IJPH | SSPH+ SWISS SCHOOL OF PUBLIC HEALTH

TYPE Original Article
PUBLISHED 16 November 2026
DOI 10.3389/ijph.2026.1609854

Check for updates

OPEN ACCESS

EDITED BY
Katherine Robsky,
Georgetown University, United States

REVIEWED BY
Felix Rigoli,
Nucleo de Bioetica, Brazil
One reviewer who chose to remain anonymous

*CORRESPONDENCE
Ying Hu,
huying-2009@scu.edu.cn

†These authors have contributed equally to this work and share first authorship

This Original Article is part of the IJPH Special Issue "Artificial Intelligence (AI) And Public Health"

RECEIVED 07 April 2026
REVISED 29 September 2026
ACCEPTED 29 October 2026
PUBLISHED 16 November 2026

CITATION
Zhong F, Tang X, Yang Z and Hu Y (2026) Platform-integrated AI assistance and online medical consultation service quality: evidence from physician-month panel data in China.
Int. J. Public Health 71:1609854.
doi: 10.3389/ijph.2026.1609854

COPYRIGHT
© 2026 Zhong, Tang, Yang and Hu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Platform-integrated AI assistance and online medical consultation service quality: evidence from physician-month panel data in China

Fangyu Zhong[1,2†], Xiumei Tang[3,4†], Zhilin Yang[2] and Ying Hu[5,6*]

[1]Business School, Sichuan University, Chengdu, China, [2]College of Business, City University of Hong Kong, Hong Kong SAR, China, [3]State Key Laboratory of Respiratory Health and Multimorbidity, Sichuan University, Chengdu, China, [4]Institute of Respiratory Health, Frontiers Science Center for Disease-related Molecular Network, West China Hospital, Sichuan University, Chengdu, China, [5]Department of Obstetrics and Gynecology; Key Laboratory of Birth Defects and Related Diseases of Women and Children; Ministry of Education, West China Second University Hospital, Sichuan University, Chengdu, China, [6]Reproductive Endocrinology and Regulation Laboratory, West China Second University Hospital, Sichuan University, Chengdu, China

Objectives: To examine the association between platform-integrated artificial intelligence (AI) assistance and online medical consultation service quality, considering consultation efficiency, conversation depth, and physician seniority.
Methods: We analyzed a 6-year panel of 77 physicians (5,544 physician-month observations) using fixed-effects regressions. Platform-observed online consultation service quality and efficiency are measured using entropy-weighted indices. Mediation analysis with Sobel and bootstrap tests, and conditional indirect effect models were conducted, followed by heterogeneity analysis across physician seniority.
Results: AI assistance use was positively associated with consultation quality ($\beta = 0.666$, $p < 0.001$) and consultation efficiency ($\beta = 0.857$, $p = 0.002$). An indirect association through efficiency was observed. This indirect association emerged only at medium and high levels of physician–patient conversation depth. Among experts, AI use was associated with efficiency and indirectly with quality; whereas among regular physicians, it was associated with quality but not efficiency.
Conclusion: Platform-integrated AI assistance was positively associated with online consultation service quality. These relationships vary with conversation depth and physician seniority, highlighting the importance of considering both interaction characteristics and clinician heterogeneity when evaluating AI integration.

KEYWORDS
consultation efficiency, consultation quality, online medical consultation, patient-physician interaction, platform-integrated AI assistance

Introduction

Artificial intelligence is increasingly incorporated into healthcare delivery [1, 2], and online medical consultation provides a practical setting for its integration into routine service workflows. In China, licensed internet hospitals provide regulated online medical services either as extensions of physical medical institutions or as online entities anchored to licensed facilities [3]. Public data indicate that China had 3,756 internet hospitals by 2025,

handling more than 100 million online consultations annually [4]. These figures establish the scale of online consultation as a major supplementary channel of care and an important setting for platform-integrated AI assistance.

Platform-integrated AI assistance refers here to functions embedded in consultation workflows, including pre-consultation intake and structured history-taking, symptom checking and self-assessment, and intelligent triage. It does not include autonomous diagnostic systems or external AI tools independently accessed by physicians. Prior research has primarily evaluated AI in controlled settings or isolated clinical tasks, particularly diagnostic accuracy, decision support, and time efficiency [2, 5–7]. Although these studies establish technical potential, evidence about routine platform use and multidimensional, patient-facing service quality remains limited [8–10]. The potential efficiency–quality tension, together with the roles of physician–patient interaction and physician seniority, remains insufficiently understood [11].

To address these gaps, this study examines how physicians' use of platform-integrated AI assistance is associated with platform-observed online consultation service quality and clarifies the process through which this relationship may arise. Service productivity theory suggests that, unlike manufacturing productivity, service productivity cannot be evaluated solely by output because changes in service processes and resource use may also alter the quality experienced by customers [12]. Within this framework, AI assistance may support consultation quality by improving information processing and workflow efficiency. Consultation efficiency is therefore a proposed pathway. Service co-production theory also recognizes patients as contributors whose information and effort shape service performance [13, 14]. Therefore, conversation depth may condition both stages of the indirect pathway. Deeper exchanges increase information-processing demands, potentially strengthening the efficiency value of AI assistance. They also provide context and opportunities for clarification and verification, potentially enabling efficiency gains to support service completion and favorable patient evaluations. Physician seniority may further shape these associations because AI can serve different functions depending on clinicians' existing expertise. The study therefore examines consultation efficiency as a potential pathway, conversation depth as a two-stage boundary condition, and physician seniority as a source of heterogeneity.

We investigate these relationships using a 6-year balanced panel comprising 5,544 physician-month observations from 77 physicians across 37 departments on a leading online consultation platform. Platform-observed consultation service quality and efficiency are measured using entropy-weighted composite indices derived from routinely recorded platform indicators. This data-driven approach reduces reliance on researcher-assigned weights and enables consistent analysis across a large physician-month panel [15]. However, because the weights reflect statistical informativeness rather than theoretical or clinical importance, the quality index should be interpreted as a standardized measure of platform-observed service and process quality rather than comprehensive clinical quality of care.

Our analysis proceeds in four steps. First, we estimate the overall association between AI-assistance use intensity and consultation quality. Second, we examine the indirect association through consultation efficiency. Third, we assess how patient-physician conversation depth conditions the proposed indirect pathway, thereby highlighting the importance of patient-provider interaction. Finally, building on the notion that AI serves different functions depending on the physicians' baseline knowledge [11], we conduct a heterogeneity analysis by physician seniority to examine whether the estimated relationships differ with professional experience.

This study makes three contributions. First, it provides longitudinal, real-world evidence on the relationship between physicians' use of platform-integrated AI assistance and patient-facing service outcomes in routine online consultation. Second, it examines consultation efficiency as a potential pathway while explicitly considering whether conversation depth conditions both the AI assistance–efficiency relationship and the efficiency–quality relationship. Third, it identifies differences across physician seniority, thereby showing how the role of AI assistance may vary with physicians' professional experience. Together, these contributions offer a more process-oriented and context-sensitive account of physician–AI interaction in online medical services.

Methods

Data and sample

The dataset comprises monthly records for 77 physicians from 37 departments on an online consultation platform in China, from January 2019 through December 2024, yielding 5,544 physician-month observations. Physician titles range from attending physicians, chief physicians, to 1–5 level senior experts. Records were de-identified, and consultation content and patient demographics were removed or aggregated. The research protocol received approval from the institutional review board (Ethics Committee of West China Hospital, No. 20251938, approved on 18 November 2025). All study procedures were in accordance with the 1964 Helsinki Declaration and its subsequent amendments or comparable ethical standards [16].

Variables

Dependent variable

The primary dependent variable is platform-observed online consultation service quality (Quality). It captures patient-facing evaluations and administrative indicators recorded by the online consultation platform through four indicators: positive feedback rate, monthly patient commendations, medical advice completion rate, and refund rate. The first three are positive indicators; refunds are negative. We combine them using the entropy-weighting method [17] and rescale the index to 0–100. This approach quantifies the informational value of each underlying indicator by measuring its dispersion across observations: a smaller entropy value indicates greater variability among samples, signifying that the indicator provides more discriminative information and should thus receive a higher weight in the composite measure [18]. By deriving weights from the observed dispersion of the constituent indicators, the entropy method reduces reliance on researcher-assigned weights. However, these data-

derived weights reflect the statistical information contributed by each indicator rather than its theoretical or clinical importance. Calculation details and constituent-indicator statistics are provided in Supplementary Section A and Supplementary Table S1.

Independent variable

The core explanatory variable in this study is the use rate of platform-integrated AI assistance (*lnAI)*, which is measured as the percentage of a physician's monthly online consultations that utilize the platform-integrated AI tools. The AI assistance examined in this study comprises a bundle of workflow-level functions natively integrated into the platform rather than a single diagnostic model. These functions support: (1) pre-consultation intake and structured history-taking, (2) symptom checking and self-assessment, and (3) intelligent triage. The functions are embedded within the platform's consultation process, and their use is recorded in the platform's operational database. A consultation is coded as AI-assisted when the platform record assigns it the AI-assistance indicator; each consultation contributes at most one AI-assisted consultation to the numerator. For physician i in month t, the AI-assisted consultation rate is calculated using Equation 1:

$$AIRate_{i,t} = \frac{N_{i,t}^{AI}}{N_{i,t}^{total}} \times 100\% \quad (1)$$

where $N_{i,t}^{AI}$ *is the number of consultations classified by the platform as AI-assisted and* $N_{i,t}^{total}$ is the physician's total number of consultations in that month. This continuous metric ranges from 0% to 15% in our sample. To address its zero-inflated and right-skewed distribution while capturing the likely diminishing marginal returns of technology adoption, we apply a natural logarithmic transformation after adding one [19, 20].

Mediating variable

The mediating variable in this study is consultation efficiency (*Efficiency*). Analogous to the construction of the dependent variable, we measure consultation efficiency using a composite index derived via the entropy-weighting method. The indicators include: (1) the monthly consultation volume (*Consult_volume*); (2) the average consultation duration (*Duration*); and (3) the median time to first response (*First_ Response*). Following the same calculating procedure used to construct the quality index, we generate a composite efficiency score for each observation.

Moderating variable

The moderating variable in this study is conversation depth (*Depth*), operationalized as the natural logarithm of the median number of message exchanges per consultation. This measure reflects the intensity and elaborateness of physician–patient interaction, where a higher number of message rounds typically indicates more thorough inquiry, detailed explanation, and iterative clarification between the physician and the patient.

Control variables

To isolate the relationship between AI assistance and consultation outcomes, we incorporate a comprehensive set of control variables across physician, patient, and platform levels. At the physician level, we control for factors affecting workload and availability, including (1) the number of online days per month (*Online*) and (2) the ratio of consultations conducted during nighttime hours (*Night*). At the patient level, we control for case-mix differences by including (3) the ratio of urgent cases (*Urgent*) and (4) the ratio of first consultations (*First).* At the platform level, we account for operational and promotional factors by controlling for (5) months with promotional campaigns (*Promo_month*), (6) coupon usage rate (*Coupon*), and (7) intelligent triage accuracy (*Triage*). Detailed variable definitions are provided in Supplementary Table S2.

Estimation models

Our baseline model estimates the direct association between Platform-Integrated AI Assistance and Platform-observed consultation service quality as specified in Equation 2:

$$\text{Quality}_{it} = \beta_1 + \beta_2 \ln\text{AI}_{it} + \mathbf{X}_{it}\boldsymbol{\gamma} + \mu_i + \epsilon_{it} \quad (2)$$

Where Quality_{it} is the entropy-weighted consultation service quality score for physician i in month t. $\ln \text{AI}_{it}$ is the natural logarithm of the AI assistance use rate. β_1 and β_2 are the parameters to be estimated. $\mathbf{X}_{it}$ is a vector of control variables at the physician, patient, and platform levels. μ_i represents physician fixed effects, controlling for physician characteristics. ε_{it} is the random error term.

To explore the potential mediating role of efficiency in the relationship between AI assistance and medical consultation quality, this paper utilizes a three-step mechanism test model outlined by Judd and Kenny [21] and Baron and Kenny [22], and conducts both Sobel and bootstrap sampling tests (with 2,000 bootstrap replications). The mediation model is expressed in Equation 3:

$$\begin{cases} \text{Quality}_{it} = \beta_{10} + \beta_{11}\ln\text{AI}_{it} + \mathbf{X}_{it}\boldsymbol{\gamma} + \mu_i + \epsilon_{it}^{(1)} \\ \text{Efficiency}_{it} = \beta_{20} + \beta_{21}\ln\text{AI}_{it} + \mathbf{X}_{it}\boldsymbol{\delta} + \mu_i + \epsilon_{it}^{(2)} \\ \text{Quality}_{it} = \beta_{10}' + \beta_{11}'\ln\text{AI}_{it} + \beta_{12}\text{Efficiency}_{it} + \mathbf{X}_{it}\boldsymbol{\zeta} + \mu_i + \epsilon_{it}^{(3)} \end{cases} \quad (3)$$

The product of the coefficients β_{21} and β_{12} represents the estimated indirect association of lnAI with Quality through Efficiency.

To examine whether conversation depth conditions the indirect association between AI assistance and consultation quality, following the methodology of Preacher et al. [23], we extend the mediation model by incorporating interaction terms into both stages of the proposed pathway as specified in Equation 4:

$$\begin{cases} \text{Quality}_{it} = \beta_{30} + \beta_{31}\ln AI_{it} + \beta_{32}Depth_{it} + \beta_{33}(\ln AI_{it} \times Depth_{it}) + X_{it}\Phi + \mu_i + \epsilon_{it}^{(4)} \\ \text{Efficiency}_{it} = \beta_{40} + \beta_{41}\ln AI_{it} + \beta_{42}Depth_{it} + \beta_{43}(\ln AI_{it} \times Depth_{it}) + X_{it}\Psi + \mu_i + \epsilon_{it}^{(5)} \\ \text{Quality}_{it} = \beta_{30}' + \beta_{31}'\ln AI_{it} + \beta_{32}'Depth_{it} + \beta_{33}'(\ln AI_{it} \times Depth_{it}) \\ \quad + \beta_{34}\text{Efficiency}_{it} + \beta_{35}(\text{Efficiency}_{it} \times Depth_{it}) + X_{it}\Omega + \mu_i + \epsilon_{it}^{(6)} \end{cases} \quad (4)$$

This specification allows the indirect association between AI assistance and consultation quality through efficiency to vary across different levels of conversation depth. This conditional indirect effect at a given value of $Depth$ can be expressed as $f(\hat{\theta}|Depth) = (\beta_{34} + \beta_{35}Depth)(\beta_{41} + \beta_{43}Depth)$.

TABLE 1 Descriptive statistics (China, 2019–2024).

Variable	N	Mean	SD	Min	Max
Quality	5,544	41.251	19.178	0.000	100.000
Efficiency	5,544	44.912	15.237	0.000	100.000
lnAI	5,544	1.423	0.716	0.000	2.773
Depth	5,544	2.096	0.215	1.545	2.635
Online	5,544	8.592	1.246	8	18
Night (%)	5,544	19.547	4.584	8.000	36.360
Urgent (%)	5,544	6.711	1.894	2.740	15.250
First (%)	5,544	78.783	0.611	76.200	80.900
Triage (%)	5,544	82.355	4.553	73.440	95.000
Coupon (%)	5,544	7.487	2.396	3.000	17.640
Promo_month	5,544	0.083	0.276	0	1
Department	5,544	20.117	10.641	1	37
Seniority	5,544	1.468	0.499	1	2
Expert = 1	2,952				
Regular physician = 2	2,592				

Quality: Consultation service quality; Efficiency: Consultation service efficiency; lnAI: The log-transformed usage rate of the platform-integrated AI assistance; Depth: Patient-physician conversation depth; Online: Physician online days; Night (%): Night consultation ratio; Urgent (%): Urgent case ratio; First (%): First consultation ratio; Triage (%): Intelligent triage accuracy rate; Coupon (%): Discount coupon use rate; Promo_month: promotional month indicator (0-1).
Data is rounded to three decimal places.

Results

Descriptive statistics

Table 1 presents the descriptive statistics for the key variables used in the analysis based on a balanced physician-month panel of 5,544 observations (77 physicians over 72 months from 2019-2024). More than half of the physicians hold senior expert titles. The sample covers 37 different departments, representing a wide range of clinical contexts.

The primary outcome, the composite consultation service quality score, averages 41.25 (SD = 19.18) on a 0–100 scale, indicating moderate overall consultation quality with substantial variation across observations. Its four constituent indicators generally exhibit high averages but limited dispersion (as shown in Supplementary Table S1). This limited variance in the underlying metrics highlights the value of the entropy-weighting method in constructing a sensitive composite score.

The core independent variable, the natural logarithm of the AI-assistance ratio (lnAI), has a mean of 1.42 (SD = 0.72), corresponding to a raw AI usage ratio of approximately 4.14% on average (SD = 3.7%), indicating a right-skewed distribution in which a small number of physicians use AI substantially more often than the rest.

Additional descriptive statistics are reported in Table 1.

Baseline regression

Table 2 reports the baseline estimation results. Among them, column (1) reports the results of a pooled OLS regression with no fixed effects or control variables. The coefficient of lnAI is positive and significant ($\beta = 0.907$, $p = 0.012$). Fixed Effects (FE) are introduced in column (2) and retained in subsequent models. As shown in columns (2) through (5), with the sequential inclusion of fixed effects and control variables, the coefficient of lnAI gradually declines from 0.801 ($p < 0.001$) to 0.666 ($p < 0.001$), while remaining statistically significant.

Robustness test

This study also conducts a series of robustness checks (Table 3). First, column (1) introduces a two-way fixed-effects model with physician and month fixed effects, controlling for time trends. The coefficient for lnAI remains stable and significant at 0.684 ($p = 0.043$), nearly identical to our baseline estimate, suggesting that the main finding is not solely driven by the specific parametrization of time trends.

Recognizing that unobserved factors may correlate within medical departments, we re-estimate the model with department-clustered standard errors in Column (2). The estimated coefficient for lnAI remains statistically significant ($\beta = 0.666$, $p = 0.002$), demonstrating that our inference is robust to potential intra-departmental error correlation.

In Column (3), we replace the log-transformed measure with the raw AI usage rate (AI_rate). The coefficient is 0.105 ($p = 0.005$), supporting a positive and significant effect. The magnitude of this linear effect is economically consistent with our semi-elasticity estimate from the log-linear model.

To mitigate the potential issue of endogeneity stemming from reverse causality or omitted variable bias, we employ a Two-Stage Least Squares (2SLS) instrumental variable approach. We use the average usage rate of platform-integrated AI assistance among other physicians in the same department (excluding the focal physician) as a benchmark. This captures peer effects in technology adoption—physicians are more likely to use AI when their colleagues do, while colleagues' AI usage is unlikely to directly affect a physician's consultation quality except through their own AI use behavior. The first-stage regression in Column (4) suggests a positive and significant relationship between the instrument peer_AI and the endogenous lnAI ($\beta = 0.148$, $p < 0.001$). The second-stage result in Column (5) yields a coefficient for the instrumented lnAI of 0.508 ($p = 0.088$).

To further assess whether the estimated association was driven by changes in the broader AI environment during the study period, we conducted several additional temporal sensitivity analyses, as reported in Columns (6)–(8) of Table 3 Panel B.

Column (6) restricts the sample to January 2019–November 2022, before the widespread public availability of large language models (LLM). The coefficient of AI-assisted consultation use remains positive ($\beta = 0.693$, $p = 0.076$), indicating that the main finding is not attributable exclusively to observations from the public-LLM era.

TABLE 2 Baseline results (China, 2019–2024).

Variables	(1)		(2)		(3)		(4)		(5)	
	β	p	β	p	β	p	β	p	β	p
lnAI	0.907	0.012	0.801	<0.001	0.690	<0.001	0.696	<0.001	0.666	<0.001
	(0.359)		(0.146)		(0.144)		(0.147)		(0.185)	
Controls										
Online					1.027	<0.001	1.027	<0.001	1.024	<0.001
					(0.114)		(0.114)		(0.114)	
Night					0.036	0.353	0.036	0.350	0.036	0.351
					(0.038)		(0.038)		(0.038)	
Urgent							0.050	0.385	0.047	0.406
							(0.057)		(0.057)	
First							0.036	0.836	0.037	0.832
							(0.173)		(0.173)	
Triage									0.026	0.335
									(0.026)	
Coupon									−0.041	0.432
									(0.052)	
Promo_month									0.479	0.248
									(0.414)	
Constant	39.960	<0.001	40.111	<0.001	30.751	<0.001	27.585	0.044	25.760	0.065
	(0.565)		(0.229)		(1.229)		(13.693)		(13.971)	
Physician FE	NO		YES		YES		YES		YES	
Observations	5,544		5,544		5,544		5,544		5,544	
R-squared	0.001		0.840		0.843		0.843		0.843	

Robust standard errors in parentheses.

Column (7) examines whether the association changed after December 2022 by interacting AI-assisted consultation use with an indicator for the public-LLM period. The estimated post–pre difference was small and statistically insignificant (β = 0.038, p = 0.961), providing no evidence of a discrete change in the association across this temporal boundary.

Column (8) excludes November and December 2024, when the medical large language model developed by West China Hospital had been publicly announced but had not yet been fully deployed on the consultation platform. The estimated coefficient remains positive and statistically significant (β = 0.785, p = 0.021).

Finally, Columns (9) and (10) address additional time-varying confounding. The result remains positive after allowing for physician-specific linear time trends (β = 0.654, p = 0.055) and after absorbing department-by-year fixed effects (β = 0.692, p = 0.045). Taken together, these analyses indicate that the main association is not explained solely by the emergence of public LLMs, gradual physician-specific trends, or department-level changes over time. Nevertheless, these specifications cannot eliminate all possible unobserved time-varying confounding.

Mediation analysis

To explore a potential pathway underlying the association between AI assistance and consultation quality, we examine the mediating role of consultation efficiency. Panel A of Table 4 presents the results of our mediation analysis based on the stepwise mediation approach and the Sobel test. As shown in column (1), the coefficient for lnAI on Efficiency is 0.857 (p = 0.002), indicating that a one-unit increase in lnAI is associated with a 0.857-point increase in the consultation efficiency index.

As shown in column (2), the coefficient for Efficiency is 0.046 (p < 0.001), indicating that consultation efficiency is positively associated with consultation quality after accounting for AI assistance and the included controls. Meanwhile, the direct association between lnAI and Quality remains positive and significant at 0.627 (p = 0.001), though its magnitude is attenuated compared to the coefficient of 0.666 from the baseline specification. This pattern is compatible with partial mediation through consultation efficiency. The Sobel Z-statistic is 2.624 (p = 0.009), indicating that the mediation effect via efficiency is

10.3389/ijph.2026.1609854

TABLE 3 Robustness tests (China, 2019–2024).

Panel A. Alternative specifications and endogeneity checks										
	(1)		(2)		(3)		(4)		(5)	
	Calendar-month FE		Dept.-clustered SE		Raw AI rate		2SLS first stage		2SLS second stage	
	β	p	β	p	β	p	β	p	β	p
lnAI	0.684	0.043	0.666	0.002					0.508	0.088
	(0.339)		(0.200)						(0.298)	
AI rate					0.105	0.005				
					(0.038)					
Peer AI use (instrument)							0.148	<0.001		
							(0.002)			
Controls	Yes		Yes		Yes		Yes		Yes	
Physician FE	Yes		Yes		Yes		Yes		Yes	
Calendar-month FE	Yes		-		-		-		-	
Physician × continuous time	-		-		-		-		-	
Department × year FE	-		-		-		-		-	
Observations	5,544		5,544		5,544		4,248		4,248	
R-squared	0.847		0.843		0.843		0.690		0.853	
Panel B. Temporal sensitivity and time-varying confounding										
	(6)		(7)		(8)		(9)		(10)	
	Pre-public LLM		Pre/post LLM		Excluding Nov–Dec 2024		Physician-specific linear trends		Department × year FE	
	β	p	β	p	β	p	β	p	β	p
lnAI	0.693	0.076	0.674	0.084	0.785	0.021	0.654	0.055	0.692	0.045
	(0.391)		(0.390)		(0.340)		(0.341)		(0.345)	
Post-public-LLM × lnAI			0.038	0.961						
			(0.787)							
Controls	Yes		Yes		Yes		Yes		Yes	
Physician FE	Yes		Yes		Yes		Yes		Yes	
Calendar-month FE	Yes		Yes		Yes		Yes		Yes	
Physician × continuous time	-		-		-		Yes		-	
Department × year FE	-		-		-		-		Yes	
Observations	3,619		5,544		5,390		5,544		5,544	
R-squared	0.852		0.847		0.848		0.850		0.853	

Robust standard errors are reported in parentheses, except in Column (2), where standard errors are clustered at the department level. Column (1) includes physician and month fixed effects. Column (2) uses department-clustered standard errors. Column (3) replaces the log-transformed AI, measure with the raw AI-assisted consultation rate. Columns (4) and (5) report the first- and second-stage estimates from the 2SLS analysis using department peers' average AI usage as the instrument. Column (6) restricts the sample to January 2019–November 2022 (before the public-LLM era). Column (7) tests whether the coefficient differs between the pre- and post-public-LLM periods. Column (8) excludes November–December 2024, when the medical LLM developed by West China Hospital had been publicly announced but had not yet been fully deployed on the consultation platform. Column (9) includes physician-specific linear time trends. Column (10) absorbs department-by-year fixed effects. All models include the full set of controls.

10.3389/ijph.2026.1609854

TABLE 4 Mediating analysis and bootstrap mediation analysis (China, 2019–2024).

Panel A				
Variables	(1)		(2)	
	Me: Efficiency		Y: Quality	
	β	p	β	p
X: lnAI	0.857	0.002	0.627	0.001
	(0.282)		(0.185)	
Me: Efficiency			0.046	<0.001
			(0.009)	
Online	3.205	<0.001	0.877	<0.001
	(0.147)		(0.118)	
Night	−0.181	0.002	0.044	0.249
	(0.058)		(0.038)	
Urgent	−0.215	0.016	0.057	0.314
	(0.089)		(0.057)	
First	−0.501	0.057	0.060	0.730
	(0.263)		(0.173)	
Triage	0.075	0.071	0.022	0.403
	(0.042)		(0.026)	
Coupon	0.112	0.166	−0.046	0.376
	(0.081)		(0.052)	
Promo_month	−0.686	0.277	0.510	0.217
	(0.631)		(0.413)	
Constant	53.617	0.012	23.301	0.094
	(21.222)		(13.932)	
Physician FE	YES		YES	
Observations	5,544		5,544	
R-squared	0.414		0.844	
Sobel Z			2.624	0.009
Proportion of the mediating effect			0.0590	

Panel B						
Path	β	Bootstrap SE	z	95% confidence interval		
				Lower	Upper	CI type
Indirect effect lnAI→Efficiency→Quality	0.039	0.015	2.61	0.012	0.073	(P)
				0.014	0.077	(BC)
Direct effect lnAI→Quality	0.627	0.179	3.49	0.270	0.989	(P)
				0.272	0.994	(BC)

Estimates use the same sample, controls, and physician fixed effects as Panel A. P, percentile 95% bootstrap CI; BC, bias-corrected 95% bootstrap CI.

TABLE 5 Conditional indirect effect analysis, 2,000 replications (China, 2019–2024).

Moderating variable	β	Bootstrap SE	z	95% confidence interval		
				Lower	Upper	CI type
For low depth (−1 SD)	0.025	0.020	1.29	−0.011	0.067	(P)
				−0.010	0.069	(BC)
For medium depth (mean)	0.040	0.015	2.59	0.013	0.073	(P)
				0.015	0.075	(BC)
For high depth (+1 SD)	0.051	0.023	2.21	0.015	0.104	(P)
				0.017	0.110	(BC)

Conditional indirect associations are estimated from the two-stage specification in Equation 4. P, percentile 95% bootstrap CI; BC, bias-corrected 95% bootstrap CI.

statistically significant. The estimated proportion of the total effect that is mediated is about 5.90%.

To provide a more robust assessment of the indirect effect and mitigate potential Type I error associated with the Sobel test [24], we further conduct a nonparametric bootstrapping analysis with 2,000 repetitions [25]. Table 4 Panel B presents the bootstrap test results. Using the same estimation sample and model specification as the stepwise mediation model, the analysis yields an estimated indirect association of 0.039, closely corresponding to the product of the estimated component paths ($0.857 \times 0.046 \approx 0.039$). The bias-corrected 95% CI is [0.014, 0.077], indicating that the indirect association through consultation efficiency is statistically distinguishable from zero. The estimated direct association is 0.627, with a bias-corrected 95% CI of [0.272, 0.994], indicating that a significant direct association remains after accounting for the indirect pathway. Overall, the Sobel and bootstrap results are consistent with a partial mediating role of consultation efficiency.

Conditional indirect effect

In this part, we estimate a two-stage conditional indirect-effect model in which conversation depth is allowed to moderate both the association between AI assistance and consultation efficiency and the association between consultation efficiency and service quality. The resulting conditional indirect association is calculated as the product of the two conditional component paths at each specified level of conversation depth.

We estimate the conditional indirect effects at three representative levels of conversation depth: one standard deviation below the mean (Low Depth, −1 SD), the mean (Medium Depth), and one standard deviation above the mean (High Depth, +1 SD). We use 2,000 bootstrap replications to construct the corresponding confidence intervals. The results are presented in Table 5. Setting Depth at a low level (−1 SD) yielded an estimated conditional indirect effect of 0.025, with a bias-corrected 95% bootstrap CI of [−0.010, 0.069]. Because this interval includes zero, the conditional indirect effect at low depth is not statistically significant. In contrast, the conditional indirect effects at medium (Mean) and high (+1 SD) levels of Depth were 0.040 and 0.051, with bias-corrected 95% bootstrap CIs of [0.015, 0.075] and [0.017, 0.110], respectively, both excluding zero. The point estimates of the conditional indirect effect increased from 0.040 at medium depth to 0.051 at high depth. Overall, the bootstrap results indicate a depth-dependent pattern in the estimated indirect association. The increase in the point estimates from 0.025 at low depth to 0.051 at high depth further suggests that the efficiency-related pathway is more evident when physician–patient conversations are deeper.

Heterogeneity analysis

To assess whether the estimated associations vary across physician seniority, we conduct a heterogeneity analysis based on physician seniority. Physicians are categorized into two broad groups: Experts (including "First-Level Expert" to "Fifth-Level Expert," totaling 2,952 observations) and Regular Physicians (including titles like "Chief Physician," "Associate Chief Physician," "Associate Professor," "Attending Physician," and "Physician," totaling 2,592 observations). We re-estimate our mediation models separately for each subgroup to examine potential differences in the direct and indirect association between AI assistance and consultation quality. The results, presented in Table 6, reveal a distinct pattern between the two groups.

For Experts (Columns 1–3), AI assistance was not significantly associated with consultation quality ($\beta = 0.382$, $p = 0.126$), but was positively associated with consultation efficiency ($\beta = 1.108$, $p = 0.006$), and consultation efficiency is positively associated with consultation quality after accounting for AI assistance and the control variables ($\beta = 0.046$, $p < 0.001$). The supplementary Sobel result is statistically significant (Sobel $Z = 2.275$, $p = 0.023$), with efficiency mediating approximately 13.4% of AI's total effect on quality for this group. This suggests that for experts, AI functions primarily as an efficiency-enhancing tool; its value lies not in a direct quality boost, but in streamlining their workflow, which indirectly supports their consultation performance.

For Regular Physicians (Columns 4–6), the pattern differs. AI assistance is positively associated with consultation quality ($\beta = 0.994$, $p < 0.001$), whereas its association with consultation efficiency is not statistically significant ($\beta = 0.590$, $p = 0.136$). The estimated association between AI assistance and consultation quality remains positive after efficiency is included in the model ($\beta = 0.967$, $p < 0.001$), and the

10.3389/ijph.2026.1609854

TABLE 6 Heterogeneity analysis by physician seniority (China, 2019–2024).

Variables	Experts						Regular physicians					
	(1)		(2)		(3)		(4)		(5)		(6)	
	Y: quality		Me: efficiency		Y: quality		Y: quality		Me: efficiency		Y: quality	
	β	p	β	p	β	p	β	p	β	p	β	p
X: lnAI	0.382	0.126	1.108	0.006	0.331	0.184	0.994	<0.001	0.590	0.136	0.967	<0.001
	(0.249)		(0.399)		(0.249)		(0.277)		(0.396)		(0.276)	
Me: efficiency					0.046	<0.001					0.046	0.001
					(0.011)						(0.014)	
Online	1.086	<0.001	3.244	<0.001	0.937	<0.001	0.987	<0.001	3.181	<0.001	0.842	<0.001
	(0.177)		(0.237)		(0.180)		(0.149)		(0.187)		(0.158)	
Night	0.094	0.071	−0.169	0.047	0.102	0.050	−0.022	0.691	−0.187	0.017	−0.014	0.806
	(0.052)		(0.085)		(0.052)		(0.056)		(0.079)		(0.056)	
Urgent	0.031	0.687	−0.240	0.061	0.042	0.584	0.072	0.396	−0.181	0.144	0.080	0.343
	(0.077)		(0.128)		(0.077)		(0.085)		(0.124)		(0.085)	
First	−0.156	0.505	−0.168	0.657	−0.149	0.525	0.256	0.320	−0.875	0.015	0.296	0.250
	(0.235)		(0.379)		(0.234)		(0.257)		(0.361)		(0.257)	
Triage	0.062	0.084	0.085	0.161	0.058	0.104	−0.015	0.708	0.063	0.269	−0.018	0.655
	(0.036)		(0.061)		(0.036)		(0.039)		(0.057)		(0.039)	
Coupon	0.018	0.794	0.184	0.114	0.010	0.890	−0.112	0.146	0.021	0.853	−0.113	0.142
	(0.070)		(0.116)		(0.070)		(0.077)		(0.112)		(0.077)	
Promo_month	0.553	0.326	−1.527	0.086	0.624	0.266	0.437	0.475	0.291	0.743	0.424	0.488
	(0.563)		(0.890)		(0.560)		(0.612)		(0.889)		(0.611)	
Constant	36.871	0.052	28.214	0.355	35.568	0.060	12.706	0.539	82.068	0.005	8.946	0.665
	(18.987)		(30.520)		(18.913)		(20.703)		(29.089)		(20.683)	
Physician FE	YES		YES		YES		YES		YES		YES	
Observations	2,952		2,952		2,952		2,592		2,592		2,592	
R-squared	0.845		0.317		0.846		0.841		0.484		0.842	
Sobel Z					2.275	0.023					1.371	0.170
Proportion of the mediating effect					0.1340						0.0272	

Robust standard errors in parentheses.

supplementary Sobel result is not statistically significant (Sobel Z = 1.371, p = 0.170). Thus, the subgroup estimates do not provide evidence of a statistically detectable efficiency-based indirect association for regular physicians.

Discussion

Main findings

Online medical consultation has become a scaled service channel in which diagnostic work, communication, and operational constraints jointly shape consultation service quality. This study examines the relationship between AI assistance and online medical consultation service quality, examines consultation efficiency as a process pathway, and tests physician-patient conversation depth as a boundary condition while also assessing heterogeneity across physician seniority. Drawing on service productivity theory, the study explains how AI-supported information processing and workflow support can influence both consultation efficiency and quality. It therefore offers insights into the complex dynamics of human-AI collaboration in professional healthcare services, where communication is central to value creation, and where the same tool may generate different benefits across segments of the physician workforce.

AI assistance is positively associated with consultation service quality. The result is consistent with the idea that AI can facilitate medical processes by supporting information processing, reducing decision errors and speeding consultation work. AI-assisted consultation tools not only improve the informational inputs and workflow conditions under which medical services are delivered, but also enhance cognitive capabilities and complement physicians' professional judgment by synthesizing medical information and knowledge. This augmentation effect suggests a complementarity model of human-AI collaboration.

Our mediation analysis supports consultation efficiency as a statistically significant but partial pathway linking AI assistance to consultation quality, implying faster response, task completion and improved allocation of clinician attention can bring benefits. From a service productivity perspective, effective use of AI-assisted tools can reduce the time physicians spend collecting, reorganizing, and initially screening patient information, enabling them to process cases more efficiently and manage larger consultation volumes, thereby creating cognitive bandwidth for more complex diagnostic reasoning and personalized treatment planning.

Patient–physician interaction depth appears to provide an important context for the efficiency-related indirect pathway. Although prior research has suggested that additional verification processes may counterproductively slow down diagnostic procedures [11], our findings suggest that the efficiency-related pathway is more evident when physician–patient interaction is sufficiently substantive. It also reinforces the view that AI assistance should support, rather than replace, meaningful physician–patient communication [26–28].

Finally, our results reveal potential different patterns of human-AI collaboration for experts versus regular physicians. For senior expert physicians, AI assistance significantly improves consultation efficiency, therefore enhancing quality outcomes, suggesting AI functions primarily as a productivity tool. Conversely, for regular physicians, AI assistance significantly enhances consultation quality through direct effects, with no significant mediation through efficiency improvements. This asymmetric pattern is not viewed as a superior or inferior utilization. It reflects distinct patterns of human-AI collaboration and resonates with the broader view that AI changes workflow and expertise in different ways across roles [11].

Practical implications

This study provides practical guidance for online consultation platforms on leveraging AI technology to enhance consultation quality. Healthcare resource constraints have placed increasing pressure on online consultation services. Healthcare organizations and platforms should consider the careful deployment of validated AI support systems. Efficiency is a meaningful but partial mechanism, and it does not universally convert into quality. Stakeholders should adopt balanced performance evaluation systems that weight quality (such as diagnostic accuracy, patient outcomes, and treatment effectiveness) equally with consultation speed, thereby discouraging superficial acceleration in favor of substantive healthcare enhancement. Conversation depth appears to be an important boundary condition for the efficiency-related pathway. Hospitals should ensure that AI-enabled service redesign remains grounded in meaningful physician-patient communication while also helping to manage rising patient demand. Platform design should pair AI tools with workflow and interface designs that foster iterative physician-patient exchanges. The heterogeneity results further suggest differentiated implementation strategies. Senior experts require streamlined AI tools prioritizing workflow efficiency, while regular physicians may benefit from comprehensive clinical decision support systems.

Limitations

Several limitations shape interpretation and point to future work. Our entropy-weighted quality index primarily captures platform-observed patient feedback and administrative outcomes rather than the full clinical quality of care or long-term health outcomes. A valuable extension would be to combine large-scale platform data with an in-depth assessment of a stratified sample of clinical interactions and examine whether the macro-level relationships identified by the entropy-weighted index are also evident in the clinical content of individual encounters.

Our variable measurement also suggests refinements. The AI-assisted consultation rate measures the monthly frequency of platform-recorded use of integrated AI functions rather than the technical capability or version of the underlying tools. Although the same record-based coding rule was applied throughout the observation period, the available AI functions evolved gradually between 2019 and 2024, and version-level deployment records were unavailable in our study. Future research with function- and version-level usage logs could distinguish the effects of different AI functions and technological generations more precisely.

Residual endogeneity may remain, as physicians' use of platform-integrated AI assistance is self-selected rather than randomly assigned. Future research could pursue stronger causal identification by exploiting observable staggered rollouts, randomized implementation, or other exogenous changes in access to platform-integrated AI assistance.

Conclusion

AI assistance demonstrates a robust positive association with online consultation service quality, and consultation service efficiency explains part of this relationship. The efficiency-related pathway to consultation quality varies with physician-patient conversation depth and becomes evident mainly in deeper interactive dialogues. The heterogeneity patterns further indicate that AI plays different roles across physician seniority, serving primarily as an efficiency-supporting tool for experts and as a potential capability augmenter for generalists. The findings extend evidence on AI-assisted support from isolated technical performance to integrated service quality in real platform workflows. They also suggest that effective

governance should couple validated AI tools with interaction-oriented design and differentiated implementation strategies that account for clinicians' experience and consultation complexity.

Author contributions

FZ: conceptualization, methodology, formal analysis, writing - original draft; XT: data curation, conceptualization, writing - review and editing; ZY: supervision, validation; YH: conceptualization, supervision, project administration. FZ and XT contributed equally to this article. All authors contributed to the article and approved the submitted version.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This work was supported by the 2024 Key Research and Development Program of the Department of Science and Technology (Grant No. XZ202401YD0022).

Conflict of interest

The authors declare that they do not have any conflicts of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.ssph-journal.org/articles/10.3389/ijph.2026.1609854/full#supplementary-material

References

1. Chen C, Cui Z. Impact of AI-Assisted diagnosis on American patients' trust in and intention to seek help from health care professionals: randomized, web-based survey experiment. *J Med Internet Res* (2025) 27:e66083. doi: 10.2196/66083

2. Wang X, Wang Y. Analysis of trust factors for AI-Assisted diagnosis in intelligent healthcare: personalized management strategies in chronic disease management. *Expert Syst Appl* (2024) 255:124499. doi: 10.1016/j.eswa.2024.124499

3. National Health Commission of the People's Republic of China. *Material Distributed at the Press Conference Held by the National Health Commission on September 14, 2018: Regarding the Implementation of Three Documents, Including the Internet Medical Treatment Management Measures (Trial)* (2018). Available online at: https://www.nhc.gov.cn/wjw/zccl/201809/410ce29d17494b54bfc07dacd2832f31.shtml (Accessed August 24, 2026).

4. Shen S., Where are the bottlenecks for internet hospitals?. *People's Daily [Internet]* p. 19. (2026). Available online at: https://paper.people.com.cn/rmrb/pc/content/202605/15/content_30156916.html (Accessed August 24, 2026).

5. Gomez C, Smith BL, Zayas A, Unberath M, Canares T. Explainable AI decision support improves accuracy during telehealth strep throat screening. *Commun Med* (2024) 4(1):149. doi: 10.1038/s43856-024-00568-x

6. Vodrahalli K, Ko J, Chiou AS, Novoa R, Abid A, Phung M, et al. Development and clinical evaluation of an artificial intelligence support tool for improving telemedicine photo quality. *JAMA Dermatol* (2023) 159(5):496–503. doi: 10.1001/jamadermatol.2023.0091

7. Yakar D, Ongena YP, Kwee TC, Haan M. Do people favor artificial intelligence over physicians? A survey among the general population and their view on artificial intelligence in medicine. *Value in Health* (2022) 25(3):374–81. doi: 10.1016/j.jval.2021.09.004

8. Jiang X, Xie H, Tang R, Du Y, Li T, Gao J, et al. Characteristics of online health care services from china's largest online medical platform: cross-sectional survey study. *J Med Internet Res* (2021) 23(4):e25817. doi: 10.2196/25817

9. Yang M, Yan Y, Xu Z, Liu H, Ran J, Zheng Y, et al. The status and challenges of online consultation service in internet hospitals operated by physical hospitals in China: a large-scale pooled analysis of multicenter data. *BMC Health Serv Res* (2025) 25(1):611. doi: 10.1186/s12913-025-12787-6

10. Kiani A, Uyumazturk B, Rajpurkar P, Wang A, Gao R, Jones E, et al. Impact of a deep learning assistant on the histopathologic classification of liver cancer. *Npj Digit Med* (2020) 3(1):23. doi: 10.1038/s41746-020-0232-8

11. D'Adderio L, Bates DW. Transforming diagnosis through artificial intelligence. *Npj Digit Med* (2025) 8(1):54. doi: 10.1038/s41746-025-01460-1

12. Grönroos C, Ojasalo K. Service productivity: toward a conceptualization of the transformation of inputs into economic results in services. *J Bus Res*. (2004) 57(4):414–23. doi: 10.1016/S0148-2963(02)00275-8

13. Bitner MJ, Faranda WT, Hubbert AR, Zeithaml VA. Customer contributions and roles in service delivery. *Int J Serv Industry Management* (1997) 8(3):193–205. doi: 10.1108/09564239710185398

14. Dong B, Sivakumar K. Customer participation in services: domain, scope, and boundaries. *J Acad Mark Sci* (2017) 45(6):944–65. doi: 10.1007/s11747-017-0524-y

15. Yang H, Wang S, Wang F. Online health community—an empirical analysis based on grounded theory and entropy weight TOPSIS method to evaluate the service quality. *Digital Health* (2023) 9: 1–16. doi: 10.1177/20552076231207201

16. World Medical Association. Declaration of helsinki: ethical principles for medical research involving human subjects. *JAMA* (2013) 310(20):2191–4. doi: 10.1001/jama.2013.281053

17. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J* (1948) 27(3):379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x

18. Wang L, Li F, Jiang S, Xia J. Comprehensive evaluation of primary health care service capacity in China based on TOPSIS and RSR methods. *Front Public Health* (2025) 13:1657961. doi: 10.3389/fpubh.2025.1657961

19. Angrist J, Pischke JS. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton (NJ): Princeton University Press (2009).

20. Wooldridge JM. *Introductory Econometrics: A Modern Approach*. 5th ed. Mason, OH: South-Western, Cengage Learning (2013).

21. Judd CM, Kenny DA. Process analysis: estimating mediation in treatment evaluations. *Eval Rev* (1981) 5(5):602–19. doi: 10.1177/0193841X8100500502

22. Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Personal Social Psychol* (1986) 51(6):1173–82. doi: 10.1037/0022-3514.51.6.1173

23. Preacher KJ, Rucker DD, Hayes AF. Addressing moderated mediation hypotheses: theory, methods, and prescriptions. *Multivariate Behav Res* (2007) 42(1):185–227. doi: 10.1080/00273170701341316

24. MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V. A comparison of methods to test mediation and other intervening variable effects. *Psychol Methods* (2002) 7(1):83–104. doi: 10.1037/1082-989x.7.1.83

25. Preacher KJ, Hayes AF. SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behav Res Methods Instr & Comput* (2004) 36(4):717–31. doi: 10.3758/BF03206553

26. Chiu EKY, Sridhar S, Sai-Yin Wong S, Tam AR, Choi MH, Lau AWT, et al. Generative artificial intelligence models in clinical infectious disease consultations: a cross-sectional analysis among specialists and resident trainees. *Healthcare* (2025) 13(7):744. doi: 10.3390/healthcare13070744

27. Nelson CA, Pérez-Chada LM, Creadore A, Li SJ, Lo K, Manjaly P, et al. Patient perspectives on the use of artificial intelligence for skin cancer screening: a qualitative study. *JAMA Dermatol* (2020) 156(5):501–12. doi: 10.1001/jamadermatol.2019.5014

28. Young AT, Amara D, Bhattacharya A, Wei ML. Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review. *The Lancet Digital Health* (2021) 3(9):e599–611. doi: 10.1016/S2589-7500(21)00132-1