Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Integrating AI Chatbot in Medical Education: A Technology Acceptance Study of ChatGPT using UTAUT2

Hazwani Mohd Mohadis[1,*], Irnani Ismail[2], Ismail Ibrahim[2], Zulfikar Ahmed Maher[3], Suraya Yaacob[4], Abdul Muizz Che Azmi[5]

[1] Department of Information Systems, Kulliyyah of Information and Communication Technology, International Islamic University Malaysia, 53100 Gombak, Selangor, Malaysia
[2] Department of Internal Medicine, Kulliyyah of Medicine, International Islamic University Malaysia, Jalan Sultan Ahmad Shah, Bandar Indera Mahkota, 25200 Kuantan, Pahang, Malaysia
[3] Information Technology Centre, Sindh Agriculture University, Tando Jam, Pakistan
[4] Department of Business Intelligence, Humanities and Governance, Faculty of Artificial Intelligence, Universiti Teknologi Malaysia, Malaysia
[5] Access Corporate Ventures, Taman Perwira Dua, 68000 Ampang, Selangor, Malaysia

ARTICLE INFO

Article history:
Received 15 July 2025
Received in revised form 24 July 2025
Accepted 19 August 2025
Available online 26 August 2025

Keywords:
Artificial Intelligence in education; AI Chatbot; ChatGPT; medical education; technology acceptance; UTAUT2; PLS-SEM

ABSTRACT

The integration of Artificial Intelligence (AI) chatbots into medical education represents a transformative shift, offering new opportunities for personalized learning while also presenting unique challenges related to user acceptance and institutional readiness. Adapted based on the Extended Unified Theory of Acceptance and Use of Technology (UTAUT2), this study examines the factors influencing the adoption of ChatGPT among undergraduate medical students. Using a convenience sampling method, data were collected from 70 undergraduate medical students via an online questionnaire. The responses were analyzed using Partial Least Squares Structural Equation Modelling (PLS-SEM) to test the hypothesized relationships within the UTAUT2 framework. Findings reveal that Habit and Facilitating Conditions are significant predictors of students' behavioural intention to use ChatGPT, emphasizing the role of prior behavioural usage patterns and institutional support in influencing ChatGPT adoption. In contrast, Performance Expectancy, Effort Expectancy, Hedonic Motivation, and Price Value did not exhibit significant influence on behavioural intention, suggesting that perceived usefulness, ease of use, enjoyment, and cost considerations are less influential in this context. These results contribute to the growing discourse on AI integration in medical education, The study highlights the need for acceptance and adoption strategies that go beyond enhancing functionality or user experience, and instead focus on cultivating habitual use and ensuring institutional support.

1. Introduction

The rapid rise of artificial intelligence (AI) in higher education has prompted a re-evaluation of how learning technologies are integrated into curricula, particularly in fields requiring high levels of cognitive complexity and ethical sensitivity. Among the most prominent AI tools is ChatGPT,

* *Corresponding author.*
E-mail address: hazwanimohadis@iium.edu.my

https://doi.org/10.37934/araset.56.1.215239

developed by OpenAI, which has been adopted widely by students for academic tasks such as writing, summarization, and information retrieval. Its adoption in medical education, a discipline characterized by intensive learning demands, clinical application, and professional accountability is especially noteworthy.

ChatGPT offers numerous benefits in medical education, including personalized learning experiences, support for clinical reasoning, and enhanced communication with patients [7,15]. The tool can simplify complex medical concepts, generate mnemonics, produce practice questions, and simulate clinical dialogues, thereby supporting both theoretical learning and practical skill development [7,15].

Despite these advantages, the adoption of ChatGPT in medical education raises concerns about information accuracy, ethical integrity, and student overreliance. Moreover, its use in medicine is not yet as well understood as in other disciplines. Given the critical importance of accuracy, ethical responsibility, and independent clinical thinking in medical training, it is essential to assess how medical students accept and integrate such AI tools into their learning. This study investigates the behavioral intention to use ChatGPT among undergraduate medical students, applying the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) to identify key predictors of adoption. By focusing on medical students, a group for whom the stakes of misinformation and academic integrity are particularly high, this study addresses an important and underexplored area of educational AI research.

2. Literature Review

2.1 ChatGPT in Higher Education

Since its release in late 2022, ChatGPT has become a widely adopted tool across higher education institutions globally. Surveys indicate high levels of familiarity and usage among students, with reports of enhanced learning efficiency, engagement, and comprehension of complex topics [4]. Students commonly use the AI tool for content generation, summarizing, academic writing, and conceptual clarification [2,11,16].

The popularity of ChatGPT is partly driven by its ability to provide immediate, interactive, and context-sensitive feedback, which enhances personalized learning experiences. In many cases, students view ChatGPT as an “intelligent tutoring assistant” that can support both autonomous and collaborative learning tasks [4,7]. ChatGPT can provide adaptive support and resources, offering a personalized learning experience that caters to students’ individual needs [1]. However, concerns regarding academic dishonesty, dependence, and the accuracy of responses have led to calls for clearer ethical guidelines and institutional policies to govern its use [2,6].

2.2 ChatGPT in Medical Education

In the domain of medical education, ChatGPT is being adopted not just for academic support but also to facilitate clinical reasoning and simulate patient communication [7,15]. It supports learning by simplifying medical content, creating mnemonics, generating exam questions, and offering automated Q&A systems. Wu *et al.,* (2024) [15] highlighted its application in intelligent tutoring and simulation-based learning, aligning well with competency-based medical education.

Empirical evidence reflects growing usage among medical students. A U.S.-based survey found that 48.9% of medical students had used ChatGPT, with 43.7% using it weekly or more frequently [16]. In Egypt, 78.5% of students reported using ChatGPT, and 64% found it useful for learning medical content [2]. Interestingly, Tangadulrat *et al.,* (2023) [14] reported that current medical students tend to view ChatGPT more positively than practicing physicians, suggesting generational or experiential differences in acceptance.

Despite the benefits, some limitations remain. Medical students have expressed concerns about the reliability of ChatGPT responses, potential ethical violations, and reduced critical thinking due to overreliance [2,6]. The tool's general-purpose design, while versatile, can fall short of domain-specific rigor needed in medicine [13].

2.3 Determinants of AI Adoption in Education: A UTAUT2 Perspective

Technology adoption in educational settings is commonly explored through theoretical models such as the Technology Acceptance Model (TAM), UTAUT, and its extended version UTAUT2. The UTAUT2 framework incorporates constructs such as Performance Expectancy, Effort Expectancy, Social Influence, Facilitating Conditions, Hedonic Motivation, Habit, and Price Value [12].

In studies involving ChatGPT, Performance Expectancy and Effort Expectancy consistently emerge as significant predictors of behavioral intention [3,4]. Habit and hedonic motivation have also been found to be strong drivers of continued use, particularly among students accustomed to digital tools [19]. Facilitating conditions, including institutional support and access, play a critical role in enabling actual system use [17].

In the medical education context, where accuracy, ethical usage, and long-term knowledge retention are key, constructs such as trust, prior usage behavior, and academic culture may moderate these relationships, warranting further investigation. Although the literature reflects increasing acceptance and adoption of ChatGPT in higher education, there is a marked lack of discipline-specific studies focused on medical education. Most existing research addresses general student populations or disciplines like programming, engineering, and the social sciences [4,6], which may not share the same constraints, values, or learning demands as medical training.

Furthermore, while many studies have applied UTAUT and TAM to study ChatGPT adoption, relatively few have employed UTAUT2 to examine how variables such as Habit, Hedonic Motivation, and Facilitating Conditions operate in a medical education setting. Given the critical importance of trust in information accuracy, the risk of academic misconduct, and the emphasis on clinical reasoning, it is likely that medical students weigh AI usage differently compared to their peers in other fields.

This study addresses this gap by using UTAUT2 to explore the determinants of ChatGPT acceptance and use among undergraduate medical students, which would provide empirical evidence on how future medical professionals engage with AI tools and what factors may encourage or hinder responsible adoption in a high-stakes learning environment.

Accordingly, in light of the growing integration of artificial intelligence in higher education and the unique demands of medical education, this study aims to investigate the factors influencing medical students' acceptance and use of ChatGPT as a learning tool. Grounded in the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2), the study seeks to understand how various technological, psychological, and contextual factors shape students' behavioural intention and actual usage of ChatGPT.

The specific research objectives of this study are as follows:

i. To assess the impact of UTAUT2 constructs, including Performance Expectancy (PE), Effort Expectancy (EE), Social Influence (SI), Facilitating Conditions (FC), Hedonic Motivation (HM), Price Value (PV), and Habit (HT), on medical students' behavioural intention to use ChatGPT in their academic learning.

ii. To determine the effect of behavioral intention on the actual use behavior of ChatGPT among medical students.

By addressing these objectives, the study aims to offer empirical insights that can inform educators, curriculum designers, and policymakers on how best to integrate AI-based tools like ChatGPT into medical education in a pedagogically sound and ethically responsible manner.

2. Theoretical Framework

The application of theoretical models such as the Technology Acceptance Model (TAM), the Diffusion of Innovation Theory, and the Unified Theory of Acceptance and Use of Technology (UTAUT) is essential for providing a structured and systematic understanding of how individuals adopt and use new technologies, including AI chatbots such as ChatGPT, in complex educational contexts [3]. These models enable researchers to identify the factors that either facilitate or hinder adoption, offering a framework for examining how technological, psychological, and contextual influences shape user perceptions and behaviours [17]. Moreover, they help address existing knowledge gaps by providing an analytical lens through which the interactions between users and technology can be understood within specific contexts, such as medical education. In doing so, theoretical models not only advance academic understanding but also inform the development of practical strategies and policies for integrating technology in ways that are pedagogically sound and contextually relevant [3].

The Unified Theory of Acceptance and Use of Technology (UTAUT), developed by Venkatesh *et al.,* (2012) [12], was created to unify and extend existing knowledge on technology acceptance by synthesizing eight prominent theoretical models. These models include the Theory of Reasoned Action (TRA), which explains how beliefs and attitudes influence behavioural intentions; the Technology Acceptance Model (TAM), which focuses on perceived usefulness and ease of use; the Motivational Model (MM), which examines intrinsic and extrinsic motivation; Social Cognitive Theory (SCT), which highlights the interaction of personal, behavioural, and environmental factors; the Model of Personal Computer Use (MPCU), which investigates predictors of computer adoption; the Diffusion of Innovation Theory (IDT), which explains how innovations spread over time; the Theory of Planned Behavior (TPB), which adds perceived behavioural control to TRA; and the Combined TAM and TPB (C-TAM-TPB), which integrates constructs from both TAM and TPB. By drawing together the strengths of these eight models, Venkatesh *et al.,* (2012) [12] developed UTAUT as a more comprehensive and unified framework for understanding technology acceptance and usage.

The original UTAUT identifies four core constructs which are Performance Expectancy, Effort Expectancy, Social Influence, and Facilitating Conditions, that would influence behavioural intention and technology use. Performance Expectancy refers to the degree to which an individual believes that using a system will improve performance, Effort Expectancy relates to the perceived ease of use, Social Influence considers the impact of important others on the decision to use technology, and Facilitating Conditions capture the availability of resources and support required for use. While UTAUT has been widely validated, it was initially developed for organizational contexts in which technology use is often mandatory. This limits its suitability for settings where adoption is voluntary, such as in the case of ChatGPT use among students.

To address this limitation, Venkatesh *et al.,* (2012) [12] extended the original model to create the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2), which retains the four original constructs and incorporates three additional factors particularly relevant in consumer and voluntary-use contexts: Hedonic Motivation, Price Value, and Habit. Hedonic Motivation captures the enjoyment or pleasure derived from using a technology; Price Value assesses the trade-off between perceived benefits and monetary costs; and Habit reflects the extent to which a behaviour has become automatic through prior learning and repeated use.

The selection of UTAUT2 for the present study is justified by the voluntary and self-directed nature of ChatGPT adoption in medical education. Unlike workplace systems where adoption is mandated, medical students choose whether or not to use ChatGPT, aligning more closely with the consumer-like contexts for which UTAUT2 was designed. Furthermore, previous studies applying UTAUT2 in educational settings have demonstrated its effectiveness in identifying the multifaceted factors influencing technology adoption by students and educators [19]. Figure 1 illustrates the adapted UTAUT2 model applied in this study.

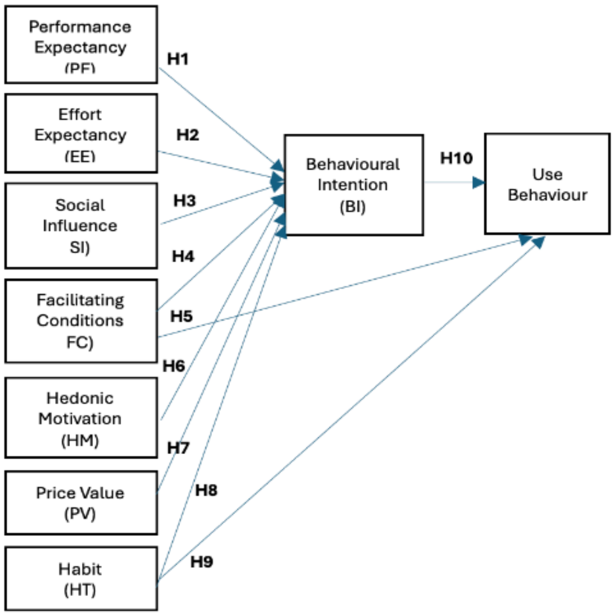

Fig. 1. The research model

Given that ChatGPT use in medical education is not universally mandated, and considering the interplay of utilitarian, experiential, and habitual factors in students' technology adoption behaviour, UTAUT2 offers a comprehensive framework to capture these dynamics. Thus, in this study, all seven (7) constructs of UTAUT2 are applied to examine the determinants of behavioural intention to use ChatGPT, as well as the relationship between intention and actual use. The following Table 1 summarize the definition of each construct presented in our research model.

Table 1
Definitions of UTAUT2 constructs used in the research model

Determinants	Description
Performance Expectancy (PE)	the extent to which individuals believe that using a system will enhance their performance in learning processes
Effort Expectancy (EE)	the degree of ease or effort associated with the use of technology
Social Influence (SI)	the extent to which important others, such as family and friends, believe that an individual should use a particular technology
Facilitating Conditions (FC)	the level of accessibility to resources and support needed to accomplish a task
Hedonic Motivation (HM)	to the pleasure or enjoyment derived from using a technology
Price Value (PV)	an individual's trade-off between the perceived benefits of using the system and its monetary cost
Habit (HT)	the extent to which an individual tends to perform behaviours automatically because of prior learning and experiences with the technology

Our review of previous studies investigating the acceptance and use of ChatGPT in higher education identifies Performance Expectancy (PE) as a consistent and significant predictor of behavioural intention, with students and educators more likely to adopt the tool if they believe it enhances academic performance and productivity [3,4,19]. Effort Expectancy (EE) generally demonstrates a positive relationship with behavioural intention, highlighting the importance of user-friendly interfaces, although some studies report insignificant or context-dependent effects [3]. Social Influence (SI) has produced mixed findings: while it often serves as a significant driver of adoption when endorsed by peers or instructors, in certain contexts its effect appears less pronounced [4].

Similarly, the influence of Facilitating Conditions (FC) including access to resources and technical support also shows mixed results, though positive associations with adoption are frequently reported [3,4,19]. Hedonic Motivation (HM), which refers to the enjoyment derived from using ChatGPT, has been shown to positively influence behavioural intention, suggesting that engaging and pleasurable experiences encourage continued use [3,4,19]. Price Value (PV), which reflects the trade-off between perceived benefits and monetary cost, may negatively affect adoption when perceived costs are high [3,19]. Habit (HT), defined as the extent to which use becomes an automatic behaviour based on prior experience, emerges as a strong positive predictor of both behavioural intention and actual use [19].

Drawing on these insights, this study formulates ten hypotheses to examine ChatGPT acceptance and use within the medical education context. While prior research has investigated ChatGPT adoption in higher education broadly, no empirical studies to date have specifically examined its acceptance and use among undergraduate medical students. The research hypotheses are presented in Table 2.

Table 2
The research hypothesis

H#	Hypothesis
H1	PE will have significant positive influence on medical students BI to use ChatGPT for medical education.
H2	EE will have significant positive influence on medical students BI to use ChatGPT for medical education.
H3	SI will have significant positive influence on medical students BI to use ChatGPT for medical education.
H4	FC will have significant positive influence on medical students BI to use ChatGPT for medical education.
H5	FC will have significant positive influence on medical students ChatGPT Use Behaviour for medical education.
H6	HM will have significant positive influence on medical students BI to use ChatGPT for medical education.
H7	PV will have significant positive influence on medical students BI to use ChatGPT for medical education.
H8	HT will have significant positive influence on medical students BI to use ChatGPT for medical education.
H9	HT will have significant positive influence on medical students Use Behaviour to use ChatGPT for medical education.
H10	BI will have significant positive influence on medical students ChatGPT Use Behaviour for medical education.

3. Methodology

3.1 Study Populations and Data Collection Procedures

The study population involved undergraduate medical students currently undertaking Bachelor of Medicine and Bachelor of Surgery (MBBS). A non-probability sampling strategy was used to collect data. The online survey was prepared and distributed through widely used messaging platforms such as WhatsApp and Telegram. A total of 70 valid responses were received. While modest, the sample size was evaluated against established criteria for PLS-SEM using 10-times rule. 10-Times Rule a minimum sample size in PLS-SEM suggests that researchers should collect at least 10 times the number of indicators pointing to the most complex construct in the model [9]. In this study, the construct Behavioural Intention is predicted by seven latent variables (PE, EE, SI, FC, HM, PV, HT). Thus: 10 x 7 = 70 respondents required. Thus, this criterion is satisfied.

3.2 Construct Operationalization and Instrument Development

The survey instrument consisted of two main Section A captured participant demographic information, including gender, age, year of study, prior ChatGPT usage experience, current physical activity level, and smartphone usage patterns.

Section B focused on the acceptance and use of ChatGPT, using a total of 25 items adapted from validated scales in prior study by Venkatesh *et al.,* (2012) [12]. This section measured seven (7) independent variables which comprised of items measuring Performance Expectancy (PE), Effort Expectancy (EE), Social Influence (SI), Facilitating Conditions (FC), Hedonic Motivation (HM), Price Value (PV), and Habit (HT). The dependent variable, Behavioural Intention (BI), was measured using three items. All items were rated on a five-point Likert scale ranging from strongly disagree (1) to strongly agree (5). The following Table 3 presents the adapted UTAUT2 items.

Table 3
The adapted UTAUT2 items [12]

Performance Expectancy (PE)	
I believe that ChatGPT is useful for my studies.	(PE1)
Using ChatGPT increases my chances of achieving important things in my studies	(PE2)
Using ChatGPT helps me to get tasks and assignments done faster in my studies.	(PE3)
Using ChatGPT increases my productivity in studying medicine	(PE4)
Effort Expectancy (EE)	
Learning how to use ChatGPT for my studies is easy for me.	(EE1)
My interaction with ChatGPT is clear and understandable	(EE2)
I find ChatGPT easy to use	(EE3)
It is easy for me to become skilful at using ChatGPT for my studies	(EE4)
Social Influence (SI)	
People who are important to me (such as friends, lecturer) think I should use ChatGPT for my studies.	(SI1)
People who influence my Behaviour believe that I should use ChatGPT for my studies.	(SI2)
People whose opinions I value prefer me to use ChatGPT for my studies.	(S3)
Facilitating Conditions (FC)	
I have the resources necessary to use ChatGPT for my studies.	(FC1)
I have the knowledge necessary to use ChatGPT for my studies.	(FC2)
ChatGPT is compatible with technologies I am currently using.	(FC3)
I can get help from others when I have difficulties using ChatGPT for my medical studies.	(FC4)
Hedonic Motivation (HM)	
Using ChatGPT to study medicine is fun.	(HM1)
Using ChatGPT to study medicine is enjoyable.	(HM2)
Using ChatGPT for to study medicine is very entertaining and informative.	(HM3)
Price Value (PV)	
ChatGPT is reasonably priced.	(PV1)
ChatGPT is good value for the money.	(PV2)
Using ChatGPT for to study medicine is very entertaining and informative.	(PV3)
Habit (HT)	
Using ChatGPT to study medicine has become a habit for me.	(HT1)

I am addicted to using ChatGPT for studying medicine.	(HT2)
I feel that I must use ChatGPT for my studies.	(HT3)
Using ChatGPT to study medicine has become natural for me.	(HT4)
Behavioural Intention (BI)	
I intend to continue using ChatGPT for my studies in the future.	(BI1)
I will always try to use ChatGPT in my studies.	(BI2)
I plan to continue using ChatGPT frequently for my studies.	(BI3)

3.3 Data Analysis Procedures

To identify the factors influencing medical students' intention to use ChatGPT for educational purposes, this study employed Partial Least Squares Structural Equation Modelling (PLS-SEM). PLS-SEM is widely used in social sciences, business, and management research for examining complex relationships among latent constructs, particularly in predictive and exploratory contexts [9]. Unlike covariance-based SEM (CB-SEM), which relies on a covariance-driven approach, PLS-SEM adopts a variance-based method aimed at maximizing the explained variance of dependent variables rather than reproducing the observed covariance matrix [18].

The method was selected for several reasons. First, it is well suited for small sample sizes and data that deviate from normality, which are applicable to the present study. Second, PLS-SEM can handle both formative and reflective measurement models, offering greater flexibility in model specification [9]. The PLS-SEM analysis method consists of a two-step procedure; 1) the assessment of the measurement model; and 2) the evaluation of the structural model as shown in Figure 2 below.

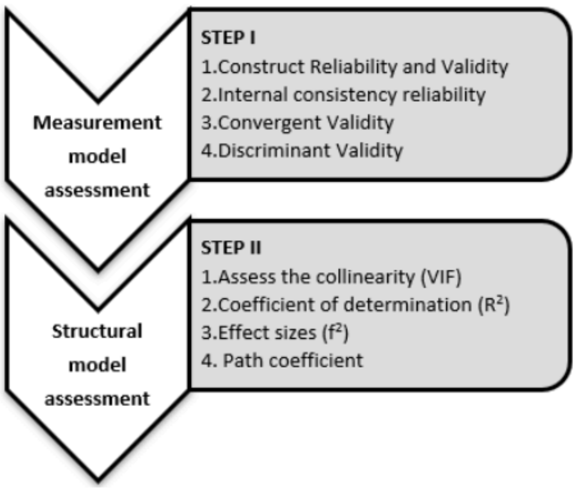

Fig. 2. A Two-Step process of PLS Path model assessment based on Henseler *et al.,* (2015) [10]

The first step of the analysis focused on assessing the measurement model, which involved evaluating construct reliability, internal consistency, convergent validity, and discriminant validity to ensure that each construct was measured accurately. This was followed by the second step, the structural model assessment, which examined multicollinearity, the coefficient of determination (R^2), effect sizes (f^2), and the significance of the hypothesized relationships among latent variables. Although PLS-SEM offers a robust and flexible analytical framework, it is important to recognize its limitations, such as potential biases in parameter estimation and the lack of a universally accepted global goodness-of-fit index [9,18]. All analyses in this study were conducted using SmartPLS version 4.

4. Results
4.1 Participant Demographics

The study sample comprised of 70 undergraduate medical students enrolled in the Bachelor of Medicine and Bachelor of Surgery (MBBS) program in Malaysia. As shown in Table 4, the majority of respondents (65.7%) were between 21 and 23 years of age, followed by 17.1% aged 24-26 years, 12.9% aged 18-20 years, and 4.3% aged 27-29 years. The predominance of young adults, who are likely to be digital natives, suggests a participant group well-acquainted with digital technologies, potentially facilitating openness to adopting AI-based tools such as ChatGPT.

In terms of gender, 54.3% of respondents were male and 45.7% were female, indicating a relatively balanced distribution that reduces the risk of gender-related bias in the results. In terms of academic progression, most participants were in Year 3 (45.7%) or Year 2 (41.4%), with smaller proportions in Year 1 (7.1%), Year 4 (4.3%), and Year 5 (1.4%).

Table 4
Demographic characteristics of participants (N = 70)

Item	Options	N	%
Age	18 - 20 years old	9	12.9
	21 – 23 years old	46	65.7
	24 - 26 years old	12	17.1
	27 - 29 years old	3	4.3
Gender	Male	38	54.3
	Female	32	45.7
Year of Study	Year 1	5	7.1
	Year 2	29	41.4
	Year 3	32	45.7
	Year 4	3	4.3
	Year 5	1	1.4

With respect to their ChatGPT experience, Table 5 indicates that a substantial majority (60.0%) had been using ChatGPT for more than one year, followed by 21.4% with 7-12 months of experience, 12.9% with 1-6 months of experience, and only 5.7% with less than one month of use. This suggests that the majority of our participants has substantial exposure to ChatGPT, which could affect their perceptions of usefulness, ease of use, and habit formation.

Table 5
Experience of using ChatGPT

Usage Experience	N	%
Less than a month	4	5.7
1 - 6 months	9	12.9
7 – 12 months	15	21.4
More than 1 year	42	60.0

In terms of frequency of use as shown in Table 6, half of the respondents (50.0%) reported using ChatGPT several times a week, while 20.0% used it several times a day and 7.1% once a day. Less frequent usage patterns included several times a month (17.1%) and once a month (5.7%). These patterns suggest that ChatGPT has already become a regular academic or personal tool for many students, aligning with UTAUT2's Habit construct as a predictor of continued usage.

Table 6
Frequency of ChatGPT usage among participants

Frequency of Use	N	%
Several times a day	14	20.0
Once a day	5	7.14
Several times a week	35	50.0
Several times a month	12	17.14
Once a month	4	5.71

Overall, this demographic and usage data indicate that the study sample consists predominantly of tech-savvy, medical students with significant prior exposure to ChatGPT. This context is important for interpreting the subsequent analysis, as it may amplify the influence of constructs such as Performance Expectancy, Hedonic Motivation, and Habit, while potentially diminishing the impact of Effort Expectancy due to participants' established familiarity with digital tools.

4.2 Measurement Model Assessment

The evaluation of the measurement model centers on establishing the reliability and validity of the constructs and their corresponding indicators. The first step is evaluating the dependability of the indicators to confirm that the items precisely reflect the structures they want to evaluate. Outer loadings above 0.7 are typically deemed acceptable, although those ranging from 0.4 to 0.7 may be preserved if construct validity standards are satisfied [8].

Secondly, the evaluation of internal consistency reliability is conducted utilizing both Cronbach's alpha (CA) and composite reliability (CR). Cronbach's alpha is commonly used; however, composite reliability is favored in PLS-SEM because of its responsiveness to indicator loadings. Composite reliability (CR) values ranging from 0.7 to 0.95 are considered satisfactory, signifying that the construct items consistently evaluate the same fundamental notion.

Third, convergent validity is evaluated by analyzing the average variance extracted (AVE), which indicates the ratio of variation ascribed to a concept in comparison to measurement error. An Average variation Extracted (AVE) of 0.5 or above signifies adequate convergent validity, indicating that a minimum of 50% of the variation in the indicators is accounted for by the construct.

The evaluation of discriminant validity is conducted to verify that the notions are empirically different. The Fornell-Larcker criteria is commonly employed for this assessment, requiring that the square root of each construct's AVE surpass its correlations with other constructs. The Heterotrait-Monotrait ratio (HTMT) serves as an additional metric, with values under 0.85 (or 0.90 in more lenient contexts) indicating sufficient discriminant validity.

4.2.1 Construct reliability and validity

The following Table 7 shows the construct reliability and validity of the model. Based on the Cronbach's Alpha (CA) and Composite Reliability (CR) values presented, our model exhibit strong internal consistency and reliability, as all constructs surpassing the recommended threshold of 0.70. Behavioural Intention demonstrates high reliability, with a CA value of 0.867 and a CR value of 0.919, while its Average Variance Extracted (AVE) of 0.791 ensures sufficient convergent validity. The factor loadings for all its items exceeds 0.7 with B1 (0.848), B2 (0.922), and B3 (0.896), confirm their substantial contributions.

Table 7
Construct reliability and validity

Construct/Item	Indicator Loading (IL)	Cronbach Alpha (CA)	Average Variance Extracted (AVE)	Composite Reliability (CR)
Behavioural Intention (BI)		0.867	0.919	0.791
B1	0.848			
B2	0.922			
B3	0.896			
Effort Expectancy (EE)		0.870	0.911	0.720
EE1	0.841			
EE2	0.870			
EE3	0.845			
EE4	0.838			
Facilitating Conditions (FC)		0.886	0.921	0.746
FC1	0.902			
FC2	0.889			
FC3	0.880			
FC4	0.780			
Hedonic Motivation (HM)		0.911	0.944	0.849
HM1	0.926			
HM2	0.934			
HM3	0.903			
Habit (HT)				
HT1	0.893			
HT2	0.909			
HT3	0.856			
HT4	0.907			
Performance Expectancy (PE)		0.870	0.911	0.719
PE1	0.801			
PE2	0.856			
PE3	0.885			
PE4	0.848			
Price Value		0.747	0.817	0.602
PV1	0.688			
PV2	0.706			
PV3	0.913			
Social Influence		0.904	0.940	0.839
SI1	0.924			
SI2	0.913			
SI3	0.911			

Effort Expectancy (EE) also meets reliability standards, with a CA of 0.870, a CR of 0.911, and an AVE of 0.720. Besides, all its items, EE1 (0.841), EE2 (0.870), EE3 (0.845), and EE4 (0.838), show strong factor loadings. Facilitating Conditions (FC) similarly exhibits robust internal consistency (CA = 0.886, CR = 0.921, AVE = 0.746), with all items, FC1 (0.902), FC2 (0.889), FC3 (0.880), and FC4 (0.780) contributing significantly to the construct.

Hedonic Motivation (HM) also reflects excellent reliability, with a CA of 0.911, a CR of 0.944, and an AVE of 0.849. The item loadings, HM1 (0.926), HM2 (0.934), and HM3 (0.903) indicate strong contributions. Habit (HT) also maintains strong factor loadings across items with HT1 (0.893), HT2 (0.909), HT3 (0.856), and HT4 (0.907), suggesting high internal consistency.

Performance Expectancy (PE) is well-supported, with a CA of 0.870, a CR of 0.911, and an AVE of 0.719. Its items, PE1 (0.801), PE2 (0.856), PE3 (0.885), and PE4 (0.848) all demonstrate significant contributions to the construct. Price Value (PV), though slightly lower in reliability (CA = 0.747, CR = 0.817, AVE = 0.602), maintains acceptable factor loadings with PV1 (0.688), PV2 (0.706), and PV3 (0.913), with PV3 displaying the highest contribution.

Finally, Social Influence (SI) exhibits strong reliability, with a CA of 0.904, a CR of 0.940, and an AVE of 0.839. The factor loadings for its items, SI1 (0.924), SI2 (0.913), and SI3 (0.911) all exceeds 0.7 which illustrate their solid contributions. Overall, all the constructs in our model meet the necessary reliability and validity criteria, reinforcing their robustness in the study.

4.2.2 Discriminant validity

Discriminant validity is a crucial aspect of construct validity in PLS-SEM that assures the construct is using a unique empirical measure from the other constructs present in the model and thus is measuring a separate phenomenon [8]. Confirming that constructs have been appropriately operationalised, discriminant validity determination is imperative to enabling interpretation of the relationships accurately in the structural model [18]. In PLS-SEM, three primary methods are commonly utilised to evaluate discriminant validity: the Fornell-Larcker criterion, the Heterotrait-Monotrait ratio of correlations (HTMT), and cross loading.

According to the Fornell-Larcker criterion the square root of the average variance extracted (AVE) for each of the constructs has to be greater than its greatest correlation with any other construct in the model. As such, this approach ensures that a construct has more variance with its indicators than other constructs. This approach has been criticized for its potential vulnerability to measurement error and potential lack of sensitivity to discriminant validity issues in certain situations [10]. Table 8 below show the result of Fornell Lacker criterion.

Table 8
Fornell Lacker criterion

Fornell-Lacker	BI	EE	FC	HT	HM	PE	PV	SI
Behavioural Intention (BI)	**0.889**							
Effort Expectancy (EE)	0.585	**0.848**						
Facilitating Conditions (FC)	0.611	0.689	**0.864**					
Habit (HT)	0.653	0.541	0.426	**0.891**				
Hedonic Motivation (HM)	0.627	0.712	0.710	0.557	**0.921**			
Performance Expectancy (PE)	0.527	0.528	0.551	0.534	0.632	**0.848**		
Price Value (PV)	0.459	0.547	0.591	0.491	0.558	0.403	**0.776**	
Social Influence (SI)	0.628	0.583	0.568	0.469	0.700	0.546	0.437	**0.916**
Use Behaviour (UB)	-0.156	-0.279	-0.222	-0.354	-0.334	-0.392	-0.361	-0.262

Based on the above table, Behavioural Intention (BI) has a square root of AVE of 0.889, which is higher than its correlations with other constructs (e.g., 0.585 with Effort Expectancy, 0.611 with Facilitating Conditions, etc.), indicating strong discriminant validity. Effort Expectancy (EE) has a square root of AVE of 0.848, which is greater than its correlations with other constructs (e.g., 0.689 with Facilitating Conditions, 0.712 with Hedonic Motivation, etc.), confirming a clear distinction.

Facilitating Conditions (FC) has a square root of AVE of 0.864, which exceeds its correlations with other constructs supporting discriminant validity. Habit (HT) has a square root of AVE of 0.891, which is greater than its correlations with other constructs (e.g., 0.653 with Behavioural Intention, 0.541 with Effort Expectancy), ensuring a clear distinction. Hedonic Motivation (HM) has a square root of AVE of 0.921, which is higher than its correlations (e.g., 0.627 with Behavioural Intention, 0.710 with

Facilitating Conditions), confirming discriminant validity. Performance Expectancy (PE) has a square root of AVE of 0.848, which is greater than its correlations with other constructs (e.g., 0.527 with Behavioural Intention, 0.528 with Effort Expectancy), indicating it is a distinct construct. Price Value (PV) has a square root of AVE of 0.776, which exceeds its correlations (e.g., 0.459 with Behavioural Intention, 0.547 with Effort Expectancy), confirming discriminant validity. Social Influence (SI) has a square root of AVE of 0.916, which is higher than its correlations (e.g., 0.628 with Behavioural Intention, 0.583 with Effort Expectancy), ensuring a distinct construct.

Besides Fornell Lacker criterion, Henseler *et al.,* (2015) [10] developed the Heterotrait-Monotrait ratio of correlations (HTMT). Procedurally it is a more rigorous and dependable metric to assess the discriminant validity. The analysis is based on a comparison of the proportion of correlations between a variety of constructs (heterotrait correlations) to correlations between measures of the same construct (monotrait correlations). HTMT value should generally be less than 0.85 for distinct constructs, or below 0.90 in more relaxed circumstances [8]. The HTMT has become increasingly significant in recent years because of its superior sensitivity in detecting issues related to discriminant validity when compared to the Fornell-Larcker criterion [10]. If the results indicate that the HTMT value is below 0.85, it confirms that discriminant validity exists between the constructs. The following Table 9 presents the result of HTMT criterion.

Table 9
Heterotrait-Monotrait ratio of correlations (HTMT) criterion

HTMT	BI	EE	FC	HT	HM	PE	PV	SI
Behavioural Intention (BI)								
Effort Expectancy (EE)	0.670							
Facilitating Conditions (FC)	0.693	0.783						
Habit (HT)	0.732	0.602	0.470					
Hedonic Motivation (HM)	0.699	0.797	0.787	0.603				
Performance Expectancy (PE)	0.594	0.613	0.614	0.600	0.712			
Price Value (PV)	0.445	0.544	0.659	0.448	0.562	0.391		
Social Influence (SI)	0.703	0.652	0.626	0.509	0.766	0.604	0.413	
Use Behaviour (UB)	0.167	0.303	0.229	0.368	0.347	0.426	0.342	0.278

Based on the above table, Behavioural Intention (BI) shows moderate correlations with other constructs, with the highest being 0.732 (with Habit). This suggests a relationship but maintains discriminant validity. Effort Expectancy (EE) has an HTMT value of 0.797 with Hedonic Motivation (HM) and 0.783 with Facilitating Conditions (FC). While below 0.85, these values indicate a strong relationship, suggesting possible conceptual overlap. Facilitating Conditions (FC) shows 0.787 with Hedonic Motivation (HM) and 0.693 with Behavioural Intention (BI). These values remain within acceptable limits but indicate a close association. Habit (HT) has a relatively high 0.732 correlation with Behavioural Intention (BI) and 0.603 with Hedonic Motivation (HM). This suggests that individuals' habitual tendencies might influence their intention and enjoyment of using ChatGPT for medical education.

Meanwhile, Hedonic Motivation (HM) exhibits high correlations with several constructs i.e. 0.797 (Effort Expectancy), 0.787 (Facilitating Conditions), and 0.766 (Social Influence). These values suggest strong interrelations, possibly requiring further analysis to confirm that they are distinct. Performance Expectancy (PE) shows moderate correlations with other constructs, with the highest being 0.712 (with Hedonic Motivation) suggesting that perceived usefulness and enjoyment may be closely linked. Price Value (PV) has a 0.659 correlation with Facilitating Conditions, which remains within acceptable limits but suggests some overlap in how affordability and ease of use are perceived. Social Influence (SI) exhibits a relatively high correlation of 0.766 with Hedonic Motivation, implying that enjoyment and social factors may be interconnected in this context.

Another approach of assessing discriminant validity is through assessing cross loading. Cross-loading indicates the degree to which an indicator is associated with multiple constructs instead of primarily aligning with its designated construct. Elevated cross-loadings may signal problems with discriminant validity, implying that the indicator might not distinctly represent the construct it aims to measure [8]. The evaluation typically involves comparing the loading of the indicator on its designated construct against its loadings on alternative constructs, where a robust loading on the intended construct is preferred [5]. The findings from the cross-loading analysis demonstrated that the items align correctly with their designated constructs, thereby confirming discriminant validity through this criterion. The following Table 10 presents the result of Cross Loading.

Table 10
Cross Loading criterion

	BI	EE	FC	HM	HT	PE	PV	SI
B1	**0.848**	0.469	0.520	0.534	0.572	0.427	0.572	0.492
B2	**0.922**	0.551	0.570	0.625	0.626	0.568	0.367	0.637
B3	**0.896**	0.539	0.538	0.505	0.541	0.398	0.295	0.536
EE1	0.459	**0.841**	0.715	0.605	0.398	0.472	0.505	0.378
EE2	0.481	**0.870**	0.492	0.590	0.522	0.435	0.431	0.524
EE3	0.535	**0.845**	0.626	0.580	0.366	0.417	0.450	0.456
EE4	0.504	**0.838**	0.509	0.642	0.552	0.471	0.474	0.615
FC1	0.574	0.677	**0.902**	0.672	0.421	0.540	0.513	0.553
FC2	0.579	0.557	**0.889**	0.614	0.346	0.576	0.536	0.515
FC3	0.479	0.616	**0.880**	0.575	0.357	0.472	0.531	0.482
FC4	0.466	0.520	**0.780**	0.588	0.345	0.275	0.463	0.394
HM1	0.522	0.599	0.616	**0.926**	0.466	0.555	0.394	0.625
HM2	0.593	0.709	0.669	**0.934**	0.468	0.600	0.510	0.618
HM3	0.610	0.653	0.671	**0.903**	0.596	0.588	0.621	0.686
HT1	0.586	0.530	0.349	0.496	**0.893**	0.513	0.439	0.426
HT2	0.583	0.476	0.395	0.428	**0.909**	0.511	0.398	0.393
HT3	0.549	0.346	0.309	0.434	**0.856**	0.422	0.385	0.287
HT4	0.609	0.560	0.460	0.615	**0.907**	0.455	0.520	0.548
PE1	0.361	0.563	0.521	0.579	0.466	**0.801**	0.475	0.374
PE2	0.434	0.320	0.417	0.507	0.454	**0.856**	0.249	0.352
PE3	0.517	0.521	0.574	0.553	0.481	**0.885**	0.346	0.539
PE4	0.455	0.403	0.358	0.518	0.416	**0.848**	0.326	0.560
PV1	0.178	0.178	0.352	0.310	0.129	0.182	**0.688**	0.242
PV2	0.168	0.290	0.390	0.258	0.218	0.075	**0.706**	0.108
PV3	0.517	0.600	0.567	0.580	0.562	0.469	**0.913**	0.484
SI1	0.617	0.607	0.577	0.704	0.446	0.536	0.459	**0.924**
SI2	0.542	0.486	0.458	0.576	0.402	0.434	0.372	**0.913**
SI3	0.562	0.502	0.518	0.636	0.438	0.525	0.364	**0.911**

Based on the above Table 10, Behavioural Intention (BI) demonstrates strong loadings ranging from 0.848 to 0.922, with B2 (0.922) being the strongest indicator of this construct. Effort Expectancy (EE) also exhibits high reliability, with all items loading above 0.83, and EE2 (0.870) being the most significant item. Facilitating Conditions (FC) show consistently strong loadings above 0.78, although FC4 (0.780) is slightly weaker than the others but still within an acceptable range. Hedonic Motivation (HM) presents very strong loadings between 0.903 and 0.934, with HM2 (0.934) emerging as the strongest, indicating high internal consistency. Habit (HT) is well-supported, with all items loading above 0.85, and HT2 (0.909) and HT4 (0.907) being the strongest indicators. Performance Expectancy (PE) also shows high reliability, with loadings ranging from 0.801 to 0.885, and PE3 (0.885) being the most dominant item. For Price Value (PV), PV3 (0.913) is the strongest, while PV1 (0.688) and PV2 (0.706) are relatively moderate. The lower loading of PV1 (0.688) may indicate some potential concerns. Lastly, Social Influence (SI) exhibits very strong loadings, all exceeding 0.91, with SI1 (0.924) being the highest indicator.

4.3 Structural Model Assessment

The structural model is evaluated to determine whether the suggested links between constructs are valid after the measurement model meets the required standards. First, to avoid bias in path estimations, it is crucial to assess the collinearity across predictor variables. This is typically done using the variance inflation factor (VIF), with values less than 5 being considered acceptable. Then, the endogenous constructs' explanatory power is assessed using the coefficient of determination (R^2). Weak, moderate, and strong explanatory power are shown by R^2 values of 0.25, 0.50, and 0.75, respectively. Besides, to assess the impact of external constructions on endogenous constructs, effect sizes (f^2) are calculated; small, medium, and large impacts are represented by values of 0.02, 0.15, and 0.35, respectively. The strength and significance of the suggested associations are then evaluated by analyzing the path coefficients, using bootstrapping to calculate p-values and standard errors. These comprehensive assessment of the measurement and structural models in PLS-SEM ensures the validity and reliability of the constructs and the empirical backing of the suggested connections between them. This comprehensive evaluation improves the validity and comprehensibility of the findings in academic research.

4.3.1 VIF @ Multicollinearity

In partial least squares structural equation modelling (PLS-SEM), multicollinearity denotes a strong connection between several predictor variables in a model. This behaviour may result in inaccurate calculations of path coefficients and reduce the findings' dependability. Standard errors may be inflated by multicollinearity, making it more difficult to evaluate each variable's unique contribution to the model [8]. The Variance Inflation Factor (VIF) is a technique used in PLS-SEM that provides insights into the presence of multicollinearity among predictor variables, ensuring that independent variables maintain distinct contributions to the model. A VIF below 5 is generally considered acceptable, indicating that multicollinearity is not a major issue [18]. The following Table 11 presents the results of VIF analysis.

Table 11
Variance Inflation Factor (VIF) criterion

Structural Path	VIF
Performance Expectancy → Behavioural Intention	1.913
Effort Expectancy → Behavioural Intention	2.552
Social Influence → Behavioural Intention	2.092
Facilitating Conditions → Behavioural Intention	2.632
Facilitating Conditions → Use Behaviour	1.600

Hedonic Motivation → Behavioural Intention	3.409
Price Value →Behavioural Intention	1.775
Habit → Behavioural Intention	1.772
Habit → Use Behaviour	1.748
Behavioural Intention → Use Behaviour	2.285
Effort Expectancy → Behavioural Intention	2.552

Based on the above table, the relationship between Behavioural Intention (BI) and Use Behaviour (UB) has a VIF of 2.285, suggesting a moderate level of collinearity, reinforcing that Behavioural Intention is a strong predictor of Use Behaviour while maintaining an independent effect. Effort Expectancy (EE) → Behavioural Intention (BI) has a VIF of 2.552, indicating that Effort Expectancy significantly contributes to shaping Behavioural Intention, although it might share some variance with related constructs such as Performance Expectancy and Facilitating Conditions. Facilitating Conditions (FC) influences both Behavioural Intention (VIF = 2.632) and Use Behaviour (VIF = 1.600). These values suggest that Facilitating Conditions plays a meaningful role in predicting both Behavioural Intention and Use Behaviour, though it does not introduce serious multicollinearity concerns.

Habit (HT) influences Behavioural Intention (VIF = 1.772) and Use Behaviour (VIF = 1.748), indicating that Habit is a relevant predictor but remains relatively independent from other constructs. Hedonic Motivation (HM) → Behavioural Intention has the highest VIF (3.409) in this model, emphasizing its strong role in predicting Behavioural Intention. However, its relatively higher VIF suggests that it shares some explanatory power with other constructs, particularly Performance Expectancy or Effort Expectancy. Performance Expectancy (PE) → Behavioural Intention shows a VIF of 1.913, reflecting a reasonable level of collinearity while still being an important predictor. Price Value (PV) → Behavioural Intention has a VIF of 1.775, suggesting that Price Value contributes to Behavioural Intention without major collinearity concerns.

Lastly, Social Influence (SI) → Behavioural Intention has a VIF of 2.092, demonstrating that Social Influence is a notable factor in shaping Behavioural Intention but does not significantly overlap with other constructs. Our model does not exhibit severe multicollinearity, as all VIF values remain below the critical threshold of 5. The strongest predictors of Behavioural Intention include Hedonic Motivation (VIF = 3.409), Facilitating Conditions (VIF = 2.632), and Effort Expectancy (VIF = 2.552), indicating their substantial influence. Meanwhile, Facilitating Conditions → Use Behaviour (VIF = 1.600) and Habit → Use Behaviour (VIF = 1.748) show lower VIF values, indicating that these predictors contribute more independently. Overall, the model appears robust, though Hedonic Motivation's slightly higher VIF suggests it may share some explanatory power with other related constructs. Further investigation into potential overlapping influences could refine the model.

4.3.2 Coefficient of determination (R^2)

The coefficient of determination (R^2) values indicates the proportion of variance in each endogenous construct explained by its predictors. It reflects the model's ability to capture key determinants that shape an individual's intention to use the system or engage in the behaviour. According to Henseler et al. (2015), R^2 values between 0.50 and 0.75 are considered moderate, while values below 0.25 are weak. The following Table 12 presents the R^2 value for our research.

Table 12
Coefficient of determination (R^2)

Endogenous construct	R-square	Interpretation (Henseler et al., 2015)
Behavioural Intention	0.612	Moderate
Use Behaviour	0.156	Weak

Based on the above table, the R^2 value for Behavioural Intention (BI) is 0.612, indicating a moderate level of explanatory power. This suggests that approximately 61.2% of the variance in Behavioural Intention is explained by its predictors: Performance Expectancy, Effort Expectancy, Social Influence, Price Value, Habit, Hedonic Motivation, and Facilitating Conditions.

In contrast, the R^2 value for Use Behaviour (UB) is 0.156, representing a weak level of explanatory power, which implies that only 15.6% of the variance in actual system usage is explained by Behavioural Intention. While Behavioural Intention to use ChatGPT plays a role in influencing Use Behaviour, this relatively low R^2 value suggests that other external factors such as institutional policies, curriculum integration, usability constraints, and available training, may substantially influence actual usage beyond what the model currently captures.

Overall, the research model demonstrates a moderate predictive capability for Behavioural Intention, affirming the relevance of the included constructs in understanding intention formation. However, its limited ability to predict actual Use Behaviour indicates the need to explore more additional variables to enhance the model's explanatory power regarding actual ChatGPT usage for medical education.

4.3.3 The effect size (f^2)

The effect size, f-square (f^2) analysis of the structural model provides insight into the relative contribution of each predictor variable in explaining medical students' acceptance and use of ChatGPT for medical education. Based on Cohen's (1988) classification, an f^2 value of 0.02 represents a small effect, 0.15 a medium effect, and 0.35 or higher a large effect. The following Table 13 presents the effect sizes (f^2) of various predictors on Behavioural Intention and Use Behaviour.

Table 13
Effect size (f^2)

Structural Path	f^2	Interpretation (Cohen, 1988)
Performance Expectancy → Behavioural Intention	0.000	No effect
Effort Expectancy → Behavioural Intention	0.001	Small
Social Influence → Behavioural Intention	0.089	Small
Facilitating Conditions → Behavioural Intention	0.077	Small
Facilitating Conditions → Use Behaviour	0.025	Small
Hedonic Motivation → Behavioural Intention	0.000	No effect
Price Value →Behavioural Intention	0.004	Small
Habit → Behavioural Intention	0.239	Medium
Habit → Use Behaviour	0.126	Small
Behavioural Intention → Use Behaviour	0.030	Small

Based on the above table, only one structural path demonstrates a medium effect size i.e. the effect of Habit on Behavioural Intention (f^2 = 0.239). This indicates that habitual use of ChatGPT plays a significant role in shaping medical students' intention to continue using ChatGPT for their studies. Additionally, the effect of Habit on Use Behaviour (f^2 = 0.126) reflects a small-to-moderate effect, suggesting that students who are already accustomed to using ChatGPT are more likely to actively engage with it in their learning activities.

Several other predictors exhibit small effect sizes, including the effect of Behavioural Intention on Use Behaviour ($f^2 = 0.030$), Facilitating Conditions on Behavioural Intention ($f^2 = 0.077$), Facilitating Conditions on Use Behaviour ($f^2 = 0.025$), and Social Influence on Behavioural Intention ($f^2 = 0.089$). These findings imply that access to enabling infrastructure (e.g., institutional support, reliable internet access) and encouragement from peers or educators contribute modestly to both students' intentions and actual usage of ChatGPT.

In contrast, the effects of Effort Expectancy ($f^2 = 0.001$), Price Value ($f^2 = 0.004$), Hedonic Motivation ($f^2 = 0.000$), and Performance Expectancy ($f^2 = 0.000$) on Behavioural Intention are negligible or non-existent. This suggests that, in this context, factors such as ease of use, cost considerations, enjoyment, and perceived usefulness do not significantly influence medical students' intention to use ChatGPT for medical education. Overall, the findings highlight the central role of habit in influencing both intention and usage, while indicating that more traditional technology acceptance variables may have limited relevance in this specific educational setting.

4.3.4 Direct effect

The path coefficient analysis provides insights into the relationships between variables influencing Behavioural Intention and Use Behaviour. The path coefficient values (β) indicate the strength and direction of the relationships, while the T-statistic (T) and p-value (P) determine the statistical significance of each path. A p-value below 0.05 indicates a significant relationship. The following Figure 3 and Table 14 present the significance testing results of structural model path coefficients.

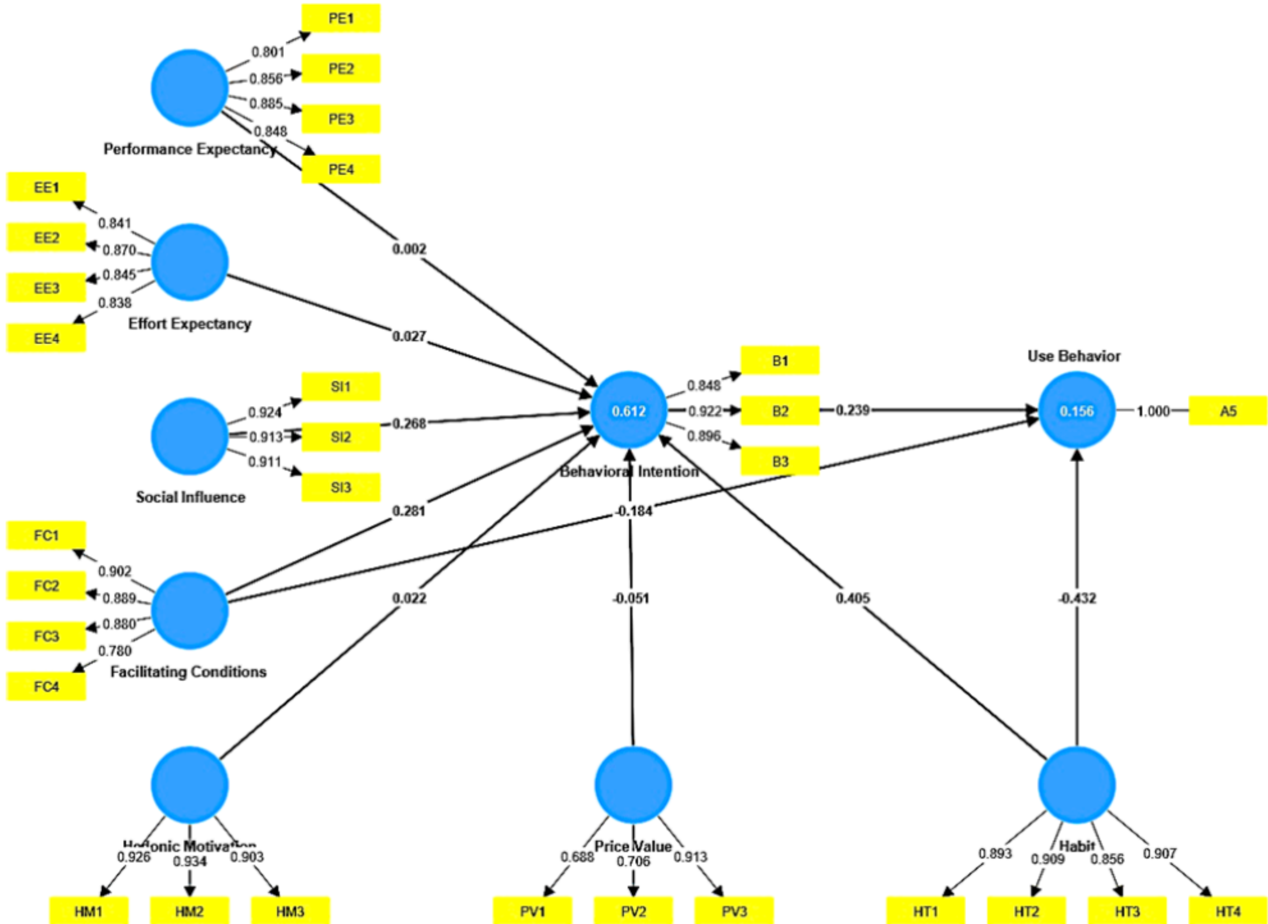

Fig. 3. Path coefficient estimation results of the structural model

Table 14
Significance test of the path coefficients corresponding to each hypothesis

Hypothesis	Structural Path	β	T	p	Supported?
H1	Performance Expectancy → Behavioural Intention	0.002	0.013	0.990	No
H2	Effort Expectancy → Behavioural Intention	0.027	0.186	0.853	No
H3	Social Influence → Behavioural Intention	0.268	1.791	0.073	No
H4	Facilitating Conditions → Behavioural Intention	0.281	2.149	0.032	Yes
H5	Facilitating Conditions → Use Behaviour	-0.184	1.393	0.164	No
H6	Hedonic Motivation → Behavioural Intention	0.022	0.140	0.889	No
H7	Price Value →Behavioural Intention	-0.051	0.445	0.656	No
H8	Habit → Behavioural Intention	0.405	2.542	0.011	Yes
H9	Habit → Use Behaviour	-0.432	3.232	0.001	Yes (negative)
H10	Behavioural Intention → Use Behaviour	0.239	1.604	0.109	No

β: Path Coefficient value, T: T-value; P: Significance Level (p-value). A p-value less than 0.05 indicates a statistically significant relationship.

Based on the above table, Habit emerges as the strongest positive predictor of Behavioural Intention (β = 0.405, T = 2.542, p = 0.011), indicating that users' established routines play a critical role in shaping their intention to use the system. Additionally, Facilitating Conditions positively influence Behavioural Intention (β = 0.281, T = 2.149, p = 0.032), suggesting that the availability of resources, infrastructure, and technical support encourages users to adopt the system. However, Habit has a significant negative effect on Use Behaviour (β = -0.432, T = 3.232, p = 0.001), implying that while users may intend to use the system due to habitual Behaviour, it does not always translate into actual usage, possibly due to resistance to change or reliance on existing practices.

Conversely, several relationships are not statistically significant, indicating that certain factors have minimal influence on users' decision-making processes. Behavioural Intention does not significantly predict Use Behaviour (β = 0.239, T = 1.604, p = 0.109), highlighting a gap between users' intention and their actual Behaviour. Similarly, Facilitating Conditions do not directly impact Use Behaviour (β = -0.184, T = 1.393, p = 0.164), suggesting that while resources and support systems promote intention, they do not necessarily lead to usage. External influences such as Social Influence (β = 0.268, T = 1.791, p = 0.073) show moderate but statistically insignificant effects on Behavioural Intention, indicating that peer pressure or organizational norms may not play a decisive role in system adoption.

Furthermore, factors like Effort Expectancy (β = 0.027, p = 0.853), Hedonic Motivation (β = 0.022, p = 0.889), Performance Expectancy (β = 0.002, p = 0.990), and Price Value (β = -0.051, p = 0.656) have negligible impacts on Behavioural Intention, implying that ease of use, enjoyment, perceived benefits, and cost considerations are not primary motivators in this context. The findings underscore the pivotal role of Habit and Facilitating Conditions in shaping users' intentions, while also revealing that intention alone does not guarantee actual system adoption. To bridge the gap between intention and Behaviour, organizations should focus on integrating new systems into users' daily routines, offering comprehensive support, and implementing Behavioural reinforcement strategies. Additionally, fostering an environment that encourages consistent system use through training and incentive programs could help translate users' intentions into sustained usage.

4.3.5 Mediating effect

The mediating effect analysis investigates whether Behavioural Intention serves as a bridge between various determinants, such as Effort Expectancy, Habit, Hedonic Motivation, and Social Influence and actual Use Behaviour. The result is presented in the following Table 15.

Table 15
The result of mediating effect analysis

Structural Path	β	T	p	Mediation Supported?
Performance Expectancy → Behavioural Intention → Use Behaviour	0.000	0.011	0.991	No
Effort Expectancy → Behavioural Intention → Use Behaviour	0.007	0.158	0.875	No
Social Influence → Behavioural Intention→Use Behaviour	0.064	1.246	0.213	No
Facilitating Conditions → Behavioural Intention → Use Behaviour	0.067	1.008	0.313	No
Hedonic Motivation → Behavioural Intention → Use Behaviour	0.005	0.118	0.906	No
Price Value → Behavioural Intention -→ Use Behaviour	-0.012	0.359	0.719	No
Habit → Behavioural Intention → Use Behaviour	0.097	1.202	0.229	No

Based on the above table, the mediating effect analysis on the relationship between various independent variables and Use Behaviour through Behavioural Intention is found to be statistically insignificant across all tested relationships. Effort Expectancy (β = 0.007, p = 0.875), Hedonic Motivation (β = 0.005, p = 0.906), and Performance Expectancy (β = 0.000, p = 0.991) demonstrate negligible mediation effects, indicating that these factors do not significantly influence Use Behaviour through Behavioural Intention. Similarly, Facilitating Conditions (β = 0.067, p = 0.313) and Social Influence (β = 0.064, p = 0.213) exhibit slightly higher mediation effects but remain statistically insignificant. Among all variables, Habit (β = 0.097, p = 0.229) shows the highest mediation effect, yet it fails to reach statistical significance. Notably, Price Value (β = -0.012, p = 0.719) exhibits a negative mediation effect, suggesting that pricing considerations may not contribute positively to Use Behaviour through Behavioural Intention. The findings indicate that Behavioural Intention does not significantly mediate the relationship between these factors and Use Behaviour in the medical context, suggesting the need to explore alternative mediating mechanisms or direct influences on Use Behaviour.

5. Discussions

The present study yields several notable findings that both corroborate and diverge from extant research on the adoption of ChatGPT and related AI technologies within higher education, particularly through the lens of the UTAUT2 framework.

Firstly, the significant influence of Facilitating Conditions on Behavioural Intention observed in this study aligns with a subset of empirical investigations emphasizing the critical role of institutional infrastructure and resource availability in enabling technology uptake among students [4,17]. However, this finding must be contextualized within a broader body of literature reporting heterogeneous effects of facilitating conditions, which vary according to disciplinary context and technological complexity [3]. Our results therefore contribute to this nuanced understanding by affirming the importance of facilitating factors in the medical education setting, where resource adequacy and technical support are paramount.

Secondly, Habit emerged as a particularly salient construct, exerting a medium-sized positive effect on Behavioural Intention, consistent with theoretical propositions and empirical evidence in technology acceptance research [12,19]. Habit's influence on intention corroborates findings from recent studies of ChatGPT use, which highlight the centrality of prior usage patterns in shaping continued engagement [6,16]. Notably, our data revealed a counterintuitive negative direct effect of Habit on actual Use Behaviour, a divergence from established literature where habit typically reinforces ongoing use. This anomaly may reflect context-specific factors prevalent in medical education, such as heightened ethical concerns, accuracy-related apprehensions, or institutional restrictions that temper habitual use despite strong intentions. Such complexities underscore the necessity of integrating contextually relevant moderators in acceptance models applied within professional educational domains.

Thirdly, contrary to extensive prior research underscoring the pivotal roles of Performance Expectancy and Effort Expectancy as primary drivers of behavioural intention [3,17,19], our findings demonstrate their limited impact in the current medical student sample. This departure suggests that in this context, considerations of ease of use and perceived utility are superseded by other determinants, potentially including trustworthiness, ethical integrity, and institutional endorsement. This interpretation aligns with studies documenting medical students' pronounced concerns regarding the reliability of AI-generated information and the implications of misuse [2,11], which may diminish the salience of conventional TAM and UTAUT2 constructs.

Similarly, Hedonic Motivation and Price Value did not exert significant influence on intention, corroborating research suggesting that when AI tools are deployed primarily for goal-oriented academic purposes and provided free of charge, enjoyment and cost factors become less relevant [2,11]. This finding diverges, however, from studies identifying hedonic motivation as a salient predictor among younger or less experienced users [12,19], underscoring the potential moderating effect of user maturity and context specificity in technology acceptance.

The marginal role of Social Influence detected here mirrors the heterogeneous findings in the literature regarding peer and instructor impact on AI adoption [3,17]. This near-significant result may reflect the predominantly self-directed learning environment of medical education, wherein social pressures are attenuated relative to other educational contexts [16].

Further, the mediating effect analysis revealed that Behavioural Intention does not significantly mediate the relationships between antecedent variables and Use Behaviour, a finding that challenges a core tenet of the UTAUT2 model which posits intention as a proximal determinant of behaviour [12]. This discrepancy resonates with emergent literature cautioning that in domains characterized by ethical sensitivities and institutional oversight such as medical education, intentions may not readily translate into usage due to external constraints [2,13].

Finally, the model's explanatory power was moderate for Behavioural Intention (R squared = 0.612) but weak for Use Behaviour (R squared = 0.156), a pattern consistent with prior research indicating that while intentions can be robustly predicted, actual technology use is often influenced by additional unmeasured contextual variables including organizational policies, trust in system outputs, and task demands [6,11]. This differential underscores the complexity of modelling actual behaviour in professional educational settings and points to the necessity of expanding theoretical frameworks to encompass situational and institutional moderators.

In sum, the findings both affirm the applicability of UTAUT2 constructs, particularly Habit and Facilitating Conditions, in the context of AI tool adoption in medical education and highlight critical divergences that invite the integration of context-specific factors such as ethical considerations, trust, and institutional constraints. These insights advance understanding of technology acceptance in high-stakes professional learning environments and suggest directions for future research that more fully account for the complexities of AI adoption in such domains.

6. Theoretical and Practical Implications

From theoretical perspective, the findings reported in this study contribute to the ongoing development and contextual adaptation of technology acceptance theories, particularly the UTAUT2 model. The diminished influence of traditionally robust predictors such as Performance Expectancy, Effort Expectancy, and Hedonic Motivation on Behavioural Intention suggests that the standard model may not fully capture the nuances of technology adoption in specialized, high-stakes contexts like medical education. This underscores the importance of integrating context-specific variables such as trust in the accuracy of AI-generated information, ethical considerations, and institutional constraints that are likely to shape acceptance decisions within professional education environments.

Furthermore, the complex role of Habit observed here, wherein it positively predicts Behavioural Intention but exhibits a negative direct effect on actual Use Behaviour, calls for a deeper examination of mediating or moderating factors, such as ethical concerns and organizational policies, that influence the transition from intention to action. This nuanced relationship challenges the commonly assumed positive habit-use link reported in broader technology acceptance literature.

Additionally, the lack of significant mediation by Behavioural Intention in the relationship between antecedents and Use Behaviour highlights an intention-behaviour gap, particularly pronounced in medical education settings. This gap suggests that acceptance models should be expanded to account for external moderating variables such as institutional policies and risk perceptions that may disrupt the conventional pathway from intention to usage. Lastly, the model's stronger explanatory power for Behavioural Intention than for Use Behaviour indicates that future theoretical advancements must incorporate situational and contextual factors to enhance the predictive validity of technology use models in professional education.

From a practical perspective, the results emphasize the critical role of institutional support in facilitating the adoption of AI tools such as ChatGPT among medical students. Educational institutions should therefore invest in strengthening technical infrastructure, ensuring reliable access, and providing dedicated IT support tailored to the specific needs of their learners. Equally important is the need to address ethical and accuracy concerns that may hinder usage, particularly given the complex role Habit plays in shaping intention but not necessarily behaviour. Institutions are encouraged to develop clear ethical guidelines and deliver targeted training programs that promote responsible use of AI, including critical evaluation of outputs to mitigate risks related to misinformation and academic dishonesty. Integrating AI literacy into medical curricula is essential, enabling students not only to use ChatGPT effectively but also to understand its limitations and maintain critical thinking skills.

Furthermore, the discrepancy between intention and actual use observed in this study highlights the importance of proactive strategies to encourage habitual, sustained engagement with ChatGPT. This could involve incorporating supervised AI-assisted assignments and reflective activities that support students in transitioning from intention to consistent, appropriate use. Recognizing the intention-use gap also calls for institutional efforts to identify and alleviate barriers such as restrictive policies or trust issues that may prevent students from fully leveraging AI tools. Lastly, given the variability in acceptance influenced by demographic and contextual factors identified in prior research, tailored support initiatives that consider students' digital literacy, prior experience, and academic backgrounds could further enhance ChatGPT adoption and effective utilization in medical education.

7. Limitations and Future Work

While this study provides valuable insights into medical students' acceptance and use of ChatGPT through the UTAUT2 framework, several limitations should be acknowledged. First, the cross-sectional research design restricts the ability to infer causal relationships or examine changes in behavioural intention and use over time. Longitudinal studies are needed to capture dynamic adoption processes, including how habitual use evolves and how intentions translate into sustained usage in different educational phases.

Second, the sample was limited to medical students from a cultural context. This potentially limits the generalizability of findings to other disciplines, educational settings, or geographical regions. Future research should replicate and extend this investigation across diverse populations and academic fields to enhance external validity and explore contextual variability.

Third, although the study integrated key UTAUT2 constructs, it did not incorporate potentially important moderators and mediators such as trust in AI accuracy, perceived ethical risks, institutional policies, or curriculum integration. These factors likely play a critical role in shaping both intention and use in medical education, as suggested by our comparative analysis. Subsequent studies should include these variables to develop a more comprehensive and contextually grounded understanding of AI adoption.

Fourth, the relatively low explanatory power for actual Use Behaviour indicates that additional external and situational factors such as organizational constraints, time availability, or assessment requirements, may influence ChatGPT usage but were not captured in the current model. Future research would benefit from mixed-methods approaches, combining quantitative modelling with qualitative inquiry to uncover these nuanced influences.

Finally, as ChatGPT and similar AI tools rapidly evolve, continuous investigation is warranted to assess how technological advancements, shifting educational policies, and changing student perceptions impact adoption patterns. Experimental or intervention-based research could also evaluate the effectiveness of educational strategies aimed at promoting responsible and effective AI use. Addressing these limitations and expanding upon the present findings will contribute to a more robust and nuanced understanding of AI technology acceptance in medical education and beyond, ultimately informing the design of supportive learning environments and ethical frameworks for AI integration.

8. Conclusions

This study examined the factors influencing medical students' acceptance and use of ChatGPT for educational purposes through the lens of the UTAUT2 framework. Our findings highlight the significant role of Habit and Facilitating Conditions in shaping behavioural intention, while traditional predictors such as Performance Expectancy and Effort Expectancy appeared less influential in this context. Notably, the unexpected negative direct effect of Habit on actual use and the lack of mediation by behavioural intention underscore the complex and context-specific nature of AI adoption in medical education.

The model demonstrated moderate explanatory power for behavioural intention but limited ability to predict actual use, suggesting that additional contextual and situational factors merit further investigation. By comparing these results with existing literature, this study contributes to a deeper understanding of how emerging AI technologies are perceived and utilized by future healthcare professionals.

Overall, this research underscores the importance of institutional support, habitual engagement, and the nuanced evaluation of AI tools within rigorous academic settings. These insights can guide educators, policymakers, and developers in fostering responsible, effective, and ethically informed integration of AI in medical education, ultimately enhancing learning outcomes and preparing students for the evolving digital landscape.

Acknowledgement
This research was self-funded by corresponding author, and not funded by any research grant.

References
[1] Abas, Mohd Azman, Sathi Eswari Arumugam, Melor Md Yunus, and Karmila Rafiqah M. Rafiq. "ChatGPT and personalized learning: Opportunities and challenges in higher education." *International Journal of Academic Research in Business and Social Sciences* 13, no. 12 (2023): 3536-3545. https://doi.org/10.6007/IJARBSS/v13-i12/20240

[2] Abdelhafiz, Ahmed Samir, Maysa I. Farghly, Eman Anwar Sultan, Moaz Elsayed Abouelmagd, Youssef Ashmawy, and Eman Hany Elsebaie. "Medical students and ChatGPT: analyzing attitudes, practices, and academic perceptions." *BMC Medical Education* 25, no. 1 (2025): 187. https://doi.org/10.1186/s12909-025-06731-9

[3] Al-kfairy, Mousa. "Factors impacting the adoption and acceptance of ChatGPT in educational settings: A narrative review of empirical studies." *Applied System Innovation* 7, no. 6 (2024): 110. https://doi.org/10.3390/asi7060110

[4] Almogren, Abeer S., Waleed Mugahed Al-Rahmi, and Nisar Ahmed Dahri. "Exploring factors influencing the acceptance of ChatGPT in higher education: A smart education perspective." *Heliyon* 10, no. 11 (2024). https://doi.org/10.1016/j.heliyon.2024.e31887

[5] Fornell, Claes, and David F. Larcker. "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error." *Journal of Marketing Research* 18, no. 1 (1981): 39–50. https://doi.org/10.1177/002224378101800104

[6] Groothuijsen, Suzanne, Antoine van den Beemt, Joris C. Remmers, and Ludo W. van Meeuwen. "AI chatbots in programming education: Students' use in a scientific computing course and consequences for learning." *Computers and Education: Artificial Intelligence* 7 (2024): 100290. https://doi.org/10.1016/j.caeai.2024.100290

[7] Guo, Alyssa A., and Jason Li. "Harnessing the power of ChatGPT in medical education." *Medical Teacher* 45, no. 9 (2023): 1063-1063. https://doi.org/10.1080/0142159X.2023.2198094

[8] Hair, Joseph F., William C. Black, Barry J. Babin, and Rolph E. Anderson. *Multivariate Data Analysis*. 8th ed. Boston, MA: Cengage Learning, 2019.

[9] Hair, Joseph F. *A primer on partial least squares structural equation modeling (PLS-SEM)*. sage, 2014.

[10] Henseler, Jörg, Christian M. Ringle, and Marko Sarstedt. "A new criterion for assessing discriminant validity in variance-based structural equation modeling." *Journal of the academy of marketing science* 43, no. 1 (2015): 115-135. https://doi.org/10.1007/s11747-014-0403-8

[11] Kieu, Luu Thi Thu, Phan Huu Vinh, Nguyen Thi Mai, Trinh Khanh Linh, Tran Xuan Dai, Tran Ngan Hanh, Nguyen Ngoc Linh et al. "ChatGPT Usage for Academic Purposes and Influencing Factors Among Students at Hanoi Medical University, Vietnam." *Health Dynamics* 2, no. 5 (2025): 183-192. https://doi.org/10.33846/hd20502

[12] Venkatesh, Viswanath, James Y. L. Thong, and Xin Xu. “Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology.” *MIS Quarterly* 36, no. 1 (2012): 157–178. https://doi.org/10.2307/41410412

[13] Mohammad, Bushra, Turjana Supti, Mahmood Alzubaidi, Hurmat Shah, Tanvir Alam, Zubair Shah, and Mowafa Househ. "The pros and cons of using ChatGPT in medical education: a scoping review." *Healthcare Transformation with Informatics and Artificial Intelligence* (2023): 644-647. https://doi.org/10.3233/SHTI230580

[14] Tangadulrat, Pasin, Supinya Sono, and Boonsin Tangtrakulwanich. “Using ChatGPT for Clinical Practice and Medical Education: Cross-Sectional Survey of Medical Students’ and Physicians’ Perceptions.” *JMIR Medical Education* 9 (2023): e50658. https://doi.org/10.2196/50658

[15] Wu, Y., Y. Zheng, B. Feng, Y. Yang, K. Kang, and A. Zhao. “Embracing ChatGPT for Medical Education: Exploring Its Impact on Doctors and Medical Students.” *JMIR Medical Education* 10 (2024): e52483. https://doi.org/10.2196/52483

[16] Zhang, Janice S., Christine Yoon, Darnell K. Adrian Williams, and Adi Pinkas. "Exploring the usage of ChatGPT among medical students in the United States." *Journal of Medical Education and Curricular Development* 11 (2024): 23821205241264695. https://doi.org/10.1177/23821205241264695

[17] Widyaningrum, Retno, Fitri Wulandari, M. Zainudin, Arina Athiyallah, and Miftahir Rizqa. "Exploring the factors affecting ChatGPT acceptance among university students." *Multidisciplinary Science Journal* 6, no. 12 (2024): 2024273-2024273. https://doi.org/10.31893/multiscience.2024273

[18] Sarstedt, Marko, Christian M. Ringle, and Joseph F. Hair. "Partial least squares structural equation modeling." In *Handbook of market research*, pp. 587-632. Cham: Springer International Publishing, 2021. https://doi.org/10.1007/978-3-319-57413-4_15

[19] Strzelecki, Artur. "Students’ acceptance of ChatGPT in higher education: An extended unified theory of acceptance and use of technology." *Innovative higher education* 49, no. 2 (2024): 223-245. https://doi.org/10.1007/s10755-023-09686-1