the parity and superiority levels and for mental health counseling primarily at the parity level, with no difference for diet and exercise planning. The sex gap in AI trust for health is among the most consistent findings, with women reporting lower trust or comfort across multiple nationally representative and international samples [10,16,23,29]. The presence of this association for health diagnosis and mental health counseling but its absence for diet and exercise planning suggests that the difference may be tied to perceived clinical stakes rather than to overarching skepticism toward AI.

Implications for Health AI Design, Implementation, Evaluation, and Governance

These patterns have implications for how health AI is designed and overseen. Trust was lowest for mental health counseling, a domain in which AI-based tools are proliferating rapidly [3]. Given this, tools for counseling warrant human oversight, including clear routes to a clinician and clear notification when a patient is not interacting with a human professional [28,46]. Comparative trust also differed by race and ethnicity. Because AI models can still produce biased or inadequate responses in health contexts when they are trained on biased input [38,39], the groups reporting the highest trust may be the groups most affected if a tool performs poorly, which could widen existing disparities. Reporting tool performance by racial and ethnic group, rather than in aggregate alone, is therefore warranted. Uncertainty regarding comparative trust was also commonly endorsed in every domain, and more commonly among racial and ethnic minority groups; developers would do well to seek subgroup-specific concerns during evaluation and before deployment.

Hypotheses for Future Research

Given these findings, we identify 3 directions for future research. Future work should determine whether the racial and ethnic differences observed in this study arise from barriers to mental health care access. We would hypothesize that differences would be larger among adults who report unmet mental health care needs, cost barriers, or long waits, and would be smaller once those measures are taken into account [31,32]. Work is also needed to determine whether the differences arise from past experiences with clinicians, where we would hypothesize that they track with reported discrimination in care, perceived

XSL•FO
RenderX

cultural match with human professionals, and satisfaction with past treatment [16]. Finally, studies should test whether cultural norms around help-seeking are involved. We would hypothesize larger differences among adults who report stronger norms of emotional restraint or greater stigma toward professional help [35-37]. Studies that collect these experiences should consider including preference and intended use alongside trust, since an adult who trusts AI more than a human professional for a task may still choose a human clinician. Future work would benefit from validated multi-item measures that specify the referent, the task, and the conditions of use when it comes to using AI in health care.

Limitations

There are several limitations. The cross-sectional design precludes causal inference, and the associations may be confounded by unmeasured factors, such as mental health status, health care experiences, or attitudes toward technology. The sex variable was binary, and the gender identity and sexual orientation of respondents were unavailable. Research suggests that LGBTQ+ (lesbian, gay, bisexual, transgender, queer, and other sexual orientations and gender identities) individuals may hold different attitudes toward AI [47]; the inability to evaluate these differences using this nationally representative dataset is a notable gap. Several constructs that likely bear on trust in AI for mental health were unavailable in the UAS, including prior mental health service use, unmet need, perceived discrimination in health care, and satisfaction with human clinicians; their absence limits our ability to determine the specific explanations above and may confound the observed associations. Because the design is cross-sectional, all associations, including that between generative AI use and trust, are noncausal and potentially bidirectional. Adults who trust AI as much as or more than a human professional for a task may still prefer a human clinician; the UAS did not ask about preference or intended use. The trust items were general and hypothetical and did not specify a particular AI system or describe its accuracy, privacy safeguards, or degree of human oversight, so responses may not correspond to trust in any specific deployed tool. The "not sure" category is unlikely to represent a single psychological state. Finally, estimates for the NH other and NH multiracial groups, which had small samples, are imprecise and warrant particular caution. The dichotomization of education, while it preserved statistical power across all 6 race and ethnicity groups, may likewise obscure finer gradations in the association between education and AI trust.

Conclusions

Trust in AI for health varies substantially across domains and demographic groups, with mental health counseling representing the domain of lowest overall trust but showing the largest racial and ethnic differences. NH Black, Hispanic, NH Asian, and NH multiracial adults are more likely than NH White adults to trust AI more than a human professional for mental health counseling. Racial and ethnic minority groups are also more likely to express uncertainty about AI, representing a broader departure from the pattern of lower reported trust that predominates among NH White adults. These data cannot distinguish whether higher trust leads to greater use of AI mental health tools. However, should trust be linked to greater use of AI for mental health, the groups reporting higher trust would gain most from tools that work well and would bear a greater share of the risk from tools that do not. Toward this end, equitable design, validation, and monitoring of these tools are a priority [46]. Clinicians, policymakers, and AI developers should attend to these differences as AI is integrated into health care delivery.

Acknowledgments

The project described in this paper used data from surveys administered by the Understanding America Study (UAS), which is maintained by the Center for Economic and Social Research (CESR) at the University of Southern California. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the University of Southern California (USC) or UAS.

Data Availability

The data used in this study are publicly available through the Understanding America Study (UAS) [48]. Access requires registration and completion of a Data User Agreement. To support reproducibility, the analytic code has been deposited in a public repository (OSF [49]) and is also available from the corresponding author. The code can be run by any investigator who has obtained access to the UAS data in accordance with the Data User Agreement.

Funding

The authors declare that no financial support was received for this work.

Generative AI Use

In the first phase, an AI-assisted programming tool, Claude Code (Claude Opus 4.6; Anthropic PBC), generated draft Python scripts to implement the prespecified analytic models. In a subsequent phase, a separate large language model, ChatGPT (GPT-5; OpenAI), was used to support code inspection by identifying potential logical inconsistencies or deviations from the analytic plan, again without exposure to study data. Generative AI use was limited to the development and inspection of analytic code.

Authors' Contributions

Conceptualization: CHL, TY

XSL•FO
RenderX

Data interpretation: CHL, AK, TY
Statistical analysis: CHL
Study design: CHL, TY
Writing—original draft: CHL
Writing—review and editing: CHL, AK, TY
CHL had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of Interest

CHL is a scientific advisor on youth mental health for Surgo Health and The Asian American Foundation and for a Harvard Medical School youth mental health project funded by the Manton Foundation. All other authors have no conflicts of interest to declare.

References

1. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. Jan 2019;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
2. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med. Jan 2022;28(1):31-38. [doi: 10.1038/s41591-021-01614-0] [Medline: 35058619]
3. Cruz-Gonzalez P, He AW, Lam EP, Ng IM, Li MW, Hou R, et al. Artificial intelligence in mental health care: a systematic review of diagnosis, monitoring, and intervention applications. Psychol Med. Feb 06, 2025;55:e18. [doi: 10.1017/S0033291724003295] [Medline: 39911020]
4. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. J Med Internet Res. Jun 19, 2020;22(6):e15154. [FREE Full text] [doi: 10.2196/15154] [Medline: 32558657]
5. Steerling E, Siira E, Nilsen P, Svedberg P, Nygren J. Implementing AI in healthcare-the relevance of trust: a scoping review. Front Health Serv. Aug 24, 2023;3:1211150. [FREE Full text] [doi: 10.3389/frhs.2023.1211150] [Medline: 37693234]
6. Cadario R, Longoni C, Morewedge CK. Understanding, explaining, and utilizing medical artificial intelligence. Nat Hum Behav. Dec 2021;5(12):1636-1642. [doi: 10.1038/s41562-021-01146-0] [Medline: 34183800]
7. Witkowski K, Dougherty RB, Neely SR. Public perceptions of artificial intelligence in healthcare: ethical concerns and opportunities for patient-centered care. BMC Med Ethics. Jun 22, 2024;25(1):74. [FREE Full text] [doi: 10.1186/s12910-024-01066-4] [Medline: 38909180]
8. Castelo N, Bos MW, Lehmann DR. Task-dependent algorithm aversion. J Mark Res. Jul 15, 2019;56(5):809-825. [doi: 10.1177/0022243719851788]
9. Longoni C, Bonezzi A, Morewedge CK. Resistance to medical artificial intelligence. J Consum Res. May 3, 2019;46(4):629-650. [doi: 10.1093/jcr/ucz013]
10. Kauttonen J, Rousi R, Alamäki A. Trust and acceptance challenges in the adoption of AI applications in health care: quantitative survey analysis. J Med Internet Res. Mar 21, 2025;27:e65567. [FREE Full text] [doi: 10.2196/65567] [Medline: 40116853]
11. Kühne S, Jacobsen J, Legewie N, Dollmann J. Attitudes toward AI usage in patient health care: evidence from a population survey vignette experiment. J Med Internet Res. May 27, 2025;27:e70179. [FREE Full text] [doi: 10.2196/70179] [Medline: 40424613]
12. Tyson A, Kikuchi E. Growing public concern about the role of artificial intelligence in daily life. Pew Research Center. Aug 28, 2023. URL: https://www.pewresearch.org/short-reads/2023/08/28/growing-public-concern-about-the-role-of-artificial-intelligence-in-daily-life/ [accessed 2026-03-05]
13. Li Y, Wu B, Huang Y, Luan S. Developing trustworthy artificial intelligence: insights from research on interpersonal, human-automation, and human-AI trust. Front Psychol. Apr 17, 2024;15:1382693. [FREE Full text] [doi: 10.3389/fpsyg.2024.1382693] [Medline: 38694439]
14. Alasmari AA, Alruwaili RF, Alotaibi RF, Youssef IK, Asklany SA. Demographic influences on trust in artificial intelligence across cognitive domains: a statistical perspective. PLoS One. Nov 5, 2025;20(11):e0331003. [FREE Full text] [doi: 10.1371/journal.pone.0331003] [Medline: 41191587]
15. Institute of Medicine, Board on Health Sciences Policy, Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care. Nelson AR, Stith AY, Smedley BD, editors. Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care. Washington, DC. National Academies Press; 2002.
16. Nong P, Platt J. Patients' trust in health systems to use artificial intelligence. JAMA Netw Open. Feb 03, 2025;8(2):e2460628. [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.60628] [Medline: 39951270]
17. Cook BL, Trinh NH, Li Z, Hou SS, Progovac AM. Trends in racial-ethnic disparities in access to mental health care, 2004-2012. Psychiatr Serv. Jan 01, 2017;68(1):9-16. [FREE Full text] [doi: 10.1176/appi.ps.201500453] [Medline: 27476805]
18. Sheehan A, Walsh R, Liu R. Racial and ethnic trends in mental health service utilisation and perceived unmet need in the USA. J Epidemiol Community Health. Mar 08, 2024;78(4):228-234. [doi: 10.1136/jech-2023-220683] [Medline: 38242681]

XSL•FO
RenderX

19. Lee JD, See KA. Trust in automation: designing for appropriate reliance. Hum Factors. 2004;46(1):50-80. [doi: 10.1518/hfes.46.1.50_30392] [Medline: 15151155]
20. Shekar S, Pataranutaporn P, Sarabu C, Cecchi GA, Maes P. People overtrust AI-generated medical advice despite low accuracy. NEJM AI. Jun 2025;2(6):AIoa2300015. [doi: 10.1056/AIoa2300015] [Medline: 42932543]
21. Beets B, Newman TP, Howell EL, Bao L, Yang S. Surveying public perceptions of artificial intelligence in health care in the United States: systematic review. J Med Internet Res. Apr 04, 2023;25:e40337. [FREE Full text] [doi: 10.2196/40337] [Medline: 37014676]
22. Korn EL, Graubard BI. Epidemiologic studies utilizing surveys: accounting for the sampling design. Am J Public Health. Sep 1991;81(9):1166-1173. [doi: 10.2105/ajph.81.9.1166] [Medline: 1951829]
23. Tyson A, Pasquini G, Spencer A, Funk C. 60% of Americans would be uncomfortable with provider relying on AI in their own health care. Pew Research Center. Feb 22, 2023. URL: https://www.pewresearch.org/science/2023/02/22/60-of-americans-would-be-uncomfortable-with-provider-relying-on-ai-in-their-own-health-care/ [accessed 2026-03-05]
24. Rackoff GN, Zhang ZZ, Newman MG. Chatbot-delivered mental health support: attitudes and utilization in a sample of U.S. college students. Digit Health. Jan 17, 2025;11:20552076241313401. [FREE Full text] [doi: 10.1177/20552076241313401] [Medline: 39839954]
25. Liu CH, Yip T. College students' trust in generative AI for mental health. Psychiatr Serv. Nov 01, 2026;77(11):1089-1093. [doi: 10.1176/appi.ps.20260116] [Medline: 42527952]
26. Liu CH, Zhang W, Lou F, Zhao C, Chow A, Yip T. Clinical and sociodemographic predictors of AI use for mental health among college students. J Affect Disord. Nov 01, 2026;412:122058. [doi: 10.1016/j.jad.2026.122058] [Medline: 42217639]
27. Varghese MA, Sharma P, Patwardhan M. Public perception on artificial intelligence-driven mental health interventions: survey research. JMIR Form Res. Nov 28, 2024;8:e64380. [FREE Full text] [doi: 10.2196/64380] [Medline: 39607994]
28. NEDA suspends AI chatbot for giving harmful eating disorder advice. Psychiatrist.com. URL: https://www.psychiatrist.com/news/neda-suspends-ai-chatbot-for-giving-harmful-eating-disorder-advice/ [accessed 2026-03-05]
29. Nong P, Ji M. Expectations of healthcare AI and the role of trust: understanding patient views on how AI will impact cost, access, and patient-provider relationships. J Am Med Inform Assoc. May 01, 2025;32(5):795-799. [FREE Full text] [doi: 10.1093/jamia/ocaf031] [Medline: 40036944]
30. Benda N, Desai P, Reza Z, Zheng A, Kumar S, Harkins S, et al. Patient perspectives on AI for mental health care: cross-sectional survey study. JMIR Ment Health. Sep 18, 2024;11:e58462. [FREE Full text] [doi: 10.2196/58462] [Medline: 39293056]
31. Alegría M, Alvarez K, Ishikawa RZ, DiMarzio K, McPeck S. Removing obstacles to eliminating racial and ethnic disparities in behavioral health care. Health Aff (Millwood). Jun 01, 2016;35(6):991-999. [FREE Full text] [doi: 10.1377/hlthaff.2016.0029] [Medline: 27269014]
32. Eylem O, de Wit L, van Straten A, Steubl L, Melissourgaki Z, Danışman GT, et al. Stigma for common mental disorders in racial minorities and majorities a systematic review and meta-analysis. BMC Public Health. Jun 08, 2020;20(1):879. [FREE Full text] [doi: 10.1186/s12889-020-08964-3] [Medline: 32513215]
33. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR Mhealth Uhealth. Nov 23, 2018;6(11):e12106. [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]
34. Friis-Healy EA, Nagy GA, Kollins SH. It is time to REACT: opportunities for digital mental health apps to reduce mental health disparities in racially and ethnically minoritized groups. JMIR Ment Health. Jan 26, 2021;8(1):e25456. [FREE Full text] [doi: 10.2196/25456] [Medline: 33406050]
35. Kramer EJ, Kwong K, Lee E, Chung H. Cultural factors influencing the mental health of Asian Americans. West J Med. Sep 2002;176(4):227-231. [FREE Full text] [Medline: 12208826]
36. Masuda A, Boone MS. Mental health stigma, self-concealment, and help-seeking attitudes among Asian American and European American college students with no help-seeking experience. Int J Adv Couns. Sep 21, 2011;33:266-279. [doi: 10.1007/s10447-011-9129-1]
37. Shahid M, Weiss NH, Stoner G, Dewsbury B. Asian Americans' mental health help-seeking attitudes: the relative and unique roles of cultural values and ethnic identity. Asian Am J Psychol. 2021;12(2):138-146. [doi: 10.1037/aap0000230]
38. Hofmann V, Kalluri PR, Jurafsky D, King S. AI generates covertly racist decisions about people based on their dialect. Nature. Sep 2024;633(8028):147-154. [doi: 10.1038/s41586-024-07856-5] [Medline: 39198640]
39. Gabriel S, Puri I, Xu X, Malgaroli M, Ghassemi M. Can AI relate: testing large language model response for mental health support. In: Findings of the Association for Computational Linguistics: EMNLP 2024. New York, NY. Association for Computational Linguistics; 2024:2206-2221.
40. Young CC, Enichen E, Rao A, Succi MD. Racial, ethnic, and sex bias in large language model opioid recommendations for pain management. Pain. Mar 01, 2025;166(3):511-517. [doi: 10.1097/j.pain.0000000000003388] [Medline: 39283333]
41. Waters EA, Kiviniemi MT, Hay JL, Orom H. Dismissing "don't know" responses to perceived risk survey items threatens the validity of theoretical and empirical behavior-change research. Perspect Psychol Sci. May 2022;17(3):841-851. [FREE Full text] [doi: 10.1177/17456916211017860] [Medline: 34813719]

XSL•FO
RenderX

42. Latkin C, Dayton LA, Yi G, Konstantopoulos A, Park J, Maulsby C, et al. COVID-19 vaccine intentions in the United States, a social-ecological framework. Vaccine. Apr 15, 2021;39(16):2288-2294. [FREE Full text] [doi: 10.1016/j.vaccine.2021.02.058] [Medline: 33771392]
43. Morehead D. Artificial intelligence will never replace psychiatry—but it will try to destroy it. Psychiatric Times. 2025. URL: https://www.psychiatrictimes.com/view/artificial-intelligence-will-never-replace-psychiatry-but-it-will-try-to-destroy-it [accessed 2026-03-05]
44. Tsai TI, Yu WR, Lee SY. Is health literacy associated with greater medical care trust? Int J Qual Health Care. Aug 01, 2018;30(7):514-519. [doi: 10.1093/intqhc/mzy043] [Medline: 29608676]
45. White RO, Osborn CY, Gebretsadik T, Kripalani S, Rothman RL. Health literacy, physician trust, and diabetes-related self-care activities in Hispanics with limited resources. J Health Care Poor Underserved. Nov 2013;24(4):1756-1768. [FREE Full text] [doi: 10.1353/hpu.2013.0177] [Medline: 24185168]
46. Scholich T, Barr M, Wiltsey Stirman S, Raj S. A comparison of responses from human therapists and large language model-based chatbots to assess therapeutic communication: mixed methods study. JMIR Ment Health. May 21, 2025;12:e69709. [FREE Full text] [doi: 10.2196/69709] [Medline: 40397927]
47. Haimson OL, Mayworm SR, Ingber AS, Andalibi N. AI attitudes among marginalized populations in the U.S.: nonbinary, transgender, and disabled individuals report more negative AI attitudes. In: FAccT '25: Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency. New York, NY. Association for Computing Machinery; 2025:1224-1237.
48. Understanding America Study. URL: https://uasdata.usc.edu/index.php [accessed 2026-11-10]
49. UAS AI Trust. OSF. 2026. URL: https://osf.io/thfn2/overview [accessed 2026-11-10]

Abbreviations

BRANY: Biomedical Research Alliance of New York
CESR: Center for Economic and Social Research
LGBTQ+: lesbian, gay, bisexual, transgender, queer, and other sexual orientations and gender identities
NH: non-Hispanic
RRR: relative risk ratio
RUCA: Rural-Urban Commuting Area
UAS: Understanding America Study

Edited by K Williams; submitted 14.Apr.2026; peer-reviewed by SS Sansgiry, TAK Manne; comments to author 24.Jul.2026; revised version received 30.Sep.2026; accepted 04.Nov.2026; published 23.Nov.2026

Please cite as:
Liu CH, Koire A, Yip T
Racial and Ethnic Differences in Trust in AI Relative to Human Professionals for Health Care: Nationally Representative Cross-Sectional Study
J Med Internet Res 2026;28:e98253
URL: https://www.jmir.org/2026/1/e98253
doi: 10.2196/98253
PMID:

©Cindy H Liu, Amanda Koire, Tiffany Yip. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 23.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX