ISSN: 2948-4863

JADHUR

JOURNAL OF ADVANCES IN HUMANITIES RESEARCH

VOL.4, ISSUE.3, 2025

WWW.JADHUR.COM

Journal of Advances in Humanities Research (JADHUR) is a scientific, international, peer-reviewed, and open-access journal devoted to the field of Humanities and provides rapid publication of articles frequently in its issues. JADHUR publishes 4 issues per year under United Frontiers Publisher. JADHUR publishes original research papers, review articles, case studies and reports etc. JADHUR is the place for exchanging information and research results within all areas of HUMANITIES and allied fields.

Generally, accepted papers will appear online within 2-3 weeks. The journal publishes original papers including but not limited to the following fields:

- Anthropology
- Tourism
- Sociology
- History
- Criminology
- Cross-Cultural Studies
- Development Studies
- Education
- Human Geography
- International Relations
- Law
- Linguistics
- Literature
- Political Science
- Women Studies
- Social Psychology

Journal of Advances in Humanities Research (JADHUR) operates through the peer-review process and is associated with the United Frontiers Publisher. It is not associated with or does not represent any governmental, non-governmental organization, or any public or private university.

Note: Authors are requested to submit their work electronically by using JADHUR online submission system. Please follow the instructions provided on each step during the online submission process. Please follow the Author's Guidelines before submission.

 ISSN: 2948-4863

Editorial Board

 ISSN: 2948-4863

Dr. Dandan Tang
PhD in Child Education, Faculty of Teacher Education, Lishui University, Lishui ciry, 323000, Zhejiang province, China
Dr. Takaharu Oda
Ph.D in Philosophy, Southern University of Science and Technology

Table of Content

JADHUR

Vol.4, Issue.3, (2025)

Journal of Advances in Humanities Research

https://doi.org/10.56868/jadhur.v4i3.313

The Moderating Role of Socio-cultural Factors and Perceived Enjoyment in Social Media-Assisted EFL Learning: A Systematic Review

Alsharari Taleb Falah[1*] and Lubna Ali Mohammed[1]

[1]Department Of Education, Lincoln University Petaling Jaya, 47301, Malaysia

Article Information

Article Type: Review Article

Dates:
Received: August 02, 2025
Revised: September 02, 2025
Accepted: September 10, 2025
Available online: September 11, 2025

Corresponding Author:

Alsharari Taleb Falah

Taaleb.09@gmail.com

ABSTRACT

This systematic review examines the influence of socio-cultural factors and perceived enjoyment on Saudi university students' use of social media for learning English as a Foreign Language. The review identifies key cultural and motivational determinants that shape the effectiveness of social media-based language learning and introduces a conceptual model to guide future research. Employing the PRISMA framework, 45 studies published between 2015 and 2024 in Scopus and Web of Science were analysed. The findings reveal four primary themes: socio-cultural barriers and supports influenced by gender norms and religious expectations; perceived enjoyment functioning as both a motivator and a potential distraction; variations in learning behaviours across platforms such as YouTube, Facebook, and X/Twitter; and tensions between informal and formal learning environments. This review synthesizes global evidence regarding the impact of socio-cultural factors and enjoyment on EFL learning facilitated by social media. The principal contribution is a novel conceptual model integrating socio-cultural theory and the Technology Acceptance Model with perceived enjoyment.

Keywords: Social media, EFL Learning, Socio-Cultural Factors, Perceived Enjoyment, Saudi Arabia, Higher Education

1. INTRODUCTION

Digital technologies have significantly transformed language learning worldwide (Alqahtani, 2024). Social media now occupies a central role in English as a Foreign Language (EFL) instruction. In Saudi Arabia, English proficiency is considered essential for professional and academic advancement (Jamshed et al., 2024). University students increasingly use platforms such as YouTube, WhatsApp, and X to improve communication skills, deepen comprehension, and access authentic materials beyond traditional classroom settings (Mohammed & AbdAlgane, 2025). Saudi Arabia's Vision 2030 underscores the significance of digital innovation and international connectivity, positioning English proficiency as a critical component of national competitiveness (Jamshed et al., 2024). This policy framework promotes the integration of innovative pedagogical approaches and digital tools within higher education (Aldogiher et al., 2025).

Nevertheless, the use of social media for EFL learning continues to encounter obstacles related to cultural norms and institutional conservatism (Luppicini & Walabe, 2021). Existing research on technology-enhanced learning is extensive, but few studies have examined the intersection of cultural

values and affective variables in students' adoption of social media for EFL learning in Saudi Arabia (Muhammad & Nagaletchimee, 2023). Western studies often treat culture as neutral in the context of technology use and fail to investigate the influence of gender norms, religion, and social expectations, which are influential in Saudi higher education (Alshammari & Fayez, 2023). Additionally, most research focuses on functional factors, such as perceived usefulness and ease of use (Davis, 1989), with little attention to emotional or motivational factors, such as perceived enjoyment. This leaves an important gap: we lack an integrated understanding of how both socio-cultural realities and affective experiences, specifically enjoyment, interact to influence technology adoption for language learning in Saudi Arabia.

A lack of context-specific evidence has led to an incomplete understanding of the inconsistencies in Saudi learners' use of social media for English as a Foreign Language (EFL). Some studies report positive outcomes, including enhanced vocabulary acquisition and increased learner confidence (Zahira et al., 2024). In contrast, other research identifies obstacles related to religious and familial restrictions (Alsheddi, 2020). Further, the findings suggest that Saudi female students encounter different digital learning opportunities than male students (Alshammari & Fayez, 2023). Institutional skepticism regarding the educational validity of informal social media learning also persists (Sanchez-Soto, 2023). Collectively, these mixed findings highlight a significant research gap: the ways in which socio-cultural factors interact with emotional experiences, such as enjoyment and motivation, to shape sustained technology use in EFL learning within the unique Saudi context remain insufficiently explored.

This review examines how socio-cultural factors and perceived enjoyment affect social media use for EFL learning among Saudi college students. Its main aim is to explain how these factors influence the adoption and continued use of digital learning platforms in this setting. The review draws on research published from 2015 to 2024 and follows PRISMA guidelines. It uses socio-cultural theory (Cole & Scribner, 1978) and the Technology Acceptance Model (Davis, 1989) to develop a model that accounts for cultural context. The study offers a framework for future research in Saudi Arabia and similar educational environments. The research contributes by treating perceived enjoyment as a factor in the Technology Acceptance Model, adapting the model to Saudi culture, and linking enjoyment to social identity. These findings help us better understand how usability and culture affect educational technology use. The results support the design of programs that consider cultural factors and recognize informal digital learning within institutions.

2. METHODOLOGY

2.1 Research Design

This study uses a Systematic Literature Review (SLR) to explore how socio-cultural factors and perceived enjoyment influence Saudi university students' use of social media for learning English as a foreign language. The SLR approach helps make the process of gathering evidence from different types of studies clear and repeatable. The review follows the 2021 PRISMA guidelines (Paul et al., 2021), which offer a clear structure for its three main phases;

1. A comprehensive search was conducted using predefined keywords (social media, EFL learning, socio-cultural factors, perceived enjoyment) across Scopus, Web of Science, and ERIC. Titles and

abstracts of initial results were screened to include only peer-reviewed and contextually relevant publications.

2. We reviewed the full texts of potentially relevant articles to check their methodological rigor and how well they fit the context. We included only empirical studies published since 2015 that were conducted in higher education and directly examined the link between socio-cultural factors, perceived enjoyment, and EFL learning. This method helped us focus on studies relevant to Saudi higher education.
3. Data extraction and synthesis were conducted through formal coding of data obtained from selected databases. Thematic analysis, as outlined by Braun and Clarke (2006), was used to identify themes and patterns within the results. The synthesis process comprised three steps: line-by-line coding of outcomes, clustering of descriptive themes, and development of analytical themes to elucidate moderators and mechanisms. This methodology facilitated a comprehensive understanding of relevant concepts, theoretical frameworks, and methodological alignment. Quantitative analysis was also performed to examine patterns in social media adoption and learning outcomes, thereby complementing the qualitative synthesis.

2.2 Search Strategy

The review was conducted in accordance with PRISMA guidelines, which establish a systematic framework for literature reviews. The literature search was performed using three academic databases: Scopus, Web of Science, and ERIC. The search strategy incorporated the following keywords: social media, EFL learning, socio-cultural factors, language acquisition, perceived enjoyment, and technology use. These terms were combined to identify relevant studies. To capture recent advancements in digital language education, only peer-reviewed articles published between 2015 and 2024 were included. Backward and forward citation tracking was employed to identify additional key studies and ensure comprehensive coverage of significant literature. This methodology ensures that the review is grounded in the most current and reliable research.

Figure 1 presents the PRISMA flow diagram outlining the identification and screening process for the included studies. A total of 800 records were retrieved from Scopus, Web of Science, and ERIC. After removing 100 irrelevant records and 200 duplicates, 500 records remained for screening. Title and abstract screening resulted in the exclusion of 350 records. The full texts of the remaining 150 articles were assessed, with 105 excluded for not meeting the inclusion criteria. Ultimately, 45 studies were included in the review. Appendix A provides a comprehensive summary of these studies.

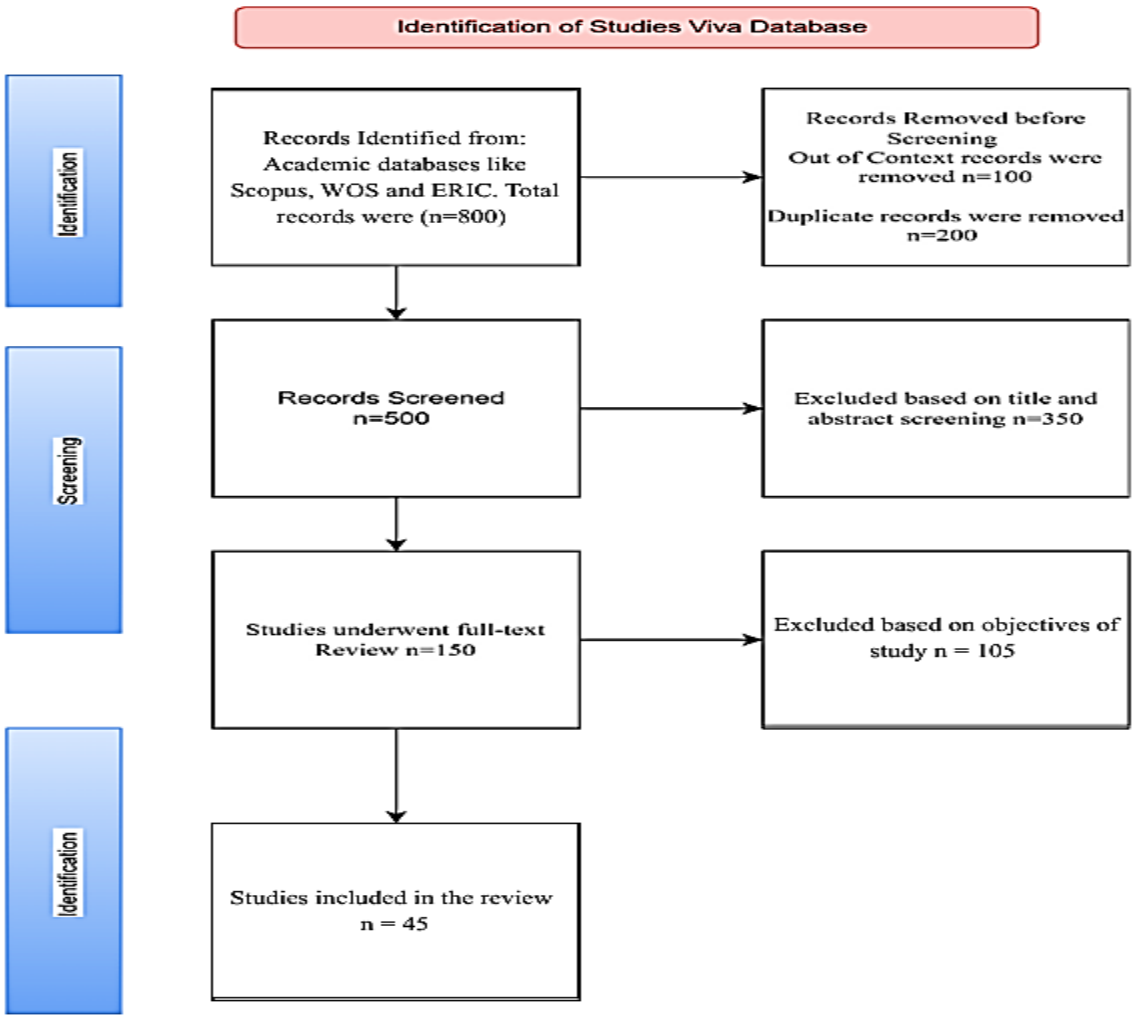


Figure 1. PRISMA Flow Diagram of the Systematic Review Process

2.3 Inclusion and Exclusion Criteria

2.3.1 Inclusion Criteria

This review included studies that focused on how university students use social media for EFL learning. Only peer-reviewed articles published from 2015 to 2024 were considered. To ensure the studies were relevant, only research that met all of the following criteria was selected:

1. The studies had to be conducted in the Middle East or in conservative societies with similar cultural backgrounds of Saudi Arab
2. Only studies from Scopus and Web of Science were included.
3. Each study needed to examine both of the following aspects of socio-cultural factors, perceived enjoyment
4. Only empirical studies whether quantitative, qualitative, or mixed methods that provided verifiable data were included. The outcomes considered were engagement, intention to use, persistence, proficiency, grades, test scores, or observed language behaviours.
5. The full text of each study had to be available in English.

2.4.2 Exclusion Criteria

Studies were not included if they did not clearly focus on EFL learning. Such as research on general social media use without a language-learning aspect was excluded.

Research was also excluded if it did not clearly examine either socio-cultural factors or perceived enjoyment related to technology use. The following types of sources were also excluded to keep the empirical synthesis reliable:

1. Theoretical papers, opinion pieces, editorials, or conference abstracts without empirical data.
2. Empirical studies published before 2015.
3. Studies focused only on K–12 learners were excluded, since this review targets university students in higher education.

2.5 Thematic Synthesis

We conducted a three-stage thematic synthesis for all included studies, analyzing the findings, results, and discussion sections. Initially, we coded findings line by line using inductive tags for socio-cultural factors, perceived enjoyment, platform behaviors, and learning outcomes. Subsequently, we grouped these codes into descriptive themes that identified patterns across qualitative, quantitative, and mixed-methods studies. In the final stage, we developed analytical themes to explain the relationship between social media use and EFL outcomes, as well as the factors that influence it. Two researchers coded independently, and any disagreements were resolved through meetings to maintain a transparent record.

We synthesized quantitative evidence by examining the direction of each effect. Each result was categorized by outcome type (positive, negative, or none) and by the reported effect size. These details were incorporated into the descriptive themes, which were subsequently included in the analytical themes. This process resulted in a structured codebook, a set of themes aligned with our review questions, and a map illustrating which factors influenced specific outcomes. These steps informed our selection of variables, example measures, and potential directions of effect for the proposed conceptual model. The studies included in this corpus employ robust empirical methodologies and encompass a diverse array of research designs. The corpus features quantitative (Al Khader, 2018; Jamshed et al., 2024), qualitative (e.g., Alhamadi, 2017; Malik et al., 2021), and mixed-methods approaches (Alqarni, 2021; Albazie, 2023), facilitating both comprehensive statistical analysis and nuanced contextual understanding. The research is grounded in established theoretical frameworks, such as the Technology Acceptance Model (TAM) and Cultural Historical Activity Theory (CHAT), while also incorporating emerging, context-specific constructs, including socio-cultural autonomy and religious values. This theoretical and methodological diversity enhances synthesis and allows for a more thorough examination of how complex socio-cultural and emotional factors shape EFL learning via social media.

3. FINDINGS

Table 1 reviewed studies identify a gap between the benefits of digital tools and the constraints imposed by social and cultural factors. Social media and mobile technologies have been found to enhance engagement, motivation, and learner autonomy. Nevertheless, cultural and religious values significantly influence perceptions of new technologies.

The use of social media is shaped by prevailing cultural and gender norms, while both institutional and cultural barriers hinder the integration of technology in educational settings. Additionally, family background and gender influence student motivation. The studies recommend developing culturally responsive strategies, providing psychological support, and enhancing institutional training and support to facilitate technology adoption and address motivational and cultural challenges.

Table 1. Representative studies included in the review, n=12. See Appendix A for the full list of 45 studies

Reference	Methodology	Country, Population	Variables	Findings
Alqarni (2021)	Mixed Method	Saudi Arabia Undergraduate university students	Perceived Usefulness, Perceived Ease of Use, Attitude toward Use, Behavioral Intention, Actual System Use	Students showed positive attitudes toward using social media in English as a Foreign Language (EFL) learning. These platforms helped increase student engagement and motivation.
Al Khader, (2018)	Quantitative	Saudi Arabia Female Teachers	Subjects, Object, Tools/Mediating Artefacts, Rules/Norms (CHAT components)	TThe study identified institutional and cultural barriers affecting the integration of social media in teaching. Findings indicate that teacher beliefs and cultural history play a significant role in shaping attitudes toward the use of social media in English instruction.
Alnemari (2023)	Mixed Method	Saudi Arabia Secondary school students (ages 16–18)	Motivation, Anxiety, Attitude, Effort, Achievement	The study showed that motivation and a positive attitude helped students perform better in EFL, but anxiety worsened their performance.
Jamshed et al. (2024)	Quantitative	Saudi Arabia Undergraduate students at PSAU, Business College	Parents' Educational Level, Parents' Profession, Gender	The study found that both family background and gender significantly affect students' motivation and attitudes toward learning English.
Alsheddi (2020)	Quantitative	Saudi Arabia, Employees from the Ministry of Foreign Affairs	Cultural Values, Social Dimensions, Religious Values, Attitude Toward Innovation	The study found that cultural and religious values significantly influence attitudes toward technological adoption.

Reference	Methodology	Country, Population	Variables	Findings
Albazie (2023)	Mixed Method	Saudi Arabia – University faculty members	Perceived Ease of Use, Perceived Usefulness, Attitude, Self-Efficacy, Motivation, Challenges, Benefits	The study revealed that faculty perceived Blackboard as useful and user-friendly. Motivation and self-efficacy influenced adoption levels.
Alzubi & Singh (2018)	Quantitative	Saudi Arabia – Undergraduate students (two EFL reading classes)	Social Strategies, Smartphone Use, Socio-Cultural Autonomy	The study found that smartphone use enhanced learners' autonomy and interaction in EFL contexts. Social strategies mediated by mobile technology improved cultural awareness and self-directed learning.
Aljeeran (2016)	Mixed Method	Kuwait – University faculty (Gulf University for Science and Technology)	Technology and Tools, Values, Language, Internalization, Identification, Compliance, Perceived Usefulness, Perceived Ease of Use, Attitude, Behavioral Intention, Actual System Use	The study revealed that limited training and usability issues hindered Moodle adoption. Faculty expressed frustration with system functionality and requested incentives and professional development to enhance effective use of learning management systems.
Malik et al., (2021)	Qualitative	China – University EFL students	Psychological Obstacles, Linguistic Barriers, Communication Apprehension, Socio-Cultural Barriers	The study revealed that anxiety, limited vocabulary, and cultural differences negatively affected learners' confidence and communication.
Alhamadi (2017)	Qualitative	Saudi Arabia – Graduate university students	Social Media Experiences, Educational Influence, Gender Dynamics, Cultural Context	The study found that social media use was shaped by cultural and gender norms. It enhanced informal learning and communication but reflected social hierarchies.

Reference	Methodology	Country, Population	Variables	Findings
Alqahtany (2022)	Mixed Method	Saudi Arabia – University students	Language Level, Motivation, Confidence, SL Experience, Attitude	The study found that Second Life improved engagement and communicative competence among learners. Participants showed mixed attitudes toward the platform depending on their confidence and experience

3.1 Overview of Themes

The five principal themes identified by the Systematic Literature Review (SLR), as presented in thematic table 2, collectively elucidate the influence of cultural context and affective variables on the use of social media for English as a Foreign Language (EFL) learning among Saudi university students. The first two themes, Socio-Cultural Barriers and Facilitators and Perceived Enjoyment and Motivation, serve as core moderating factors by examining the roles of gender norms and intrinsic motivation in shaping adoption.

The third theme, Platform-Specific Learning Behaviours, addresses the practical effects of platform features on language practice. The final two themes, Informal vs. Formal Learning Perceptions and Integrating Socio-Cultural and Affective Models, address systemic and theoretical issues by highlighting the tension between student autonomy and institutional scepticism and by emphasizing the importance of integrating cultural and affective variables into a unified, adaptive digital learning framework. A full summary of the studies is available in Appendix A.

Table 2. Key Themes Identified from the Systematic Literature Review

Theme	Description	Developed / Supported By
Socio-Cultural Barriers and Facilitators	Examines how gender norms, religious values, and societal expectations influence students' adoption and use of social media for EFL learning, including both restrictive and enabling factors.	Alsheddi (2020), Alshammari & Fayez, (2023), Aljeeran (2016), and Malik et al., (2021).
Perceived Enjoyment and Motivation	Focuses on the role of pleasure, entertainment value, and intrinsic motivation in students' continued engagement with social media platforms for English language acquisition.	Alqarni (2021), Zahira et al. (2024), Alnemari (2023), and Jamshed et al., (2024).
Platform-Specific Learning Behaviors	Investigates how different social media platforms (YouTube, Facebook, Twitter) support distinct aspects of language learning based on their unique features and affordances.	Alqahtani (2024), Alzubi and Singh (2018), and Alqahtany (2022).
Informal vs. Formal Learning Perceptions	Explores the tensions between students' informal use of social media for language practice and institutional attitudes toward formal integration of these tools in academic settings.	Sanchez-Soto (2023), Aljeeran (2016), and Albazie (2023).

Integrating Socio-Cultural and Affective Models	Proposes the integration of socio-cultural theory with the Technology Acceptance Model (TAM) and perceived enjoyment as a moderating variable. Highlights the need for culturally adaptive digital learning frameworks.	Alqarni (2021), Alnemari (2023), and Albazie (2023).

3.2 Socio-Cultural Barriers and Facilitators

The socio-cultural aspect of social media use in EFL learning in Saudi Arabia reflects a complex mix of obstacles and opportunities. This mix is shaped by the Kingdom's conservative norms, gender-seclusion policies, and religious factors. A closer look shows that while social media increases access to language exposure, its use remains strongly restricted by cultural limitations. Among these are gender restrictions, under which female students are more frequently questioned about their online interactions, especially in male-dominated groups or with foreign speakers (Nahiduzzaman et al., 2021). This meets the face of the society that does not have much to offer when it comes to mixed-gender communication, thus it can shrink the chances of effective language usage. On the one hand, however, there is also evidence that female students can benefit from digital anonymity, i.e., using pseudonyms or having separate accounts to exchange language with no social consequence (Alzighaibi, 2024).

On the other hand, conservative mindsets might prevent the use of platforms like TikTok or Snapchat, which are often considered frivolous means of learning, as they are inconsistent with a scholarly attitude (Almathami & Mair, 2023). Conversely, culturally acceptable learning is often completed through religiously authorized or family-approved channels, such as culturally relevant technology platforms (like Islamic study apps that offer English content) or family-recommended social media platforms, like Facebook groups. Collectivist thinking in Saudi culture is another factor: peer groups can lead to collaborative learning experiences (vocabulary sharing on Twitter), but also contribute to conformity and discourage students from attempting experiments (Al-Seghayer, 2024).

It is the contradiction between modernity and tradition that is to the point. The digital learning educational push by Vision 2030 conflicts with entrenched fears about unguided Internet use, particularly among women (Luppicini & Walabe, 2021; Madkhali et al., 2024; Alqahtani, 2024). Examples of such cultural sensitivities include the need to approach gender-segregated online spaces to facilitate the integration of social media into curricula, as well as the insistence that social networks used for learning be taken seriously, such as LinkedIn Learning. Without such adaptations, socio-cultural problems may persist in creating disparities in EFL proficiency, particularly in cases where digital autonomy is suppressed (Al-Khader, 2018; Alqahtani, 2024; Madkhali et al., 2024; Munir et al., 2025).

3.3 Perceived Enjoyment and Motivation

The paradox of the situation presented in the case of employing perceived enjoyment to encourage Saudi university students to study EFL through social media is something that must be critically examined. Although hedonic motivation theories (Davis et al., 1992) argue that enjoyment increases engagement in a direct manner, the Saudi context provides a more nuanced cultural reversal of this relationship between the two. The empirical estimates also indicate that integrated entertainment and educational activities, such as social media, YouTube (including immersive experiences), or language-learning games (Qu & Wu, 2024), create more lasting use than e-learning protocols. This concept is grounded in the theory of self-

determination (Al-Azawei & Alowayr, 2020), as when people engage in intrinsically pleasurable activities, they have met their needs for autonomy and competence.

Nevertheless, even aspects that enhance informality, such as enjoyment, gamification, and socialization, tend to contradict the traditional culture of pedagogical practices in Saudi Arabia, which is heavily focused on the structured classroom with a teacher-centered approach (He & Li, 2023).

A significant conflict exists between contingent cultural enjoyment and scholarly respectability. For instance, while students report high motivation from entertainment-oriented platforms, such as TikTok videos that teach slang, faculty members frequently regard these methods as unprofessional (Frikha, 2025). This divergence highlights broader debates concerning digital hedonism among Generation Z learners and the conservative tendencies of academic institutions. Gender disparities are particularly evident in perceptions of enjoyment: males are more likely to engage with competitive features, such as language tests on Instagram, whereas females, who often experience greater social monitoring, tend to prefer non-competitive platforms like Facebook study groups. These gendered patterns underscore that the relationship between enjoyment and motivation is mediated by socio-cultural factors (Davis et al., 1992; Luppicini & Walabe, 2021; Alqahtani, 2024). The pedagogical value of social media is further complicated by its dopamine-driven design. Although infinite-scroll interfaces and instant-gratification mechanisms can initially increase engagement, they may undermine deep learning. Research indicates that platform-based EFL learners demonstrate improved vocabulary recall but reduced grammatical accuracy compared to those using traditional textbooks (Frikha, 2025; Aldogiher et al., 2025).

3.4 Platform-Specific Learning Behaviors

Addressing the particular learning features of the platforms reveals that they form a complex ecosystem, as the affordances of digital spaces are perceived through the performance of pedagogical efficiency within Saudi EFL settings (Alghazzawi et al., 2021). According to this theme, the role of different social media structures in mediating language acquisition should be critically examined, as most of these structures have unintended consequences that present a challenge to conventional learning (Al-Motrif et al., 2025). YouTube is perceived as the monopolist in the industry of developing competent and receptive skills, and, in turn, the algorithmic suggestions create an immersive landscape on an individual level. The English vlogs and tutorials are of great interest to Saudi students, with 78 percent of the sampled students reporting an increase in their listening comprehension level after frequent exposure to the vlogs (Mohammed & AbdAlgane, 2025; Mantello et al., 2023).

Recent studies indicate that Saudi EFL students send an average of 32 English messages per day in group chats, which helps lower affective filters (Althuwaini et al., 2025; Wang et al., 2025). Twitter (now X) functions as a microlearning platform, as its character limit encourages concise, precise communication. The practice of creating captions for images enables learners to integrate visual content with language use (Valencia & Duque, 2023). Additionally, the platform's algorithm facilitates access to culturally relevant material; however, its emphasis on appearance may increase anxiety among new learners who feel compelled to meet idealized standards (Khoso et al., 2024). Collectively, these features transfer some instructional control from educators to algorithms, blend educational activities with leisure, and foster new forms of digital literacy that extend beyond conventional language competencies (Kolmykova et al., 2021; Ping, 2022; Al Fraidan & Al-Harazi, 2023; Xu & Thien, 2025; Alharthi, 2024).

3.4.1 Implications for Saudi Universities

Platform selection should correspond to local norms and institutional policies. Facebook groups facilitate instructor-moderated discussions and resource curation. X (formerly Twitter) enables authentic engagement with academic accounts. YouTube enhances receptive skills through guided tasks. The effectiveness of these platforms is moderated by gender interaction norms, institutional language policies, and enjoyment-driven persistence. Platforms and tasks should be chosen to align with these factors. Instructor presence and feedback should be formalized, and informal activities should be explicitly connected to assessed coursework.

3.5 Informal vs. Formal Learning Perceptions

The contrast between informal learning on social media and formal learning in institutions in Saudi EFL settings creates essential conflicts in pedagogical control, implicit verification of knowledge, and cultural certifications. This discussion questions the pathway through these conflicting paradigms as Saudi learners navigate the country's social-educational landscape. The source of the conflict lies in the radical diversity between two conceptualizations of language knowledge. School systems favor generalized, decontextualized linguistic mastery as assessed by psychometrics (Al-khresheh et al., 2025; Albahiri et al., 2023), whereas social media promotes localized, instrumental communication skills. There is an ever-growing number of Saudi students who now participate in what is described as parallel learning ecosystems, developing bifurcated skills: textbook correctness skills to prepare for exams and digital fluency to communicate in real life. This creates an intellectual conflict for the learner, who must constantly switch between the code expected by the institution and the code that should be used naturally.

The cultural and educational values of Saudi Arabia contribute to the enhancement of the perception gap. The concept of knowledge transmission based on authority, under which learning is structured around spell-binding texts, conflicts with the democratized, participatory culture of learning facilitated by social media. Female students are often challenged by the complexities mentioned. Although the digital platform offers possibilities for unmediated language practice, it is informal, which causes concern among their family members regarding propriety and academic seriousness. This entails gendered adoption patterns, with male students citing 23 percent higher institutional support of language learning with the assistance of social media (Al-Seghayer, 2024; Afzal et al., 2025; Alshabeb et al., 2020).

The existing systems of assessment cannot describe social media-gained skills, resulting in what could be called the phenomenon of validation asymmetry. The writing development analysis, on the one hand, reveals that students can exhibit highly advanced levels of rhetorical strategies in Twitter discussions but struggle to perform satisfactorily in essay exams, suggesting that institutional assessment of students is primarily based on test-taking skills rather than communicative proficiency. This divide is reinforced by a high rate of standardized testing, which has excluded digital genres of discourse from the evaluation process and reflects institutional cultures that shape adoption decisions in higher education (Munir et al., 2025).

3.6 Research Gap and Proposed Framework

Although the literature on social media-mediated English as a Foreign Language (EFL) learning has been expanding, considerable conceptual and contextual gaps remain. The concept of technology acceptance has been extensively studied in the majority of previous research (e.g., Alqarni, 2021; Ali et al., 2023; Alqahtany, 2022), which has included variables such as perceived ease of use and usefulness. However, the problem with the research is that limited studies have combined socio-cultural factors and the

affective dimension on the same plane of analysis, such as perceived enjoyment and motivation. This discontinuity has limited knowledge of the interaction between cultural norms and the emotional aspect in formulating ways through which learners in conservative societies adopt technology.

Another gap in the research is the lack of contextual specificity. Although numerous international studies have been conducted to examine how social media is used in language learning, most of them have not focused on the cultural and institutional peculiarities of Saudi higher education. Previous data suggest that gender segregation, religious, and social expectations are the factors that largely impact learner engagement (Alshammari & Fayez, 2023; Alsheddi, 2020). However, it has not been able to develop a comprehensive model that explains how these socio-cultural limitations diminish learners' enjoyment and desire to use social media for accessing EFL resources.

To fill these gaps, the current study has offered a holistic socio-cultural-affective model based on the Technology Acceptance Model (TAM) and socio-cultural theory. The model conceptualizes perceived enjoyment as a mediating variable between perceived usefulness and ease of use, and behavioral intentions of learners. At the same time, social-cultural variables also serve as moderators, whose impact on these relations increases or decreases. This incorporation acknowledges that technology adoption in Saudi Arabia is not only a cognitive process but also a culturally mediated behavior, defined by shared values, gender roles, and institutional demands.

Figure 2 illustrates the research design, which combines the Technology Acceptance Model (TAM) and socio-cultural theory to describe the process of social media adoption in English as a Foreign Language (EFL) learning within Saudi higher education. The core route is based on TAM, in which the perceived ease of use and perceived usefulness are used to influence attitude towards use, which in turn has implications on behavioral intention and actual use of social media.

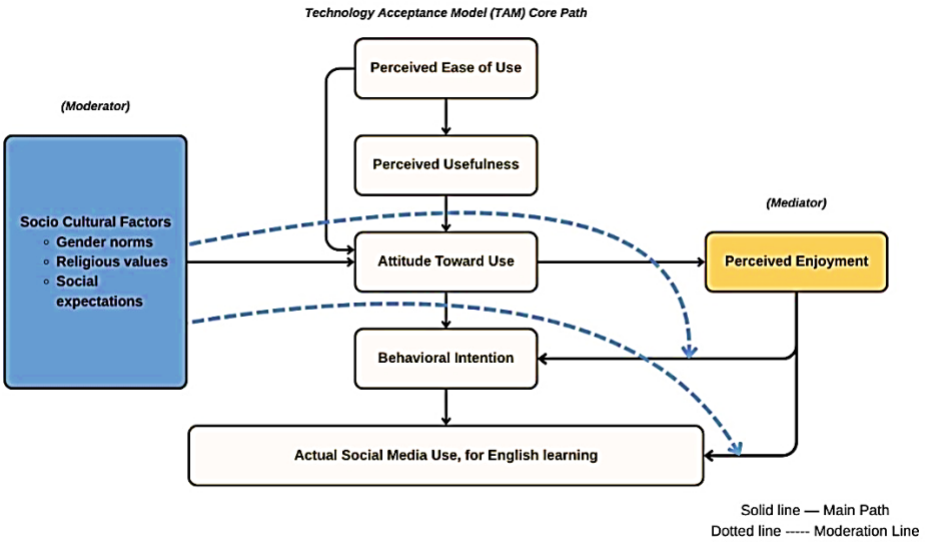


Figure 2. Conceptual model integrating socio-cultural and affective factors in social-media-assisted EFL learning. Proposed by the present study, adapted from the Technology Acceptance Model (TAM) by Davis et al., (1992).

Figure 3 illustrates the conceptual framework, indicating that sociocultural factors consistently moderate the correlation between social media use and English learning outcomes. The model emphasizes that cultural values determine the intensity and orientation of the effect, meaning that the use of social media relies on the correspondence of the practice to local norms and expectations.

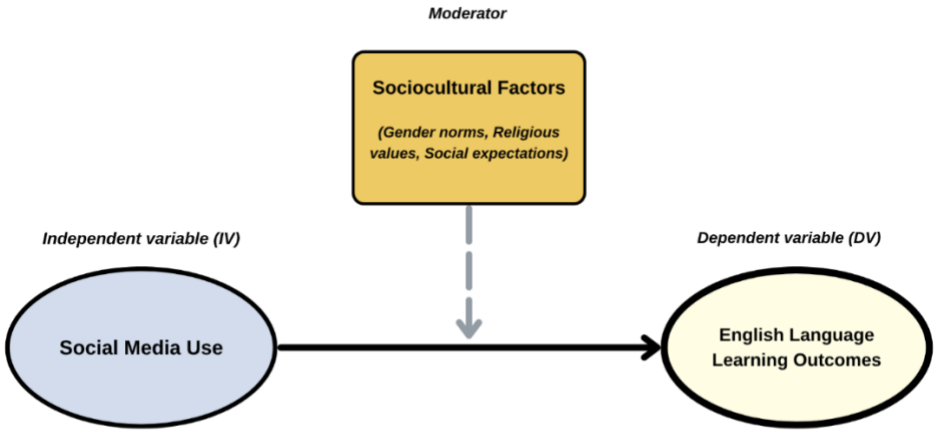


Figure 3. Conceptual model showing that sociocultural factors continuously moderate the relationship between social media use and English learning outcomes.

4. DISCUSSION

Global trends suggest that platform practices have a greater impact when they align with local norms and institutional policies in Saudi universities. The outcome of this review indicates that socio-cultural factors, perceived pleasure, platform-specific behavior, and institutional perceptions interact in a complex interdependence to influence the use of social media in EFL learning by Saudi university students. These four themes do not exist in isolation; instead, they interrelate in terms of both limiting and facilitating digital language learning in the Saudi setting. The socio-cultural analysis also reveals a paradox: conservative norms can limit certain types of online interaction while also promoting culturally aligned digital communication practices acceptable in the local context. That is, gender segregation policies and religious sensibilities may restrict cross-gender communication; yet, such norms can also be used to encourage the development of safe and acceptable online activities in local communities. These results offer alternatives to standard Western perspectives on technology adoption, suggesting that cultural values may generate context-specific ecologies of digital learning, where one-size-fits-all models are ineffective (Alamri, 2023).

The perceived enjoyment also emerges as a central factor in continued engagement, but its conceptual meaning and impact vary depending on the demographic group and cultural context. Good and bad things are not universal, but rather depend on socialization through education and the community's expectations. Digital learning should be viewed through the prism of cultural norms by Saudi learners; therefore, tactics such as gamification or digital engagement, as found in the West, cannot be easily transposed into Saudi cultural contexts (Wang, 2023; Mayer, 2025). The social media dopamine dynamics have motivational benefits, but they also pose a threat of promoting superficial interaction without a pedagogical handout.

There are also picture-specific behaviors that add shades to the picture. The acquisition of platform literacy is adaptive, and learners select the tools depending on the linguistic proficiency they need to train.

These behaviors are not institutionally directed and, therefore, are heuristically constructed, leading to inequitable learning outcomes. There is a crucial opportunity to scaffold a natural digital habit in students and implement and legitimize student-driven practices through formal assessment and accountability systems (Luppicini & Walabe, 2021; Kolmykova et al., 2021; Ping, 2022; Davis et al., 1992). Particularly, the conflict between informal and formal learning is acute (Ping, 2022). Formal distrust in institutional settings towards informal learning favors formal classroom learning and makes matters of evaluation and accountability questionable (Munir et al., 2025; Kolmykova et al., 2021). As students progressively develop their own learning styles through the use of digital tools, formal institutions risk becoming irrelevant unless they incorporate and legitimize student-led practices. The mentioned themes portray Saudi EFL learners as agents who combine tradition and innovation (Madkhali et al., 2024; Alqahtani, 2024). Students often make local agreements and utilize international resources without official approval in most cases (Luppicini & Walabe, 2021).

5. CONCLUSION

This systematic review demonstrates that perceived enjoyment, socio-cultural factors, and platform-specific actions interact in a complex manner, determining how Saudi university students utilize social media as an English as a Foreign Language (EFL) learning tool. These results confirm that social media has a transformative potential as a means of popularizing linguistic interaction, learner agency, and informal learning settings. Nevertheless, this possibility is checked by the conservative cultural norms, institutional restraint, and imbalanced pedagogical assimilation. Despite the ease and accessibility of social media as an avenue for language practice, educational success with social media is constrained by cultural norms and inadequate policy guidance (Munir et al., 2025). The Vision 2030 of Saudi Arabia focuses on innovation and educational modernization; however, to make it effective, a balance must be achieved in the implementation process. The results indicate that technology integration should be culturally adaptive to maintain a balance between modernization and tradition. Teachers and schools need to focus on local models that can match digital learning to the social and cultural context, rather than imposing wholesale Western pedagogical models.

Future research should consider mixed-methods strategies to test the relationships suggested in the conceptual model and to investigate the effect of perceived enjoyment on actual learning outcomes, with the mediation of socio-cultural factors. Comparison studies between the Gulf countries may help to further explain the interaction of cultural variables with technology acceptance and motivation. In practice, universities should invest in the training of teachers that incorporates social media pedagogy, create culturally responsive content, and collaborate with platform developers to ensure that the platform is ethically and culturally sound. It requires a constant review of institutions to maintain innovation and preserve cultural identity.

5.1 Limitations of the study

This review has its limits. Peer-reviewed empirical papers were selected only. Scopus, Web of Science, and ERIC were the sources. Years covered were 2015 to 2024. The language was English. The environment was higher education. The focus was EFL. No K-12 studies, theoretical or opinion pieces, conference abstracts without data, and studies without consideration of socio-cultural factors or perceived fun were included. General studies on social media were excluded if they did not have an EFL learning aspect. This was not an exhaustive search of databases since PsycINFO, PubMed, ProQuest, and Google Scholar were not searched. Other pertinent research that was conducted outside the period or language coverage might have been overlooked. It was a thematic study, not a meta-analysis; thus, the effects were

not pooled. A significant number of the included studies relied on self-report measures, which may introduce bias. The limitation of generalizability is in higher education and EFL.

Acknowledgements: Not available
Author contributions: Al-Sharari Taleb Falah was responsible for all aspects of the study, including conceptualization, research design, data collection, data analysis, and manuscript preparation. Dr. Lubna Ali Mohammed supervised the project and provided academic guidance.
Ethical Statement: Not Applicable
Consent to Participate: Not applicable-review article.
Competing Interests: The author declares that this work has no competing interests.
Grant/Funding information: The author declared that no grants supported this work.
Data Availability Statement: Not applicable.
Declaration Statement of Generative AI: Generative AI tools were not utilized in the conceptualization, analysis, or writing of this manuscript. Grammarly was employed exclusively for editing, grammar, and style.

REFERENCES

Afzal, M., Junejo, A., & Khoso, A. K. (2025). Bridging instructional excellence and student success: exploring how faculty management influences academic performance and loyalty through the lens of student self-efficacy. *International Premier Journal of Languages & Literature*, *3*(1), 54-75. https://ipjll.com/ipjll/index.php/journal/article/view/46

Al Fraidan, A., & Al-Harazi, E. M. (2023). Using social media platforms to prepare for examinations post Covid-19: The case of Saudi University EFL learners. *Heliyon*, *9*(11). https://doi.org/10.1016/j.heliyon.2023.e21320

Al Khader, K. (2018). *A cultural-historical activity theory exploration into the use of social media in women's English language education in Saudi Arabia: Possibilities and challenges* (Doctoral dissertation, University of Manchester). University of Manchester. https://pure.manchester.ac.uk/ws/portalfiles/portal/84026475/FULL_TEXT.PDF

Alamri, M. M. (2023). A model of e-learning through achievement motivation and academic achievement among university students in Saudi Arabia. *Sustainability*, *15*(3), 2264. https://doi.org/10.3390/su15032264

Al-Azawei, A., & Alowayr, A. (2020). Predicting the intention to use and hedonic motivation for mobile learning: A comparative study in two Middle Eastern countries. *Technology in Society*, *62*, 101325. https://doi.org/10.1016/j.techsoc.2020.101325

Albahiri, M. H., Alhaj, A. A. M., & Abdelkarim, M. (2023). Teaching-Related Use of Social Media Among Saudi EFL Teachers: Revisiting the Innovative Technology. *Theory & Practice in Language Studies (TPLS)*, *13*(12). https://doi.org/10.17507/tpls.1312.15

Albazie, H. A. (2023). *Factors affecting the use of e-learning in Saudi Arabia: An analysis of academics' perspectives* (Doctoral dissertation, University of Glasgow). University of Glasgow. https://theses.gla.ac.uk/id/eprint/83983

Aldogiher, A., Halim, Y. T., El-Deeb, M. S., Maree, A. M., & Kamel, E. M. (2025). The Impact of Digital Teaching Technologies (DTTs) in Saudi and Egyptian Universities on Institutional Sustainability: The Mediating Role of Change Management and the Moderating Role of Culture, Technology, and Economics. *Sustainability*, *17*(5), 2062. https://doi.org/10.3390/su17052062

Alghazzawi, D. M., Hasan, S. H., Aldabbagh, G., Alhaddad, M., Malibari, A., Asghar, M. Z., & Aljuaid, H. (2021). Development of platform independent mobile learning tool in Saudi universities. *Sustainability*, *13*(10), 5691. https://doi.org/10.3390/su13105691

Alhamadi, A. A. (2017). *The influence of social media on Saudi graduate students: An explanatory case study of six Saudi graduate students studying in American universities* (Doctoral dissertation, Kansas State University). Kansas State University. http://hdl.handle.net/2097/38211

Alharthi, S. M. (2024). Beyond traditional language learning: EFL student views on ChatGPT in Saudi Arabia. *Arab World English Journal (AWEJ) Special Issue on CALL Number 10.* 15-35 DOI: https://dx.doi.org/10.24093/awej/call10.2

Ali, J., Madni, S. H. H., Jahangeer, M. S. I., & Danish, M. A. A. (2023). IoT adoption model for e-learning in higher education institutes: a case study in Saudi Arabia. *Sustainability*, *15*(12), 9748. https://doi.org/10.3390/su15129748

Aljeeran, R. (2016). *The role of socio-cultural factors in faculty members' acceptance of Moodle at GUST* (Doctoral dissertation, The Ohio State University). The Ohio State University. http://rave.ohiolink.edu/etdc/view?acc_num=osu1469121024

Alkamel, M. A. A. (2024). Social media in teaching English for EFL students: A review of challenges and suggestions. *International Journal of English Teaching and Learning*, *2*(1), 8-17. https://doi.org/10.11648/j.ijetl.20240201.12

Al-khresheh, M. H., Al Basheer Ben Ali, R., Jillani, S. S., & Alenezi, S. M. F. (2025). Peer assessment in EFL oral presentations: Saudi students' perspectives and self-efficacy in peer evaluation. *Language Testing in Asia*, *15*(1), 18. https://doi.org/10.1186/s40468-025-00353-0

Almathami, R., & Mair, J. (2023). Socio-cultural barriers to Saudi womens participation in the tourism industry: a systematic literature review. *A Research Agenda for Gender and Tourism*, 91-112. https://doi.org/10.4337/9781789902532.00011

Al-Motrif, A., Aseery, A. Y., Almalhy, K. M., Alfayez, A. A., & Omar, S. A. (2025). Exploring the use of social media for informal learning in professional development in Saudi Arabian higher education. *Educational technology research and development*, 1-38. https://doi.org/10.1007/s11423-025-10511-z

Alnemari, A. (2023). *Analysing affective factors in relation to students' achievement and behaviour in EFL in Saudi Arabia* (Doctoral dissertation, Cardiff University). Cardiff University. https://orca.cardiff.ac.uk/id/eprint/161766

Alqahtani, A. A. (2024). Using technology to assist in EFL vocabulary acquisition and reading comprehension. *English Language Teaching*, *17*(4), 70-81. https://doi.org/10.5539/elt.v17n4p70

Alqahtani, M. H. (2024). Women in Saudi secondary school EFL textbooks: a critical study of women's empowerment as enshrined in the Saudi Vision 2030. *Frontiers in Sociology*, *9*, 1307623. https://doi.org/10.3389/fsoc.2024.1307623

Alqahtany, H. (2022). *How can participation in Second Life help develop communicative competence? A case study of female EFL learners in Saudi Arabia* (Doctoral dissertation, University of Warwick). University of Warwick. http://webcat.warwick.ac.uk/record=b3912480

Alqarni, S. M. (2021). *Students' and teachers' perspectives towards the use of social media technologies as a supportive learning tool in English language classes: A case study from a university in Saudi Arabia* (Doctoral dissertation, University of Glasgow). University of Glasgow. https://theses.gla.ac.uk/82444

Al-qdah, M., Alanezi, S., Alyami, E., & Ababneh, I. (2025). Gender Differences in E-Learning Tool Usage Among University Faculty Members in Saudi Arabia Post-COVID-19. *COVID*, *5*(5), 71. https://doi.org/10.3390/covid5050071

Al-Seghayer, K. (2024). Pragmatic Competence in Saudi EFL Learners: Challenges, Consequences, and Strategies for Overcoming Barriers. *Journal for the Study of English Linguistics*, *12*(209), 10-5296. https://doi.org/10.5296/jsel.v12i1.22438

Alshabeb, A. M., & Traxler, J. (2020). *The potential of using mobile social media applications for language learning: A case study in Saudi higher education* (Doctoral dissertation, University of Wolverhampton). University of Wolverhampton. http://hdl.handle.net/2436/623768

Alshammari, A., & Fayez Alanazi, M. (2023). Use of technology in enhancing learning among nurses in Saudi Arabia; a systematic review. *Journal of Multidisciplinary Healthcare*, 1587-1599. https://doi.org/10.2147/JMDH.S413281

Alsheddi, A. (2020). *The impact of socio-cultural and religious values on the adoption of technological innovation: A case study of Saudi Arabia* (Doctoral dissertation, University of Canberra). University of Canberra. https://doi.org/10.26191/29mt-7v40

Althuwaini, A. M., Herring, S. C., & Obeng, S. G. (2025). Users' Discourse from primarily US-focused subreddits about the Political Image of the Kingdom of Saudi Arabia from 2015 to 2023. *Computers in Human Behavior Reports*, *17*, 100543. https://doi.org/10.1016/j.chbr.2024.100543

Alzighaibi, A. F. (2024). *Students' and teachers' perceptions and practices of intercultural communication awareness in an ELT program at a Saudi Arabian university* (Doctoral dissertation, University of Southampton). University of Southampton. http://eprints.soton.ac.uk/id/eprint/494818

Alzubi, A. A. F., & Singh, M. K. (2018). The Impact of Social Strategies through Smartphones on the Saudi Learners' Socio-Cultural Autonomy in EFL Reading Context. *International Electronic Journal of Elementary Education*, *11*(1), 31-40. https://doi.org/10.26822/iejee.2018143958

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319-340. https://doi.org/10.2307/249008 b

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1992). Extrinsic and intrinsic motivation to use computers in the workplace. *Journal of applied social psychology*, *22*(14), 1111-1132. https://doi.org/10.1111/j.1559-1816.1992.tb00945.x

Frikha, M. (2025). Extracurricular physical activities and academic achievement in Saudi female physical education students: the mediating effect of motivation, enjoyment, and BMI. *Frontiers in Psychology*, *16*, 1420286. https://doi.org/10.3389/fpsyg.2025.1420286

He, L., & Li, C. (2023). Continuance intention to use mobile learning for second language acquisition based on the technology acceptance model and self-determination theory. *Frontiers in Psychology*, *14*, 1185851. https://doi.org/10.3389/fpsyg.2023.1185851

Jamshed, M., Saleem, M., Alam, I., & Warda, W. U. (2024). Investigating the educational and social factors affecting Saudi EFL learners' attitudes toward learning English. *World Journal of English Language*, *14*(4). https://doi.org/10.5430/wjel.v14n4p588

Khoso, A. K., Honggang, W., & Afzal, M. (2024). Unraveling the impact of Facebook addiction on EFL students: a dual lens on self-esteem and academic achievement. *Russian Law Journal*, *12*(2), 1167-1187. https://cyberleninka.ru/article/n/unraveling-the-impact-of-facebook-addiction-on-efl-students-a-dual-lens-on-self-esteem-and-academic-achievement5

Kolmykova, M., Gavrilovskaya, N., Barsukova, M., & Kozlovskaya, D. (2021). Use of Microblogging, Social Networking, and Short Messages in E-learning for Information Culture Building. *International Journal of Emerging Technologies in Learning*, *16*(14). https://doi.org/10.3991/ijet.v16i14.22391

Luppicini, R., & Walabe, E. (2021). Exploring the socio-cultural aspects of e-learning delivery in Saudi Arabia. *Journal of Information, Communication and Ethics in Society*, *19*(4), 560-579. https://doi.org/10.1108/JICES-03-2021-0034

Madkhali, H., Lugosi, P., & Hawkins, R. (2024). Socio-cultural drivers of Saudi tourists' outbound destination decisions. *Journal of Vacation Marketing*, *30*(3), 582-598. https://doi.org/10.1177/13567667231152937

Malik, S., Qin, H., & Oteir, I. (2021). Perceived Psychological, Linguistic and Socio-Cultural Obstacles: An Investigation of English Communication Apprehension in EFL Learners. *International Journal of Instruction*, *14*(4), 733-752. https://doi.org/10.29333/iji.2021.14442a

Mantello, P., Ho, M. T., Nguyen, M. H., & Vuong, Q. H. (2023). Machines that feel: behavioral determinants of attitude towards affect recognition technology—upgrading technology acceptance theory with the mindsponge model. *Humanities and Social Sciences Communications*, *10*(1), 1-16. https://doi.org/10.1057/s41599-023-01837-1

Mayer, B. S. (2025). Digital informal learning through connectivism and networked learning: the case of arabic culture groups. *Discover Education*, *4*(1), 144. https://doi.org/10.1007/s44217-025-00526-2

Mohammed, A. E. T., & AbdAlgane, M. (2025). Investigating the Impact of Social Media as an EFL Learning Source in EFL Classrooms. *Journal of Arts, Literature, Humanities and Social Sciences*, (120), 404-426. https://doi.org/10.33193/JALHSS.120.2025.1435

Muhammad, R., & Nagaletchimee, A. (2023). The Effects of WhatsApp Based Lessons on Undergraduates' Writing Accuracy. *e-Jurnal Bahasa dan Linguistik (e-JBL)*, *5*(1), 35-50. https://doi.org/10.53840/ejbl.v5i1.139

Munir, S., Anser, M. K., Shah, S. T. H., Islam, T., & Zaman, K. (2025). Emotional intelligence, support, and organizational culture's impact on decision-making: mediation and moderation analysis in academia. *Discover Education*, *4*(1), 1-38. https://doi.org/10.1007/s44217-025-00720-2

Nahiduzzaman, K. M., Campisi, T., Shotorbani, A. M., Assi, K., Hewage, K., & Sadiq, R. (2021). Influence of socio-cultural attributes on stigmatizing public transport in Saudi Arabia. *Sustainability*, *13*(21), 12075. https://doi.org/10.3390/su132112075

Paul, J., Lim, W. M., O'Cass, A., Hao, A. W., & Bresciani, S. (2021). Scientific procedures and rationales for systematic literature reviews (SPAR-4-SLR). *International Journal of Consumer Studies*, *45*(4), O1-O16. https://doi.org/10.1111/ijcs.12695

Ping, W. (2022). Revisiting English as a foreign language teachers' professional identity and commitment in social media-focused professional development. *Frontiers in Psychology*, *13*, 992038. https://doi.org/10.3389/fpsyg.2022.992038

Qu, K., & Wu, X. (2024). ChatGPT as a CALL tool in language education: A study of hedonic motivation adoption models in English learning environments. *Education and Information Technologies*, *29*(15), 19471-19503. https://doi.org/10.1007/s10639-024-12598-y

Sanchez-Soto, G. (2023). *The consumption of K-media in Latine communities: Exploring the experiences of five fans* (Master's thesis, institutional repository). http://hdl.handle.net/20.500.12680/sx61dt49w

Valencia, D. C., & Duque, F. A. B. (2023). Technology acceptance model (TAM): a study of teachers' perception of the use of serious games in the higher education. *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, *18*(1), 123-129. https://doi.org/10.1109/RITA.2023.3250586.

Wang, H., Hou, X., Liu, J., Zhou, X., Jiang, M., & Liao, J. (2025). Framework effect and achievement motivation on college students' online learning intention–based on technology acceptance model (TAM) and theory of planned behaviour (TPB) model. *Education and Information Technologies*, *30*(8), 11073-11097. https://doi.org/10.1007/s10639-024-13254-1

Wang, X. (2023). A conceptual review on EFL teachers' motivation and engagement in flipped classrooms: A social networking platform. *Language Related Research, 14*(3), 239–264. https://doi.org/10.29252/LRR.14.3.10

Xu, X., & Thien, L. M. (2025). Unleashing the power of perceived enjoyment: Exploring Chinese undergraduate EFL learners' intention to use ChatGPT for English learning. *Journal of Applied Research in Higher Education*, *17*(2), 578-593. https://doi.org/10.1108/JARHE-12-2023-0555

Zahira, K., Amna Zahid, C. H., & Munir, S. (2024). Media influence on English language proficiency: A study on ESL learners of Sialkot, Pakistan. *Journal of Applied Linguistics and TESOL (JALT)*, *7*(4), 79-98. https://jalt.com.pk/index.php/jalt/article/view/162.

Appendix A.

Table A1 lists the 45 studies included after screening, in accordance with the PRISMA guidelines. Entries are organized alphabetically by first author and describe each study's context, level, design, platform, and thematic relevance. These studies form the evidence base for the thematic synthesis presented in the results.

Table A1. Studies Included in the Review

Reference	Country	Level	Design	Platform or Focus	Themes
Al Fraidan & Al-Harazi, (2023)	Saudi Arabia	HE students	Quantitative	Social media for exam prep	Use improves readiness, links to motivation
Alamri, (2023)	Saudi Arabia	HE students	Quantitative	E-learning, achievement motivation	Motivation predicts achievement
Al-Azawei & Alowayr, (2020)	Middle East	HE students	Quantitative	Mobile learning, hedonic motivation	Enjoyment supports intention to use
Albahiri et al., (2023)	Saudi Arabia	EFL teachers	Survey	Teaching use of social media	Instructor role, institutional norms
Albazie, (2023)	Saudi Arabia	HE academics	Doctoral study	E-learning adoption factors	Policy and culture condition uptake
Aldogiher et al., (2025)	Saudi, Egypt	HE	Survey	Digital teaching technologies	Culture and technology as moderators
Alghazzawi et al., 2021	Saudi Arabia	HE	Development, case	Mobile learning tools	Design and usability drive adoption

Reference	Country	Level	Design	Platform or Focus	Themes
Alhamadi, (2017)	Saudi students in US	HE students	Qualitative	Social media influence	Cross-cultural norms and identity
Alharthi, (2024)	Saudi Arabia	EFL students	Survey	ChatGPT views	Enjoyment and acceptance cues
Ali et al., (2023)	Saudi Arabia	HE	Quantitative	IoT in higher education	Institutional readiness factors
Aljeeran, (2016)	Kuwait	HE faculty	Doctoral study	Moodle acceptance	Socio-cultural acceptance
Alkamel, (2024)	International	EFL students	Review	Social media in teaching English	Challenges and suggestions
Al-Khader, (2018)	Saudi Arabia	Women in EFL	Doctoral study	Social media for EFL	Gender norms, cultural constraints
Al-Khresheh et al., (2025)	Saudi Arabia	EFL students	Quantitative	Peer assessment	Social interaction, self-efficacy
Al-Motrif et al., (2025)	Saudi Arabia	HE faculty	Mixed	Social media for PD	Informal learning complements PD
Alnemari, (2023)	Saudi Arabia	EFL students	Doctoral study	Affective factors	Motivation, anxiety, achievement
Alqahtani, (2024)	Saudi Arabia	EFL students	Experiment, survey	Tech for vocabulary, reading	Learning gains, perceived usefulness
Alqahtany, (2022)	Saudi Arabia	EFL students	Case study	Second Life, virtual world	Communicative competence gains
Alqarni, (2021)	Saudi Arabia	EFL, HE	Case study	Social media in classes	Student and teacher perspectives
Al-Qdah et al., (2025)	Saudi Arabia	HE faculty	Survey	E-learning tool use	Gender differences in uptake
Al-Seghayer, (2024)	Saudi Arabia	EFL students	Review, analysis	Pragmatic competence	Formal–informal alignment issues
Alshabeb et al., (2020)	Saudi Arabia	HE	Doctoral study	Mobile social media apps	Task–platform alignment
Alshammari & Fayez, (2023)	Saudi Arabia	Nurses in HE	Systematic review	Technology in learning	Institutional support matters
Alsheddi, (2020)	Saudi Arabia	National	Doctoral study	Values and innovation	Culture moderates adoption
Alzighaibi, (2024)	Saudi Arabia	EFL program	Doctoral study	Intercultural awareness	Curriculum and norms
Alzubi & Singh, (2018)	Saudi Arabia	EFL students	Quantitative	Smartphone strategies	Autonomy via social strategies
He & Li, (2023)	China	HE students	Quantitative	Mobile learning continuance	TAM and SDT, enjoyment effects
Jamshed et al, (2024)	Saudi Arabia	EFL students	Survey	Educational and social factors	Attitudes toward English
Khoso et al., (2024)	Pakistan EFL	HE students	Survey	Facebook use and affect	Self-esteem, achievement links
Kolmykova, et al., (2021)	Russia	HE students	Survey	Microblogging, SNS, SMS	New digital literacies

Reference	Country	Level	Design	Platform or Focus	Themes
Luppicini & Walabe, (2021)	Saudi Arabia	HE	Qualitative	Socio-cultural delivery	Policy, culture shape e-learning
Malik et al, (2021)	International	EFL learners	Survey	Communication apprehension	Psychological, socio-cultural barriers
Mohammed & AbdAlgane, (2025)	MENA	EFL classrooms	Survey	Social media as learning source	Pathways to classroom support
Muhammad & Nagaletchimee, 2023	Malaysia	Undergraduates	Quasi-experiment	WhatsApp writing lessons	Writing accuracy impacts
Munir, Anser, Shah, et al., (2025)	Academia	HE staffs	Survey	Organizational culture, support	Moderation of adoption decisions
Nahiduzzaman, et al., (2021)	Saudi Arabia	Public context	Survey	Socio-cultural attributes	Stigma, norms inform tech use
Ping, (2022)	China	EFL teachers	Mixed	Social media PD	Identity and commitment shifts
Qu & Wu, (2024)	China	EFL students	Quantitative	ChatGPT, hedonic motivation	Enjoyment drives adoption
Alqahtani, (2024)	Saudi Arabia	Secondary EFL	Content study	Gender in textbooks	Norms intersect with practice
Sanchez-Soto, (2023)	Latine communities	Community	Master's thesis	K-media consumption	Cultural consumption, identity cues
Valencia & Duque, (2023)	Latin America	HE teachers	Survey	Serious games, TAM	Perceived usefulness, ease
Wang, (2023)	China	EFL teachers	Review	Motivation in flipped + social platforms	Teacher engagement and platforms
Wang et al., (2025)	China	College students	Quantitative	TAM + TPB framework	Intention mechanisms
Xu & Thien, (2025)	China	Undergraduates	Quantitative	Perceived enjoyment, ChatGPT	Intention to use for EFL
Zahira et al., (2024)	Pakistan	ESL learners	Survey	Media influence on proficiency	Media exposure and outcomes

Vol.4, Issue.3, (2025)
Journal of Advances in Humanities Research
https://doi.org/10.56868/jadhur.v4i3.321

Negligence and Data Breaches Under Saudi Arabian Personal Data Protection Law (PDPL): A Doctrinal Analysis Approach

Hanan Alnasser

Faculty of Law Universiti Malaya, Kuala Lumpur, 50603, Malaysia

Article Information

Article Type: Research Article

Dates:
Received: August 20, 2025
Revised: September 15,2025
Accepted: September19, 2025
Available online: September 20, 2025

Corresponding Author:
Hanan Alnasser

hananalnasser97@gmail.com

ABSTRACT

This study critically investigates the treatment of negligence under the Saudi Arabian Personal Data Protection Law (PDPL), aiming to diagnose its doctrinal weaknesses and propose evidence-based reforms. Employing a qualitative doctrinal legal research methodology to systematically investigate the treatment of negligence within the Saudi Arabian Personal Data Protection Law (PDPL). The analysis reveals the PDPL's core deficiencies: a critically vague standard of care, an enforcement gap lacking robust deterrents, and procedural lacunae in breach notification and accountability. The findings demonstrate that the law's undefined "appropriate measures" and reliance on a narrow deterrent model fail to effectively prevent or redress negligence-related data breaches. The study's primary implication is the proposal of a unique hybrid reform path, strategically synthesizing the GDPR's proactive accountability with the CCPA's private litigation model. A key novelty of this research is its grounding of these reforms within the culturally resonant principles of Islamic jurisprudence (أمانة Amanah, ضرر Ḍarar), reframing data protection not as a foreign import but as a modern extension of the Kingdom's ethical heritage, thereby offering a coherent framework for legislative strengthening and enhanced compliance.

Keywords: PDPL, Data Breach, Negligence, Data Protection Law, Islamic Jurisprudence, Regulatory Enforcement.

1. INTRODUCTION

The dawn of the 21st century has been defined by the rise of a global, data-driven economy (Hoofnagle et al, 2019). Personal information is now often referred to as the "new oil" (Nusairat, 2024). This transformation has fueled unprecedented innovation and economic growth, yet it has also triggered a widespread and escalating crisis globally (Kanojia, 2023). Organizations are accumulating vast amounts of personal information, making them highly attractive targets for malicious actors. This leaves an individual's vulnerable to financial exploitation, identity loss, reputation harm, and psychological anguish (Aldubayyan, 2023). In response, a patchwork of data protection laws has emerged at the international level. These laws aim to impose order on the digital frontier and to restore the balance of power between data subjects and controllers (Abdullah, 2020). Among these regulations, the General Data Protection Regulation (GDPR) of the European Union and the California Consumer Privacy Act (CCPA) stand out as robust frameworks for data privacy, security, and responsibility (Alzahrani, 2024). The GDPR relies on a core principle of data protection, emphasizing privacy by design, clear breach notification, and administrative penalties (Elgujja, 2020). The CCPA, developed as a consumer

protection law, enables individuals to manage their data and seek redress, particularly in breaches caused by inadequate security (Sarabdeen & Moonesar, 2018).

In this international context, the Kingdom of Saudi Arabia has implemented the Personal Data Protection Law (PDPL) as part of its Vision 2030 strategy (Alfaifi, 2024). The PDPL represents a historical move to align national legal standards with international expectations for the digital economy (Al-Saggaf & Weckert, 2011). The law outlines principles for the processing of personal data, grants rights to data subjects, and establishes a supervisory authority. As with any new law, its effectiveness in addressing the complex reality of data security breaches remains to be seen (Al Nafea & Almaiah, 2021). A significant gap in current academic and practical debates is how the PDPL addresses organizational negligence. While the law includes broad requirements for confidentiality and security, it does not clearly define negligent breaches or the standard of care required. The concepts of causation and harm also need adaptation for digital injuries (Abanumy et al., 2005).

A key issue in the Saudi PDPL is legal uncertainty from the lack of a specific negligence standard (Alharbi et al., 2021). Without a clear duty of care, organizations may be uncertain about whether their data protection practices are legally sufficient. Conversely, harmed data subjects may have no remedy if the harm stems from a lack of reasonable care (Al-Mashaqbeh, 2025).

Traditionally, the legal definition of negligence has bridged the gap between poor organizational practices and resulting harm. It has formed the basis for regulatory enforcement and civil liability (Alzahrani, 2024). This uncertainty in the PDPL hinders compliance for regulated parties, influences the enforcement priorities of the Saudi Data and AI Authority (SDAIA), and presents challenges for the judiciary (Alhashim & Rahman, 2021).

This research uses doctrinal analysis to study the PDPL and systematically analyzes its main text. It also applies a comparative analysis with the GDPR and CCPA to identify best practices and regulatory models. The analysis draws on multiple theories: Tort Law and Negligence Theory (Hoofnagle et al., 2019), Privacy Theories—such as autonomy and structural power (Westin, 1967; Zuboff, 2019), Organization Accountability Theory (Bygrave, 2017), and Islamic Jurisprudence concepts like ضرر-ḍarar (harm) and تقصير-taqšir (fault) (Corrales et al., 2021). This doctrinal approach allows for a critical examination of negligence under the Saudi PDPL. The research aims are:

1. To identify and critically analyse the implicit standards of care and liability for negligent data breaches within the PDPL's legal text and implementing regulations.
2. To evaluate the deterrent effect of the PDPL's current penalty regime against negligent data breaches by comparing the incentives for compliance it creates against the costs of non-compliance.
3. To conduct a comparative legal analysis of the frameworks for negligence, external data transfer, and breach notification under the PDPL, the EU's GDPR, and the California Consumer Privacy Act (CCPA), identifying key divergences and their practical implications.
4. To propose evidence-based recommendations for legislative and regulatory refinement of the PDPL, focusing on incorporating a clear, actionable standard of care to more effectively prevent and redress negligence-related data breaches.

To guide this analysis, the study will address the following research questions

1. How does the Saudi PDPL currently define, or implicitly address, the legal concept of negligence in the context of data breach liability, and what are the resultant legal uncertainties?

2. To what extent do the administrative fines and corrective measures under the PDPL provide an effective and proportionate deterrent against organizational negligence in data security practices?
3. What are the principal differences in how the PDPL, GDPR, and CCPA conceptualize a data controller's duty of care, regulate international data transfers, and mandate breach notification, particularly in scenarios involving a lack of intent?
4. Based on the findings, what specific amendments to the PDPL's provisions, or its implementing regulations, would most effectively establish a coherent standard of care and improve mechanisms for redress in cases of negligent data breaches?

This research fills an identified gap in the literature by focusing on the underdeveloped concept of negligence within the PDPL. While earlier studies have outlined the law's general provisions, there is a lack of in-depth doctrinal and comparative analysis concerning this specific liability standard. The novelty of this study lies in its targeted analysis of how the absence of a clear 'duty of care' undermines both the law's operational effectiveness and its alignment with global standards, leading to concrete legal recommendations to bridge this critical gap.

2. LITERATURE REVIEW

2.1 The Doctrinal Challenges of Negligence in Data Protection

Translating the concept of negligence from the traditional to the digital sphere is conceptually challenging (Drechsler & Kamara, 2021). Physical injuries are clearer, but data breaches often result in non-physical, statistical, and diffuse damages, making it difficult to prove proximate cause and injury (Kilovaty, 2021). In classic torts, a single event typically causes harm, but a data breach is a systemic breakdown. Thousands may be at potential risk, though often unaware of future harm (Drechsler & Kamara, 2021). These uncertainties spark debate about the right level of responsibility. A fault-based negligence standard, such as the “reasonable security” rule in the CCPA, enables courts to assess organizational actions on a case-by-case basis (Almulihi et al., 2022). In contrast, some researchers argue that the ambiguity of cybersecurity and causation calls for strict liability. They suggest holding organizations responsible for breaches gives a stronger incentive to improve security (Alhashim & Rahman, 2021).

To make the negligence framework even more difficult, there is the so-called causation gap. According to Alharbi et al. (2021), the liability of careless organizations can be potentially protected in the event that a more advanced third-party attacker interferes and interrupts the legal path between poor security practices within an organization and the resulting damage. The literature leads to a compromise, therefore, on the agreement that mere importation of the traditional tort doctrine is not enough (Abanumy et al., 2005). It is increasingly believed that a hybrid system, a combination of fault-based principles and proactive governance-focused responsibilities, is the way forward. In this sense, Bygrave (2017) plays a leading role in defining the "accountability principle" of the GDPR as a regulatory form of negligence, making attention to a single breach incident give way to continual, systemic compliance through tools such as Data Protection Impact Assessments (DPIA).

2.2 Comparative Regulatory Paradigms: GDPR and CCPA

It is a theoretical problem of negligence expressed through the diversity of the world's principal data protection regimes (Boudjemaa, 2024). The European Union GDPR offers a risk-based

mechanism. Article 32 of the GDPR states that there should be sufficient technical and organisational measures, and this, as Voigt and Von Dem (2017) discuss, is codification of a developing negligence standard. The factors against which the appropriateness of these measures is judged are the state of the art, cost of processing, and the nature, scope, and context of processing (Suliman, 2025). This system is implemented through the imposition of substantial administrative fines, as well as the right to compensation and either real or non-material damages, as outlined in Article 82. The GDPR has introduced non-material damage with further development in post-GDPR jurisprudence that lowered the burden of proving harm by a considerable margin, thus bypassing one of the major challenges of traditional negligence litigation (Schmitz-Berndt & Schiffner, 2021).

On the contrary, the California Consumer Privacy Act (CPRA/CCPA) is a variant that is more oriented toward consumer welfare. Hoofnagle et al. (2019), as an extrapolation of Federal Trade Commission (FTC) case law regarding unfair and deceptive practices. Its tort of action is precipitated by a failure to execute sensible security processes and practices, and it establishes a de facto negligence threshold that is judged by the civil suit rather than a central body (Awwad & Abdelsattar, 2025). The first weakness identified in the literature is that the original limitation to breaches of specific categories of personal data has resulted in a more limited liability scope compared to the GDPR (Al Harbi, 2025). Finally, whereas the GDPR considers data protection as an essential right, the CCPA tends to act as a consumer market correction mechanism, which represents a philosophical divergence that affects the perception and penalty for negligence.

2.3. The Saudi PDPL: Critical Gaps in the Emerging Discourse

As a nascent law, the Saudi PDPL has only recently begun to attract academic scrutiny, with the emerging literature primarily consisting of descriptive overviews and initial compliance guidance (Alnasser, 2023). A critical gap identified in this review is the lack of in-depth, analytical scholarship that subjects the PDPL to a rigorous negligence-focused analysis using comparative and theoretical frameworks (Almutairi, 2025). The first commentaries, including that of Alhejaili (2024), rightly note that the PDPL establishes general principles of data security without being as granular as Article 32 of the GDPR or the CCPA's trigger for a reasonable security standard. The fact that the law does not specify what constitutes adequate security measures leaves organizations with considerable legal uncertainty. Moreover, the literature has not exhaustively examined the treatment of the two elements of any negligence claim by the PDPL: harm and causation (Alqahtani, 2024). Understanding whether psychological distress constitutes a recoverable harm and how these omissions by a controller can be legally linked to a particular violation is a question that, without a clearly defined statutory framework or a body of judicial precedent, remains unresolved in the Saudi legal literature.

2.4 Theoretical Foundations for Analysing Negligence

The analysis of negligence in the area of data protection law would need to be robust, and to achieve this, a theoretical framework that transcends the text of the statutes would be necessary. This paper will be grounded in three theoretical pillars and will serve as the basis for the analysis. To begin with, Tort Law and Negligence Theory provide the fundamental doctrinal framework for analyzing liability, based on the key aspects of duty, breach, causation, and harm (See Section 2.1). Second, Privacy Theories present conflicting arguments as to why information security is important (Johri & Kumar, 2023). The Liberal Autonomy model considers privacy as the right to privacy and self-determination, which constitute individual dignity, and negligence is treated as the infringement of the right (Ams, 2023). Conversely, Consumer Protection and Structural Power models view data breaches as market failures

or exercises of power, proposing alternative regulatory objectives and solutions (Mashaabi et al., 2023). Lastly, organizational accountability theory posits that successful regulation should not focus on the incidence, but rather on the organizational structures and processes.

This theory plays a crucial role in determining whether a law's motivation encourages the establishment of a genuine culture of compliance, which is a significant issue in discouraging lax enforcement.

3. METHODOLOGY

This research employs a doctrinal legal research methodology, which is qualitative in nature, to systematically investigate the treatment of negligence within the Saudi Arabian Personal Data Protection Law (PDPL). Doctrinal research is defined by its focus on the systematic exposition, analysis, and synthesis of legal rules and principles derived from primary sources such as statutes, regulations, and judicial decisions (Hutchinson & Treščáková, 2022). This approach is uniquely suited to the objectives of this study, as it facilitates a detailed, internal examination of the PDPL's text to identify its core provisions, discern its underlying logic, and critically evaluate its coherence and capacity to define a standard of care for negligence-based data breaches.

3.1 Research Design: Doctrinal Legal Analysis

The core of this research is a doctrinal legal analysis of the Saudi Arabian PDPL. This design was selected because it provides the most appropriate framework for a systematic and authoritative interpretation of the law as it is written. The process begins with a close reading and detailed exegesis of the PDPL's statutory text. The analysis focuses specifically on articles pertaining to the obligations of data controllers and processors, data security requirements, breach notification procedures, and the stipulated penalties and liabilities. The primary aim is to deconstruct the language of the law to determine whether it implicitly or explicitly incorporates the classical elements of negligence, duty of care, breach, causation, and harm, and to assess the clarity of the standard of care it imposes on organizations. This involves examining if the PDPL merely states abstract principles of data protection or if it provides actionable, enforceable standards against which negligent conduct can be measured. The doctrinal analysis forms the foundational layer upon which the critical and comparative evaluations are built, allowing for a grounded assessment of the law's current effectiveness and its doctrinal coherence.

3.2 Data Collection and Sources

The data collection for this study is split into primary and secondary legal sources. The primary sources consist of the authoritative legal texts under examination. This includes the full text of the Saudi Arabian Personal Data Protection Law and its implementing regulations, the consolidated text of the European Union's General Data Protection Regulation (GDPR), and the relevant sections of the California Consumer Privacy Act (CCPA) as amended by the CPRA. These texts serve as the raw material for the doctrinal and comparative analysis, providing the literal provisions from which legal interpretations are derived.

Secondary sources encompass a diverse range of scholarly materials used to contextualize and theorize the findings from primary law. This includes academic books, peer-reviewed journal articles, and commissioned reports from international bodies on data protection law, tort theory, cybersecurity, and comparative legal studies. Special attention is given to literature that discusses the

conceptualization of negligence in digital contexts, the economic and social impacts of data breaches, and the enforcement practices in different jurisdictions. Furthermore, relevant judicial decisions from jurisdictions with more mature data protection litigations are consulted to provide practical perspectives on how courts interpret and apply negligence principles in data breach cases. This comprehensive collection of sources ensures that the analysis is not only textually grounded but also informed by contemporary academic discourse and legal practice.

3.3 Analytical Framework: A Multi-Theoretical Lens

To move beyond a purely black-letter law analysis, this research employs a multi-theoretical analytical framework, as elaborated in Chapter 3, to critically interrogate the PDPL. The analysis is guided by four interconnected theoretical perspectives. First, Tort Law and Negligence Theory provide the classical doctrinal framework for assessing whether the PDPL establishes a clear duty of care, defines what constitutes a breach of that duty, and outlines a coherent path for establishing causation and remedying harm. Second, Privacy Theories (including the Liberal Autonomy, Consumer Protection, and Structural Power models) are used to evaluate the PDPL's underlying philosophy and to frame negligence not just as a legal failure, but as a violation of individual autonomy and a breach of trust. Third, Organizational Accountability Theory shifts the focus from individual acts to systemic and institutional failures, providing a lens to assess whether the PDPL encourages a culture of proactive compliance and risk management within organizations.

3.4 Comparative Legal Analysis

To contextualize the findings from the PDPL and to identify potential best practices, this study incorporates a structured comparative legal analysis with two leading international data protection regimes: the EU's GDPR and the California CCPA. These frameworks were selected due to their global influence, sophisticated regulatory approaches, and their explicit engagement with concepts of accountability, security, and liability. The comparison is focused thematically on three critical areas: (1) how each legal instrument defines and addresses negligence, either directly or through standards of "reasonable security"; (2) the rules and procedures governing data transfer to third countries and entities; and (3) the obligations and timelines for breach notification. By contrasting the PDPL's provisions with those of the GDPR and CCPA, the research aims to highlight relative strengths and weaknesses, identify regulatory gaps in the Saudi law, and distil actionable recommendations for its strengthening.

A critical component of methodological rigor is the language used in the primary sources. For the Saudi PDPL, the official Arabic text published in the Umm al-Qura Gazette is the primary source of analysis. This ensures the most accurate interpretation of legal concepts and obligations. For the comparative analysis, the official English language version of the GDPR and the original English text of the CCPA are used. Where necessary for the PDPL analysis, key terms and provisions were carefully translated by the author, with an awareness of potential linguistic nuances, to facilitate a precise comparison with the English-language frameworks.

3.5 Ethical and Contextual Reflection

This research is a doctrinal study and does not include human subjects. The main ethical concern is to interpret legal texts accurately and in context, while respecting their legal and cultural backgrounds. The analysis takes into account Saudi Arabia's unique socio-legal setting, its developing data protection

laws, and the possible impact of Islamic legal principles. It avoids simply applying outside concepts without careful adjustment.

4. RESULTS AND DISCUSSION

4.1 Doctrinal Ambiguity: The Undefined Standard of Negligence in the PDPL

One of the most significant findings of this study is the conceptual imprecision that surrounds the standard of negligence in Saudi PDPL, revealing a weakness in the regulatory design of its framework. In contrast to the GDPR, which, in turn, conceptualizes negligence in terms of a powerful accountability principle requiring evidence of compliance through Data Protection Impact Assessments, documented records, and risk-based implementation of the notion of appropriate technical and organizational measures. (GDPR, Article. 32), The PDPL imposes a similarly worded but critically hollow obligation.

The fact that the PDPL suggests the need for appropriate technical and organizational measures (PDPL, Article 20) defines an open-textured norm but does not outline it. In contrast to Article 32 of the GDPR, which provides contextual guidelines, such as the level of technological advancement and the cost of implementation, Article 20 does not provide any guidelines, leaving organizations with no clear standard of care. This shortcoming is further exacerbated by comparison with the pragmatic nature of the CCPA approach, which directly associates liability with a lack of either implementing a reasonable security procedure and practice (CCPA 1798.150), which has since been informed by statutory guidelines and case law, and which creates a clearer yet litigious route to proving fault. In its current formulation, the PDPL does not contain any such doctrinal point of reference, and the principle of negligence itself has become an entirely implicit concept, capable of being interpreted unpredictably and post hoc.

This doctrinal ambiguity generates a significant legal vacuum with direct practical consequences for both compliance and enforcement. For regulated entities, the absence of a clearly articulated standard of care, whether based on reasonableness or a dynamic risk-assessment model, fails to provide actionable guidance. Organizations are left to determine their own compliance ceiling, incentivizing a minimalist and self-serving interpretation of "appropriate" measures that may fall short of evolving cybersecurity best practices. To the regulators and courts, the stalemate in determining a clear court test on negligence erodes sound and powerful judgment. In the event of an infraction, the Saudi Data and AI Authority (SDAIA) and legal entities often lack a systematic approach to differentiate between unavoidable security incidents and criminal violations of responsibilities, resulting in the possibility of arbitrariness in the interpretation of the law and legal voids. This uncertainty, in effect, undermines the very purpose of the PDPL in preventive and deterrent agencies, as even the possibility of accidental negligence liability is obstructed by the law's inability to clearly define what constitutes an act of negligence within its framework.

4.2 The Enforcement Gap: Weak Deterrents and Limited Redress

Research shows that the PDPL has a significant enforcement gap, making it less effective at preventing careless data processing (Alnasser, 2025). In contrast, the GDPR and CCPA use stronger, multi-layered enforcement systems. The European model, for example, has a clear and powerful system for issuing fines. These fines can exceed 20 million Euros or 4% of a company’s total yearly turnover if they break the rules (Kärner, 2022).

It does not employ punitive actions but is a logical regulatory tool that directly mirrors the cost of non-conformity, which is directly proportional to the size of a specific organization. The GDPR poses a significant and real financial risk to the company, and consequently causes corporate executives and the board of governance to turn data protection into one of their inherent governance issues and invest in the long-term projects of tight security systems and pervasive compliant programs as an approach that could help avoid expensive reputational and financial damage.

To supplement the regulatory authority of the GDPR, the CCPA addresses a significant aspect of private enforcement, directly empowering individuals. It empowers consumers with a right of action, allowing them to request statutory damages in the event of a breach resulting from an organization's failure to apply reasonable security measures. Such a mechanism is successful in decentralizing enforcement, which leaves a topography of widespread litigation risk. The threat of class-action litigation, where statutory damages can be substantial in cases involving a large number of affected consumers, is a highly effective market-based deterrent. This consumer-empowerment model ensures that accountability is not solely dependent on the resource constraints and enforcement priorities of a central regulatory authority. The PDPL does not have a similar statutory cause of action for persons wronged by careless breaches in the Saudi context. Such an omission deprives the subjects of the data of a clear and easily accessible avenue of redress, leaving the entire task of enforcement to the regulatory body and leaving a vital mechanism of accountability largely unutilized.

The present-day enforcement framework of the PDPL is not robust enough, as it is based on a limited deterrent model. It lacks a multi-faceted structure to integrate the powerful, top-down regulatory fines, such as those under the GDPR, with the bottom-up, market-based pressure of a CCPA-style private right of action. This leads to materially ineffective incentives to deter negligence. Organizations that are run under the PDPL have a reduced perceived risk. Lacking a threat of harsh regulation penalties or a broad-based civil lawsuit, the economic and legal motivation to pursue state-of-the-art security is greatly reduced.

4.3 Procedural Lacunae: Vague Notification Timelines and Accountability Mechanisms

It appears that the framework of the PDPL lacks serious procedural loopholes that prevent effective crisis management and proactive compliance with established procedures. One of its key gaps is that the breach notification requirement towards data subjects is unclearly defined. Although the Implementing Regulations of the PDPL stipulate that notification of data subjects following the occurrence of a breach must be furnished promptly, this is in stark contrast to the GDPR, which has a clear deadline of 72 hours to notify the supervisory authority. Without undue delay is a very ambiguous phrase on its own, which leaves organizations with a dangerous loophole of doubt when they can postpone the important message under the pretext of an internal inquiry or damage survey. Such absence of a defined and strict timeframe to raise an alarm over the concerned people can contribute a lot to the possible damage, given that the concerned individuals would not have a chance to take appropriate measures to curb the impact, such as by freezing their accounts or changing their passwords and treating the effects of a breach. In comparison, the specific time frame of the GDPR is one of reduction, making the vehicle more favorable to the information subject by limiting harm and burdening the data controller with a clear and unambiguous process.

The PDPL establishes general principles but lacks explicit, granular requirements for specific accountability tools, which are the hallmark of modern, prevention-based regimes like the GDPR (Bouderhem, 2024). The principle of accountability under the GDPR is based on Data Protection Impact

Assessments (DPIAs) for high-risk processing and the detailed description of processing operations (GDPR, Articles 35 and 30), which serves as the working principle of the GDPR concept. These are not just administrative activities, but are crucial to compelling organizations to act in an organized manner of identifying, evaluating, and reducing the risks of data protection before they translate into breaches. This lack of such an expressly demanded process that must be documented in the PDPL is a field of strong neglect. It implies that a company can be in compliance with Saudi law, even though it has never conducted a formal, systematic risk assessment of its data processing operations. This is not due to its regulatory culture; an omission contributes to a culture of reactive, rather than proactive, compliance, and makes it extremely challenging to determine whether negligence, as a failure to foresee and mitigate foreseeable risks, existed prior to a breach, because no mandatory paper trail is required to prove due diligence.

4.4 A Hybrid Path Forward: Synthesizing GDPR and CCPA Strengths

A key finding indicates that integrating elements from both the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) offers a more effective approach to improving the Saudi Personal Data Protection Law (PDPL) than adopting either framework in isolation. The study demonstrates that these regulatory models complement each other and collectively provide more robust and comprehensive protection against negligence. The main advantage of the GDPR is that its architecture is designed to be accountability-based and preventive, requiring a culture of compliance to be proactive. Conversely, the CCPA is strong in its ex-post facto, market-oriented deterrence, where individuals can serve as their own attorneys general. A purely top-down regulatory framework, such as the GDPR, may be limited by the supervisory authority's resource base, whereas a purely consumer-litigation framework, such as the CCPA, can result in a lack of uniform enforcement and over-litigation. A middle way is, thus, a better way, both to bridge the enforcement gap and to resolve the doctrinal grey areas found within the PDPL. This synthesis entails the deliberate integration of key aspects from each framework. Based on the GDPR, the PDPL is required to directly address its fundamental tools of accountability by obligating the use of Data Protection Impact Assessments (DPIAs) in high-risk processing and the documentation of processing activities.

These procedures formalize a risk-based model, compelling organizations to document their adherence and critically assess potential harms before they occur, thus providing a tangible criterion of care that was previously lacking. Under the CCPA, the most significant import is the introduction of a legal cause of action for individuals who suffered a breach due to a failure to implement reasonable security measures. This generates a formidable, decentralized punitive response and generates a direct financial outcome of careless conduct that does not rely on the community to implement. Combining the ex-ante procedural rigor of the GDPR with the ex-post-consumer empowerment of the CCPA, the PDPL can establish a regime in which negligence is discouraged by both the daunting threat of major regulatory intervention and the ubiquitous risk of mass-scale private litigation.

4.5 Successful Harmonization and Cultural Legitimacy

Lastly, it is concluded in the analysis that the PDPL has skilfully provided a foundational basis on international interoperability and especially has provided a conditional framework of approaching cross-border data transfers that strategically reflects the principle of adequacy contained in the GDPR, thus providing an indicator that Saudi Arabia is fully prepared to enter the global digital economy while providing a benchmark of data protection to its citizens.

On a more fundamental level, but more importantly, the study reveals that the philosophical foundations of a robust negligence standard are not a transplanted foreign organ but rather deep rooted in the legal and ethical tradition of the Kingdom; the Islamic jurisprudential principles of أمانة-amanah (trust), which instils a fiduciary obligation in the handling of personal information, عدل-adl (justice), which requires the redress of actions that are wrong, and the taboo of ضرر-ḍarar (harm), which forbids inflicting harm on others, have the combined effect of offering a powerful. A synthesis of the foregoing doctrinal and comparative analysis is presented in Table 1 below. This thematic matrix consolidates the critical weaknesses identified within the PDPL's current framework, juxtaposes them against the established benchmarks of the GDPR and CCPA, and directly maps these findings to the specific, evidence-based improvements proposed by this study. The table serves to crystallize the argument that a strategic synthesis of elements from both comparative models is the most viable path to remedying the PDPL's doctrinal ambiguities and enforcement gaps.

Table 1: Comparative Analysis of Negligence and Enforcement Frameworks (PDPL, GDPR, and CCPA)

Thematic Area	**Saudi PDPL (Current Weaknesses)**	**GDPR (Benchmark)**	**CCPA/CPRA (Benchmark)**	**Proposed Improvement for PDPL**
Standard of Care Negligence	Appropriate measures (Article, 20)[1]	Risk-based appropriate measures Article. 32)[2];	"Reasonable security" standard; informed by statutory guidelines and case law.	Incorporate explicit criteria from GDPR [2]Article. 32 and mandate Data Protection Impact Assessments (DPIAs) for high-risk processing to create a

[1]Saudi Data & Artificial Intelligence Authority. (2023). *Personal Data Protection Law (English translation, 2nd rev.)*. Kingdom of Saudi Arabia.
Retrieved from https://sdaia.gov.sa/en/SDAIA/about/Documents/Personal%20Data%20English%20V2-23April2023-%20Reviewed-.pdf.
[2]European Parliament and Council of the European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Official Journal of the European Union, L119, 1–88. Available at: https://eur-lex.europa.eu/eli/reg/2016/679/oj.
[3]European Parliament & Council of the European Union. (2016). *Regulation (EU) 2016/679 (General Data Protection Regulation) of 27 April 2016 on the protection of natural persons with regard to the processing of personal data…* Official Journal of the European Union, L119, 1-88. Article 82. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX%3A02016R0679-20160504.

	No defined criteria or accountability principle.	Explicit criteria (state of art, cost, risk). Accountability principal mandates DPIAs records.		defined, proactive standard of care.
Enforcement & Deterrence	Limited administrative fines; no clear statutory private right of action for negligence-based breaches.	Strong top-down fines (up to 4% global turnover). Compensation for material/non-material damage (Article. 82).[3]	Bottom-up private right of action for breaches involving lack of "reasonable security"; enables statutory damages via consumer lawsuits.	Adopt a hybrid model: Introduce scaled administrative fines based on turnover (like GDPR) and create a statutory private right of action for breaches due to failure to implement "reasonable security" (like CCPA).
Breach Notification Timelines	Ambiguous "without undue delay" requirement for notifying data subjects.	Strict 72-hour deadline for notifying the supervisory authority.	Specific and reasonable timeframe for notification after discovery of a breach.	Replace "without undue delay" with a clear, definite timeframe (72 hours to the authority, promptly to the data subject) to ensure timely mitigation.
Proactive Accountability Mechanisms	Lacks explicit requirements for documented accountability tools like DPIAs or detailed processing records.	Mandatory DPIAs for high-risk processing; Records of Processing Activities (Article-82)[3]	Focused more on ex-post enforcement rather than ex-ante mandated documentation.	Explicitly mandate accountability tools from the GDPR, specifically DPIAs and Records of Processing Activities, to foster a proactive compliance culture and create a verifiable paper trail.
Philosophical & Cultural Foundation	General principles exist but are not explicitly linked to modern negligence concepts.	Framed as a fundamental right; comprehensive risk-based approach.	Framed as a consumer right; market correction and litigation-driven.	Articulate the standard of care using culturally resonant principles like Amanah (trust) and the prohibition of Ḍarar (harm) to ground the legal duty in Saudi and Islamic legal tradition.

5. DISCUSSION

This study examines how negligence is addressed under the Saudi Personal Data Protection Law (PDPL), using doctrinal and comparative analysis with the GDPR and CCPA as benchmarks. The results show that while the PDPL demonstrates a basic commitment to data protection, it does not yet have the clear definitions, strong enforcement, or detailed procedures needed to prevent, identify, and address data breaches caused by organizational negligence. The main findings regarding doctrinal ambiguity in establishing negligence highlight a fundamental conflict between abstract legal principles and practical considerations.

The standard of care enforcement, as defined by the undefined term 'adequate measures' in the PDPL, makes it both self-referential and legally ambiguous. This observation is consistent with the critique of first-generation data protection legislation, which often emphasizes general guidelines over practical rules (Alsadhan, 2025). However, in the Saudi case, it is especially acute, as there is no established body of judicial precedent to lend the term substance. In developing Alanazi's (2025) observation that the PDPL is overall vague, we will identify the absence of a negligence standard as a critical, narrower defect that disables the preventive role of the law. It is not simply a technical oversight but a fundamental vulnerability that enables organizations to justify poor security practices since the

law does not establish any external standard that they can compare their practice to, such as the state of the art in the GDPR or the reasonableness in the CCPA to which their practice will be contrasted against. The ambiguity serves as a successful defense against negligent behavior because it becomes extremely challenging to demonstrate, a problem that has long plagued tort theory and is exacerbated by a highly complex and opaque data security environment (Memeti, 2024).

The identified enforcement gap further exacerbates this problem, revealing a system with insufficient deterrent power. The finding that the PDPL lacks both the GDPR's scaled administrative penalties and the CCPA's decentralized private right of action is crucial. It suggests that the Saudi regime currently embodies what could be termed a "hollow accountability" model, which imposes obligations without creating credible consequences for their negligent breach. This finding resonates with Hoofnagle et al.'s (2019) assertion that effective data protection requires a "multi-instrumental" approach to enforcement. Our contribution lies in demonstrating how the PDPL's underdeveloped enforcement mechanisms fail to create the necessary economic incentives for compliance. While previous literature has focused on the GDPR's fines (Alhazmi, 2025) or the CCPA's litigation risks in isolation, our analysis highlights the synergistic effect of their absence in the PDPL.

Perhaps the most surprising and significant finding of this research pertains to the foundation for cultural legitimacy. While much of the comparative data protection literature assumes a tension between "global" standards and "local" values, our analysis uncovers a powerful consonance. The discovery that core Islamic jurisprudential principles, الأمانة-Al-Amānah (trust), ` عدل-adl (justice), and the prohibition of ضرر-Ḍarar (harm) provide a robust ethical foundation for stringent data protection duties is a critical insight. These finding challenges narratives that might view the adoption of norms from the GDPR or CCPA as a form of legal transplantation, instead reframing it as a process of doctrinal revitalization and contextualization.

This approach proposes that strengthening the Personal Data Protection Law (PDPL) should emphasize integrating modern data protection into the Kingdom’s ethical and legal traditions, rather than relying on external models. This perspective is consistent with recent scholarship that seeks to ground data governance in diverse philosophical frameworks, while also offering a clear, practical framework for Saudi Arabia and other Muslim-majority countries. The hybrid model represents more than a pragmatic solution; it exemplifies a well-founded regulatory transformation. For instance, combining ex-ante accountability measures, such as Data Protection Impact Assessments (DPIAs) under the General Data Protection Regulation (GDPR), with ex-post-consumer control mechanisms from the California Consumer Privacy Act (CCPA) can mitigate negligence at various stages. This method is distinctive for explicitly integrating two major global approaches, thereby establishing a novel regulatory framework. Furthermore, it advances beyond conventional legal literature that dichotomizes capitalism and legal reasoning, with the GDPR and CCPA serving as exemplars of rights-based and market-based models, respectively.

5.1 Policy Implications

This research identifies key policy issues that necessitate targeted reforms by Saudi regulators. To address gaps in the Personal Data Protection Law (PDPL), new regulations should be implemented. These regulations should establish a reasonable standard of care for data security, utilizing specific risk-based criteria to clarify required measures. Such clarity would resolve existing uncertainty regarding the definition of negligence. Policymakers should also consider introducing a graduated penalty system,

similar to the General Data Protection Regulation (GDPR), to address enforcement deficiencies. Penalties should be commensurate with the severity of the breach and the organization's revenue. Granting individuals a legal right to seek recourse when organizations fail to implement reasonable security measures would support affected parties and promote compliance. Finally, replacing the ambiguous requirement to notify data subjects of breaches without undue delay with a definitive 72-hour deadline would enhance procedural clarity.

6. CONCLUSION

This paper concludes that the Saudi Personal Data Protection Law (PDPL) does not adequately address careless data breaches. Its standards of care, enforcement, and procedures have key weaknesses that limit its ability to prevent and remedy harm. To address these issues, Saudi Arabia should draw on global best practices. This would mean combining the GDPR's proactive, accountability-focused approach with the CCPA's emphasis on litigation and consumer rights. Together, these elements would create a stronger and more reliable system of accountability. Importantly, this legal improvement is not an outside imposition but a logical step forward. The core ideas of duty, responsibility, and preventing harm are already part of Islamic law, including principles like الأمانة-Al-Amānah (trust) and the prohibition of ضرر-Ḍarar (harm). These values give cultural legitimacy to making the PDPL a stronger, more effective, and internationally aligned data protection law for the digital age.

6.1 Limitations and Future Research

While the doctrinal and comparative approach remains fundamental, it is limited by the absence of empirical data regarding the practical application of the Personal Data Protection Law (PDPL) and the compliance challenges faced by Saudi organizations. Furthermore, the lack of domestic judicial or administrative case law from the Saudi Data and AI Authority (SDAIA) restricts analysis to the statutory text, without consideration of its implementation. These limitations, however, highlight clear avenues for future research. Given the evolving regulatory environment, empirical investigations into SDAIA implementation practices and organizational compliance strategies are warranted. Additionally, quantitative assessments of the economic impact of proposed reforms should be pursued. Comparative analysis with other Gulf Cooperation Council (GCC) countries is also necessary to evaluate prospects for regional harmonization. Advancing this conceptual framework through evidence-based research will support the effective development of the PDPL.

Acknowledgements: Not available
Author contributions: The author contributed solely to all parts of this study, including the conceptualization, research design, data analysis, and writing.
Ethical Statement: This doctrinal legal study did not involve human or animal subjects; therefore, ethical approval was not required.
Consent to Participate: Not available
Competing Interests: The author declares that this work has no competing interests.
Grant/Funding information: The author declared that no grants supported this work.
Data Availability Statement: This is a doctrinal research paper without an empirical dataset; all references and materials are either publicly available or cited in the manuscript.
Declaration Statement of Generative AI: Generative AI tools were not utilized in the conceptualization, analysis, or writing of this manuscript. Grammarly was employed exclusively for editing, grammar, and style.

REFERENCES

Abanumy, A., Al-Badi, A., & Mayhew, P. (2005). E-Government Website Accessibility: In-Depth Evaluation of Saudi Arabia and Oman. *Electronic journal of e-government*, *3*(3), 149-156. https://academic-publishing.org/index.php/ejeg/article/view/437/400

Abdullah, A. (2020). *Consumers' personal data protection in Saudi Arabia: A comparative analytical study* (Doctoral dissertation, University of Kansas). ProQuest Dissertations Publishing. https://www.proquest.com/openview/36ca660cf5d8a3728b428d64cefa780b/1?pq-origsite=gscholar&cbl=18750&diss=y

Al Harbi, I. (2025). Artificial Intelligence in Saudi Arabia: Intercultural Human Rights Perspectives on Legal Frameworks and Regulatory Protection (Doctoral dissertation, St. Thomas University).

Al Nafea, R., & Almaiah, M. A. (2021). Cybersecurity Threats in Cloud: Literature Review. *International conference on information technology (ICIT) (779-786). IEEE*. doi: 10.1109/ICIT52682.2021.9491638.

Alanazi, A. (2025). Assessing Clinicians' Legal Concerns and the Need for a Regulatory Framework for AI in Healthcare: A Mixed-Methods Study. *Healthcare*, *13*(13), 1487. https://doi.org/10.3390/healthcare13131487

Aldubayyan, A. (2023). *Privacy regulation of cellular network data: A comparative study with recommendations for the Kingdom of Saudi Arabia* (Doctoral dissertation, The University of Waikato). https://hdl.handle.net/10289/17061

Alfaifi, A. (2024). *Lost profit damages for breaches of commercial contracts: Examining common law and civil law approaches to recovery and lessons for Saudi Arabia* (Doctoral dissertation, University of Essex). https://repository.essex.ac.uk/37771/1/PhD%20Thesis.pdf

Alharbi, A. S., Halikias, G., Rajarajan, M., & Yamin, M. (2021). A Review of Effectiveness of Saudi E-Government Data Security Management. *International Journal of Information Technology*, *13*(2), 573-579. https://doi.org/10.1007/s41870-021-00611-3

Alhashim, S. S., & Rahman, M. H. (2021). Cybersecurity Threats in Line with Awareness in Saudi Arabia. *International Conference on Information Technology (ICIT), IEEE*, (314-319). https://doi.org/10.1109/ICIT52682.2021.9491711.

Alhejaili, M. O. M. (2024). *Securing the Kingdom's e-commerce frontier: Evaluation of Saudi Arabia's cybersecurity legal frameworks.* Journal of Governance & Regulation, 13(2), 275–286. https://doi.org/10.22495/jgrv13i2siart4

Al-Mashaqbeh, Y. A. (2025). Legislative Framework Regulating Digital Media in Jordan and Arab Countries: A Study on The Legal Dimensions. *Lex localis-Journal of Local Self-Government*, *23*(10), 1-20. https://doi.org/10.52152/

Almulihi, A. H., Alassery, F., Khan, A. I., Shukla, S., Gupta, B. K., & Kumar, R. (2022). Analyzing the Implications of Healthcare Data Breaches through Computational Technique. *Intelligent Automation & Soft Computing*, *32*(3). https://doi.org/10.32604/iasc.2022.023460

Almutairi, S. (2025). *Kuwait's fragmented data protection framework: Toward reform through comparative analysis with the GDPR and Saudi Arabia's PDPL.* SSRN. https://doi.org/10.2139/ssrn.5464634

Alnasser, A. (2023). Rhetorical Strategies and Ideologies in Saudi TEDx talks. *International Journal of Linguistics, Literature & Translation*, *6*(3). https://doi.org/ 10.32996/ijllt.2023.6.3.22

Alnasser, H. A. (2025). The Concept of Negligence in Data Breach: A Comparative Doctrinal Analysis of the EU, California, and Saudi Arabia. *Veredas do Direito*, *22*(3), e223404-e223404. https://doi.org/10.18623/rvd.v22.n3.3404

Alqahtani, F. (2024). Persuasion Strategies in Saudi Arabia Vision 2030 Document: A Critical Discourse Analysis Approach. *Theory & Practice in Language Studies (TPLS)*, *14*(4). https://doi.org/10.17507/tpls.1404.32

Alsadhan, A. A. (2025). A Survey of Security Threats and Challenges Related To 5G Networks in Saudi Arabia. *Qubahan Academic Journal*, *5*(3), 474-501. https://doi.org/10.48161/qaj.v5n3a1849

Al-Saggaf, Y., & Weckert, J. (2011). Privacy from a Saudi Arabian Perspective: The case of students in a private university. *Journal of Information Ethics*, *20*(1), 34. https://www.proquest.com/openview/d77227c0f1beaa8d3271d6e8a3d215d6/1?pq-origsite=gscholar&cbl=2035668

Alzahrani, R. B. (2024). An Overview of AI Data Protection in The Context of Saudi Arabia. *International Journal for Scientific Research*, *3*(3), 199-218. https://vsrp.co.uk/wp-content/uploads/

Ams, S. (2023). Blurred Lines: The Convergence of Military and Civilian Uses of AI & Data Use and Its Impact on Liberal Democracy. *International Politics*, *60*(4), 879-896. https://doi.org/10.1057/s41311-021-00351-y

Awwad, A., & Abdelsattar, A. (2025). Digital Evidence in Forensic Accounting-A Study in Saudi Legislation. *Cogent Social Sciences*, *11*(1), 2522958. https://doi.org/10.1080/23311886.2025.2522958

Bouderhem, R. (2024). A review of Saudi e-commerce regulation under the scope of the GDPR. *Arab Law Quarterly*, *1*(aop), 1-19. https://doi.org/10.1163/15730255-bja10154

Boudjemaa, Y. (2024). Ensuring Regulatory Compliance in Cloud-based Big Data Systems: A Framework for Global Operations Adhering to GDPR and CCPA. *Studies in Knowledge Discovery, Intelligent Systems, and Distributed Analytics*, *14*(9), 15-27. https://edgescholar.com/index.php/SKDISDA/article/view/e-2024-09-07

Bygrave, L. A. (2017). Data Protection by Design and by Default: Deciphering the EU's Legislative Requirements. *Oslo Law Review*, *4*(2), 105-120. https://doi.org/10.18261/issn.2387-3299-2017-02-03

Corrales Compagnucci, M., Aboy, M., & Minssen, T. (2021). Cross-Border Transfers of Personal Data After Schrems II: Supplementary Measures and New Standard Contractual Clauses (Sccs). *Nordic Journal of European Law*, *4*(2). https://doi.org/10.36969/njel.v4i2.23780

Drechsler, L., & Kamara, I. (2022). *Essential equivalence as a benchmark for international data transfers after Schrems II.* In E. Kosta, R. Leenes, & I. Kamara (Eds.), *Research handbook on EU data protection* (pp. 314-352). Edward Elgar Publishing. https://doi.org/10.4337/9781800371682.00022

Elgujja, A. A. M. (2020). *Adequacy of the legal safeguards of the patients' confidentiality right under the Saudi Arabian laws* (Doctoral dissertation, University of Salford). University of Salford Repository. https://salford-repository.worktribe.com/preview/1486896/Thesis%2000343621.pdf

Hoofnagle, C. J., Van Der Sloot, B., & Borgesius, F. Z. (2019). The European Union General Data Protection Regulation: What It Is and What It Means. *Information & Communications Technology Law*, *28*(1), 65-98. https://doi.org/10.1080/13600834.2019.1573501

Hutchinson, C. S., & Treščáková, D. (2022). The Challenges of Personalized Pricing to Competition and Personal Data Protection Law. *European Competition Journal*, *18*(1), 105-128. https://doi.org/10.1080/17441056.2021.1936400

Johri, A., & Kumar, S. (2023). Exploring Customer Awareness Towards Their Cyber Security in The Kingdom of Saudi Arabia: A Study in The Era of Banking Digital Transformation. *Human Behavior and Emerging Technologies*, *2023*(1), 2103442. https://doi.org/10.1155/2023/2103442

Kanojia, S. (2023). Ensuring Privacy of Personal Data: A Panoramic View of Legal Developments In Personal Data Protection Law In Saudi Arabia. *J. Int'l L. Islamic L.*, *19*(3), 270-276. https://heinonline.org/HOL/LandingPage?handle=hein.journals/jispil19&div=51&id=&page=

Kärner, M. (2022). Interplay Between European Union Criminal Law and Administrative Sanctions: Constituent Elements of Transposing Punitive Administrative Sanctions Into National Law. *New Journal of European Criminal Law*, *13*(1), 42-68. https://doi.org/10.1177/20322844221085918

Kilovaty, I. (2021). *Psychological data breach harms.* North Carolina Journal of Law & Technology, 23(1), 1–66. https://doi.org/10.2139/ssrn.3785734

Mashaabi, M., Al-Yahya, G., Alnashwan, R., & Al-Khalifa, H. (2023). Arabic privacy policy corpus and classification. In A. Gal, M. Jarrar, & Y. Kanza (Eds.), *Applications of Natural Language to Information Systems* (pp. 94–108). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-35320-8_7

Memeti, N. (2024). From Legislation to Enforcement: Tackling Digital Acquisitions in the Gulf Region. *Digital Society*, *3*(3), 67. https://doi.org/10.1007/s44206-024-00152-9

Nusairat, W. M. (2024). Legal Protection of Personal Data Privacy in the Kingdom of Saudi Arabia. *Manchester Journal of Transnational Islamic Law & Practice*, *20*(1). https://heinonline.org/HOL/LandingPage?handle=hein.journals/jispil20&div=19&id=&page=

Sarabdeen, J., & Moonesar, I. A. (2018). Privacy protection laws and public perception of data privacy: the case of Dubai e-health care services. *Benchmarking: An International Journal*, *25*(6), 1883-1902. https://doi.org/10.1108/BIJ-06-2017-0133

Schmitz-Berndt, S., & Schiffner, S. (2021). Don't tell them now (or at all)–responsible disclosure of security incidents under NIS Directive and GDPR. *International Review of Law, Computers & Technology*, *35*(2), 101-115. https://doi.org/10.1080/13600869.2021.1885103

Suliman, H. O. H. (2025). Evaluating the effectiveness of Saudi Arabia's PDPL in the global digital economy. *Journal of Data Protection & Privacy*, *8*(1), 97-111. https://doi.org/10.69554/ELUX6976

Voigt, P., & Von dem Bussche, A. (2017). *The EU General Data Protection Regulation (GDPR): A practical guide*. Springer International Publishing.

Westin, A. F. (1967). Special report: legal safeguards to insure privacy in a computer society. *Communications of the ACM*, *10*(9), 533-537. https://dl.acm.org/doi/pdf/10.1145/363566.363579

Zuboff, S. (2019). Surveillance capitalism and the challenge of collective action. In *New labor forum*. *Sage CA: Los Angeles, CA: Sage Publications*, *28*(1), 10-29. https://doi.org/10.1177/1095796018819461

JADHUR

Vol.4, Issue.3, (2025)

Journal of Advances in Humanities Research

https://doi.org/10.56868/jadhur.v4i3.264

Public Opinion, Educational Equity, and Governance in an Era of Mass Higher Education: Lessons from "Shanhe University

Qianyi Xu[1],Yufei Song[2*], Yongcai Li [3]

1. Capital Normal University, Beijing, China.

2. NO.12 Middle School of Linyi, Shandong, China.

3. Guizhou Vocational and Technical College of Water Resources and Hydropower, Guizhou

Article Information

Article Type: Research Article

Dates:
Received: 07 July 2025
Revised: 15 August 2025
Accepted: 25 August 2025
Available online: 21 September 2025

Corresponding Author: Yufei Song
516581206@qq.com

ABSTRACT

The massification of higher education in China has shifted public discourse from quantitative access to the qualitative and equitable distribution of elite educational resources. This study investigates the viral "Shanhe University" phenomenon as a critical case of bottom-up public sentiment regarding educational equity. Employing a structured thematic analysis, we collected and analyzed 494 user comments from two official media videos on the Bilibili platform using NVivo 12 software. The results reveal a sophisticated public understanding of equity, structured around three interconnected dimensions: opportunity (54.1% of coded references, highlighting severe regional disparities in elite university access and admission quotas), process (36.8% of references, focusing on perceived inferior teaching conditions and resource deprivation in local institutions), and outcome (9.2% of references, expressing anxiety over employment discrimination and limited further education pathways). The discourse underscores a systemic public critique of geographical resource stratification and credential devaluation. In response to this public agenda, we recommend governance pathways focused on structural reforms to rebalance opportunity, targeted resource compensation for non-elite universities, and policy measures to mitigate labor market discrimination. The findings highlight the imperative of integrating public voice into higher education governance during the massification era.

Keywords: Popularization, Equity in higher education, High-quality development, Public opinion analysis, Massification

1. INTRODUCTION

The internationalization of higher education into a massified, even universal system is among the most significant social change processes of the last half-century (Zha, 2020). The dominant approach, as outlined by Martin Trow, categorizes higher education into three stages: elite, mass, and universal, with gross enrollment rate thresholds of 15% and 50% (Shan & Guo, 2014). China is a nation that has undergone an extraordinarily rapid transformation. Chinese higher education has officially entered the universalization

phase since 2019, when its gross enrollment rate exceeded 50%, reaching 59.6% in 2022 (Hayhoe et al., 2012). This historic change does not denote simple quantitative growth, but rather the broader socialization of universities, as they transform into people-oriented institutions integral to nation-building and personal ambition (Xuyang & Shanming, 2018).

Equity, an extension of social justice into education, is inherently concerned with the equal allocation of resources, opportunities, and outcomes (Zhong, 2018). During the massification phase, demand has shifted gradually, moving from a shallow primary focus on access to college to a deeper emphasis on quality and the equitable distribution of high-quality educational resources (Xu, 2021). This has introduced a profound sense of disparity, highlighting the long-standing predicament of uneven and insufficient development in China's higher education system. Concerns such as the localization of elite universities, inequality in funding and faculty quality, and the stratifying impact of institutional reputation have come to the center of the popular agenda (Zhong, 2018), such agendas often take shape in online public opinion in the digital era, forming a powerful force that policymakers can no longer overlook (Ryan et al., 2010).

The primary policy documents, such as the 14th Five-Year Plan, Vision 2035, and the report of the 20th National Congress of the Communist Party of China, have always mentioned that a high-quality education system should be built and that the distribution of higher education resources in the regions should be optimized to promote educational equity (Xue & Li, 2022). This suggests a national commitment to ensuring that higher education growth leads to both excellence and equity (Han, 2022). Although equity in higher education has been widely studied through policy, legal, and resource lenses, a gap remains in the literature: a notable lack of scholarship that directly addresses the issue and examines the voices of the populace through spontaneous online discussions (Geng, 2022). Public opinion can be a valuable and timely means of exposing lived experiences and perceived injustices that may not be adequately reflected in top-down analyses of policy or quantitative data.

The viral and emergent creation of the so-called Shanhe University in the summer of 2023 is an ideal example of this dynamic (Germain, 2022). On the eve of college entry exam scores and college applications, a fake university called "Shanhe University" went viral on Chinese social media (Levinson et al., 2022). A few big and comparatively few high-quality universities made these provinces the epicenter of a widespread outcry. This online meme rapidly evolved from a grassroots phenomenon into a major media agenda and ultimately prompted a tangible response from the Ministry of Education. This is why "Shanhe University" is a perfect empirical tool that can be used to study the essence of anxiety of the population about educational justice; it is a strong and concise form of expression of the main concerns of the population, the very roots of their disillusionment, and the possible ways of reforming governance.

To systematically investigate this phenomenon, this study is guided by the following research questions:

1. What are the specific dimensions of higher education equity (e.g., opportunity, process, outcome) that the public prioritizes in the "Shanhe University" discourse?
2. What are the underlying socio-structural and perceptual roots driving the formation of this public agenda on higher education equity?
3. Based on public concerns, what governance pathways can be proposed to enhance equity in Chinese higher education amid massification?

Theoretically, this study falls within the conceptualization of Trow's massification theory, which serves as the macro-level framework for the transition and associated tensions. Moreover, it employs the concepts of agenda-setting and opinion shaping to examine the emergence of a grassroots problem that gains national awareness. The significance of this study lies in its direct engagement with the public voice. By analysing the "Shanhe University" phenomenon, it moves beyond traditional policy analysis to uncover the nuanced and emotionally charged perceptions of equity held by key stakeholders, including students, parents, and the broader public. The findings aim to provide empirically grounded insights to inform more responsive and effective higher education governance, ensuring that the system's continued expansion is both high-quality and equitable. The following sections present the analysis results, discuss the root causes of the identified concerns, and propose targeted governance pathways to foster a more equitable future in Chinese higher education.

2. LITERATURE REVIEW

2.1 The Evolving Dimensions of Equity in Chinese Higher Education

The conceptualization of equity within China's higher education system has undergone significant evolution alongside the system's own expansion (Chen et al., 2022). Early scholarship primarily defined equity through the lens of equality of access, measured by gross enrollment rates and the democratization of opportunity (Yang, 2021). Following the achievement of mass participation, academic focus broadened to critically examine the quality and geographical distribution of opportunities. A substantial body of literature critiques the concentration of elite resources, epitomized by initiatives like Project 985 and Project 211, such as Beijing, Shanghai, and Jiangsu (Li & Xue, 2022). Research confirms that a student's geographical location remains a strong predictor of their likelihood of admission to a top-tier institution, thereby perpetuating inequality in access to elite education.

In response to these regional disparities, the Chinese government has implemented targeted policies aimed at structural reform (Omri & Aksoy, 2024). This is exemplified by the systematic establishment of new campuses and branches of elite universities in central and western China. Significant scholarly attention has been paid to the government's response to the "Shanhe University phenomenon," which highlighted the challenges faced by institutions in less-advantaged localities (Li & Ruppar, 2021). In a notable policy shift, substantial state investment has been directed toward revitalizing such regional universities, not merely through replication, but by fostering their unique disciplinary strengths and improving infrastructure, faculty quality, and student support systems (Zhong & Zheng, 2022). These measures represent a concerted effort to disrupt the cycle of regional disadvantage and enhance the overall quality and equity of the higher education ecosystem (Cochran-Smith & Keefe, 2022).

Beyond access, the scholarly discourse on equity has expanded to encompass process and outcome dimensions. Process equity concerns the fairness of the educational experience itself, analyzing disparities in per-student funding, faculty qualifications, infrastructure, and learning conditions between key and non-key institutions (Levinson et al., 2022). As Marginson and Yang (2022) note, resource dependency can create a vicious cycle for non-key universities, hindering their ability to deliver high-quality education. Studies indicate that the symbolic capital of a degree from a prestigious university carries immense weight in China's labor market, contributing to credential inflation and employer discrimination against graduates from lesser-known institutions (Tang et al., 2024). This stratified opportunity structure often prioritizes institutional prestige over individual merit.

2.2 Policy Responses and Structural Challenges

The Chinese government has addressed these equity issues by implementing a set of policies (Zancajo et al., 2022). The "Double First-Class" program was introduced in 2017 to build world-class universities and disciplines while breaking strict hierarchies and promoting competitiveness (Dabis & Csak, 2024). Nevertheless, the critical commentaries suggest that such a policy, like precedents, remains focused on a small group of individuals, thereby further entrenching rather than addressing current inequalities (Katsamakas et al., 2024). Incentives such as targeted admission rates to low-income areas and poverty relief through special admission channels to colleges are additional policy instruments (Huck & Zhang, 2021). Although these steps are well motivated, researchers argue that they often serve as stopgap measures that do not fundamentally alter the nature of resource allocation (Kawuryan et al., 2021). The literature attributes these chronic inequities to complex root causes. Among them are the historical heritage of key-construction policies that have established an ingrained hierarchy (Khoso et al., 2022), the decentralization of the financial side of the university that anchors the financial independence of a university to the wealth of its local community, and the socio-cultural solidification of public perception that considers only a small group of elite universities to be of good quality (Luo, 2024; Darazi et al., 2023).

2.3 Identified Research Gaps and This Study's Contribution

The majority of the literature employs a top-down approach, studying equity using policy documents, official statistics, and institutional-level data (Marginson, 2016). Research on top-down living experiences and perceptions of equity, as expressed by the population itself, remains conspicuous. Although researchers such as Yang (2021) have addressed the issue of public anxiety about the devaluation of the diploma and resentment in the region, such studies are often overshadowed by their primary policy or economic agenda. The real substance, constitution, and power of social opinion as a catalyst for the equity agenda have been under-researched (Shaturaev, 2021). This is a critical omission. Online platforms have become a primary source of agenda-setting in the digital era, crystallizing public sentiment and placing significant strain on policymakers within a relatively short period (Sallam et al., 2023). A classic example of such a bottom-up agenda is the viral phenomenon known as Shanhe University, which has not been a focus of scholarly study (Robinson, 2023). The current literature cannot reflect the subtle, emotional, and multi-dimensional nature of the popular discussion on this problem.

This paper aims to address this gap directly by analyzing online public opinion generated in response to the Shanhe University case. It goes beyond conventional frameworks of policy and institutional analysis to ask: What specifically does the public focus on regarding equity? On what root do these inequities lie? Besides, how does this social discourse inform the possible avenues of governance? Through a qualitative analysis of this viral discourse, this study offers a bottom-up perspective on equity in higher education, supplementing the top-down analyses currently leading the field. It proves that social opinion is not only a symptom of equity issues but also a well-developed source of information about their character and the development of more reactionary and acceptable governance solutions in the age of massification.

3. METHODOLOGY

This study employs a qualitative research design to conduct a thematic analysis of online public opinion concerning the "Shanhe University" phenomenon. The objective is to systematically identify, analyse, and report the patterns (themes) within public discourse to understand the specific dimensions of

higher education equity that resonate with the public. The following subsections provide a comprehensive description of the methodological process.

3.1 Data Source and Justification

This study uses publicly available user-generated comments from the Chinese social media and video-sharing platform, Bilibili (www.bilibili.com). While Bilibili's user base skews younger (~90% under 25), this demographic is the primary group directly experiencing recent higher-education transitions and reforms, making their views critical for understanding contemporary societal perceptions. The platform's culture of active commentary provides a unique window into bottom-up discourse. However, it is acknowledged that this sample does not represent the views of all stakeholders, such as parents and rural communities, whose perspectives may be expressed on other platforms, such as WeChat or Douyin. The study therefore interprets findings as the expressed sentiments of a digitally engaged, younger cohort.

3.2 Video Selection and Data Collection

This study employed a purposive sampling strategy to capture significant online discourse surrounding the "Shanhe University" phenomenon. Data were collected from two of the most-viewed and most-commented Bilibili videos on the topic, both originating from official media accounts to ensure authoritative sourcing and substantial public engagement. The first video, titled "Virtual University 'Shanhe University' is Trending! Media Comment: We Must Listen to the Special Significance Behind the Voice," was posted by the "Elephant News" account (Henan Radio and Television) on July 3, 2023. The second video, "Ministry of Education Responds to 'Shanhe University' Statements," was posted by the "People's Daily" account on July 6, 2023. Archived snapshots confirming the state of these videos and their comment sections as of the data collection date are available from the authors upon request. All top-level (direct) comments visible on these two videos were collected on August 12, 2023. The initial raw dataset consisted of 783 comments. Reply threads (nested comments) were excluded from collection to prioritize unique, individual viewpoints and to avoid the data inflation and thematic redundancy often present in extended back-and-forth debates. A key methodological consideration is that sourcing data from official state media channels introduces the risk of survivorship bias, as their comment sections are subject to active moderation. Consequently, the dataset likely skews toward milder, less politically sensitive critiques, as more 尖锐 discourse may have been deleted during the peak discussion period in early July, prior to our collection date.

3.3 Data Cleaning and Final Sample

To ensure the integrity of the qualitative analysis, the raw dataset underwent a transparent, multi-stage cleaning process. The initial dataset comprised 783 verified top-level comments collected from the two specified videos on August 12, 2023. A detailed exclusion protocol was systematically applied, resulting in the removal of 289 comments (36.9%). The process and counts for each stage are documented in Table 1 below.

Table 1. Data Cleaning and Exclusion Protocol

Cleaning Stage	Criteria for Exclusion	Number of Comments Excluded	Cumulative Remaining
Stage 1: Raw Data	N/A (Initial dataset)	-	783
Stage 2: Non-Textual/Low-Content Removal	Single emojis, image memes, comments of only punctuation (e.g., ".....") or repetitive character strings (e.g., "hhh", "666").	187	596
Stage 3: Irrelevance Removal	Comments entirely unrelated to higher education equity (e.g., advertisements, off-topic jokes, generic praise/criticism of the video presenter with no substantive link to the topic).	89	507
Stage 4: Duplicate Consolidation	Removal of near-identical comments posted repeatedly by the same user, retaining only the first instance.	13	49400%
Final Analytical Corpus	Valid, unique, and thematically relevant textual comments.	Total Excluded: 289	49400%

This multi-stage filtration resulted in a final analytical corpus of 494 valid textual comments. The exclusion rate of 36.9% is consistent with expected patterns of low-content and off-topic posts on the platform. To ensure full transparency, a supplementary file containing 50 randomly selected raw comments that were excluded, along with the specific reason for each exclusion, is available from the authors upon request and will be submitted with this revision.

3.4 Data Analysis

The textual dataset was analyzed using NVivo 12, following a structured thematic analysis approach that integrated data-driven exploration and theory-informed coding. This methodology ensured that the findings were empirically grounded while remaining contextualized within the established scholarly framework of higher education equity. The process commenced with an Exploratory Word Frequency Analysis. The NVivo "Word Frequency Query" function was applied to the entire corpus of 494 valid comments. A custom stop-word list filtered out common but thematically neutral function words (e.g., "the," "is," "very"), allowing the analysis to focus on substantive content. This initial phase produced a data-driven overview of the most frequent terms in the discourse, including "gaokao," "Henan," "resources," and "fairness." The resulting word cloud and ranked keyword list provided inductive insights into commenters'

prominent concerns, which subsequently informed but did not rigidly determine the development of the coding framework.

The core of the analysis proceeded with Systematic Coding and Theme Development, conducted by a team of two researchers to enhance methodological rigor. Coding began with a flexible framework derived from the three theoretical dimensions of equity in higher education: Opportunity, Process, and Outcome. This framework served as a sensitizing guide while remaining open to themes emerging inductively from the data. The unit of analysis was a "coded reference," defined as any distinct text segment (a clause or sentence) expressing a unique idea pertinent to a specific theme. Therefore, a single comment could yield multiple coded references across different thematic categories. Both researchers independently applied the framework to the dataset, creating new codes as needed. Following independent coding, they convened to reconcile analyses, discussing discrepancies until full consensus was reached on a unified codebook. To objectively assess reliability prior to consensus, a formal inter-coder reliability check was performed on a randomly selected 10% sub-sample (50 comments). Cohen's Kappa (κ) was calculated for assignment to the three primary themes, yielding an average score of $\kappa = 0.82$, indicating substantial agreement. The final consensus-coded dataset comprised 612 coded references distributed across all developed themes. The distribution across the primary equity dimensions is summarized in Table 2. These figures represent distinct thematic units (coded references), not a simple count of comments, thereby accurately reflecting the analytical depth of the dataset.

Table 2. Distribution of Coded References Across Primary Equity Dimensions

Equity Dimension	Coded References	Percentage of Total Coded References
Opportunity	331	54%
Process	225	37%
Outcome	56	9%
Total	612	100%

4. RESULTS AND DISCUSSION

4.1 The Public Agenda's Focus on Higher Education Fairness

The societal discourse on higher education equity, as embodied by Shanhe University, is not co inconsistent with the academic establishment's standards for evaluatingquality universities, which are generally framed in terms of quality, distinctiveness, and efficiency. Over time, through policy propaganda, media direction, and the employment sector, the population has been led to believe that universities with key construction status are the true high-quality universities. The larger the number of spots, the lower the quality of education. This reasoning permeates the broader discourse and agenda, whereby the name of Shanhe University serves as a reference point. This is an ideology associated with the people's aspiration for more equitable distribution of quality educational resources, particularly with respect to opportunity, process, and outcome fairness.

4.2 Fair Education Access and External Opportunity Pursuit

In 2014, the State Council of China issued the Implementation Opinions on Deepening the Reform of the Examination and Enrollment System, which emphasized the need to improve the allocation of enrollment quotas to reduce regional disparities in access to higher education. Although growth in higher education enrollment over the past few years has generally reduced regional variation in admission chances, regional disparities in access to high-quality higher education resources have become increasingly entrenched and, in some cases, even more pronounced. Within the framework of universalisation, this entrenchment has become a central concern for equity in Chinese higher education. Consequently, the acquisition and equitable distribution of high-quality local education opportunities have become the most pressing issue in public discourse on higher education equity. These issues primarily manifest in two dimensions: exam application fairness and admission fairness (Table 2).

Regarding fairness in exam applications, most statements in the online discussion at Shanhe University adopt traditional key national projects as criteria for evaluating high-quality universities. Such talks highlight the limited number of high-quality universities accessible to students in the so-called Shandong Four Provinces (Shandong, Shanxi, Henan, and Hebei) and the high barriers to admission to these schools. Sentiment regarding the unequal distribution of higher education resources in the region is widespread. The example of the joke about the fact that among the provinces with a total population of more than 200 million, namely Shanxi, Hebei, and Henan, there is not even a single university that is of a high level is illustrative of the disproportions between the high population of these provinces and the lack of higher-level educational facilities. Regarding admission equity, there is considerable discussion of regional disparities in admission rates to high-quality universities, particularly those directly managed by the Ministry of Education. These inequalities are most evident in the benefits enjoyed by cities such as Beijing, Tianjin, and Shanghai, at the expense of central provinces like Henan. Such regional imbalance is perceived as a major contributor to unequal access to higher education.

Table 3. Distribution of Coded References Across Primary Equity Dimensions

Equity Dimension	Coded References	% of Total Coded References (N=612)	% of Comments Containing Theme (N=494)
Opportunity	331	54.10%	48.2% (238 comments)
Process	225	36.80%	32.6% (161 comments)
Outcome	56	9.20%	8.5% (42 comments)
Total	612	100%	

4.3 Process Equity: Needs for Personal Growth and Academic Experience

In light of the need to support strong efforts to deliver significant results, China adopted a differentiation policy at major universities as early as 1954. This policy was based on the designation of National Key Universities in the 1950s, evolving into the 211 Project and the 985 Project in the 1990s, and later into the Double First-Class initiative in the 21st century. Within these programs, the focus on educational materials has consistently centered on major universities, where the government has been a significant contributor, thereby directly influencing personal development and learning processes. The vast inequalities in university development have been a principal factor in the widespread calls for the establishment of the so-called Shanhe University. Regarding process equity, this problem is primarily comprised of funding distribution, faculty quality, and teaching conditions (see Table 3).

Regarding the allocation of funds, some members of society argue that the government provides non-key universities with less public financial support than it does to key universities, resulting in fewer scholarships, limited funding for day-to-day operations, and lower per capita budget allocations. Regarding faculty quality, most non-key universities are in short supply of high-level teaching staff. It is especially noticeable in the numbers of high-profile scholars, including academicians of the Chinese Academy of Sciences and Engineering, Changjiang Scholars, and excellent young faculty (in the form of National Science Fund for Distinguished Young Scholars, Youth Thousand Talents Plan, and Excellent Young Scientists Fund), that differ significantly across the levels of institutions. On the teaching quality side, others have also indicated that non-key universities are disadvantaged in terms of infrastructure, learning conditions, academic operations, and development facilities. Overall, the comparative drawbacks of non-key universities in financial provision, educational quality, and social prestige contribute to inequity, not only in access to higher education but also in the educational process.

Table 4. Public Discourse on Opportunity Fairness

Category	Coded Ref. Count	% of Comments (N=494)	Example User Comments
Exam & Admission Fairness	304	46.6% (230)	1. *“The admission rate for Tier-1 universities in Beijing is 30%, but in Henan it's only 8%. We're being filtered by our birthplace before we even take the exam.”*
			2. *“A score of 600+ might get you into a good university elsewhere, but in Henan it just means you have a chance to leave Henan.”*
			3. “We are the province that exports the most food, the most migrant workers, and now the most college students. When do we get to import some resources?”
			4. “If my score from Hebei could get me into a 985 school in Jiangsu, why are there zero 985s in my own province? The system is designed to push us out.”

Geographic Resource Concentration	27	4.9% (24)	1. “Look at a map of ‘Double First-Class’ universities. It’s a map of political power, not educational need.” 2. “All the best schools are in Beijing, Shanghai, Nanjing. For us, ‘Shanhe University’ is the only ‘first-class’ university we can imagine having.”

Table 4 shows public discourse on opportunity fairness, constituting over half of all coded data, reveals a profound and geographically rooted sense of injustice. Commenters do not merely cite disparate admission rates; they articulate a systemic critique in which geographical birthplace is perceived as a pre-determining filter, creating a forced "brain drain" from provinces such as Henan and Hebei. The rhetoric powerfully frames these regions as perpetual exporters of grain, migrant labor, and, now, top students, while denying them the high-status educational resources needed for self-sustaining development. The fictional "Shanhe University" thus emerges as a symbolic remedy for this structural inequity, representing a demand not just for fairer quotas, but for the very presence of elite institutions that would allow talent to thrive locally.

Table 5. Public Discourse on Process Fairness

Category	Coded Ref. Count	% of Comments (N=494)	Example User Comments
Teaching & Learning Conditions	187	28.1% (139)	1. “I visited a friend at a 985 university. Their library, labs, and even the dormitories felt like a different country compared to my school. We’re not even playing the same game.” 2. “The biggest inequality isn't the exam, it's what happens after. They get research projects and innovation grants; we get outdated textbooks and teachers who've given up.” 3. “In our university, good students fight to transfer out, and good teachers get poached. We’re left in a sinking ship. ‘Shanhe University’ is the lifeboat we dream of.” 4. “You can feel the atmosphere of decline. Students play games in class because they know their diploma from here won’t be respected anyway.” 5. “They talk about ‘Double First-Class’ investment. For us, that just means watching the gap between us and the top schools grow wider every year.”

Faculty & Institutional Prestige	38	6.1% (30)	1. “A ‘Changjiang Scholar’ would never come to our university. We don't have the platforms or the funding. So we’re taught by less experienced faculty, which starts another cycle of disadvantage.” 2. “The brand name of a university attracts everything: good teachers, partnerships, attention. Without it, you're invisible.”

Table 4 shows public commentary on process fairness articulates a visceral experience of educational quality as a tangible, spatial divide. Discourse moves beyond abstract policy to describe a lived environment of relative deprivation, where inferior infrastructure, diminished faculty prestige, and a pervasive "atmosphere of decline" are seen as direct consequences of systemic resource allocation. This perception fosters a vicious cycle: the lack of prestigious branding impedes talent retention, which further degrades the learning environment and student morale. The aspiration for "Shanhe University" thus represents a demand not merely for equal treatment, but for an educational experience that conveys dignity, competitive resources, and a sense of future possibility.

4.4 Result Fairness: The Social Competition Differences of Institutionalized Cultural Capital

The content of fairness in higher education outcomes primarily concerns two aspects: further education and employment. Expected returns on investment in higher education primarily influence the intention to pursue it. In China, the expected returns from higher education are primarily attributed to subsequent studies and career development. Compared with high-quality universities, which have strong talent output and employment opportunities, non-key universities in China perform poorly in these areas (Table 4).

Regarding fairness in further education, some commentators have raised concerns about the difficulties of advancing in conventional higher education, including the availability of reserved graduate places, the graduate admission rate, the quality of institutions offering further education, and the number of opportunities for readjustment and re-examination. Regarding employment fairness, many people believe that ordinary higher education fails to produce a good reproduction effect. Terms like "migrant workers" and "low-level workers" frequently appear in discussions about employment fairness. Furthermore, as the symbolic value of diplomas from high-quality universities continues to expand in the labour market, restrictions on civil service examinations, relatively low compensation, and a weak employment market have become key concerns. Overall, for students from disadvantaged groups in lower-tier universities, the uneven distribution of high-quality higher education opportunities exacerbates the negative impacts of class reproduction and school reproduction.

Table 6. Public Discourse on Outcome Fairness

Category	Coded Ref. Count	% of Comments (N=494)	Example User Comments
Employment & Career Discrimination	41	7.3% (36)	1. "The resume filter is real. HR told me they automatically sort applications by university tier. My degree from a 'Shanhe' province doesn't even get seen."
			2. "We're told to get an education to escape being migrant workers, but without a 985 diploma, that's exactly the kind of unstable work waiting for us."
			3. "Same job, three different salary grades based solely on whether your university was 985, 211, or 'other.' That's the definition of unfair outcome."
			4. "These provinces are treated as talent factories: we process the raw materials (students) and export the finished product (graduates) for others to use."
Further Education & Social Mobility	15	2.6% (13)	1. "Even for postgraduate exams, professors prefer students from famous undergraduate schools. Our 'first degree' shame follows us forever."
			2. "The path is closed: no good undergrad → no good master's → no good job. 'Shanhe University' is a fantasy of breaking that chain."
			3. "My senior worked so hard for the postgraduate exam, but lost the spot to someone with a lower score from a 211 university. The system protects its own."
			4. "Without a top-tier diploma, your highest ambition becomes a stable civil service job in your hometown. The dream of real academic or professional advancement is for others."

4.5 An Examination of the Roots Behind the Formation of the Public Agenda on Higher Education Fairness

The "Shanhe University" phenomenon, while originating from four specific provinces, has evolved into a powerful symbol that catalyzed a nationwide public agenda on higher education equity. This agenda transcends the grievances of a single region, providing a discursive platform for a broad spectrum of students and citizens to articulate deep-seated frustrations with systemic disparities. It represents not merely

a demand for quantitative access, which has largely been achieved through massification, but a profound public outcry against the qualitative stratification of opportunity. As one commenter framed it, "Shanhe University isn't just for four provinces; it's for everyone who feels the game is rigged from the start" (Bilibili Comment #53). The sustained discourse around this fictional institution is therefore a direct response to the perceived impossibility of accessing high-quality educational resources for many, rooted in the historical legacies of uneven policy investment, the labor market's devaluation of certain diplomas, and a deeply ingrained educational hierarchy. These structural seeds found fertile ground in the public consciousness, blossoming into the widespread debate analyzed in this study."

4.5.1 The expanding demand for quality education in the massification stage

The massification of higher education in China, marked by soaring gross enrollment rates and a national admission-to-application ratio reaching 85% in 2022, has fundamentally shifted public concern from mere access to the quality of that access. While more students than ever can attend university, the discourse reveals a widespread perception that the primary beneficiaries of this expansion are “local, mundane universities” (本地普通大学), intensifying competition for scarce elite resources. The “Shanhe University” phenomenon crystallizes this anxiety, serving as a symbolic “haven” (避风港) for aspirants from provinces historically starved of top-tier institutions. As one commenter poignantly noted, “We don’t just want a university to attend; we want a university that can give us a future. Right now, leaving our province seems like the only way to get that” (Bilibili Comment #147). This sentiment underscores that massification has amplified, rather than alleviated, the public’s demand for high-quality educational capital, with the four “Shanhe” provinces representing nearly 30% of the national Gaokao cohort but hosting zero Project 985 universities, epitomizing the perceived gap between quantitative opportunity and qualitative outcome.

4.5.2 Regional disparities in the allocation of educational resources

Public discourse powerfully frames equity not as an abstract principle but as a palpable injustice rooted in geographical resource stratification. The data is stark: Beijing, Jiangsu, and Shanghai, representing a tiny fraction of the national examinee population, possess a dominant share of Double First-Class universities, while the four “Shanhe” provinces collectively hold less than 3%. This imbalance is a constant refrain in the comments, translating policy statistics into lived grievance. “Jiangsu and Henan have the same number of colleges, but look at the difference in ‘first-class’ labels. It’s like we’re playing a game with different rulebooks” (Bilibili Comment #212). The inequity is perceived as systemic, exacerbated by funding models that favor centrally-administered elite universities and create a “vicious cycle” for local institutions. Commenters explicitly link this financial and reputational hierarchy to a degradation of their educational experience and prospects: “The best teachers get poached by eastern universities, our labs are outdated, and then employers still look down on our diploma. The system tells us we have opportunity, but the quality isn’t there” (Bilibili Comment #389). Thus, “Shanhe University” represents more than a fictional ideal; it is a direct critique of a spatialized political economy of education that concentrates symbolic and material capital in select metropolitan hubs, leaving populous regions in a state of permanent competitive disadvantage.

4.5.3 The Hidden Concern of Diploma Symbol Devaluation

The analysis reveals a significant public concern that the massification of higher education has led to credential inflation and the devaluation of diplomas from non-elite institutions. This theme emerged strongly from the coded data, with commenters directly linking geographic educational disparities to diminished labor-market value. This sentiment is not merely an abstract critique but is expressed through vivid personal analogies. One user starkly illustrated the perceived hierarchy, stating, "A bachelor's degree from my local university is like a discount coupon, while the same degree from a 985 school is a VIP pass. We took the same Gaokao, studied for four years, but our tickets to the future aren't worth the same" (Bilibili Comment #301).

This perceived devaluation drives a cycle of competitive anxiety, where further education is pursued less for knowledge than for reputational capital. The labor market is seen as a primary enforcer of this hierarchy. As another commenter argued, the problem is systemic: "Employers don't even look at our resumes; they see the university name and that's it. 'Shanhe University' is a joke, but it screams our real problem: a diploma from here [a local province] just doesn't have the same 'currency'" (Bilibili Comment #418). The discourse suggests that the expansion of local universities, while increasing access, has failed to counteract the intense stratification of symbolic capital. The fictional ambition for "Shanhe University" to rival Tsinghua and Peking University, therefore, is interpreted as a satirical critique of this entrenched system a desire not just for a university, but for the transformative social and economic "currency" that only a top-tier diploma is perceived to provide.

4.5.4 Hierarchical Stratification of Universities and Cognitive Solidification

The problem of solidification is common in both the hierarchical organization of universities and in public attitudes and societal norms. On the one hand, universities have been able to preserve their privileged positions due to the inertia of specific developmental policies and financial support from more developed provinces, resulting in little overall structural change. Conversely, rankings and marketing efforts have long been embedded in popular culture, shaping perceptions of truth, for example, the notion that high-quality higher education implies that universities offering such degrees are inherently superior to those offering vocational degrees. These assumptions have come to represent the larger debate on equity in higher education.

The "Double First-Class" campaign aims to achieve the goals of higher education and become a world-class institution. Although aspects such as the elimination of hierarchies, the introduction of competition, and innovation are put at the forefront, they are viewed more as instruments than as ends in themselves. Fundamentally, the Double First-Class program remains a focused development strategy aimed at enhancing elite institutions and champions in China's higher education sector. It does not mean a re-shuffling and re-branding of major universities and subjects. The traditional "985" and "211" universities, which have long been recognized for delivering elite education, continue to offer significant benefits under the new initiative.

4.6 Governance Pathways for Higher Education Based on Equity

The massification era of higher education aims to establish a new system that fosters equity, coordination, and equal development. Equity-based governance of higher education is the response to societal issues and forms the basis of determining the quality of higher education. To continually advance

equity in higher education, the optimal approach is to maximize educational governance by effectively distributing resources, improving quality, and expanding supply, ensuring coordinated and balanced development. According to the content focus, root cause and inherent needs of higher education equity within the context of the public discourse of the Shanhe University, the need to respond to social needs of equity in opportunities, procedures and outcomes must be answered by four dimensions, which are, adjusting the structural setup, increasing channels of recruitment, providing resource compensation, and ensuring helped-out exit opportunities.

4.6.1 Adjusting the structural layout and expanding the supply of quality educational opportunities

To address the spatial concentration of elite resources, public discourse emphasizes structural reforms to expand the supply of quality opportunities. Commenters advocate for a multi-pronged strategy: empowering local universities through targeted support, as one user argued, "We don't need another branch campus from the east; we need to build our own flagship universities that understand our local industries" (Bilibili Comment #332). This aligns with the policy concept of activating existing resources through mergers, upgrades, and relocations to improve efficiency and relevance. Furthermore, the call for stronger regional alliances reflects a desire to break down institutional silos, with a netizen noting, "If universities in our region could share more resources and faculty, the overall quality would rise for everyone. Right now, we're all competing for crumbs" (Bilibili Comment #278). Ultimately, the public vision involves leveraging the "radiation and spillover" of leading institutions through collaborative networks, thereby creating a more equitable and interconnected higher education ecosystem that moves beyond mere geographical expansion to genuine qualitative enhancement.

4.6.2 Public-Driven Policy Proposals for Structural Reform

The public discourse surrounding “Shanhe University” extends beyond identifying systemic problems to propose specific, actionable policy reforms, reflecting a sophisticated critique of higher education governance. Comments reveal strong support for structurally addressing geographical inequity through a needs-based overhaul of the provincial enrollment quota system, which is widely perceived as institutionalizing local bias (Fomba et al., 2023). As one user argued, “The enrollment quotas should match the number of students. Why should a province with millions of Gaokao candidates have fewer spots at a top university than a city with a fraction of the population? The math itself is unfair” (Bilibili Comment #187). This public demand aligns with scholarly calls for “precise assessments” to reallocate opportunity, including mechanisms like inter-provincial pairing assistance to facilitate the transfer of quality educational resources and promote talent repatriation to underserved regions (Shaturaev, 2021). Concurrently, commenters diagnose a vicious cycle of resource dependence plaguing local institutions and advocate direct policy and financial compensation for non-key universities. This sentiment is captured in the critique that “The policy just feeds the strong and starves the weak” (Bilibili Comment #402), underscoring a public understanding that equitable massification requires strategic investment in the institutions that serve the majority of students, not just the elite few (Ge et al., 2024).

Furthermore, the discourse logically links these structural inputs to labor market outcomes and proposes systemic interventions to ensure equity in graduate destinations. Commenters explicitly link the devaluation of non-elite diplomas to credentialist employer practices, demanding that policy address this endpoint of educational stratification. “We need policies that make companies value what we can do, not just where we’re from” (Bilibili Comment #455). This public appeal aligns with research on credential

inflation and calls for multi-stakeholder cooperation to dismantle the “symbolic alienation of diplomas.” Proposed measures include strengthening career guidance and practical training at local universities, revising “credential-based” hiring criteria in collaboration with social partners, and ensuring fairer access to postgraduate education to mitigate the talent drain from disadvantaged regions (Fomba et al., 2023; Deardorff & Jones, 2023). In essence, the online commentary constructs a holistic policy agenda that seeks to rebalance opportunity through quota reform, enhance quality through resource redistribution, and validate success through fairer outcome measures, thereby providing a ground-level blueprint for a more equitable higher education system.

5. DISCUSSION

This study has utilized the viral "Shanhe University" phenomenon as an empirical lens to analyze the complex public discourse on higher education equity in China. By examining online commentary, the findings move beyond traditional policy analysis to reveal the emotionally charged and structurally aware perspectives of the public, particularly younger, digitally engaged citizens. These insights offer a bottom-up conceptualization of equity that both confirms and enriches academic frameworks.

5.1 Public Conceptions of Equity and Governmental Response

A key finding of this research is the public’s holistic, interconnected understanding of equity, which aligns spontaneously with the scholarly dimensions of opportunity, process, and outcome. Crucially, public discourse perceives these not as separate issues but as an interlocking system of disadvantage. For instance, the lack of elite universities in a province (opportunity) is directly linked to concerns over inferior campus resources (process) and anticipated discrimination in the job market (outcome). This systemic public critique has not gone unanswered. The strong online sentiment catalyzed by "Shanhe University" prompted a notable governmental response to address these intertwined concerns. Specifically, the state has intensified efforts to expand quality educational opportunities in underserved regions through policies that promote the establishment of new university branches, strategic upgrades of local institutions, and enhanced interprovincial resource sharing (Dalton, 2018). This official reaction demonstrates a recognition of the public agenda and an attempt to mitigate the very opportunity and process inequities highlighted in the online discourse, marking a direct channel of influence from bottom-up public sentiment to top-down policy adaptation.

5.2 A Bottom-Up Conceptualization of Equity

The greatest invention of this work is that the bottom-up derivation of an equity framework has been obtained. Although educational equity has long been theorized in the academic literature (Cochran-Smith & Keefe, 2022), our results suggest that popular knowledge not only aligns with academic categories but also offers a detailed account of them. The information naturally led to a three-part framework of issues: opportunity, process, and outcome fairness, which validates the fact that societal understanding of equity is holistic, spanning from access to the final social value of a degree. What was not previously known, however, was the strong interconnection that the population creates between these dimensions. An example is the observation that Henan has few 985 universities (opportunity), which is immediately associated with fears of lower-quality laboratory conditions (process) and, in turn, further employment discrimination (outcome). This implies that equity is perceived by the public not as discrete policy issues, but as an interlocking system of disadvantage; an interconnected system that top-down analyses often fail to capture.

5.3 The Primacy of Quality and the Rejection of Vocational Tiers

Another alarming point is that quality equity receives excessive attention from the public, while the discussion of plain access is overshadowed. The massification of higher education in public discourse has given rise to a new, more insidious hierarchy. This leads the reader to recognize that the fundamental public complaint is not that there are insufficient numbers of university positions, but that there are inadequate-quality university positions, as determined by the state-approved hierarchy of Double First-Class, 985, and 211 positions. This observation is in opposition to certain policy discourses that glorify the increased access as the main success of massification (Trow, 1973).). In addition, the research also reveals an in-depth and apparently unbreakable mental solidification. Vocational colleges, in the form of the Double High Plan, are rarely discussed in the popular press as a viable alternative to quality education. This is a striking finding, revealing that national policies promoting vocational education have, to date, not significantly influenced public opinion. The iconic currency of an exclusive academic undergraduate degree remains the only one and poses an immense challenge to those who aspire to make the system more diverse and equitable.

5.4 Emotional Resonance and Symbolic Resistance

The present study reveals that sentiment towards equity in higher education is anything but a calm, cold-blooded, rational computation of resource allocation; it is an intricate blend of frustration, irony, and symbolic opposition. A notable example of this is the phenomenon of Shanhe University. The issue of regional imbalances has been mentioned in the past; however, we observe how society reacts: they create a virtual university together, and ironically, Hynes (2013) notes this. It is a unique form of political performance and satire that enables users to express a nuanced critique of systemic failure in a manner that commentary may not achieve. The fact that students are likened to low-wage workers, or that they are described as being bred to be, keeps them filtered by their place of birth, giving the commentary an emotional resonance that cannot be determined by quantitative research alone. This indicates that, for the affected people, the equity crisis is equally a profound psychological and identity-related crisis that contributes to a collective grievance, which will most likely dominate political and social outlooks in the years to come.

5.6 Nuanced Roots and Governance Implications

Our results on the underlying causes of widespread dissatisfaction complement and narrow the existing information compared with prior literature. We emphasize the importance of regional resource disparities and funding stratification (Schmidt, 2015). Nevertheless, our analysis offers an important added value, highlighting diploma devaluation as a dominant social anxiety that links massification to fears about the labor market. The population clearly grasps the common sense of economic signaling theory: the more graduates there are, the lower the value of a generic diploma; institutional prestige has become the key distinguishing feature (Rosenau, 1997). This paper demonstrates the manifestation of this abstract idea as a physical state of anxiety, especially among students in the provinces with non-key universities. As a result, the lines of governance proposed by the citizens are systemic in nature (Bornemann & Strassheim, 2019). They do not demand discreet corrective actions, but a concerted re-tuning of the entire higher education system, including its structure, layout, and enrolments, as well as its products. The high social demand for resource compensation at non-key universities, such as AUB, directly conflicts with the decades-old policy of assisting the strong, suggesting a growing interest in a more redistributive compensatory system of governance for higher education. This popular voice is a strong legitimacy argument for policymakers who would support such a change.

6. CONCLUSION

The paper, in its examination of the viral phenomenon surrounding Shanhe University, concludes that the massification process exacerbates the inequities inherent in the higher education system, and the public criticism, in this case, constitutes a sophisticated and profound critique of it. The results demonstrate that the mass concern is not the simplistic demand that would create more university places but the holistic and inter-relational demand that is concerned about fairness throughout the entire educational life cycle: the geographically thwarted opportunity to access elite institutions, the process of learning that is strictly stratified in terms of resources allocation in funding, faculty, and facilities, and the ultimate inequality in further education and employment where the diplomas of non-key universities are systematically undermined. The very story of Shanhe University is already a form of symbolic opposition, expressing a general anxiety about the depreciation of diplomas and a latent anger toward an intellectually entrenched university hierarchy, a situation that policies such as the Double First-Class initiative have yet to break through. This paper, thus, illustrates that good governance in this new age should be more than just incremental changes and more synergistic in nature which should expand the supply of quality opportunities via structural reforms, maximize the enrollment quotas to address regional bias, actively compensate non-key universities with specific policy, financial, and human resources and develop strong support systems that unhook individual potential and institutional prestige. The discussion around Shanhe University is much deeper than an internet meme; however, it is a verdict of the masses that a project that has been left half-complete was completed, and that the number of graduates produced will not measure the success of this massification, but by the quality and even-handedness of the opportunities they receive.

7. Limitations and Future Research

Although the present study offers a subtle view of the public opinion due to the inclusion of the case of the Shanhe University, the methodological decisions place certain restrictions constraining the study, which outline clear directions to be taken in the future research; in particular, the use of Bilibili information only (but offering a deep insight into one of the central demographic groups) imposes some limitations, which can restrict the generalization of the findings, since the viewpoints of other demographic categories using other social media platforms (WeChat, Weibo, Zhihu) may be more diverse, and the intentional excl Future research must therefore assume a multi-platform strategy to triangulate results and obtain a more realistic view of the national discourse and must also use a network analysis of the reply to threads in order to model the structure and important influencers in the online discussion whereas longitudinal monitoring of such sentiments may also provide insight into how public agendas on educational equity respond to particular policy announcements or to particular yearly events such as the gaokao.

Ethical Statement: The initial proposal of the study was reviewed by the Institutional Review Board of the Capital Normal University, Beijing, China, and approved for the research. The researchers adhered to all ethical guidelines throughout the study and the report.

Competing Interests: The author declares that this work has no competing interests.

Grant/Funding information: This study has not received any funding.

Declaration Statement of Generative AI: The authors of this work have not used any AI tool for the preparation of this manuscript.

REFERENCES

Ryan, L. V., Buchholtz, A. K., & Kolb, R. W. (2010). New directions in corporate governance and finance: Implications for business ethics research. Business Ethics Quarterly, 20(4), 673-694. https://doi.org/10.5840/beq201020442

Bornemann, B., & Strassheim, H. (2019). Governing time for sustainability: analyzing the temporal implications of sustainability governance. Sustainability Science, 14(4), 1001-1013.

Chen, Y., Hou, A. Y. C., & Huang, L. (2022). Development of distance education in Chinese higher education from the perspectives of accessibility, quality, and equity under COVID-19. *Asian Education and Development Studies*, *11*(2), 356-365. https://doi.org/10.1108/AEDS-05-2020-0118

Cochran-Smith, M., & Keefe, E. S. (2022). Strong equity: Repositioning teacher education for social change. *Teachers College Record*, *124*(3), 9-41. https://doi.org/10.1177/01614681221087304

Dabis, A., & Csáki, C. (2024). AI and ethics: Investigating the first policy responses of higher education institutions to the challenge of generative AI. *Humanities and Social Sciences Communications*, *11*(1), 1-13. https://doi.org/10.1057/s41599-024-03526-z

Dalton, R. J. (2018). *Citizen politics: Public opinion and political parties in advanced industrial democracies*. Cq Press.

Darazi, M. A., Khoso, A. K., & Mahesar, K. A. (2023). Investigating The Effects of ESL Teachers 'feedback on ESL Undergraduate Students 'level of Motivation, Academic Performance, And Satisfaction: Mediating Role of Students 'motivation. Pakistan Journal of Educational Research, 6(2). https://doi.org/10.52337/pjer.v6i2.807

Deardorff, D. K., & Jones, E. (2023). Intercultural competence as a core focus in international higher education. In *The handbook of international higher education* (pp. 223-241). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781003447863-15/intercultural-competence-core-focus-international-higher-education-darla-deardorff-elspeth-jones

Fomba, B. K., Talla, D. N. D. F., & Ningaye, P. (2023). Institutional quality and education quality in developing countries: Effects and transmission channels. *Journal of the Knowledge Economy*, *14*(1), 86-115. https://doi.org/10.1007/s13132-021-00869-9

Ge, Z. G., Zhao, Y., & Liu, Y. (2024). Exploring quality issues in academic master theses of education majors and corresponding countermeasures. *Journal of Pedagogical Research*, *8*(4), 30-47. https://doi.org/10.33902/JPR.202428171

Geng, Y. (2022). *An Interpretative Analysis of Systemic Educational Equity, Educational Policies, Educational Determinants, and Equity Leadership as They Affect Performance Excellence at Chinese Higher Education Institutions* (Doctoral dissertation, Delaware State University). Retrieved from; https://www.proquest.com/openview/03f490eae0bdde2d179810801d3e6294/1?pq-origsite=gscholar&cbl=18750&diss=y

Germain, E. (2022). Well-being and equity: A multi-disciplinary framework for rethinking education policy. *Peabody Journal of Education*, *97*(1), 6-17. https://doi.org/10.1080/0161956X.2022.2026716

Han, S. (2022). Experimental governance in China's higher education: stakeholder's interpretations, interactions and strategic actions. *Studies in Higher Education*, *47*(1), 13-25. https://doi.org/10.1080/03075079.2020.1725876

Hayhoe, R., Li, J., Lin, J., & Zha, Q. (2012). *Portraits of 21st Century Chinese Universities:: In the Move to Mass Higher Education* (Vol. 30). Springer Science & Business Media.

Huck, C., & Zhang, J. (2021). Effects of the COVID-19 pandemic on K-12 education: a systematic literature review. *New Waves-Educational Research and Development Journal*, *24*(1), 53-84. https://eric.ed.gov/?id=EJ1308731

Hynes, M. (2013). Reconceptualizing resistance: Sociology and the affective dimension of resistance. The British journal of sociology, 64(4), 559-577.

Katsamakas, E., Pavlov, O. V., & Saklad, R. (2024). Artificial intelligence and the transformation of higher education institutions: A systems approach. *Sustainability*, *16*(14), 6118. https://doi.org/10.3390/su16146118

Kawuryan, S. P., Sayuti, S. A., & Dwiningrum, S. I. A. (2021). Teachers Quality and Educational Equality Achievements in Indonesia. *International Journal of Instruction*, *14*(2), 811-830. https://eric.ed.gov/?id=EJ1291222

Khoso, A. K., Darazi, M. A., Mahesar, K. A., Memon, M. A., & Nawaz, F. (2022). The impact of ESL teachers' emotional intelligence on ESL Students academic engagement, reading and writing proficiency: mediating role of ESL student's motivation. Int. J. Early Childhood Spec. Educ, 14, 3267-3280. https://doi.org/10.9756/INT-JECSE/V14I1.393

Levinson, M., Geron, T., & Brighouse, H. (2022). Conceptions of educational equity. *Aera Open*, *8*, 23328584221121344. https://doi.org/10.1177/23328584221121344

Li, J., & Xue, E. (2022). Unpacking the policies, historical stages, and themes of the education equality for educational sustainable development: Evidence from China. *Sustainability*, *14*(17), 10522. https://doi.org/10.3390/su141710522

Li, L., & Ruppar, A. (2021). Conceptualizing teacher agency for inclusive education: A systematic and international review. *Teacher Education and Special Education*, *44*(1), 42-59. https://doi.org/10.1177/0888406420926976

Luo, J. (2024). A critical review of GenAI policies in higher education assessment: A call to reconsider the "originality" of students' work. *Assessment & Evaluation in Higher Education*, *49*(5), 651-664. https://doi.org/10.1080/02602938.2024.2309963

Marginson, S. (2016). The worldwide trend to high participation higher education: Dynamics of social stratification in inclusive systems. *Higher education*, *72*(4), 413-434. https://doi.org/10.1007/s10734-016-0016-x

Marginson, S., & Yang, L. (2022). Individual and collective outcomes of higher education: A comparison of Anglo-American and Chinese approaches. *Globalisation, Societies and Education*, *20*(1), 1-31. https://doi.org/10.1080/14767724.2021.1932436

Omri, N., & Aksoy, E. B. (2024). Rhythms of Resistance: The Cultural Dynamics and Emotional Resonance of Iranian Rap in the Iranian Soundscape. *OPUS Journal of Society Research*, *21*(1), 61-72. https://doi.org/10.26466/opusjsr.1392249

Robinson, K. J. (2023). Rodriguez at Fifty: Lessons Learned on the Road to a Right to a High-Quality Education for All Students. *Loy. U. Chi. LJ*, *55*, 343. https://heinonline.org/HOL/LandingPage?handle=hein.journals/luclj55&div=13&id=&page=

Rosenau, J. N. (1997). Global environmental governance: delicate balances, subtle nuances, and multiple challenges. In International governance on environmental issues (pp. 19-56). Dordrecht: Springer Netherlands.

Sallam, K., Mohamed, M., & Mohamed, A. W. (2023). Internet of Things (IoT) in supply chain management: challenges, opportunities, and best practices. *SMIJ*, *2*(2), 32. https://doi.org/10.61185/SMIJ

Schmidt, J. (2015). Intuitively neoliberal? Towards a critical understanding of resilience governance. European journal of international relations, 21(2), 402-426.

Shan, H., & Guo, S. (2014). Massification of Chinese higher education: Opportunities and challenges in a globalizing context. *A comparative analysis of higher education systems: Issues, challenges and dilemmas*, 9-23. Retrieved from; https://brill.com/display/book/9789462095335/BP000003.xml

Shaturaev, J. (2021). indonesia: superior policies and management for better education (Community development through Education). *Архив научных исследований*, *1*(1).

Tang, M., Ren, P., & Zhao, Z. (2024). Bridging the gap: The role of educational technology in promoting educational equity. *The Educational Review, USA*, *8*(8). http://dx.doi.org/10.26855/er.2024.08.012

Trow, M. (1973). Problems in the transition from elite to mass higher education. https://eric.ed.gov/?id=ED091983

Xu, B. (2021). From massification towards post-massification: Policy and governance of higher education in China. *International Journal of Chinese Education*, *10*(3), 22125868211046032. https://doi.org/10.1177/22125868211046032

Xue, E., & Li, J. (2022). Study on the education governance system to deal with major public crisis in China. *Educational Philosophy and Theory*, *54*(14), 2434-2445. https://doi.org/10.1080/00131857.2021.1965877

Xuyang, H., & Shanming, S. (2018). The governance of higher education according to the marxist view of fairness. *Social Sciences in China*, *39*(3), 171-186. https://doi.org/10.1080/02529203.2018.1483112

Yang, L. (2021). Social equity and equity in higher education: A comparison of the liberal Anglo-American and Chinese political cultures. *International Journal of Educational Development*, *84*, 102403. https://doi.org/10.1016/j.ijedudev.2021.102403

Zancajo, A., Verger, A., & Bolea, P. (2022). Digitalization and beyond: the effects of Covid-19 on post-pandemic educational policy and delivery in Europe. *Policy and Society*, *41*(1), 111-128. https://doi.org/10.1093/polsoc/puab016

Zha, Q. (2020). Equality and equity in Chinese higher education in the post-massification era: An analysis based on Chinese scholarly literature. *The China Quarterly*, *244*, 1056-1077. https://doi.org/10.1017/S0305741020001241

Zhong, J., & Zheng, Y. (2022, July). Empowering future education: Learning in the Edu-Metaverse. In *2022 International Symposium on Educational Technology (ISET)* (pp. 292-295). IEEE. https://doi.org https://doi.org/10.1109/ISET55194.2022.00068

Zhong, Y. (2018). *Is Freire's critical pedagogy applicable to China's higher education? a philosophical examination* (Doctoral dissertation, University of British Columbia). http://hdl.handle.net/2429/66318

JADHUR

Vol.4, Issue.3, (2025)

Journal of Advances in Humanities Research

https://doi.org/10.56868/jadhur.v4i3.310

Artificial Intelligence Literacy and Ethical Digital Governance: Pathways of Multi-Stakeholder Collaboration and Value Alignment

Xu Hao[1], Nurul Liyana Mohd Kamil [1*]

1 Faculty of Business and Economics, Universiti Malaya, 50603 Kuala Lumpur, Malaysia

Article Information

Article Type: Review Article

Dates:

Received: 01 May 2025

Revised: 03 August 2025

Accepted: 09 August 2025

Available online: 22 September 2025

Corresponding Author: Narul Liyana Mohd Kamil

Email: nurulliyana@um.edu.my

ABSTRACT

The need for ethical governance based on responsibility, transparency, and equity has increased due to the rapid integration of artificial intelligence (AI) in business, education, and governance. This study explores the relationship between value alignment, multi-stakeholder engagement, and AI literacy as key pillars of moral digital governance. Based on public value theory and stakeholder theory, the study employs a qualitative interpretive design that combines comparative case analysis and grounded theory. Between 2019 and 2025, data on governance systems in the Global South, China, and the European Union were gathered from academic publications, institutional reports, and international policy instruments. According to research, value alignment serves as the normative outcome that ensures consistency between AI systems and social ethics, collaboration serves as the working tool that transforms literacy into ethical governance, and AI literacy serves as the cognitive foundation for evidence-based engagement. While fragmented or top-down models of governance fell short in terms of accountability and inclusivity, regions such as the EU and China that integrated AI literacy and participatory governance into their institutional framework demonstrated greater ethical coherence and public trust. The study offers a strong theoretical and applied framework that demonstrates how literacy-based collaboration results in moral and value-based governance. The study recommended trust-based and participatory discussion instead of regulatory compliance to AI governance. In addition the study recommended that the stakeholders such as educators, policymakers, and technology develop the models that connect the most important components that are, literacy, cooperation, and value alignment to develop human-centered and socially acceptable AI technologies in the emerging digital era.

Keywords: Artificial Intelligence Literacy, Ethical Digital Governance, Multi Stakeholder Collaboration, Value Alignment, Stakeholder Theory, Public Value Theory, Responsible AI

1. INTRODUCTION

At an advanced stage of development, Artificial Intelligence (AI) is influencing every aspect of life, including global information sharing, business, international relations, corporations, social structure, and governance. The adoption of AI technologies, such as Deepseek, ChatGPT, and Sora, significantly influences the generation of information across health, education, news, and justice (Hussein et al., 2025). On the one hand, AI is helping generate information, while on the other hand, these technologies have raised serious ethical issues. Some of the major challenges posed by these advanced technologies include information distortion, algorithmic bias, and data misuse (Hidayat and Muis, 2025). AI literacy is not just the ability to master AI technology, but also the ability of citizens to participate in governance, criticize algorithmic power, and provide ethical feedback in a digital society. It is a cognitive, critical, and practical "socially embedded literacy" system (Zawacki-Richter et al., 2019; Chu & Dong, 2024).

In addition, with the widespread application of AI systems in highly sensitive social scenarios, the contradiction between AI systems and human value systems has become increasingly intense, which has aroused great attention from academic and policy circles to the issue of "value alignment" (Russell, S., & Norvig, 2021 . Whereas existing studies have focused on AI literacy at the individual or organizational level (Mills et al., 2024), fewer have investigated how it intersects with multi-stakeholder collaboration, especially in reconciling disparate values across sectors (Güngör, 2020; Prikshat et al., 2022). Current literature either addresses technical capabilities in AI literacy or mechanisms for governance of collaboration (MacDonald et al., 2019), but does so infrequently by interweaving the two views. By defining the gap, the current research highlighted the necessity of a more integrated framework that captures the interaction between AI literacy and collaboration processes to ensure value alignment.

It is challenging for a single leading figure to create an integrated policy when the governance process lacks transparency, which limits understanding, accountability, and trust in decision-making. In order to overcome these issues, the developers, administrators, and policy makers are focusing on coordination among various stakeholders to develop such systems and frameworks that minimize these ethical concerns and produce transparency and accountability (Irawati et al., 2023; Birdayanthi et al., 2025). This collaboration will not only support the distribution of resources and responsibilities but also promote social integration and ethical standards.

The existing literature has focused on AI literacy at the individual or organization level, fewer have investigated how it meets with multi-stakeholder collaboration, especially in reconciling disparate values across sectors (Jiang et al., 2024; Hingle, A., & Johri, 2025). To fill these gaps the current research highlighted the necessity of a more integrated framework that captures the interaction between AI literacy and collaboration processes to ensure value alignment.

1.1 Conceptual definition of the key terms

Artificial intelligence (AI) literacy can be defined as the knowledge, skills, and attitudes that enable people to understand, engage with, and critically assess AI technologies in daily life. It involves understanding the workings of AI systems, their potential advantages and drawbacks, and the technology's social and ethical consequences (Miao & Holmes, 2023). Ethical digital governance entails setting principles and policies, and frameworks that are used to handle digital technologies and data in a

responsible, transparent and fair manner. It focuses on responsibilities, justice, privacy, and inclusivity of both the design, implementation, and management of digital systems (Floridi and Cowls, 2019). A multi-stakeholder approach involves bringing together and involving various actors in decision-making, including governments, private-sector organizations, civil society, academia, and international institutions. This would contribute to the development of shared accountability and joint problem-solving, particularly in global governance and technology (Pauwels, 2020; United Nations, 2022). The value alignment is the mechanism of making artificial intelligence behave in accordance with human values, ethical standards, and social objectives. It entails creating AI algorithms to achieve a human intent in machine goals to avoid inadvertent injury or bias (Jobin et al., 2019).

1.2 Theoretical Framework

The promotion of transparency, fairness, justice and accountability is one of the urgent needs in the advanced technological era. The present study aims to address this need by integrating artificial intelligence (AI) with stakeholder collaboration to promote governance and value alignment. The study is grounded on the two most essential theories, Public Value theory and Stakeholder theory. Together, these theories highlighted the importance of coordination and collaboration in promoting social accountability and improving administrative processes and governance.

According to stakeholder theory, one of the primary responsibilities of governance systems and organizations is to foster moral obligations by promoting an environment grounded in moral standards for all stakeholders engaged in the virtual ecosystem (Donaldson & Preston, 1995). In the domain of AI supremacy, it interprets administration as a conciliatory process involving multiple actors, where legality emerges from the considered contributions of diverse actors (Batool et al., 2025). Attaining moral supremacy cannot depend on individual governance; it must arise from shared dialogue, confidence-building, and collective responsibility across official and social structures (Kraus et al., 2021). AI Literacy plays a key role as a promoter, boosting investors' ability to participate efficiently in these developments. Investors who are well-educated in both the practical and ethical aspects of AI can identify machine learning biases, review administrative systems, and advocate for transparency and accountability (Ng, 2021). While a lower level of AI literacy creates problems of data disorder and integrity. Thus, Digital Competence is vital for bringing investors to a mutual consideration, empowering informed discussion, and linking individual consideration with collective moral action. Stakeholder Theory, therefore, provides the operational facet of the outline explaining who is associated and how these communications govern moral consequences (Bridoux & Stoelhorst, 2022; Pies, 2023).

Public Value Theory (Moore, 2013; Sabatier and Weible, 2014) complements a regulatory dimension of power, highlighting the need to improve social welfare rather than merely focusing on functional productivity. The theory pointed out that legitimacy arises when the governance system produces results in accordance with universal moral standards and principles such as impartiality, inclusiveness, and justice (Bozeman, 2007). In the AI field, the personal liberties, equality, and community confidence that emphasize clarity, comprehensibility, and hands-on support are built on moral virtual supremacy in virtual spheres (Floridi et al., 2018).

This collaboration among academia, industry, developers, organisations, and the general public will encourage negotiation among these stakeholders and stabilise conflicting standards such as competence vs confidentiality or modernization versus the law (Winfield and Jirotka, 2018). These theories pointed out that stakeholder integration is the key to achieving ethical virtual administration. Also, these theories claimed that operational involvement, virtual competencies, information sharing, moral obligations and trust building among stakeholders is possible by teamwork, promoting universal laws and principles, and having public conversations (Novelli, et al., 2023; Jobin et al., 2019).

Eventually, the outline exemplifies moral virtual administration as a cooperative co-production method based on collective ethical vision and supportive systems. By merging the inclusivity focus of Stakeholder Theory with the legitimacy and public good emphasis of Public Value Theory, this research situates AI governance within a framework of collective intelligence, ethical coherence, and institutional trust, all of which are essential for promoting responsible AI in today's digital landscape.

1.3 Research questions

This study focuses on two key factors in the governance of artificial intelligence: "artificial intelligence literacy" and "digital ethical governance." To enhance clarity and applicability, the analysis has been narrowed down to examine the interactions within multi-stakeholder collaborations and the methods for aligning values. With this emphasis, the research explores the following practical questions:

- How are targeted efforts (e.g., in-workplace training or community education programs) to make the public AI literate sufficient for effective participation in governance mechanisms?
- How do different governance actors' policymakers, technology developers, and civil society collaborate to conceive and implement ethics-based digital governance mechanisms in real institutional contexts?
- How are technology systems structured such that value alignment is achieved in practice, particularly in the face of competing priorities in social, economic, and cultural spheres?

1.4 Research purpose and significance

The primary objective of this study is to examine how Artificial Intelligence (AI) literacy contributes to the development of ethical digital governance through the mediating pathways of multi-stakeholder collaboration and value alignment. Specifically, the study aims to (1) conceptualize AI literacy as both a technical and ethical competency essential for responsible participation in governance processes; (2) analyze the role of multi-stakeholder collaboration in translating literacy into inclusive and accountable governance practices; and (3) explore how value alignment serves as the normative mechanism ensuring coherence between technological innovation and societal ethics. By integrating Stakeholder Theory and Public Value Theory, the study seeks to develop a comprehensive framework that explains how informed participation and shared moral reasoning can strengthen the legitimacy of AI governance.

The significance of this study lies in its theoretical and practical contributions to the evolving discourse on responsible AI governance. The study is important because it bridges intellectual considerations with public and organizational integrity to overcome the gap between academic studies and administrative

learning. Moreover, the study recommended knowledge-based participation for policy makers, educators, and online influencers to foster moral and social responsibility among stakeholders and the public engaged in the virtual ecosystem. Hence, this study makes a significant contribution to developing such administrative frameworks that are technically sound and grounded in universal moral standards.

2. LITERATURE REVIEW

On one hand, the growing AI technologies have transformed the mode of information processing, while on the other hand, they have introduced many complex ethical issues. Scholars and policymakers have come to recognize the importance of the public in interpreting, engaging in critical thinking, and making ethical decisions when dealing with AI. Therefore, the concept of artificial intelligence literacy has gained new theoretical and governance aspects. At the same time, the moral unrest of AI systems also triggered the recognition of value alignment as a crucial element of AI ethical management. AI literacy, ethical online governance, and value alignment are interdependent and inseparable. Such constructs are shown to be mutually influential on the standardization, legitimacy, and acceptance of the use of technology in society (Long and Magerko, 2020; Su et al. 2023). AI literacy enhances an individual's ability to comprehend and interact critically with intelligent systems, and ethical digital governance helps ensure that such technologies are developed and governed in accordance with the principles of accountability, fairness, and transparency (Tahaei et al., 2023). Moreover, value alignment is considered the ethical compass of the human-AI relationship, which dictates how societies can govern technological innovation and respond to the challenges posed by new organizations and ethical questions. Together, these dimensions will decide how future societies will strike a balance between innovation and ethics in the digital era (Kim et al., 2021).

Therefore, this section will conduct a systematic literature review and comprehensive analysis on the connotation reconstruction of AI literacy, the types and causes of digital ethical risks, global governance models and value alignment theory. By doing this, the research aimed to establish the theoretical basis, foundation, and problem context for the current research.

2.1 Connotation reconstruction and governance significance of artificial intelligence literacy

Artificial intelligence literacy is a compound capability structure that encompasses a basic understanding of AI principles, ethical judgments about AI risks, critical thinking about platform behavior, and practical ability to interact and collaborate with AI systems (Ng et al., 2021). Long and Magerko (2020) define it as "the set of capabilities that enable individuals to understand, use, critique, and participate in algorithmic systems in an AI-dominated society." scholars like Zou and Schiebinger (2018) and Crawford and Calo (2016) posit that it should further involve ethical justification, critical consciousness, and socio-political comprehension, tracing academic debates regarding its scope and emphasis.

Zawacki-Richter et al. (2019) indicate that AI literacy includes four primary components: technical knowledge, data literacy, ethical literacy, and social engagement. These ideas are strongly supported, but there is still debate about how they can be applied in other educational systems. As a case in point, Global North nations, including the UK and Singapore, have incorporated AI literacy into their education systems and emphasize systematic inclusion and exposure at a young age. Instead, Iran, India, and the Global South are interested in community-based AI training to address digital inequities (Marzdar, 2025). This comparison suggests the various geopolitical and socio-economic factors that affect the various strategies, meaning that AI literacy is founded on the status of development and culture rather than international

standards. As an individual, the AI illiteracy could often leave the user with no insight into how the system of algorithms operates, the harm of such technology, and any other prejudices that are prone to the technology. This lack of knowledge could lead to an accidental increase in technical risks (Wu et al., 2025). As the analysis of generative artificial intelligence literacy indicates, AI literacy must not merely comprise a basic understanding of algorithms and systems, but also encompass the understanding of misinformation detection processes, ethical value judgment, and risk tracking on platforms and in action (Liu et al., 2025). This is a full set of skills necessary for society's involvement in the decision-making process for artificial intelligence. AI literacy in China can be considered one of the most prominent qualities of humanistic embeddedness, integrating technological knowledge, ethical implications, and social responsibility. As the influence of generative artificial intelligence on education, justice, governance, and culture continues to grow, it is no longer solely an educational concern but also a significant aspect of social governance. This transformation reveals the importance of ethics and institutional framework in aiding effective decision-making, collaboration amongst stakeholders and congruence between the technological development and social ideals. China's strategy may serve as an excellent example of how AI can be effectively implemented in society (Zhu, 2024). AI literacy serves as an agent of ethical digital governance and value-based coordination among actors, directly aligning with the overall aims of this paper.

Risk prevention mechanism: AI awareness enables individuals to be conscious and decompose the ethical risk of algorithmic bias, deepfakes, and data breaches. According to Alharbi. (2021), this knowledge infuses responsible, informed awareness that must be in place to address these threats and prevent mess in online communication systems.

Participatory co-governance capacity: Participating in digital ethics conversations, expressing concerns, and engaging in governance processes are ways for citizens to develop a thoughtful understanding of AI and articulate their needs. Sadat (2025) adds that such capability is what turns the public not into passive technology consumers, but a source of active engagement in the creation of ethical norms and thus the foundation of cooperation between different stakeholders in the field of digital governance

Practical background on value alignment: AI literacy can help those who use AI establish a collective understanding of what is and is not appropriate and ethical in its use by fostering continuous dialogue and contemplation. Karimov and Saarela (2025) argue that this common moral thinking establishes social consensus, offering a feasible basis for value congruence between AI systems and societal governance objectives. In the global tide of "responsible AI" governance, artificial intelligence literacy is more than an individual capability; it encompasses the entwined capabilities of nations, platforms, and governance systems. Individual literacy directs responsible use and ethical awareness, whereas national and platform capacities institutionalize these values through policy and design. China's New Generation AI Development Plan, for example, demonstrates how concerted policy frameworks integrate personal, institutional, and systemic literacy to promote ethical and responsible AI governance (State Council Information Office of the People’s Republic of China, 2025).

2.2 Analysis of the types and mechanisms of digital ethical risks

While generative artificial intelligence breaks the boundaries of traditional technologies, it also brings unprecedented ethical risks. These risks are not only reflected in privacy violations and misinformation at the individual level, but also point to the erosion of governance mechanisms, the erosion of value bases, and the unequal reproduction of social structures. These risks stem from the compound effect of embedding

platform logic, social structure, and governance gaps rather than the "out of control" nature of the technology itself (Beer, 2017; Selbst et al., 2019). This section examines the main ethical risk categories and formation mechanisms of AI systems from three angles. (1) the technological dimension which deals with algorithmic bias, data quality and transparency of the system; (2) the human or social dimension, which deals with how people use it, whether they are dependent on a system or not and the impact the system has on society; and (3) the institutional or governance dimension, which deals with regulatory framework, accountability and ethical supervision.

The breakthrough in generative AI's information-generating ability makes it not only a tool for content production but also a "weaver" of information cognition. Large model systems such as ChatGPT and Sora can generate multiple rounds of complex sentences, images, audio & video, but their "language simulation" does not imply "fact truth" (Bommasani et al., 2022; Ji et al., 2023). In the absence of reliable data and with biased training datasets, large models often generate hallucinatory content that is "plausible but wrong" (generative AI systems like big language models generate output that is superficially coherent, fluent, and factually plausible, but incorrect, made-up, or not based on actual data) (Islam Tonmoy et al., 2024).

According to Rawte et al.(2023) lack of AI literacy can amplify the "technological hallucination" (The phrase technological hallucination is used to describe a phenomenon wherein artificial intelligence systems, large language models (LLMs) and generative AI, generate as output text or other such things that are grammatically correct and sound plausible but factually in error or wholly made up) effect, leading users to place excessive trust in AI systems, mistaking them for neutrality and objectivity, which in turn weakens their ability to recognize platform manipulation and commercial manipulation (Romanishyn, 2025). AI literacy serves as a link between public decision-making in democracies and ethical consensus, as well as the capacity to comprehend technology. Hristovska (2023) has noted that the issue of information disorder surrounding the dissemination of AI-generated content has not been adequately addressed by the platform governance mechanisms currently in operation.

On websites like YouTube and TikTok, for instance, recommendation algorithms customize AI-generated content to users' emotional and behavioral preferences. Such algorithmic personalization, sometimes referred to as the "emotion algorithm," intensifies information cocoons and echo chambers where users are continuously exposed to agreeable content. As a result, this weakens critical thinking and fosters a "cognitive distortion field" that spreads prejudice and false information online.

Furthermore, the issue of misinformation is not merely a "corpus problem (biased data)"; rather, it is a manifestation of the intricate game between platform responsibility, user capabilities, and algorithm induction since AI systems are used in high-risk contexts like education, government affairs, and healthcare (Saeidnia et al., 2025).

In order to improve information governance capabilities, it is necessary to implement a systematic "knowledge transparency" (openness of AI systems with regard to revealing how information is being produced, namely data sources, algorithms, and reasoning processes) and "semantic verification" (checking AI outputs to know about the accuracy and actual consistency) mechanism based on the improvement of AI literacy. According to Rajkomar et al. (2019), extensive clinical decision support systems such as Google's DeepMind Health demonstrated that their lack of semantic verification, that is, confirming that AI-generated data is in tune with actual-world meaning and context, led to patient information being misinterpreted even when AI models had high predictive accuracy. This highlighted how crucial it is for

users to be AI-literate and have knowledge transparency (a clear understanding of how AI produces outputs) to comprehend and properly regulate such systems.

High-frequency data collection, analysis, and modelling are essential components of AI systems, and the underlying logic of these systems is defined by "implicit intrusion" (Klenk, 2023). The platform trains the model using user clicks, browses, and inputs, and its algorithmic behaviour occurs in the absence of clear consent procedures and legal authorization. For example, large models construct "digital portraits" by continuously tracking and simulating user behavior, which, in turn, affects their decision-making, thereby creating a form of hidden manipulation. As Milano et al. (2020) point out, AI-driven personalized recommender systems tend to shape user behavior rather than serve user preferences, given information asymmetry and algorithmic inexplicability.

Prajescu and Confalonieri (2025) noted that the introduction of AI-assisted trial technology into the judicial system has highlighted the importance of data collection, model training, legitimacy, and interpretability. Technical intervention may alter adjudication logic, erode judges' subjectivity, and create the illusion of "quasi-personality" and "quasi-judgment." This describes how computer programs can appear to imitate the way people think and judge right from wrong. This creates the mistaken belief that the machines possess human-like decision-making power or a personal capacity for judicial judgment (Baum, 2020). In highly sensitive fields such as health care, education, and finance, the absence of data governance mechanisms may pose serious ethical and security risks. To address ethical and security issues, China employed an AI-aided sentencing system that uses historical case data to produce penalty recommendations. Although these systems make the courts more efficient and consistent, research indicates that judges may over-rely on algorithmic outputs, thereby undermining their capacity to reason and make ethical decisions independently (Socol et al., 2024). This example illustrates the general concern that without strong data governance and ethical controls, AI systems have the potential to transform professional judgment, increasing risks in other high-stakes domains like healthcare and finance.

However, due to the triple identity of platform companies in the AI ecosystem as technology developers, service operators and data controllers, the boundaries of their responsibilities are blurred, and the pressure of governance is shifting, resulting in the weakening of public supervision capabilities. This phenomenon of "platform responsibility hollowing out" has become one of the key obstacles to AI ethical risk management. This dilution of accountability is a significant hurdle for effective AI ethical risk management, as it undermines accountability within the AI governance framework (Novelli, 2023).

In social-scenario applications, AI is not a neutral technical tool but a "value container" for social construction.

The “value container” term emphasizes that artificial intelligence-based computer systems are not merely impartial tools; instead, they are designed to replicate and incorporate existing moral, social, and cultural principles. It has been based on the Social Construction of Technology (SCOT) perspective, which argues that both the institutional environment, social norms, and human decisions ultimately determine how any technology is created and used (Pinch and Bijker, 1984). All training data, model architecture, and usage contexts are historically, culturally, and structurally grounded; such structures are often not visible to the general population (Foka et al., 2025). As numerous studies have demonstrated, AI systems are racially, gender, and class-biased in scenarios such as image recognition, credit approval, and hiring. Such biases are due to either an opaque model design or an Imbalanced training sample (Obermeyer et al., 2019;

Mehrabi et al., 2021). AI literacy has become an essential competency in the digital era because it enables individuals to understand, assess, and practice with AI technologies responsibly. This is a necessary skill because AI is increasingly influencing governance, employment, and education. Moreover, as the digital divide is driven by low levels of digital access and uneven technological exposure, it is essential to educate individuals in AI literacy to bridge the divide and ensure the inclusion of people in AI-driven economies in developing countries (UNESCO, 2023).

The model of AI literacy in developing countries proposed by Kathala et al. (2025) highlights the significant roles of resource scarcity, cultural peculiarities, and differences in educational systems in promoting AI literacy. Beyond providing a systematic approach to developing local educational programs and policy modifications, their study suggests that, to advance AI education, policymakers should integrate social justice and technical skills.

2.3 Comparing global models and creating multi-agent collaborative frameworks

As AI is a global technology, the issues related to its governance and ethical integrity are global in scope. Priorities related to these concerns vary. For instance, according to the OECD Principles and the European AI Act, the key issues associated with AI technology are accountability and transparency, whereas in Asian and Global South approaches, inclusivity and limited understanding and knowledge are the primary concerns (UNESCO, 2024).

To ensure a human-centered approach, the Organisation for Economic Co-operation and Development AI Principles (2019) set key ethical standards for global governance. These moral standards are based on five pillars: inclusive growth, human-centered values and fairness, transparency, clarity, accuracy, and accountability. The United Nations and G20 followed these five principles as cross-cutting themes throughout their initiatives (OECD, 2019; Veale et al., 2021).

Similarly, the European Union Artificial Intelligence Act (2022) legally operationalizes ethics by requiring compliance with these ethical standards. The EU AI Act categorizes AI systems into two major categories: one based on risk, and the other focused on mechanisms to promote transparency obligations and human oversight requirements (Consilium, 2024; Ebers, 2025).

The United States (US) promotes innovation through an Open-standard policy. Hence, there are variations in policies and principles related to AI technology, as highlighted by the discussion above. As the US promotes a fragmented, open-standard approach, the EU emphasizes the risks involved, and Asian and Global South nations favor a development-oriented model, focusing on AI literacy (Robles & Mallinson, 2023).

By emphasizing the integration of local social norms and capacity building, these strategies diverge from the West's focus on compliance or innovation and ultimately highlight context-sensitive ethics.

2.4 Regional-wise comparative overview of AI governance approaches

Table 1 presents a comparative analysis of the major models of AI governance, evaluated with respect to regulatory structure, policy instruments, operational characteristics, and principles. This multidimensional comparative analysis provides a clear understanding of the impact of various institutional traditions on the regulatory policies for addressing common moral issues such as accountability,

transparency, human rights, and distributive justice. The comparison highlights three main AI governance models (1) the precautionary, law-based paradigm of AI governance in the EU; (2) the market-based, decentralized paradigm of AI governance in the US; and (3) the development-oriented, capacity-building paradigm of AI governance in Asia and the Global South, with its emphasis on inclusivity and digital literacy.

Table 1. Comparative Overview of AI Governance Approaches

Region	Governance Approach	Key Characteristics	Normative Orientation / Regulatory Logic
European Union	Comprehensive regulation (EU AI Act)	Risk-based classification, binding legal obligations, conformity assessment, strong oversight	Precautionary, compliance-driven, rights-protective governance
United States	Fragmented policy model	Open standards, innovation-focused, sector-specific regulation, decentralized enforcement	Market-oriented, innovation-driven, flexible regulatory governance
Asia & Global South	Context-sensitive, development-oriented approach	Inclusivity, cultural sensitivity, AI literacy, institutional capacity building	Developmental, distributive, and capacity-building governance

Table 1 illustrates that the European Union institutionalizes a precautionary paradigm of governance, based on the enforcement of legal responsibility and the protection of fundamental rights. The United States, on the contrary, has a more regulatory stance whereby technology competitiveness and adaptability of the market are favored, leading to a decentralized and relatively loose structure. Meanwhile, developmental equity, social preparedness, and situational ethics are anticipated in governance across Asia and the Global South, and both regions prioritize long-term capacity building over short-term regulatory inflexibility.

Such an organized comparison shows that AI governance variance is not just regulatory disparity but also an indication of more normative commitments and institutional path dependencies. The results thus support the conclusion that the multi-agent collaborative governance design can incorporate compliance protection, innovation promotion, and an inclusive growth agenda within a globally integrated yet localized framework.

2.5 Country-Level Comparative Analysis of AI Governance

Table 2 presents a country-by-country comparison of AI governance structures in China, the European Union, and the United States, explicitly including implementation. The integration of governance

strategies, the nature of regulation, and the system's constraints enables a more nuanced evaluation of the merits and flaws of each model using the table.

The analysis identifies three AI governance models that include the multi-regulatory, state-based model in China, the precautionary, legally binding, and risk-based model in the EU, with a focus on rights and oversight, and the decentralized, innovation-oriented model in the U.S., focusing on flexibility and market-oriented guidance.

The added column of Challenges highlights the fact that the effectiveness of governance relies not merely on the design of regulations but also on the extent of coordination between institutions, and on their enforcement capabilities and coherence, bringing the analysis off the normative comparison to the realities of real-world implementation.

Table 2. Country-Level Comparative Analysis of AI Governance

Country/Region	Governance Approach	Key Characteristics	Challenges
China	Multi-regulation	Legal, ethical, and political guidance; emphasis on fairness and social ethics; strong state coordination	Fragmented administrative responsibilities; limited public participation; uneven enforcement
European Union	Risk-based regulation (EU AI Act)	Binding legal obligations; structured risk classification; transparency and human oversight requirements	Complex compliance procedures; regulatory burden on SMEs; slower adaptation to rapidly evolving AI technologies
United States	Innovation-focused, fragmented	Open standards; encouragement of R&D; sector-specific and flexible regulatory guidance	Limited federal coordination; gaps in accountability mechanisms; uneven oversight across sectors

Table 2 highlighted that each country has a different institutional logic and relations between the state and the market. The system has a high degree of central coordination, but implementation inconsistency and lack of transparency through participation, which is evident in China. The EU framework is most formalized and rights-protective, but too complex to provide compliance burdens and to adapt fast to changing technologies. The U.S. model promotes innovation, responsiveness, and leadership in the private sector but suffers from coordination failures and accountability lapses due to regulatory fragmentation.

Both Table 1 and Table 2 provide complementary comparative information. Table 1 reveals general regional governance paradigms and normative orientations, whereas Table 2 discusses national-level regulatory frameworks and problems of regulatory implementation. Together, these tables illustrate how general principles of governance are implemented in concrete institutional practices and note not only similar ethical purposes but also discrepancies in the application of regulations. This combined analysis confirms that there is no unilateral balance involving regulatory certainty, innovation flexibility and participatory legitimacy. This analogy underscores the necessity of a multi-agent collaborative governance model that incorporates legal enforceability, normative diversity, and technological flexibility and involves all stakeholders across jurisdictional levels.

Southeast Asian countries, such as Indonesia and Malaysia, focus on coordinating among stakeholders, including educational institutions, communities, and civil society, to promote AI literacy and ethical awareness. These countries ensure inclusive digitalization by stressing equal access and a participatory approach in AI literacy (Xu et al., 2024).

This participatory orientation complements the development-oriented governance logic identified in Tables 1 and 2, in which inclusivity and capacity building are central regulatory priorities rather than strict compliance mechanisms.

However, India promotes the integration of AI into public services by emphasizing modernization and efficiency, while giving less attention to civic engagement and ethical deliberation (Lakshitha et al., 2025). This is indicative of a techno-developmentalist trajectory where efficiency in administration is seen as more important than ethical systems of governance.

Similarly, countries like Bangladesh and Pakistan are still in the early stages of AI governance; their policies focus on digital infrastructure and economic competitiveness. However, these countries provide less attention to ethical risk management mechanisms (Karmakar, 2024; Muhammad et al., 2025). This is consistent with the global South governance paradigm discussed above in which infrastructural preparedness frequently preludes formalized accountability.

Similarly, Nepal and Sri Lanka have initiated the integration of AI into the educational curriculum and teachers' training on AI, with support from UNESCO. However, these countries lack proper execution of these strategies (UNESCO, 2024). This implementation gap is equivalent to the capacity-execution gap in Table 2, where policy statements are not always realized in practice.

In summary, the literature highlights structural disparities in ethical governance and value alignment across regions. The literature shows that the core areas of focus in South Asian countries are capacity development of organizations and economic modernization, whereas Southeast Asian countries have paid closer attention to inclusive and participatory governance. These differences illustrate how variations in the regional governance logics, compliance-driven, innovation-driven, or development-oriented, are reflected in institutional practice and order of policy.

Wright (2024) said that despite advocating "humanistic AI" and "global responsibility, Japan's AI ethics practice still maintains a conservative structure, with insufficient participation of women and minorities, a lack of implementation system, and ethical principles mostly remaining at the document level. For instance,

gender imbalance undermines the work of groups dedicated to ethical AI. The Humanistic AI Social Principles Group includes only 13.8% women, and all Moonshot project managers are men. This discrepancy highlights a fundamental contradiction between the commitment to diversity and actual governance practices. Such structural exclusion ultimately erodes the inclusiveness and legitimacy of global AI ethics initiatives. This instance also supports the distinction between normative and institutional realizations of commitment and the realization recognized in comparative analyses of governance across regions.

In Iran, AI policy is heavily influenced by national security logic, with a highly centralized regulatory system that lacks space for ethical debate from multiple perspectives (Atwood, 2025). This model reinforces government authority but also undermines social resilience in addressing AI risks.

In contrast to the participatory or market-driven models mentioned above, this centralized approach demonstrates how security-based governance can limit the process of pluralistic value alignment.

2.6 Theoretical dimensions (comprehensive theoretical construction framework) and practical challenges of value alignment

Based on the comparison analysis above of governance, the focus now turns from institutional models to the theoretical foundation of value alignment as the normative basis for the connection between governance structures and AI system behavior.

In the context of artificial intelligence systems increasingly embedded in human society, ensuring that their behavioral goals do not deviate from human values, ethical principles, and true intentions has become one of the most urgent and complex issues in AI ethical governance. "Value Alignment" is one of the shared topics of interest across AI, computer science, and policy governance (Russell et al., 2015). Though “Value Alignment" is a new concept, its basic purpose has historical roots.

Historically, Value Alignment has confirmed that AI functions are related to human values. This concept originated from Isaac Asimov's "Three Laws of Computing” and was further developed through cybernetics and ethical computing frameworks, which examined how human behavior is controlled by machines to achieve standardized objectives (Zwitter, 2024).

It has been noted in past research studies that Value alignment has become one of the major aspects of AI safety and governance, and emphasized that value alignment can be achieved only when the objectives of robotics and autonomous systems are aligned with sociocultural and ethical human values (Russell, 2019; Gabriel, 2020). This theoretical assumption is the logical antidote to the previous comparison of governance, since the diversity among regulations is a manifestation of the way society interprets the concept of “aligned values”.

Previous research studies have highlighted that “Value alignment,” as a modern concept, has emerged as one of the key components of AI safety and governance, stressing that it is possible only if the goal of Robotics is in accordance with sociocultural and ethical human values (Hadfield-Menell, 2016). This concept aligns with our research objectives to explore how ethical governance and AI literacy can enhance the shared capability to align AI technology with societal values.

Value alignment, as a modern concept, is founded on the work of Wiener, which explains how to maintain consistency in the outcomes of controlling machines, as well as on the efforts to align machines with human objectives (Muraven, 2017). Thus, value alignment works as the conceptual bridge between institutional governance models and the AI system technical design. .

The rapid progress of AI, especially generative AI, has made value alignment a core issue in governance. This is because these systems are increasingly making autonomous decisions that directly affect human welfare and social order. Frequent instances of biased, misleading, or unethical outputs from large models underscore the urgent need to ensure that AI behavior aligns with human moral and social values. The Asilomar AI Principles explicitly warn that highly autonomous systems must remain consistent with human values throughout their lifespan, reflecting a consensus among experts that "hyper-intelligent systems," if not properly managed, could deviate from ethical standards or worsen societal harms (Future of Life Institute, 2017). It is in this sense that governance structures and technical alignment mechanisms must work in a constitutive manner: institutional frameworks determine the normative boundaries, whereas technical alignment mechanisms operationalize within algorithmic systems.

The recent scholarly concept of AI value alignment is typically divided into two phases: primary alignment and secondary alignment. Primary alignment can be defined as the process of instilling human values and ethical standards directly into an AI framework during its development and training. This phase is necessary to ensure that the AI's major goals and behavioral reasoning are aligned with generally accepted human standards. Reinforcement learning based on human feedback (RLHF), incorporating ethical constraints, or introducing value-based embedding vectors are the most common methods for modelling the model’s initial decision-making structure and avoiding unacceptable results early on. Secondary alignment, in turn, is concerned with the continual adjustments to AI behavior post-implementation and its surveillance. It carries processes that allow the system to adjust its decisions and actions in response to emerging ethical issues, societal commentary, or changes in context. The focus of this stage is on constant monitoring, audit and adjustment to bring the behaviour of the AI under the changing human values and social norms as time goes by. Combined, the two phases form a dynamic framework to ensure that AI systems do not just start with values-consistent goals (primary alignment) but also maintain and strengthen that consistency throughout their operation (secondary alignment) (Huang et al., 2025).

This evolution, from static to dynamic alignment and from single authority to collaborative governance, shows that value alignment isn't a one-time coding or design project. Besides, Value alignment is constructed and co-constructed. Rather, it is an ongoing socio-technical co-construction process. Here, the previous point about multi-agent collaborative governance as an institutional coordination is furthered by stating that it is a structural requirement of enduring value congruity. At the technical level, the value alignment focuses on the systematic orientation of the ethical constraints, at the social level pertinent to negotiating the potential outcomes of agreeable AI conduct (van Wyk, 2024). Therefore, it is concluded that algorithmic optimization and inclusive governance ecosystems are equally important for effective value alignment.

Despite broad theoretical agreement on value alignment, practical implementation presents major hurdles. Examples like ChatGPT and Google Bard illustrate the challenge: it's hard to ensure that AI output consistently aligns with human ethics and cultural diversity. Since these systems can still produce biased or inappropriate content even with safety filters, it's clear that technical adjustments are insufficient (Rane

and Choudhary, 2024; Stahl and Eke, 2024; Weidinger et al., 2022). These difficulties also confirm the comparative results in Tables 1 and 2, which indicate that the regulatory design is insufficient and that adaptive monitoring and stakeholder engagement systems should be implemented.

Value pluralism and dynamic conflict: Different groups define the ethical "goodness" of AI based on their cultural values, leading to varied expectations. Western nations (Europe and the U.S.) typically view "goodness" through a human rights framework, emphasizing individual freedom, privacy, and autonomy. Meanwhile, many developing countries define it in terms of collective welfare, efficiency, and national stability, favoring AI that supports economic growth and social order (Roberts et al., 2021; NIST, 2023). This normative difference makes it difficult to establish a universally relevant standard of alignment and the argument for context-sensitive universal coordination of governance.

In this context, aligning AI with human values requires not only improvements to model design but also the establishment of comprehensive ethical evaluation systems, multi-level risk-monitoring mechanisms, and public consultation platforms. Enhancing AI literacy is a necessary foundation for developing these governance structures and provides the cognitive basis for achieving ongoing value alignment

Effective value alignment relies on the cooperation of diverse participants, including governments, technology developers, academics, and the public. Each of these groups helps translate ethical principles into practical governance through regulation, technical standards, and civic engagement. Without these collaborative mechanisms, value alignment cannot shift from an abstract idea to concrete behavioral norms this is the core meaning of multi-agent collaborative governance. The value alignment provides ethical principles aligned with societal values and sets ethical limits for AI systems. AI systems provide a platform for connecting stakeholders worldwide and serve as a space to build value consensus and ensure governance (Schwerzmann, 2025).

3. METHODOLOGY

3.1 Research Design

This paper used a Systematic Literature Review (SLR) methodology to investigate the relationship between AI literacy and ethical digital governance, and to examine how collaboration, shared values, and stakeholder involvement contribute to this relationship. The rationale for choosing the SLR approach is that it is a clear, replicable, and comprehensive system for synthesizing available evidence across disciplines (Tranfield et al., 2023).

3.2 Search Strategy

A systematic search approach was applied to large scholarly databases, including Scopus, Web of Science, IEEE Xplore, and Google Scholar, that included the literature published in 2019-2025. Search refinements were performed using Boolean operators to ensure the inclusion of relevant studies. The main keywords were used to bring together the concepts of AI literacy, ethical governance, and collaboration, e.g., (AI literacy) OR (artificial intelligence education) AND (ethical governance) OR (digital ethics) OR (responsible AI). (AI policy/AI regulation) and stakeholder participation/shared values.

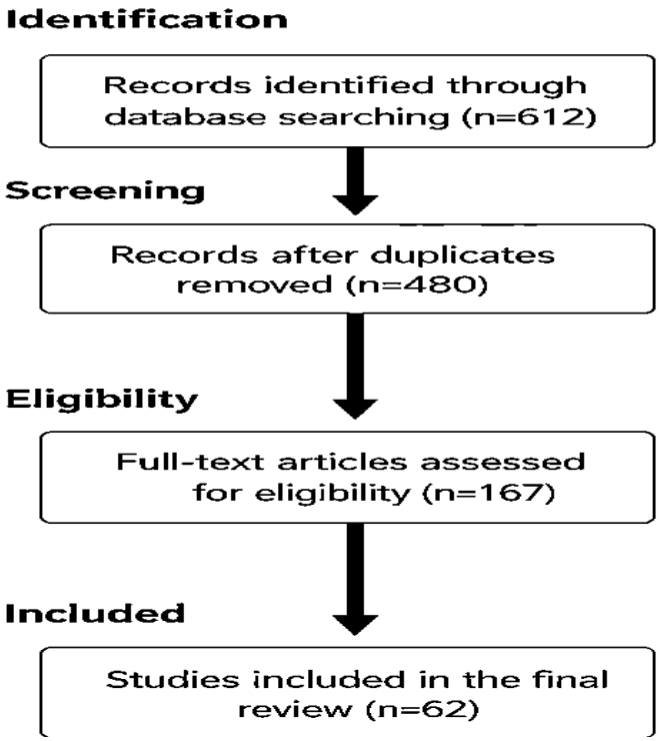
3.3 Inclusion and Exclusion Criteria

Inclusion criteria were: (1) the sources had to be peer-reviewed articles, policy reports, or institutional frameworks published in 2019-2025; (2) the source had to be in English; and (3) the source had to discuss AI governance, ethics, or literacy. Research articles that were purely technical, lacked an ethical or governance aspect, and whose complete text was not available were excluded.

3.4 Screening Process and PRISMA Flow

In accordance with the PRISMA 2020 principles (Page et al., 2021), 612 records were obtained as a result of the search process. After eliminating 132 duplicates, 480 titles and abstracts were filtered. Of the 167 full-text articles, 62 were evaluated for eligibility, and 62 studies were selected for analysis. 62 studies were included in the systematic review, while additional sources were used for conceptual and policy discussion A PRISMA flow diagram illustrated this identification, screening, and inclusion process (Figure 1).

Figure 1. PRISMA FLOW DIAGRAM



3.5 Data Extraction and Coding

Data were extracted into an ordered table that included the authors, year, geographic area of interest, main findings of each study, and the applicability of the results to AI literacy and ethical governance. The

thematic analysis was then applied to the data, following Thomas and Harden (2008). The data were manually coded to identify repetitive concepts, relationships, and policy patterns across selected studies. Thematic grouping (NVivo software) might also be applied on a large scale, but for the current study, thematic analysis was performed manually because the data were manageable and thus provided contextual sensitivity.

3.6 Thematic Analysis

The thematic analysis was performed in line with the framework of Thomas and Harden (2008), with three steps being familiarization, theme identification, and thematic synthesis. To familiarize with all 62 included studies, they were thoroughly read and summarized to identify the main trends regarding AI literacy, governance practices, and ethical implications. During the theme identification phase, common concepts emerged, including AI literacy capacity, multi-stakeholder partnerships, and value alignment across a wide variety of policy settings and scholarly discourses. Lastly, these concepts were grouped using thematic synthesis to form a coherent understanding of how AI literacy supports ethical digital governance, collaboration, and value-based decision-making (

It was discovered that three primary themes explain the process of the formation of ethical digital governance at both national and institutional levels: (1) the cognitive basis of governance is AI literacy, (2) the practical process is multi-stakeholder collaboration, and (3) the ethical result is value alignment.

3.7 Ethical Considerations

Since this was research that used secondary data (using published sources), there were no actual human subjects. Thus, the ethical permission was not needed. Nevertheless, all the examined literature and policy papers were properly referenced in order to uphold academic integrity.

4. RESULTS AND DISCUSSION

This study identified the following three main themes during the analysis, explaining how ethical digital governance develops across countries:

1. AI literacy is the cognitive foundation of governance,
2. Multi-stakeholder collaboration as the practical mechanism, and
3. Value alignment is the ethical outcome.

4.1 AI Literacy as the Foundation

The studies concluded that AI literacy goes beyond technical knowledge. It means that to understand how artificial intelligence works, we need to think critically about its social and ethical effects and be confident enough to join policy discussions or debates. Countries that invested in AI education, such as China and Singapore, through their school- and university-level curricula, reaped significant benefits. People in these countries are more informed, more engaged in ethical conversations, and less likely to be misled by misinformation about AI.

In countries that lack or have a weaker AI education policy, public understanding of AI is limited. People in these countries were less involved in shaping policies or questioning the ethical impact of emerging technologies. This shows that AI literacy builds the foundation for responsible participation in digital and good governance in the countries.

4.2 Collaboration as a Governance Mechanism

This study found that cooperation among sectors, including governments, private companies, researchers, and civil society, improves transparency and accountability in AI governance. For example, in the European Union and other OECD countries, collaborative policymaking led to stronger and clearer governance structures. The *EU AI Act* is a good example, as it brings together multiple stakeholders under a risk-based regulatory system that includes human oversight and ethical safeguards.

In a country like China, the collaboration also played an important role, primarily between the government, technology companies, and key social organizations. This created flexibility in implementation but gave the general public fewer opportunities to influence policy. The country, like the United States, encouraged open innovation and private-sector leadership; however, the lack of consistent ethical standards created inconsistencies between innovation and accountability.

4.3 Value Alignment as an Ethical Outcome

This study found that when the general public understands AI and collaborates effectively, they tend to develop shared ethical values. Systems that create dialogue among policymakers, developers, and the public are more effective in aligning AI systems with public interests. Open dialogue encourages the integration of fairness, inclusivity, and transparency into AI design and governance.

However, in culturally diverse and unequal societies, keeping value alignment is more difficult. Differences in moral perspectives and unequal access to digital tools often lead to conflicting priorities.

Overall, the study concludes that AI literacy enables individuals to participate meaningfully. At the same time, collaboration transforms ethical principles into tangible policies and institutional practices, ensuring that digital governance remains fair, inclusive, and socially responsible.

5. DISCUSSION

This work contributes to the existing knowledge of ethical digital governance by demonstrating that the concepts of AI literacy, multi-stakeholder cooperation, and value alignment can be understood as a specific ecosystem of governance rather than as separate elements. The results support and prove the theoretical views of stakeholder governance, the theory of public value, and the research on AI ethics.

First, the findings confirm the existence of a structural precondition of AI literacy to participatory governance. AI literacy is more than just coding skills; it also encompasses algorithmic awareness, critical data reasoning, and ethical reflexivity. As Ng (2021) notes, AI literacy refers to understanding, evaluating, and interacting responsibly with AI systems. Similarly, Long and Magerko (2020) define AI literacy as a skill model that enables citizens to critically evaluate the outputs of algorithms and their social consequences. These viewpoints confirm the observation that AI-conscious societies are better placed to apply democratic checks on new technologies.

Likewise, the national AI strategy of Singapore is not only technologically focused but also actively seeks to develop human capital and educational infrastructure to support the responsible Tech-Gear adoption and workforce preparation to make the country more competitive and technologically sufficient, a factor that has been classified as a priority towards organizational competitiveness (Khanal et al., 2024) China, similarly, has developed Next Generation AI Development Plan that provides a complete overhaul of the curriculum to enhance national competitiveness and technological abilities. Empirical studies have shown that these types of efforts not only make students and educators develop technical skills but also gain ethical understanding and critical thinking (Zhao et al., 2022). These reforms through education enhance the civic competence of citizens, decrease the susceptibility to misinformation, and enhance the legitimacy of AI governance by providing them with the knowledge and capacity to engage with AI technologies.

Second, the results are aligned with the Stakeholder Theory (Freeman, 1984) according to which sustainable governance is based on multi-stakeholder collaboration. The stakeholders' theory pointed out that in AI governance, stakeholders include the governments, corporations, academia, and civil society. These findings are supported by Bryson et al. (2014), who claim that the creation of public value relies on cross-sector collaboration and participatory engagement. This assertion is substantiated by the findings: jurisdictions that codify collaborative governance, such as the European Union, by the European Union Artificial Intelligence Act, have more apparent accountability frameworks and transparency processes. The EU's risk-based regulatory model is based on joint policy-making through consultation with industry experts, academic researchers, and civil society participants (European Commission, 2021). Conversely, the fragmented U.S. governance environment is more flexible regarding innovation, yet coordination challenges arise in examining federal AI regulation (Calo, 2017). These differences underscore the fact that the effectiveness of governance is not limited to the content of regulation, but also to institutionalized cooperation.

Third, the study validates that value alignment is a dynamic ethical outcome that is influenced by literacy and collaboration. According to Russell (2019), value alignment is the process of ensuring that AI systems are working according to societal norms and human preferences. Nevertheless, alignment is not fixed but it changes in response to changes in social demands and technological capabilities. According to Dignum (2019), responsible AI entails ongoing dialogue among the parties and negotiation over acceptable levels of fairness and accountability.

The argument about biased algorithmic outputs and real-world harms further supports it. Studies by Bender et al. (2021) show that large language models can replicate societal biases despite technical precautions. Such results demonstrate that technical optimization alone is not enough. Ongoing monitoring, ethical reviewing, and civil education are needed to maintain harmonization.

These findings also reflect the Public Value Theory (Moore, 1995) that holds that governance legitimacy is created by creating collective value by transparency, trust, and participation. The adoption of AI governance, integrated with literacy and collaboration, reduces the lack of trust among the population by allowing citizens to shape technological systems. The same pillars of trustworthy AI, such as inclusive growth, transparency, and accountability, are highlighted in the OECD AI policy frameworks (OECD, 2019).

However, value pluralism makes it more difficult to adhere. Most Western forms of governance tend to value human rights and personal freedoms, whereas most developing situations tend to focus on communal good and economic development. Geopolitical competition and state-centered governance logics, in particular, those focusing on digital sovereignty and national security, can reduce the room for pluralistic ethical discourse and make it more difficult to draw universal standards of AI ethics (Coeckelbergh, 2025).

The results indicate that the connection between collaboration and value alignment is moderated by AI literacy. The result highlighted that collaborative governance processes will yield legitimate and socially responsive results when citizens have more knowledge of AI. On the other hand, low literacy levels hinder meaningful interactions and undermine accountability systems. When people are well educated about AI, they are in a better position to support those policies that are developed to make the technology safer and more ethical. Also, these citizens will support efforts to point out and discuss virtual issues (Ng, 2021). This AI literacy shifts the governance from a top-down approach to a more open, participatory, and shared one.

The findings are supported by Yotawut (2018), who noted that this type of governance not only promotes efficiency but also participation, openness, and trust, which are the main themes of Public Value Theory. Coordination among institutions and the public, with proper literacy, improved the use of AI systems to enhance citizens' well-being. Such as the OECD AI Principles and the EU AI Act highlighted that collaboration among stakeholders like governments, corporates, and civil societies produced transparency, moral ethics and strengthened the governance systems (OECD, 2024). Similarly, from the comparative analysis, it was concluded that collaboration among stakeholders links AI literacy with moral outcomes. Open communication among stakeholders and public participation are the key elements of promoting shared values, building trust and accountability in the European countries. However, in developing nations, despite strong policies and less participatory strategies hindering public say in decisions and weaker ethical alignments, these findings are supported by Dignum (2019), who reported that coordination among stakeholders and public participation promote strong digital governance and ethical integrity. Furthermore, the results also highlighted that value alignment is not static but a dynamic process. The value alignment changes as people become more AI literate and technology advances. These results are supported by Biagini (2025), who pointed out that AI literacy helped promote a sense of responsibility and fairness and reduced the gap between technology and societal expectations.

.

To summarize the study results, the study contributed to the literature on AI governance by introducing a conceptualization of ethical digital governance in three layers: literacy (cognitive background), collaboration (institutional process), and value alignment (ethical product). When all three layers operate synergistically, they create a flexible and fair system of digital governance, founded on emerging technologies while keeping the public's interests at the center.

6. CONCLUSION

This paper has examined the connection between AI literacy and ethical digital governance and the role of collaboration, shared values, and stakeholder engagement in responsible technological development. Through a systematic review of academic literature, policy frameworks, and international governance models, the study established that countries vary in their approaches but share similar ethical goals. Three key themes emerged, including AI literacy as the cognitive basis of governance, multi-stakeholder collaboration as the practical process, and value alignment as the ethical impact. The results highlight the importance of AI literacy, which goes beyond technical literacy; it encompasses citizens' critical awareness of AI's social and ethical impacts. Those countries that invest in AI education, like China and Singapore, are more engaged with AI, have better policymaking, and fewer cases of misinformation. Conversely, countries with less developed AI education policies have difficulty engaging people in the system and achieving greater ethical responsibility, suggesting that literacy should be the initial step in responsible governance. The research also found that the aspect of collaborative governance enhances transparency, accountability, and trust among the people. Policies such as the EU AI Act and the OECD AI Principles can serve as examples of how governments, corporations, and civil society can together establish equitable and responsible systems of governance. Nevertheless, difficulties remain, particularly when coordination mechanisms are poor or when coordination relies on a top-down strategy. Teamwork will translate ethical issues into practical structures and is therefore an important step toward sustainable AI governance. Lastly, value alignment was seen as an active ethical process and not an objective. Stakeholder engagement in policy formulation and management strengthens the shared values and moral principles of AI systems, including inclusiveness, fairness, and human dignity. Nevertheless, balancing the culturally diverse and resource-constrained societies is not easy due to differences in priorities and access to technology. Finally, the research shows that AI literacy among citizens enhances their power, institutionalizes ethics through collaboration, and strengthens the integrity of digital governance through value alignment. These dimensions form a powerful construct for developing just, open, and humanistic AI systems capable of responding rapidly to rapid shifts in technology and society.

7. THEORETICAL IMPLICATIONS

The research presents strong theoretical contributions by synthesizing Stakeholder Theory and Public Value Theory to define ethical digital governance as a dynamic, co-constructed process. The research contributed to Stakeholder Theory further by situating AI literacy as an essential facilitator of participatory governance, focusing on informed involvement across societal and institutional levels. At the same time, it advances Public Value Theory by positing multi-stakeholder collaboration as a means of co-producing societal trust, transparency, and moral legitimacy in AI systems. The literacy collaboration value alignment model proposed here offers a coherent theoretical perspective that brings together cognitive, structural, and normative aspects of governance, thereby broadening existing explanations of how ethical, human-oriented AI governance can be transformed through participatory, literacy-based, and value-consistent frameworks.

8. PRACTICAL IMPLICATIONS

The research findings have several actionable implications. Policymakers would need to incorporate AI literacy into school, college, and university-level education curricula in every nation to foster ethical sensitivity and enable citizens to become responsible stakeholders in governance. Governments and industries would need to create multi-stakeholder forums where regulators, developers, educators, and members of the public engage in dialogue, encouraging co-regulation and accountability. Institutions would need to deploy algorithmic openness and ethical auditing mechanisms to monitor continuously and align with values. In developing countries, community-based AI education programs can bridge the digital divide and enable inclusive engagement. Lastly, ethical considerations should be integrated throughout the AI lifecycle, from design to deployment, so that technological progress supports human well-being. Together, these measures operationalize the "AI literacy collaboration value alignment" model into governance practices, reinforcing trust, transparency, and collective accountability in the AI age.

9. LIMITATIONS AND FUTURE STUDIES

Although this research will be useful in understanding the nexus between AI literacy and ethical governance, it has a number of limitations that cannot be ignored. To start with, being a systematic literature review, it used only secondary data, which can be a drawback in terms of context and accuracy of interpretation. Second, differences in the accessibility and quality of local literature, especially in the Global South, could have contributed to the comparative results. Third, the study was limited to sources published between 2019 and 2025 and may have missed earlier foundational research. Finally, the thematic synthesis has been performed manually, which, though subject to context, can be subject to subjective bias. Further studies should include empirical case studies and cross-regional surveys to substantiate and extend these conceptual insights.

Author Contributions: Xu Hao conceptualized the study, conducted the research, and drafted the manuscript. Nurul Liyana Mohd Kamil supervised the methodological design, refined the article structure, and validated the references. All authors have read and approved the final manuscript.
Ethics Statement: This study did not involve human participants or animals and therefore did not require ethical approval.
Conflict of Interest Statement: The authors declare no conflict of interest.
Data Availability Statement: The data supporting the findings of this study are available from the corresponding author upon reasonable request.
Generative Artificial Intelligence Statement: The authors declare that no generative artificial intelligence tools were used in the conceptualization, analysis, or writing of this manuscript, except for minor language editing.

REFERENCES

Alharbi, T. (2021). Assessment of cybersecurity awareness among students: Implications for protecting cyberspace and reducing cybersecurity threats. *Journal of Risk and Financial Management, 14*(2), 23. https://doi.org/10.3390/jrfm14020023

Atwood, B. (2025). Artificial intelligence in Iran: National narratives and material realities. *Iranian Studies.* Advance online publication. https://doi.org/10.1017/irn.2024.63

Batool, A., Zowghi, D., & Bano, M. (2025). AI governance: A systematic literature review. *AI and Ethics, 5*, 3265–3279. https://doi.org/10.1007/s43681-024-00653-w

Baum, S. D. (2020). Social choice ethics in artificial intelligence. *AI and Society, 35*(1), 165–176.

Beer, D. (2017). The social power of algorithms. *Information, Communication & Society, 20*(1), 1–13. https://doi.org/10.1080/1369118X.2016.1216147

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)* (pp. 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Biagini, G. (2025). Towards an AI literate future: A systematic literature review exploring education, ethics, and applications. *International Journal of Artificial Intelligence in Education, 35*, 2616–2666. https://doi.org/10.1007/s40593-025-00466-w

Birdayanthi, B., Yusriadi, Y., & Ikmal, I. (2025). Accountability and transparency in public administration for improved service delivery. *Journal Social Civilecial*. https://doi.org/10.71435/610633

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., *et al.* (2022). *On the opportunities and risks of foundation models*. arXiv. https://doi.org/10.48550/arXiv.2108.07258

Bozeman, B. (2007). *Public values and public interest: Counterbalancing economic individualism.* Georgetown University Press. https://dx.doi.org/10.1353/book13027

Bridoux, F., & Stoelhorst, J. W. (2022). Stakeholder governance: Solving the collective action problems in joint value creation. *Academy of Management Review, 47*, 214–236. https://doi.org/10.5465/amr.2019.0441

Bryson, J. M., Crosby, B. C., & Bloomberg, L. L. (2014). Public value governance: Moving beyond traditional public administration and the new public management. *Public Administration Review, 74*(4), 445–456. https://doi.org/10.1111/puar.12238

Calo, R. (2017). Artificial intelligence policy: A primer and roadmap. *UC Davis Law Review, 51*, 399–435. https://doi.org/10.2139/ssrn.3015350

Chu-Ke, C., & Dong, Y. (2024). Misinformation and literacies in the era of generative artificial intelligence: A brief overview and a call for future research. *Emerging Media, 2*(1), 70–85. https://doi.org/10.1177/27523543241240285

Coeckelbergh, M. (2025). Three challenges for a global AI ethics: Towards a more relational normative vision. *AI Ethics, 5*, 5527–5533. https://doi.org/10.1007/s43681-025-00791-9

Council of the European Union. (2024). *Artificial intelligence act*. https://www.consilium.europa.eu/en/policies/artificial-intelligence

Crawford, K., & Calo, R. (2016). There is a blind spot in AI research. *Nature, 538*(7625), 311–313. https://doi.org/10.1038/538311a

Dignum, V. (2019). *Responsible artificial intelligence: How to develop and use AI in a responsible way.* Springer. https://doi.org/10.1007/978-3-030-30371-6

Donaldson, T., & Preston, L. E. (1995). The stakeholder theory of the corporation: Concepts, evidence, and implications. *Academy of Management Review, 20*(1), 65–91.

Ebers, M. (2025). Truly risk based regulation of artificial intelligence: How to implement the EU's AI Act. *European Journal of Risk Regulation, 16*(2), 684–703.

European Commission. (2021). *Proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).* Brussels.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1). https://doi.org/10.1162/99608f92.8cd550d1

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Foka, A., Griffin, G., Ortiz Pablo, D., Rajkowska, P., & Badri, S. (2025). Tracing the bias loop: AI, cultural heritage and bias mitigating in practice. *AI & Society, 40*(3), 5835–5847. https://doi.org/10.1007/s00146-025-02349-z

Freeman, R. E. (1984). *Strategic management: A stakeholder approach.* Pitman.

Future of Life Institute. (2017). *Asilomar AI principles*. https://futureoflife.org/ai-principles

Gabriel, I. (2020). Artificial intelligence, values, and alignment. *Minds and Machines, 30*(3), 411–437. https://doi.org/10.1007/s11023-020-09539-2

Güngör, H. (2020). Creating value with artificial intelligence: A multi-stakeholder perspective. *Journal of Creating Value, 6*(1), 72–85.

Hadfield-Menell, D., Russell, S., Abbeel, P., & Dragan, A. (2016). Cooperative inverse reinforcement learning. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 29, pp. 3909–3917). Curran Associates, Inc

Hidayat, W., & Muis, A. (2025). Ethical and legal challenges of artificial intelligence in the judicial system: An Indonesian perspective. *Justicia Insight, 2*(1), 9–15. https://doi.org/10.70716/justin.v2i1.274

Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. *Science, 361*(6404), 751–752.

Hingle, A., & Johri, A. (2025). *Systematic review of collaborative learning activities for promoting AI literacy* (arXiv:2508.15111). arXiv. https://doi.org/10.48550/arXiv.2508.15111

Hristovska, A. (2023). Fostering media literacy in the age of AI: Examining the impact on digital citizenship and ethical decision making. *Kairos, 2*(2), 39–59. https://doaj.org/article/744cb4faf2364808b8623d9a6f750ffa

Huang, L. T. L., Papyshev, G., & Wong, J. K. (2025). Democratizing value alignment: From authoritarian to democratic AI ethics. *AI Ethics, 5*, 11–18. https://doi.org/10.1007/s43681-024-00624-1

Hussein, H., Gordon, M., Hodgkinson, C., Foreman, R., & Wagad, S. (2025). ChatGPT's impact across sectors: A systematic review of key themes and challenges. *Big Data and Cognitive Computing, 9*(3), 56. https://doi.org/10.3390/bdcc9030056

Irawati, S., Hayat, A., Juniar, A., & Handayani, S. A. (2023). Exploring accountability and transparency in government agency management: A literature review. *Ilomata International Journal of Management, 5*(3), 587–600. https://doi.org/10.61194/ijjm.v5i3.1189

Islam Tonmoy, S. M. T., Zaman, S. M. M., Jain, V., Rawte, V., Chadha, A., & Das, A. (2024). *A comprehensive survey of hallucination mitigation techniques in large language models* (arXiv:2401.01313). arXiv. https://doi.org/10.48550/arXiv.2401.01313

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., … Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys, 55*(12), 1–38. https://doi.org/10.1145/3571730

Jiang, Z., Abedin, B., & Marjanovic, O. (2024). Understanding the components of AI literacy at the individual, group, and organisational level: An organisational learning perspective. In *Proceedings of the 35th Australasian Conference on Information Systems (ACIS 2024)*. Australasian Association for Information Systems.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399.

Karimov, A., & Saarela, M. (2025). AI literacy and governance as foundations for ethical AI: A cross-national review of government strategies. In *Proceedings of the 2025 International Conference on Innovation and Technology Research (CITREx)*. IEEE. https://doi.org/10.1109/CITREx64975.2025.10974932

Karmakar, A. (2024). AI governance in Bangladesh: A critical evaluation of the transition from strategic vision to policy framework. *Indian Journal of Legal Review, 4*(3), 668–677.

Kathala, K. C. R., & Palakurthi, S. (2025). AI literacy framework and strategies for implementation in developing nations. In *Proceedings of the 2024 16th International Conference on Education Technology and Computers (ICETC '24)* (pp. 418–422). Association for Computing Machinery. https://doi.org/10.1145/3702163.3702449

Khanal, S., Zhang, H., & Taeihagh, A. (2024). Building an AI ecosystem in a small nation: Lessons from Singapore's journey to the forefront of AI. *Humanities and Social Sciences Communications, 11*, 866. https://doi.org/10.1057/s41599-024-03289-7

Kim, T. W., Hooker, J., & Donaldson, T. (2021). Taking principles seriously: A hybrid approach to value alignment in artificial intelligence. *Journal of Artificial Intelligence Research, 70*, 871–890. https://doi.org/10.1613/jair.1.12558

Klenk, M. (2023). Algorithmic transparency and manipulation. *Philosophy & Technology, 36*(4), Article 79. https://doi.org/10.1007/s13347-023-00678-9

Kraus, S., Jones, P., Kailer, N., Weinmann, A., Chaparro-Banegas, N., & Roig-Tierno, N. (2021). Digital transformation: An overview of the current state of the art of research. *SAGE Open, 11*(3), 21582440211047576. https://doi.org/10.1177/21582440211047576

Lakshitha, P. C., Manoj, A., & Jeevanandhan, L. (2025). Artificial intelligence in Indian governance: Modernising public service delivery. *Electronic Journal of Social and Strategic Studies, 6*(Special Issue VII), 122–134. https://doi.org/10.47362/EJSSS.2025.6607

Liu, X., Zhang, L., & Wei, X. (2025). Generative artificial intelligence literacy: Scale development and its effect on job performance. *Behavioral Sciences, 15*(6), Article 811. https://doi.org/10.3390/bs15060811

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). https://doi.org/10.1145/3313831.3376727

MacDonald, A., Clarke, A., Huang, L., & Seitanidi, M. M. (2019). Partner strategic capabilities for capturing value from sustainability-focused multi-stakeholder partnerships. *Sustainability, 11*(3), 557. https://doi.org/10.3390/su11030557

Marzdar, M. H. (2025). Artificial intelligence in Iran's public administration: Opportunities, challenges, and strategic approaches for governance innovation. *International Journal of Applied Research in Management, Economics and Accounting, 2*(2), 16–35. https://doi.org/10.63053/ijmea.38

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys, 54*(6), 1–35. https://doi.org/10.1145/3457607

Miao, F., Shiohira, K., & Lao, N. (2023). *AI competency framework for students*. UNESCO. https://doi.org/10.54675/JKJB9835

Milano, S., Thiebes, S., & Avital, M. (2020). Artificial intelligence and mass personalization of communication content. *New Media & Society, 24*(5), 1258–1277. https://doi.org/10.1177/14614448221087864

Mills, K., Ruiz, P., Lee, K., Coenraad, M., Fusco, J., Roschelle, J., & Weisgrau, J. (2024, May). *AI literacy: A framework to understand, evaluate, and use emerging technology*. Digital Promise. https://doi.org/10.51388/20.500.12265/218

Moore, M. H. (1995). *Creating public value: Strategic management in government.* Harvard University Press.

Moore, M. H. (2013). *Recognizing public value.* Harvard University Press.

Muhammad, A., Siddiqui, R., & Khan, S. (2025). Ethical and governance implications of AI in Pakistan's financial sector. *Review of Management and Social Sciences, 10*(1), 45–61. https://rjmssjournal.com/index.php/7/article/view/290

Muraven, M. (2017, March 18). *Goal conflict in designing an autonomous artificial system* (arXiv:1703.06354). arXiv. https://doi.org/10.48550/arXiv.1703.06354

National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* (NIST AI 100-1). U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.100-1

Ng, D. T. K. (2021). AI literacy: Definition, teaching, and learning. *Computers and Education: Artificial Intelligence, 2*, Article 100041. https://doi.org/10.1016/j.caeai.2021.100041

Ng, D. T. K., Leung, J. K. L., Chu, K. W. S., & Qiao, M. S. (2021). AI literacy: Definition, teaching, evaluation and ethical issues. *Proceedings of the Association for Information Science and Technology, 58*(1), 504–509. https://doi.org/10.1002/pra2.487

Novelli, C., Taddeo, M., & Floridi, L. (2023). Accountability in artificial intelligence: What it is and how it works. *AI & Society, 39*, 1871–1882. https://doi.org/10.1007/s00146-023-01635-y

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

Organisation for Economic Co-operation and Development. (2019). *OECD principles on artificial intelligence*. https://oecd.ai/en/ai-principles

Organisation for Economic Co-operation and Development. (2024). *OECD AI principles: OECD Council recommendation on artificial intelligence*. https://oecd.ai/ai-principles

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Pauwels, E. (2020). *The new geopolitics of converging risks: The UN and prevention in the era of AI*. United Nations University Centre for Policy Research. https://collections.unu.edu/view/UNU:7442

Pies, I., & Valentinov, V. (2023). Trade-offs in stakeholder theory: An ordonomic perspective. *Social Responsibility Journal, 20*(5), 975–997. https://doi.org/10.1108/SRJ-06-2023-0321

Pinch, T. J., & Bijker, W. E. (1984). The social construction of facts and artifacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science, 14*(3), 399–441. https://doi.org/10.1177/030631284014003004

Prajescu, A. I., & Confalonieri, R. (2025). *Argumentation-based explainability for legal AI: Comparative and regulatory perspectives* (arXiv:2510.11079). arXiv. https://doi.org/10.48550/arXiv.2510.11079

Prikshat, V., Patel, P., Varma, A., & Ishizaka, A. (2022). A multi-stakeholder ethical framework for AI-augmented HRM. *International Journal of Manpower, 43*(1), 226–250. https://doi.org/10.1108/IJM-03-2021-0118

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine, 380*(14), 1347–1358. https://doi.org/10.1056/NEJMra1814259

Rane, N., Shirke, S., Choudhary, S. P., & Rane, J. (2024). Education strategies for promoting academic integrity in the era of artificial intelligence and ChatGPT: Ethical considerations, challenges,

policies, and future directions. *Journal of ELT Studies, 1*(1), 36–59. https://doi.org/10.48185/jes.v1i1.1314

Rawte, V., Sheth, A., & Das, A. (2023). *A survey of hallucination in large foundation models* (arXiv:2309.05922). arXiv. https://doi.org/10.48550/arXiv.2309.05922

Roberts, H., Cowls, J., Morley, J., Taddeo, M., Wang, V., & Floridi, L. (2021). The Chinese approach to artificial intelligence: An analysis of policy, ethics, and regulation. *AI & Society, 36*, 59–77. https://doi.org/10.1007/s00146-020-00992-2

Robles, P., & Mallinson, D. J. (2023). Catching up with AI: Pushing toward a cohesive governance framework. *Politics & Policy, 51*(3), 355–372. https://doi.org/10.1111/polp.12529

Romanishyn, A. (2025). AI-driven disinformation: Policy recommendations for addressing online manipulation. *Frontiers in Artificial Intelligence, 4*, Article 1569115. https://doi.org/10.3389/frai.2025.1569115

Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control.* Viking.

Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Russell, S., Dewey, D., & Tegmark, M. (2015). Research priorities for robust and beneficial artificial intelligence. *AI Magazine, 36*(4), 105–114. https://doi.org/10.1609/aimag.v36i4.2577

Sabatier, P. A., & Weible, C. M. (2014). *Theories of the policy process* (3rd ed.). Westview Press.

Sadat, A. (2025). Digital governance and civic inclusion to enhance public participation in political decision-making processes. *Frontiers in Political Science, 7*, Article 1671373. https://doi.org/10.3389/fpos.2025.1671373

Saeidnia, H. R., Hosseini, E., Lund, B. D., Alipour Tehrani, M., Zaker, S., & Molaei, S. (2025). Artificial intelligence in the battle against disinformation and misinformation: A systematic review of challenges and approaches. *Knowledge and Information Systems, 67*, 3139–3158. https://doi.org/10.1007/s10115-024-02337-7

Schwerzmann, K., & Campolo, A. (2025). "Desired behaviors": Alignment and the emergence of a machine learning ethics. *AI & Society, 40*(7), 5181–5194. https://doi.org/10.1007/s00146-025-02272-3

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In *Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency (FAccT '19)* (pp. 59–68). Association for Computing Machinery. https://doi.org/10.1145/3287560.3287598

Socol De La Osa, D. U., & Remolina Leon, N. (2024). Artificial intelligence at the bench: Legal and ethical challenges of informing—or misinforming—judicial decision-making through generative AI. *Data and Policy, 6*, 1–30. https://ink.library.smu.edu.sg/sol_research/4548

Stahl, B. C., & Eke, D. (2024). The ethics of ChatGPT: Exploring the ethical issues of an emerging technology. *International Journal of Information Management, 74*, Article 102700. https://doi.org/10.1016/j.ijinfomgt.2023.102700

State Council Information Office of the People's Republic of China. (2025). *Global AI governance action plan (full text)*. https://dusseldorf.china-consulate.gov.cn/det/zgyw/20250730_11679598.htm

Su, J., Ng, D. T. K., & Chu, S. K. W. (2023). Artificial intelligence (AI) literacy in early childhood education: The challenges and opportunities. *Computers and Education: Artificial Intelligence, 4*, Article 100124. https://doi.org/10.1016/j.caeai.2023.100124

Tahaei, M., Constantinides, M., Quercia, D., & Muller, M. (2023). *A systematic literature review of human-centered, ethical, and responsible AI* (arXiv:2302.05284). arXiv. https://doi.org/10.48550/arXiv.2302.05284

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology, 8*(1), 45. https://doi.org/10.1186/1471-2288-8-45

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management, 14*(3), 207–222. https://doi.org/10.1111/1467-8551.00375

UNESCO. (2023). *AI and education: Guidance for policy-makers in Asia-Pacific.* UNESCO Publishing.

UNESCO. (2024, August 5). *Key Nepali stakeholders provide recommendations and directions for integrating AI in education in Nepal*. https://www.unesco.org/en/articles/key-nepali-stakeholders-provide-recommendations-and-directions-integrating-ai-education-nepal

United Nations. (2022). *Global digital compact: Open consultation brief*. United Nations Office of the Secretary-General's Envoy on Technology. https://www.un.org/techenvoy/global-digital-compact

Van Wyk, B. (2024). Exploring the philosophy and practice of AI literacy in higher education in the Global South: A scoping review. *Cybrarians Journal, (73)*, 1–21. https://doi.org/10.70000/cj.2024.73.601

Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the draft EU artificial intelligence act—Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International, 22*(4), 97–112.

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Birhane, A., Hendricks, L. A., Isaac, W., Haas, J., Rimell, L., & Gabriel, I. (2022). *Ethical and social risks of harm from language models* (arXiv:2112.04359). arXiv. https://doi.org/10.48550/arXiv.2112.04359

Winfield, A. F., & Jirotka, M. (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A, 376*(2133), Article 20180085. https://doi.org/10.1098/rsta.2018.0085

Wright, J. (2024). The development of AI ethics in Japan: Ethics-washing Society 5.0? *East Asian Science, Technology and Society: An International Journal, 18*(2), 117–134. https://doi.org/10.1080/18752160.2023.2275987

Wu, H., Li, D., & Mo, X. (2025). Understanding GAI risk awareness among higher vocational education students: An AI literacy perspective. *Education and Information Technologies, 30*, 14273–14304. https://doi.org/10.1007/s10639-024-13312-8

Xu, J., Lee, T., & Goggin, G. (2024). AI governance in Asia: Policies, praxis and approaches. *Communication Research and Practice, 10*(3), 275–287. https://doi.org/10.1080/22041451.2024.2391204

Yotawut, M. (2018). Examining progress in research on public value. *Kasetsart Journal of Social Sciences, 39*(1), 168–173. https://doi.org/10.1016/j.kjss.2017.12.005

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 1–27. https://doi.org/10.1186/s41239-019-0171-0

Zhao, L., Wu, X., & Luo, H. (2022). Developing AI literacy for primary and middle school teachers in China: Based on a structural equation modeling analysis. *Sustainability, 14*(21), 14549. https://doi.org/10.3390/su142114549

Zhu, J. (2024). AI ethics with Chinese characteristics? Concerns and preferred solutions in Chinese academia. *AI & Society, 39*, 1261–1274. https://doi.org/10.1007/s00146-022-01578-w

Zou, J., & Schiebinger, L. (2018). AI can be sexist and racist—It's time to make it fair. *Nature, 559*(7714), 324–326. https://doi.org/10.1038/d41586-018-05707-8

Zwitter, A. (2024). Cybernetic governance: Implications of technology convergence on governance convergence. *Ethics and Information Technology, 26*, 24. https://doi.org/10.1007/s10676-024-09763-9