

Culturally Situated AI Safety for Youth: Saudi Arabian Perspectives of Youth, Parents and Teachers

Aljawharah Alzahrani, Tanusree Sharma

College of Information Sciences and Technology, Pennsylvania State University
ama9555@psu.edu, tanusree.sharma@psu.edu

Abstract

Generative AI (GenAI) tools are widely used by youth and have introduced new privacy and safety challenges. While prior research has explored youth's safety in GenAI within a Western context, it often overlooks the cultural, religious, and social dimensions of technology use that strongly shape youth's digital experiences in countries like Saudi Arabia. To address this gap, this study explores youth's (aged 7-17), parents' and teachers' interactions with GenAI tools and risk perceptions through a non-Western lens. We analyzed 736 Reddit posts, and 1,262 X (Twitter) posts, and conducted interviews with 31 Saudi Arabian participants (8 youth, 13 parents, 10 teachers).

Our findings highlight context-dependent and relational privacy and safety needs of GenAI use from non-Western context, which are often shaped by communal structure and prescribed norms. We found significant risks tied to youth's disclosure of personal and family information, which conflict with culturally rooted expectations of modesty, privacy, and honor, particularly when youth seek emotional support from GenAI. These risks are further compounded by socioeconomic factors such as cost-saving practices leading to the use of shared GenAI accounts (e.g., ChatGPT) within families or even among strangers. We provide design implications reflecting parents' and teachers' expectations of how youth should use GenAI. This work lays the groundwork for inclusive, context-sensitive parental controls that adhere to cultural norms and values.

Introduction

Youth are increasingly interacting with Generative AI (GenAI) from an early age, which shapes how they learn, play, and communicate. Generative Artificial Intelligence (GenAI) tools such as ChatGPT, Gemini, and DeepSeek have become widely accessible and are rapidly transforming these experiences. Tools such as ChatGPT, which supports advanced text generation for content creation, and DALL-E, which produces highly realistic images from textual prompts, highlight the transformative role GenAI plays in digital media and design (Choi et al. 2024). Meanwhile, platforms like Character.ai expand the scope of interaction by offering character-based AI chatbots, enabling personalized and dynamic conversations that often mirror human

dialogue (Davis, Sherson, and Rafner 2025; Zhang et al. 2025b). While the global uptake of GenAI tools has been remarkable, adoption in Saudi Arabia has been especially significant, with a recent survey reporting that 80% of adults in Saudi Arabia actively engage with GenAI, which surpasses usage rates in the United States (Alhamawi 2025).

This momentum is particularly visible among younger audiences, raising critical questions about how these technologies intersect with relational and collective privacy embedded within cultural values. In Saudi Arabia, youth and young adults aged 10–19 constitute the second-largest demographic of AI users at 26.4%, closely following adults aged 20–29 at 27.3%, with ChatGPT currently ranking as the most downloaded AI application in the country (Communications, Space and Technology Commission (CST) 2024). Such widespread engagement signals that youth are not only aware of GenAI but actively engage with it, exposing them to unique risks associated with their developmental stage and psychological vulnerabilities (Neugnot-Cerioli and Laurenty 2024). These risks can be further compounded by the cultural context, where family structures and communal expectations influence both usage patterns and the potential for harm, including breaches of privacy, exposure to inappropriate content, and conflict with culturally prescribed norms (Abokhodair et al. 2017; Farooq et al. 2024).

Existing research has explored youth's interactions with GenAI in Western contexts, highlighting risks such as inappropriate content, over-reliance on AI, and misinformation (Yu et al. 2024, 2025). Yet, these studies often overlook the cultural, religious, and social dimensions of technology use that strongly shape youth's digital experiences in non-Western contexts. In Saudi Arabia, discussions around technology are not only framed in terms of individual safety but also in terms of communal trust, family honor, and adherence to religious norms (Alashwali and Alashwali 2022; Alsiyat et al. 2025; Mian and Alatawi 2023). Evidently, previous research across digital domains in Saudi Arabia, such as youths' use of social media platforms like TikTok and Snapchat, also highlights concerns regarding privacy, surveillance, and the erosion of family values (Alashwali and Alashwali 2022; Aleisa, Renaud, and Bongiovanni 2020; Alqarni et al. 2024). In particular, data collection by foreign companies has raised broader questions related to national sovereignty and identity (Alorini et al. 2025; Saudi

Gazette 2019). Privacy in Saudi Arabia is not just about individual data security; it is intertwined with broader questions of religion, morality, and family cohesion. Given GenAI's conversational nature, and potential to mediate sensitive topics, such as religion or personal relationships, as demonstrated in Western contexts, these concerns may be amplified in non-Western settings, where cultural differences may intensify sensitivities.

To address this gap, we examined two research questions from three user groups with different perceptions of GenAI from a non-Western context:

RQ1 What are the current strategies for ensuring safety of youth using GenAI in Saudi Arabia?

RQ2 What unique risks emerge from youths' interactions with GenAI in the Saudi Arabian context?

To answer these two research questions, we conducted interviews with 8 youth, 13 parents, and 10 teachers. We also analyzed 736 Reddit posts and 1,262 X (formerly Twitter) posts to identify the risks associated with youth's interactions with GenAI, and to explore the strategies employed by parents and teachers to mediate and mitigate potential harms.

Main Findings. By synthesizing our findings:

- Our culturally situated taxonomy introduces three unique risk categories: *Religious and Moral Risk*, *Cultural Norms Violation Risk*, and *Data Privacy Violation Risk*, which have not been addressed in prior AI risk taxonomies (Yu et al. 2025; Critch and Russell 2023; Slatery et al. 2024).
- We identified significant risks associated with youth disclosing personal and family information, details of romantic interest or family conflict, which often violate culturally rooted expectations of modesty, privacy, and family honor. Similarly, youth's exposure to GenAI's inconsistent religious interpretations reflects concerns related to religious appropriateness rather than universal content safety.
- Compounded risk by socio-economic factors, such as cost-saving practices leading to shared GenAI accounts (e.g., ChatGPT) within families or with strangers.

Related Work

Security and Safety for Youth in Generative AI

The rapid adoption of GenAI tools by youth has opened up many great opportunities but also significant concerns. GenAI and its impact on youth's lives have been explored widely in Western contexts. Prior work has documented youth's interactions with GenAI tools from various aspects, such as integrating GenAI in youth's education, enhancing youth creativity with GenAI, youth safety in GenAI, and youth GenAI risks taxonomies (Yu et al. 2024, 2025), and mediation of GenAI (Zhang et al. 2023; Yu et al. 2024). Prior research showed key risks in youth's GenAI interactions, such as relying on GenAI for emotional and mental support (Yu et al. 2024, 2025), over-reliance on GenAI (Yu et al. 2024, 2025), GenAI bias and discrimination (Yu et al. 2025), and privacy concerns (Yu et al. 2024). Parents

and families play a critical role in mediating and managing GenAI's risks for youth (Yu et al. 2024; Zhang et al. 2023); however, evidence shows that parents often struggle to assess the safety of GenAI tools for youth (Yu et al. 2025).

Several studies have documented risks related to youth's online privacy and security (Williams, Choong, and Buchanan 2023; Sun et al. 2021; Cao et al. 2024; Livingstone et al. 2014). Prior work (Williams, Choong, and Buchanan 2023) examined how youth understand and manage online privacy and security, and found that children as young as six understand the importance of privacy; however, their understanding is limited to interpersonal risks (e.g., "stranger danger") rather than institutional or commercial data practices. The study also revealed that parents rely heavily on restrictive monitoring and limiting use rather than having conversations about nuanced security and privacy knowledge (Williams, Choong, and Buchanan 2023). Beyond privacy, misinformation and generated content have emerged as security concerns for youth. Prior work documented youth's and parents' experiences with misinformation on social media platforms (e.g., TikTok and YouTube) (Sharevski and Vander Loop 2024). These studies highlight that existing work has largely developed in Western contexts and may not translate effectively to non-Western settings, calling for culturally appropriate approaches to GenAI safety.

While recent work has begun to examine youth's interactions with GenAI tools in non-Western contexts, such as Brazil (Fortim and Zandavalli 2025), Kenya (Kaberi, Buisa, and Kabata 2025), and India (Kasturi, Mehta, and Raman 2025), youth's use of GenAI in Saudi Arabian context specifically remains under-researched. Moreover, insights from prior work on youth's privacy and safety concerns in Western contexts may not fully translate to parental controls design for Middle Eastern users, given differences in cultural norms, regulatory environments, parental mediation styles (Dwairy et al. 2006), and socio-technical infrastructure. For example, in Western contexts, privacy concerns may focus on individual rights, while in Middle Eastern or Asian contexts, family and community norms and government surveillance may play larger roles. Moreover, cultural factors shape the way people interact, their values, and consequentially, how they use the internet (Moubarak and Afthanorhan 2024). Therefore, culture shapes how youth deal with online risks. For example, studies show that cultural barriers hinder youth from admitting to being exposed to online threats or inappropriate content and from reporting them to their parents, for fear to defying these barriers (Alqahtani et al. 2017; Teimouri et al. 2016). Saudi Arabia is a conservative society that has holds these cultural norms, where parents have expectations and trust that their youth would not see or engage in discussions about sensitive or taboo topics such as inappropriate content on the Internet.

Digital Safety of Youth in Middle Eastern Countries

Although, in 2022, a Boston Consulting Group (BCG) study (Panhans et al. 2022) reported that youth in the Middle East and Latin American countries experienced the highest lev-

els of online threats compared to youth in North America and Europe, research on youth's online safety in the Middle East region is limited. The BCG's study (Panhans et al. 2022) reported that the online threats youth are exposed to are numerous and come in many forms, such as pop-up ads, inappropriate content, cyberbullying and harassment, sexual approaches, malware and phishing. Few Middle Eastern studies have documented these online risks that youth faced, such as cyberbullying (Othman et al. 2025; Rehima, Al-Tkaynehb, and Al-Shamsc 2020; Farooq, Abdullatif, and Altheeb 2023; Alfakeh et al. 2021) and online extortion (Martin et al. 2023).

Youth's digital literacy refers to the knowledge and skills that allow youth to navigate the digital world, both effectively and safely (UNICEF 2019). Youth's cybersecurity education helps youth protect themselves by recognizing online threats (Ebrahimi et al. 2025). Youth's AI literacy is defined as the set of skills that enables youth to understand how AI works and how to use it (Su, Ng, and Chu 2023). These are important skills for Middle Eastern youth to acquire in order to protect themselves from online risks. However, in Arab nations, few studies have examined youth's digital literacy (Almethen and Alomair 2024) and cybersecurity education (AlShabibi and Al-Suqri 2021). To the best of our knowledge, there is no study that has examined youth's AI literacy nor GenAI risks in the Middle East region. Moreover, a prior study reported that Saudi parents lacked parental mediation; they did not discuss online threats with their youth nor monitored their online activities. However, parents have expressed interest in using parental control apps to monitor their children's online use (Alqahtani et al. 2017). This shows that there is a desire to protect their children, but they lack the knowledge to do so, as well as a lack of Arabic language resources to raise awareness about youth's online risks.

Current AI Safety Tools in Western and Middle Eastern Countries

The current state of AI moderation tools: OpenAI's new parental controls for ChatGPT have rolled out, which allows parents to link their own ChatGPT accounts with their teen's account (minimum age 13) using an email invitation, manage how ChatGPT interacts with their teens, manage and disable features (e.g., memory and chat history), and receive notifications when the system detects that a "teen is in a moment of acute distress" (OpenAI 2025). Google's Gemini app availability for youth under 13 is controlled by Google's Family Link parental controls, which allow parents to control their children's access to the app and manage their child's Gemini settings (Shakir 2025; Google 2025b). Google stated that youth's data will not be collected to train AI (Shakir 2025; Google 2025a). However, these AI safety tools are not culturally adaptive and are limited in understanding Arabic content.

The current state of regional and culturally-sensitive tools: In 2025, the UAE's telecoms company (e&) launched a new safety control feature that allows Emirati parents to filter content, block unwanted apps, block access to unsafe websites, and limit Internet use (e& 2025). This parental

control is designed specifically to meet the Emirati families needs. Moreover, the UAE Ministry of Possibilities and Snap Inc. developed a new parental control feature in Snapchat's Family Center; this feature allows parents to view youth's friends lists without viewing the actual conversations, in order to protect youth's privacy and to report abusive accounts (ArabNews 2022). In addition to these safety tools, there is a Saudi initiative called "Qayyem" on Game consoles (e.g., Xbox, Play Station) that aims to protect youth; it provides a rating system for popular games based on culturally-sensitive topics, such as "conflicts with Islamic values" or "decency, morality, and good behavior" (Alashwali and Alashwali 2022). However, these safety tools are limited to Internet use, device, or social media use, and there is an absence of region-specific AI safety tools tailored to youth's needs. Therefore, we aim to address youth's, parents', and teachers' needs for AI safety in these contexts.

Regulatory Environment of Saudi Arabia

Saudi Arabia currently has no dedicated law regulating AI or GenAI, and no legislation addressing the safety of youth in GenAI environment (Nemko Digital 2025; HSF Kramer 2025). Instead, AI is governed through a layered framework in which binding data protection law works alongside non-binding policies (Nemko Digital 2025). This analysis maps the existing and proposed laws and regulations that govern AI, digital safety, and cybersecurity in order to assess how far they protect youth from the specific risks of GenAI.

The first layer consists of foundational binding data and cyber laws that govern the use of technology in general, such as the Personal Data Protection Law (PDPL) and the Anti-Cyber Crime Law (Saudi Data and Artificial Intelligence Authority (SDAIA) 2021; Saudi Ministry of Interior 2007), which govern personal data and online behavior (e.g., unauthorized access, interception, privacy invasion, fraud, defamation, and e-extortion). The PDPL treats minors like any other data subject, with no sensitive consent threshold like EU GDPR's (Art.8) or the US's COPPA. Moreover, the Anti-Cyber Crime Law is not child-specific; it applies uniformly regardless of the victim's age. The law contains no provisions for AI-generated content or algorithmic harms. Both the PDPL's and Anti-Cyber Crime Law's intersections with GenAI harms are incidental rather than purposely built: AI is not explicitly named in either the PDPL or Anti-Cyber Crime Law. However, the PDPL applies to any processing of personal data, whether manual or automated (e.g., collection), and the Anti-Cyber Crime Law applies to malicious deepfakes used for defamation or extortion.

The second layer is a compilation of AI standards, ethics, and instruments issued by national authorities like the Saudi Data and Artificial Intelligence Authority (SDAIA). SDAIA regulates AI through a mix of non-binding instruments, such as the AI Ethics Principles (SDAIA 2023), the Generative AI Guidelines (for the public and for the Government) (SDAIA 2024c,b), and the AI Adoption Framework (SDAIA 2024a). These AI-specific instruments establish responsible use expectations. However, none was designed around minors in AI environments or the GenAI-specific risks to youth, such

Instrument	Coverage	Youth + GenAI Gaps	D1	D2	D3	D4	D5
Personal Data Protection Law (PDPL)	Personal-data processing, including automated processing	No dedicated youth’s-data or synthetic-data provisions	O	O	–	–	–
Anti-Cyber Crime Law (2007)	Unauthorized access, interception, privacy invasion, fraud, defamation, extortion, and harmful content	No AI-generated/deepfake provisions or child-specific protections	–	O	O	–	–
SDAIA AI Ethics Principles (2023)	Fairness, privacy, security, transparency, safety, accountability, and social benefit	Non-binding; no youth- or age-specific safeguards	O	O	O	–	–
SDAIA Generative AI Guidelines (2024)	Responsible GenAI use, disclosure, deepfakes, oversight, risk management, and watermarking	Non-binding; no child-safety-by-design provisions	O	O	Y	–	O
SDAIA AI Adoption Framework (2024)	National framework for responsible AI adoption	No consumer-GenAI or youth-specific safeguards	O	O	O	–	–
Child Protection Law (2014)	Protection of minors from abuse, neglect, sexual exploitation, and inappropriate content	Predates GenAI; no digital, AI, platform, or developer duties	–	–	–	–	–

Table 1: Saudi Arabian regulatory landscape mapped to GenAI child-safety dimensions. Key: Y = direct; O = partial/indirect; – = not addressed. D1 = youth’s data processing/consent; D2 = data privacy; D3 = AI-content/deepfake labeling; D4 = synthetic CSAM; D5 = harmful-content exposure.

as synthetic child sexual abuse material (CSAM) or harmful generated content, and none establishes accountability for youth’s-GenAI safety. Among the relevant AI parts, the AI Ethics Principles (Principal 2) require that AI system to be built and operated with respect for the privacy of collected data, and the Generative AI Guidelines require the disclosure and labeling of AI-generated content and deepfakes (SDAIA 2023, 2024c). However, they are non-binding and do not impose child-safety-by-design obligations.

The third layer contains pre-GenAI laws, such as the Child Protection Law (2014) (Saudi Ministry of Human Resources and Social Development (MHRSD) 2014), which is the only regulation in the landscape built entirely around youth. The law protects youth under age 18 from abuse and neglect. Under this law, the harms most relevant here are: sexual exploitation and exposing to inappropriate content. The Child Protection Law is more oriented toward welfare and offline contexts, meaning that it contains no digital, online, or platform provisions, and predates generative AI (Saudi Ministry of Human Resources and Social Development (MHRSD) 2014). It offers no mechanism for GenAI risks nor platform accountability.

We mapped five areas where generative AI most directly affects youth (e.g., youth’s data processing & consent, youth’s data privacy, AI-content & deepfakes labeling, synthetic CSAM, harmful content exposure) against Saudi Arabia’s laws and instruments to demonstrate that partial coverage is incidental rather than purposely designed to address therisks (Table 1).

Method

We conducted content analysis of social media (Reddit and X (Twitter)), and semi-structured interviews with youth, parents, and teachers. We used a mixed-methods approach to address our research questions. We considered online discussions (e.g., Reddit and X/Twitter) and interviews as complementary evidence.

Social Media Data. We collected and analyzed users-generated data from two social media platforms. We chose Reddit and X (formerly known as Twitter) for several reasons. First, Reddit has been used frequently in prior work (Yu et al. 2024, 2025; Ali, Arunasalam, and Farrukh 2025) as a source to investigate and understand users’ privacy perceptions, concerns, and attitudes. Second, X/Twitter is among the most used social media platforms in Saudi Arabia, with 15.7 million active users (Communications, Space and Technology Commission (CST) 2024; DataReportal - Global Digital Insights 2025). Third, both Reddit and X/Twitter allowed youth aged 13 and over to use the platforms (Reddit 2025; X 2025). We then, collected Reddit posts and comments using PRAW (Python Reddit API Wrapper) (PRAW 2025) and collected X/Twitter tweets and replies using the Apify X/Twitter Search Scraper tool (Apify 2025) from April 2025 to June 2025 to understand the current state of knowledge in Saudi Arabia on the use of GenAI tools. This analysis of Reddit and X content revealed unique insights on how the Saudi population used GenAI tools and its potential risks. Throughout this paper, we denote posts and comments from Reddit as (R) and from X/Twitter as (X).

Data Collection. Our keyword selection was motivated by previous work and contextual keywords for Saudi Arabia. We started with an initial list from prior work (Yu et al. 2024, 2025) and refined it through several manual searches on Reddit and X to build our keyword list. We noticed that Saudi users refer to GenAI tools as Artificial Intelligence or AI, we included these terms in our keyword list. Our final keyword list combined general and technology-focused keywords such as Generative AI, GenAI, Artificial Intelligence, AI, LLM, ChatGPT, Deepseek, Copilot, Gemini, and Grok paired with children or youth in English and Arabic. We selected relevant subreddits and X communities in Saudi Arabia based on the number of active users and engagements. On Reddit, we manually reviewed relevant subreddits and performed keyword searches. Then, we used PRAW to

Reddit				
Cat.	Community	Posts	Cmts	Rel.
<i>Country-specific</i>				
	r/SaudiArabia	26	97	10
	r/SaudiForSaudis	40	300	57
	r/AskSaudi	6	12	5
<i>Purpose-related</i>				
	r/saudi_gamers	3	12	2
	r/UniKSA	5	13	3
	r/SaudiProfessionals	22	200	23

X (Twitter)				
Cat.	Community	Posts	Cmts	Rel.
<i>Country-specific</i>				
	STUDY A+	288	131	271
	Universities	249	205	202
	English Teachers SA	121	118	13
<i>Purpose-related</i>				
	#Saudi.Teacher.Diary	77	23	15
	#learning_w_ai	11	0	3
	Initial search	29	10	39

Table 2: Overview of data collection statistics across Reddit and X (Twitter) datasets. *Col. headers:* *Posts* = # of Posts Pulled; *Cmts* = # of Comments Pulled; *Rel.* = # of Related Posts/Comments.

collect 736 posts and comments. On X, we conducted the same manual review and keyword search process, and collected 39 posts and replies relevant to GenAI usages and concerns. We also collected posts and replies in dedicated spaces on X known as communities, relevant to our targeted users (youth, parents, teachers), and we employed common hashtags among teachers in Saudi Arabia. We used Apify to expand our dataset and collected an additional 1,223 tweets and replies, for a total of 1,262 X posts and replies.

Data Analysis. The posts and comments collected were in Arabic and English; The Arabic content was translated into English by a native Arabic-speaking researcher prior to analysis. We reviewed each post and comment for relevance to GenAI topics and then applied thematic analysis following the same iterative coding approach used for interview data. Researchers regularly discussed emerging patterns and reached consensus on two high-level themes: Use of GenAI Tools in Saudi Arabia and Emerging Threats and Concerns.

Participant Recruitment. We recruited teachers, parents, and youth participants through word of mouth, snowball sampling, and researcher connections, while our first author is from Saudi Arabia. Eligibility criteria varied by participant group. All participants were required to (1) use GenAI platforms and (2) be Saudi citizens and fluent in Arabic. Youth participants were additionally required to be aged 7-17. Parent participants were required to have at least one child aged 7-17. Teacher participants were required to be currently employed as K-12 teachers. We focused on youth around age 7 because this marks the start of early independent technology use. In Saudi Arabia, youth typically receive their first mobile device at this age (Zawya 2018), which may expose them to GenAI. This choice aligns with global practices in youth’s cybersecurity education, where programs target ages 4-18 to build safety and digital literacy skills (Sağlam, Miller, and Franqueira 2023). We conducted interviews with 31 Saudi participants (8 youth, 13 parents, 10 teachers). Our study is approved by IRB.

Interview Procedure. All interviews are conducted online via Zoom. Consent and interviews were conducted in Arabic; recordings were transcribed and translated into English by a native Arabic-speaking researcher. We obtained the parents’ informed consent and their children’s assent for youth’s interviews. First parent-child interview was joint, conducted at the parent’s request; we observed that the

child’s responses mirrored the parent’s, indicating potential influence. To minimize this influence, we conducted all other parents and youth interviews separately, with parental consent. Throughout this paper, participant annotations (C,P,T) denote children, parents, and teachers, receptively.

Interview with Youth. The interview with youth was divided into three sections to answer our research questions. First, we examined youth’s current state of knowledge about GenAI. We began with a warm-up question to assess familiarity, “Do you know about Generative AI?”, followed by an open-ended question, “What do you know about it?” We then asked “What GenAI tools have you used?” and “How often do you use these GenAI tools?”

Second, we explored youth’s experience and use of GenAI tools. We asked, “Who do you learn about GenAI from?”, followed by “Have your parents taught you about using GenAI tools?” and a follow-up question “What did they teach you?” We then asked about their purpose of using GenAI and to provide examples of recent uses: “Can you think of some recent uses of GenAI tools?” We asked participants to elaborate more on specific uses to gain a deeper understanding: “How did you use GenAI tools for that purpose?”

Third, we examined youth’s privacy and safety practices in GenAI with questions such as, “What have your parents told you about privacy when using GenAI tools?” and “What have your teachers told you about privacy when using GenAI tools?”. We then asked a follow-up question on their understanding of what information is considered private: “What kind of information do you think you should keep private when using GenAI tools?” Finally, we asked what they would like to learn about privacy and safety in GenAI.

Interview with Parents and Teachers. First, we explored teachers’ and parents’ experience and use of GenAI tools. We began with a warm-up question to assess familiarity, “Do you know about Generative AI?”, followed by an open-ended question, “What do you know about it?” Then, we asked, “What are some purposes you use a generative AI tool for?” Based on the response, we asked, e.g., “How did you use GenAI tools for [tutoring] purposes?” Then, we asked a warm-up question about youth’s use of GenAI tools, “Does your [child/student] use GenAI tools?”, followed by a follow-up question “What was the purpose?”

Second, we investigated how teachers and parents understood and addressed privacy and safety in the context of GenAI: “Do you know about privacy in AI?”, We then explored whether they had engaged youth in conversations about these issues with a warm-up question, “Have you discussed the security and privacy risks of using GenAI tools with your [child/student]?”, followed by a follow-up question, “What did you discuss?” Finally, we asked about their specific concerns: “What safety and privacy concerns do you have regarding youth’s use of GenAI tools?”

Finally, we asked about their expectations for improving privacy and safety in GenAI for youth: “What would you like your [child/student] to learn about privacy and safety when using GenAI tools?”

Data Analysis. We used thematic and deductive analysis on the translated transcriptions (Braun and Clarke 2006). The first author, a native Arabic-speaker, translated transcriptions into English. We began independently reading through transcripts and coding the first round of interview transcripts (3 teachers’ interviews, 3 parents’ interviews, 3 youth interviews). We discussed the coding regularly, developing themes and refined them, until consistent themes emerged. Weekly, we discussed the emerging themes, resolved any differences, and reached consensus on the interpretations of the findings. This iterative process ensured a reliable coding framework (McDonald, Schoenebeck, and Forte 2019). We then coded the remaining of interview transcripts. We developed high-level themes: Usage of GenAI tools in Saudi Arabia, Emerging Threats and Concerns, and Mediation Strategies.

Findings: GenAI Risk: Non-Western Perspectives

In this section, we expand the youth-GenAI risk category from prior work (Yu et al. 2025) by identifying the risks associated with youth-GenAI interactions from a Saudi perspective. These expanded risks reflect Saudi-specific privacy and safety needs, which are essential for developing contextually appropriate interventions. Following the approach outlined in prior work (Yu et al. 2025), we analyzed data from social media and interviews to identify risk patterns. We synthesize our insights into **five high-level risk categories**, of which three are “new” and unique to non-Western (Saudi) group as shown in Table 3. We also map our proposed risk taxonomy against prior work on youth-GenAI risk taxonomy (Yu et al. 2025) in Table 3.

In this section, we present the three unique risk categories which, to the best of our knowledge, have not been explored in AI risk research (Yu et al. 2025, 2024; Critch and Russell 2023) nor in youth safety research (Livingstone et al. 2014; Sun et al. 2021; Williams, Choong, and Buchanan 2023): *Religious and Moral Risk*, *Cultural Norms Violation Risk*, and *Data Privacy Violation Risk* within cultural context.

Religious and Moral Risk

This risk refers to the potential effect of GenAI providing inaccurate or misleading interpretations of religious and moral principles which often youth learn from their parents.

Misinterpretation of Islamic Fundamentals and Values. This risk typically arises when youth seek religious interpretations or moral guidance from GenAI, leading to confusion and contradictions with family guidance. Unlike traditional search engines (e.g., Google), where youth often encounter and are overwhelmed by multiple sources, as noted by parents, and may turn to their parents to clarify religious interpretations, P1 noted:

“Matters of religion, rulings, verses [of the Qur’an], should be searched for in a religious encyclopedia, [...] I allow them to search Google [...] but not ChatGPT. As soon as youth search, they will find themselves confused about the answers with a lot of links which lead them to turn to their parents or someone more knowledgeable and older to get the correct answer.”

However, GenAI is designed to offer tailored, conversational, and often inaccurate responses. This reduces the likelihood that youth will seek parents’ guidance and clarifications, which increases the risk of relying on unverified or culturally and religiously inaccurate interpretations and moral advice. Parents and teachers expressed strong concerns and discouraged their youth from using GenAI for religious and moral inquiries.

GenAI Generating Responses that Contradict Religious Practices and Family Guidance. Parents and teachers expressed concerns about youth asking ChatGPT existential questions before they are equipped with a strong religious foundation. T5 said-

“GenAI does not have 100 percent accurate information. So when [youth] at this age, it is possible they will ask it existential questions. ‘Does Allah exist? They do not have a strong religious background and often they learn foundational concepts from us [adults]’.”

This shows the concerns around conflicting response of AI with established religious beliefs.

GenAI Offering Inappropriate Religious Advice. GenAI may fabricate religious rulings (Fatwas) or alter Qur’an verses, misleading youth seeking religious advice. Online users shared the same concerns about GenAI fabricating rulings and verses. R76 shared that ChatGPT altered a verse of the Qur’an

“True, it distorted a verse from Surat Al-Furqan [In Islam, Allah protects the Qur’an’s text from any alterations]”.

X23 mentioned that ChatGPT provided an incorrect ruling about drinking water while fasting during Ramadan, saying

“‘It’s permissible to drink water during the fast.’ Are you kidding?”.

Furthermore, GenAI has limitations in understanding Arabic, which further undermines its ability to interpret Qur’anic text accurately. These instances led users, like X24 and X26, to criticize generative AI tools for fabricating answers to religious rulings.

High-Level	Medium-Level	Low-Level	Tag
Religious & Moral Risk	Misinterpretation of Islamic Fundamentals and Values	GenAI generating responses contradicting religious practices	New
		GenAI offering inappropriate religious advice	New
	Inaccurate Depiction of Moral Values	GenAI offering inappropriate moral advice	New
		GenAI generating inaccurate moral depictions in educational materials	New
Cultural Norms Violation Risk	Replacing Family Bonds	GenAI influencing family dynamics	New
		GenAI overanalyzing family/social relationships	New
	Emotional Attachment and Romanticization	GenAI facilitating youth isolation from family	New
		GenAI facilitating emotional attachment conflicting with cultural expectations	New
	Misinterpreting Cultural Nuances	GenAI offering culturally conflicting relationship advice	New
		GenAI generating inaccurate cultural or political content	New
Data Privacy Violation Risk	Disclosure of Personal Information	GenAI confusing Arabic and Persian languages	New
	Inadvertent Metadata Disclosure	GenAI facilitating sharing personal information, photos, and locations	Privacy
		Data extraction from screenshot metadata	New
	Disclosure of Family Information	GenAI encouraging microphone/camera usage	Privacy
		Sharing family information and private family matters	Privacy
	Disclosure by Others	Teachers sharing student information for lesson planning	New
		Parents sharing children’s handwriting	New
	Account Sharing Privacy Breach	Sharing GenAI accounts with family members	New
		Sharing GenAI accounts with strangers	New
	Passive Identity Inference	Inferring family/social graph from shared accounts	New
		Inferring age, religion, or origin from essays	New
Mental Wellbeing Risk	Mental Health Handling	Inferring age from voice	New
		Ethnicity inference from linguistic style	New
	Validation of Culturally Taboo Topics	GenAI reinforcing OCD-related religious behaviors	New
	Voice-enabled Emotional Dependency	Emotionally validating responses increasing distress	Mental
Behavioral & Social Development Risk	Romantic Development Distortion	Young users treating GenAI voice as a companion	New
		GenAI substituting for religiously prohibited relationships	Behavioral

Table 3: Culturally-aware risks for non-Western youth using Generative AI. High, medium, and low-level labels denote the taxonomy hierarchy rather than severity. New indicates categories introduced in this work.

“ChatGPT and DeepSeek cannot be relied upon for religious rulings? [...] The questions about one of the most well-known religious rulings, yet these two AI fabricate imaginary answers. So how can you trust their answers in a religious matter related to your faith?”

GenAI Offering Inappropriate Moral Advice. Parents expressed concerns about youth asking GenAI about fundamental virtues before they develop a strong moral foundation. Parents mentioned their youth asking queries to GenAI, such as *“Is it fine to lie to parents? Is it fine to steal money?*

Can I delay prayers? Does my mom deserve respect?” In most cases, parents shared that GenAI responds agreeably, which could potentially shape youth’s values in harmful ways.

Cultural Norms Violation Risk

This risk refers to interactions with GenAI that potentially mislead youth or normalize behaviors conflicting with their cultural and communal expectations.

Replacing Family Bonds. This risk arises when youth turn to GenAI for advice, emotional support, and compan-

ionship, trusting GenAI more than their families. This behavior weakens family bonds and leads to isolation. Interviews and social media data reveal that youth perceive GenAI as knowledgeable and unbiased, describing it as-

“ChatGPT understands me more than my family” (R17) or “My relationship with ChatGPT is stronger than my family” (X27).

This risk type includes three low-level risks: (1) GenAI influencing family dynamics, (2) GenAI overanalyzing family and social relationships, and (3) GenAI facilitating youth isolation. Youth trust GenAI blindly because, as stated by prior work, that GenAI is designed to offer personalized attention and emotional validation, which blurs reality and creates an illusion of safety (Yu et al. 2025). Parents expressed concerns about this blind trust; P2 said, *“This lack of awareness makes my kids rely on the system and [...] and build false trust”*.

Parents and teachers reported that youth asked GenAI about sensitive topics such as family conflicts, bullying, and other issues. For example, a youth confided about her parents' separation, and GenAI responded with an emotionally validating response which is identified by teacher T4 where the youth responded -

“Thank you for appreciating me and giving me this level of trust”

. Very young youth saw GenAI as a confidant as seven-year-old C5 shared *“I asked it, ‘How do I reconcile with my siblings?’”*. We asked C5 how she communicates with ChatGPT, *“sometimes I type, sometimes I use voice.”* Other youth participants similarly described using the voice feature to communicate with ChatGPT. C8 described using voice to practice language pronunciation

“I was taking Chinese classes, and I wanted to pronounce words [...] I used generative AI, and I would say, ‘I’m gonna pronounce a certain word, and you tell me if I pronounced it correctly’.”

This highlights how using the voice feature on ChatGPT made it more accessible to younger youth and posed greater risks to this vulnerable age group. Even more troubling was the effect on youth's view of **gender roles** and its influence on boys' relationship with their sister. P3 shared,

“My child asked ChatGPT, ‘Are men in charge of women?’ ChatGPT said, ‘Yes, men are in charge of women’.”

Such interactions can replace or affect family bonds and create a blind trust in GenAI, which can lead to isolation and weaken cultural norms about seeking support from a trusted circle of family members.

Emotional Attachment and Romanticization in Conflict with Social Values. Participants shared instances where GenAI's inability to accurately interpret cultural nuances resulted in advice that contradicts cultural expectations. X31 noted GenAI giving advice to disrespect mother,

“I consulted it [ChatGPT]; I had a problem that happened with my mom, and its response was shocking. It said my mom doesn't deserve respect.”

This advice contradicts Saudi cultural values that prioritize family unity and respect for parents and the elderly. Moreover, GenAI normalizing romantic behaviors conflicts with the values of Saudi's conservative society. As X33 reported, *“Talking to an AI is not inherently forbidden, but it can be a problem if it leads to flirtation, unhealthy attachment, or anything that negatively affects your heart and mind.”*

Misinterpreting Cultural Nuances. This type includes two low-level risks: (1) GenAI generating inaccurate content about culture or politics, and (2) GenAI confusing Arabic and Persian languages. Teachers observed these risks in classrooms; T1 noted students submitting projects with errors, such as Persian text instead of Arabic. This language confusion between Arabic and Persian may further misinform youth about their language and heritage. GenAI's western-centric responses are, moreover, reshaping youth's understanding of family values and cultural identity. Additionally, youth rely on GenAI for political news, which may provide culturally insensitive responses, as when C4 asked ChatGPT, *“I asked about the Iran-Israel war: ‘Who is in the right side?’”* This demonstrates how GenAI may reshape youth's understanding and worldview.

Data Privacy Violation Risk

This risk refers to the potential effect of GenAI facilitating or encouraging youth's disclosure of personal or family-related information in ways that conflict with Saudi cultural expectations of privacy. Such disclosures can harm family reputation and trust, and violate norms around protecting face, body, and family honor, as discussed in the social media context for woman's privacy (Abokhodair et al. 2017; Farooq et al. 2024). We present the privacy and safety risks, which are uniquely positioned within their cultural context.

GenAI Account Sharing Privacy Breach. This includes lower-level risks, such as (1) youth sharing GenAI accounts with family members, and (2) users sharing or buying GenAI accounts with strangers. In Saudi families, **privacy is relational and collective**, meaning an individual's disclosures reflect on the entire family. Participants, particularly parents and youths, mentioned using shared GenAI account.

From social media posts, users stated that they purchased shared ChatGPT accounts to reduce costs. A primary way they found such accounts was through Ads on X/Twitter and online stores in Saudi Arabia. This practice evidently leads to privacy issues, where private conversations and photos are exposed to strangers through shared accounts. Online users reported finding sensitive medical details and personal photos in **shared chat history**, which causes severe privacy breaches. X14 explained,

“I bought a shared account and found a girl's pictures in the chat history, and she probably doesn't know that I am a subscriber, too. I also found medical information.”

Similarly, in the interviews, we often found youth and parents using shared account, as reported by-

“I use my mom's account.” C2; “I shared ChatGPT with my mom, with my brother, and with my sister.” C5.

Participant also mentioned youth often sharing credential of their account with classmates, prompting P3 to later discuss the privacy risks of sharing account access with her child. We also observed tension within families where parents shared accounts with their children. Parents expected transparency, while youth wanted to keep certain sensitive topics private. As reported by a parent finding out her child was bullied through ChatGPT conversations shared with the child, P3 *“They tell it [ChatGPT] everything,[...] and about their friends bullying them.”* Meanwhile, C3 expressed discomfort and fear about parents viewing their chat history, highlighting the emotional impact of compromised privacy, *“That’s why I think ChatGPT shouldn’t save chats because they could contain personal stuff.”*

Disclosure of Personal Information. This category includes privacy risks such as the disclosure of personal information, photos, and location, and GenAI accessing the microphone and camera, among others. These risks arise because youth often perceive GenAI as a trusted companion, which lowers their guard when sharing sensitive information. The risk is further compounded when youth use shared accounts, as discussed in the previous category. Interviews and social media data reveal concerns that youth would disclose personal details such as names, age, health issues, and families conflicts. Unlike traditional online platforms, where youth are taught not to share private information, GenAI’s conversational design creates an illusion of confidentiality and one-to-one interaction. This makes youth feel safe sharing information they would normally keep private. Moreover, youth use voice input or camera access for convenience, as seven-year-old C5 and other youth (C1, C6, and C8) described using the microphone instead of typing (see Replacing Family Bonds). Such permissions significantly increase the risk of privacy violations. Parents and teachers worried about ChatGPT accessing sensitive phone data, cameras, and microphones. P1 cited a trend where youth treated GenAI as a companion, giving it an Arabic male name, Abdul Malik, and asking it for style advice. P1 explained

“I saw the ‘Abdul Malik’ trend where a girl asked ChatGPT which t-shirt to wear. It told her, ‘Open the camera so I can see’, and she did. For me, that was disturbing.”

Disclosure of Family Information. This include risks, such as GenAI facilitating the sharing of family information, photos, and private family matters. In conservative cultures like Saudi Arabia, where family honor is critical and highly protected, such disclosures can carry extreme consequences. Parents and teachers expressed concerns about youth’s use of GenAI, discussing how sharing images violates core cultural norms around modesty, gender boundaries, and the protection of the household. Parents taught youth to avoid sharing family images and information, as noted by C1 *“I shouldn’t send my photos or my family photos.”* C4 similarly explained, *“Don’t share your photos with it [ChatGPT], because suddenly the photos could end up spreading everywhere.”* C6 added,

“And we don’t send photos, in case someone gets

them or something. And some are of my family [photos]. [...] Don’t say anyone’s name, and about my family.”

Discussion

While prior work has emphasized privacy and bias risks in Western contexts (Yu et al. 2024, 2025; Zhang et al. 2023, 2025a; Young et al. 2024; Sharma, Zhou, and Berisha 2025), our findings surface culturally specific concerns, such as religious misinterpretation, collective privacy violations, and shared account practices that existing regulatory and technical frameworks are ill-equipped to address.

Regulatory Gaps in Youth Protection

Our regulatory analysis demonstrates that Saudi Arabia’s governance of GenAI operates across three layers, including foundational data and cybercrime law, SDAIA’s non-binding AI instruments, and pre-digital child welfare legislation, and that none of these layers was designed with GenAI-specific child safety in mind. This is not incidental. The Personal Data Protection Law (Saudi Data and Artificial Intelligence Authority (SDAIA) 2021) treats minors as ordinary data subjects, with no specific consent thresholds comparable to the EU’s GDPR Article 8 or the U.S. COPPA framework. Furthermore, the Anti-Cyber Crime Law addresses defamation and unauthorized access but contains no provisions for AI-generated content, deepfakes, or algorithmic harms. The Child Protection Law of 2014 is the only instrument built entirely around youth, yet it is oriented toward offline abuse. SDAIA’s AI-specific instruments, such as the AI Ethics Principles, Generative AI Guidelines, and AI Adoption Framework, recently established responsible use norms but remain non-binding, which partly address the public and government sectors rather than platform developers, and impose no child-safety-by-design obligations (SDAIA 2024a). This structural gap directly shapes the risk environment our participants navigated. Parents and teachers in our study were not aware of any regulatory resources that might mandate age verification or content filtering for religiously or culturally sensitive topics. Their responses were therefore improvised by restricting access, sharing accounts to enable monitoring, or simply hoping that communal norms would compensate for absent institutional safeguards. The risks we identify empirically are, in part, the predictable product of this governance vacuum.

Religion and Cultural Norms: Risks That Regulation Does Not See

Much of the existing research on youth and GenAI is from Western contexts (Yu et al. 2024, 2025; Zhang et al. 2023, 2025a; Young et al. 2024), where concerns center on privacy, bias, and over-reliance on AI for learning. Our findings mirror some of these concerns, such as youth dependency, trust inflation, and parasocial attachment, but also reveal distinct cultural dimensions that fall entirely outside the conceptual vocabulary of current Saudi regulatory instruments.

Saudi Arabia’s AI Ethics Principles include provisions on fairness and human oversight, and the Generative AI Guide-

lines require disclosure of AI-generated content, yet neither instrument contemplates what our participants identified as a primary risks, such as GenAI providing inaccurate or contextually inappropriate interpretations of Islamic religious principles and moral norms. Parents expressed strong concern about youth querying ChatGPT for religious rulings (Fatwas), interpretations of Qur'anic verses, or guidance on moral conduct, interactions that, when mediated by a conversational AI prone to hallucination, could weaken youth's religious foundation and bypass the authority of family and religious educators.

On one hand, parents and teachers firmly rejected AI as a legitimate source of religious instruction, framing it as a threat to moral development and to the authority of families and religious leaders. On the other hand, this blanket rejection risked silencing youth's authentic questions about faith, inadvertently pushing them to seek answers elsewhere, often unsupervised and online. This highlights the delicate balance between protecting youth from misinformation and fostering space for safe, guided exploration of spiritual topics.

Shared Accounts & Socioeconomic Pressures

A novel and particularly acute finding of this study concerns shared GenAI account practices. Participants, both within families and across strangers purchasing shared subscriptions, reported that cost-saving behaviors created serious secondary privacy risks, such as sensitive medical records, personal photographs, and private family communications appearing in shared chat histories accessible to unknown parties. Online users reported purchasing shared ChatGPT subscriptions through advertisements on X/Twitter and at online local stores, and discovered intimate information from prior users.

This finding exposes a structural inadequacy in the PDPL's treatment of data subjects (Saudi Data and Artificial Intelligence Authority (SDAIA) 2021). Saudi Arabia's data protection framework, modeled broadly on international standards, presupposes a one-to-one relationship between a user and an account. Principles such as data minimization, purpose limitation (Sharma et al. 2024), and consent all operate on the assumption that personal data belongs to a single identifiable subject. Furthermore, many Saudi parents mentioned that they shared the same ChatGPT account with their children and would monitor their children's interactions with ChatGPT. Parents framed shared access as a way to monitor and control their children's use of GenAI, but this often undermined privacy and blurred accountability. This highlights, the absence or lack of awareness of parental controls, leading parents to instead rely on sharing accounts with their youth.

Design Implications

Based on Saudi parents' and teachers' expectations of control and GenAI safe use, we suggest the following implications to implement in GenAI parental controls for the safety of this population.

Culturally Adaptive Safety Controls. A primary concern of Saudi parents and teachers is the lack of culturally

adaptive GenAI controls. Parents emphasized the need to have control over child-GenAI interactions, to make sure these interactions follow expected traditional norm and religious values. Parents and teachers reported several cases where GenAI responses lacked understanding of correct religious interpretations, or cultural nuance, and misaligned with family values. Moreover, the existing GenAI mediation tools are insufficient (Yu et al. 2024), especially in religious and cultural contexts. The GenAI tools need to have content-filtering content features where parents manage what topics are allowed to be discussed with their child, to ensure that topics adhere to cultural expectations. This is also suggested by Western parents, who want the ability to filter content based on child age and development stage (Yu et al. 2024). Safety tools should take into account diverse cultural norms and values when filtering content, and families should be able to adjust them to meet their needs.

AI Literacy in Local Languages. Parents and teachers emphasized the urgent need for AI literacy programs tailored to Arabic-speaking youth. While Western countries are advancing AI literacy initiatives for youth, and extensive research has been done in this area (Williams et al. 2024; Lee et al. 2021; Ali et al. 2019; Williams et al. 2022; Ravi et al. 2023; Jia et al. 2025; Touretzky et al. 2019b,a; Hollands et al. 2024), the absence of similar efforts in the Middle East risks widening the digital divide. Teaching youth how GenAI works, and how data is collected and processed, and the consequences of sharing information is critical for mitigating GenAI's unique risks. AI literacy in Arabic, or another local language, should be a core competency integrated into school curricula to equip youth with the necessary skills for safe engagement with GenAI.

Platform-Level Responsibilities. Parent and teachers repeatedly expressed that this AI technology is "not from here", indicating that the development of GenAI is unfamiliar with local expectations and culturally disconnected. There should be platform-level features to regularly collect feedback from users from different cultures to enhance their experience while maintaining safety. For example, to supporting safe shared-use practices, parents wanted GenAI to disable the camera and microphone features for youth.

Conclusion

Our work presents one of the initial analyses of youth's interactions with generative AI in Saudi Arabia, illustrating how cultural, religious, and social factors influence both the benefits and drawbacks. While youth utilize AI tools for educational purposes, creativity, and amusement, parents and educators voice significant concerns regarding religious impact, privacy, excessive attachment, and trust. The expanded risks taxonomy and results underscore the critical necessity for culturally appropriate safeguards, such as enhanced parental controls, defined limits around sensitive subjects, and the integration of AI literacy into educational programs. By prioritizing non-Western viewpoints, this study provides insights for inclusive and context-aware privacy and safety design for youth using Generative AI.

Ethical Statement

Our study involved youth (ages 7–17) and was approved by our institution’s IRB, and we obtained informed parental consent alongside child assent prior to all youth interviews. All interview recordings, transcriptions, and translations were handled by a native Arabic-speaking research team member and stored securely to protect participant confidentiality, particularly given the sensitivity of disclosures involving religious belief, family conflict, etc.

In this paper, first author is a native Arabic speaker and Saudi national with lived experience of the cultural, religious, and family norms examined in this work which informs recruitment, interview design, and interpretation. Our second author brings expertise in AI safety and risk taxonomies, primarily from cross cultural contexts. We mitigated potential bias through iterative collaborative coding and regular discussions to reach consensus on culturally grounded interpretations.

Acknowledgments

We thank all parents, teachers, and youth who participated in this study and shared their valuable experiences of using Generative AI in Saudi Arabia. We are grateful to our colleagues and reviewers for their thoughtful feedback on earlier drafts of this paper. We also thank the broader research community whose prior work on youth-GenAI risk taxonomies informed and motivated this study.

References

- Abokhodair, N.; Abbar, S.; Vieweg, S.; and Mejova, Y. 2017. Privacy and social media use in the Arabian Gulf: Saudi Arabian & Qatari traditional values in the digital world. *The Journal of Web Science*, 3.
- Alashwali, E.; and Alashwali, F. 2022. Saudi Parents’ Privacy Concerns about Their Children’s Smart Device Applications. *International Journal of Child-Computer Interaction*, 33: 100486.
- Aleisa, N.; Renaud, K.; and Bongiovanni, I. 2020. The privacy paradox applies to IoT devices too: A Saudi Arabian study. *Computers & Security*, 96: 101897.
- Alfakeh, S. A.; Alghamdi, A.; Kouzaba, K. A.; Altaifi, M.; Abu-Alamah, S.; and Salamah, M. M. 2021. Parents’ perception of cyberbullying of their children in Saudi Arabia. *Journal of Family & Community Medicine*, 28: 117 – 124.
- Alhamawi, L. 2025. 80% Saudi adults reported using AI tools, with 1 in 3 using them regularly. Arab News.
- Ali, M.; Arunasalam, A.; and Farrukh, H. 2025. Understanding Users’ Security and Privacy Concerns and Attitudes Towards Conversational AI Platforms. In *2025 IEEE Symposium on Security and Privacy (SP)*, 298–316.
- Ali, S.; Payne, B. H.; Williams, R.; Park, H. W.; and Breazeal, C. 2019. Constructionism, Ethics, and Creativity: Developing Primary and Middle School Artificial Intelligence Education. In *International Workshop on Education in Artificial Intelligence K-12 (EDUAI’19), held in conjunction with IJCAI 2019*, 1–4. Macao, China.
- Almethen, A. A.; and Alomair, M. A. 2024. The role of age, grade level, and subject area in determining the inclusion of digital citizenship elements in elementary school curricula: Perspectives of teachers in the Kingdom of Saudi Arabia. *Heliyon*, 10.
- Alorini, A.; Wu, Y.; Bin Sawad, A.; Prasad, M.; and Kocaballi, A. B. 2025. Unplug, Mute, Avoid: Investigating smart speaker users’ privacy protection behaviours in Saudi Homes. *arXiv preprint arXiv:2508.04202*.
- Alqahtani, N.; Furnell, S.; Atkinson, S.; and Stengel, I. 2017. Internet risks for children: Parents’ perceptions and attitudes: An investigative study of the Saudi Context. *2017 Internet Technologies and Applications (ITA)*, 98–103.
- Alqarni, A. M.; et al. 2024. Perceived usefulness of Internet of Things (IoT) in the lives of Saudi users. *PLOS ONE*, 19(1): e0280521.
- AlShabibi, A.; and Al-Suqri, M. N. 2021. Cybersecurity Awareness and Its Impact on Protecting Children in Cyberspace. *2021 22nd International Arab Conference on Information Technology (ACIT)*, 1–6.
- Alsiyat, Y.; Chang, Y.; Zhang, N.; and Flechais, I. 2025. Smart Spaces, Private Lives: A Culturally Grounded Study of Privacy Behaviors in Saudi Homes. In *Proceedings of the 21st Symposium on Usable Privacy and Security (SOUPS 2025)*. USENIX Association.
- Apify. 2025. Twitter Search Scraper. <https://apify.com/epctex/twitter-search-scraper>. Accessed: 2025-04-15.
- ArabNews. 2022. Snapchat and UAE Ministry launch new parental control feature. ArabNews.
- Braun, V.; and Clarke, V. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2): 77–101.
- Cao, J.; Das, A.; Emami-Naeini, P.; et al. 2024. Understanding parents’ perceptions and practices toward children’s security and privacy in virtual reality. In *2024 IEEE Symposium on Security and Privacy (SP)*, 1554–1572. IEEE.
- Choi, D.; Hong, S.; Park, J.; Chung, J. J. Y.; and Kim, J. 2024. CreativeConnect: Supporting Reference Recombination for Graphic Design Ideation with Generative AI. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI ’24. New York, NY, USA: Association for Computing Machinery. ISBN 9798400703300.
- Communications, Space and Technology Commission (CST). 2024. Saudi Internet Report 2024. <https://www.cst.gov.sa/en/media-center/news/N2025051200>. Pp. 1–51.
- Critch, A.; and Russell, S. 2023. TASRA: a taxonomy and analysis of societal-scale risks from AI. *arXiv preprint arXiv:2306.06924*.
- DataReportal - Global Digital Insights. 2025. X Users, Stats, Data & Trends for 2025. <https://datareportal.com/essential-x-stats>.
- Davis, N.; Sherson, J.; and Rafner, J. 2025. The Co-Creative Design Framework for Hybrid Intelligence. In *Proceedings of the 2025 Conference on Creativity and Cognition, C&C ’25*, 560–572. New York, NY, USA: Association for Computing Machinery. ISBN 9798400712890.

- Dwairy, M.; Achoui, M.; Abouserie, R.; Farah, A.; Sakhleh, A. A.; Fayad, M.; and Khan, H. K. 2006. Parenting Styles in Arab Societies: A First Cross-Regional Research Study. *Journal of Cross-Cultural Psychology*, 37.
- e&. 2025. e& Puts Online Safety First with New Parental Control Service. <https://www.eand.com/en/news/23-6-25-eand-puts-online-safety-first-with-new-parental-control-service.html>. Accessed: 2025-08-29.
- Ebrahimi, E.; Pare, M.; Stoker, G.; and White, S. 2025. Cybersecurity Early Education: A Review of Current Cybersecurity Education for Young Children. In *International Conference on Computer Supported Education*.
- Farooq, A.; Salminen, J.; Martin, J. D.; Aldous, K.; Jung, S.-G.; and Jansen, B. J. 2024. Exploring social media privacy concerns: a comprehensive survey study across 16 Middle Eastern and North African countries. *IEEE Access*, 12: 147087–147105.
- Farooq, S.; Abdullatif, M.; and Altheeb, A. 2023. Prevalence of cyberbullying and associated factors among adolescents in Dubai schools: complex design survey – 2019. *International Journal of Adolescence and Youth*, 28.
- Fortim, I.; and Zandavalli, S. D. 2025. RIGHTS.AI: Children's Experiences of Generative Artificial Intelligence in Brazil. Technical report, Digital Futures for Children centre, 5Rights Foundation, London, UK.
- Google. 2025a. Help keep your family safer online. Accessed: 2025-08-29.
- Google. 2025b. Manage your child's access to Gemini Apps. Accessed: 2025-08-29.
- Hollands, F. M.; DiPaola, D.; Breazeal, C.; and Ali, S. 2024. AI Mastery May Not Be for Everyone, But AI Literacy Should Be. In *Proceedings of the 2024 on ACM Virtual Global Computing Education Conference V. 1*.
- HSF Kramer. 2025. Saudi Arabia: AI Tracker. <https://www.hsfkramer.com/insights/reports/ai-tracker/saudi-arabia>. Accessed: 2026-05-21.
- Jia, K.; Leung, T. H. M.; Cheung, N. Y. I.; Li, Y.; and Yu, J. 2025. Developing a Holistic AI Literacy Framework for Children. *ACM Transactions on Computing Education*, 25: 1–30.
- Kaberi, J.; Busisa, A.; and Kabata, D. 2025. RIGHTS.AI: Children's Experiences of Generative Artificial Intelligence in Kenya. Technical report, Digital Futures for Children centre, 5Rights Foundation, London, UK.
- Kasturi, S.; Mehta, C.; and Raman, U. 2025. RIGHTS.AI: Children's Experiences of Generative Artificial Intelligence in India. Technical report, Digital Futures for Children centre, 5Rights Foundation, London, UK.
- Lee, I. A.; Ali, S.; Zhang, H.; DiPaola, D.; and Breazeal, C. 2021. Developing Middle School Students' AI Literacy. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*.
- Livingstone, S.; Kirwil, L.; Ponte, C.; and Staksrud, E. 2014. In their own words: What bothers children online? *European Journal of Communication*, 29(3): 271–288.
- Martin, A. P.; Keskes, H.; Heni, W.; Burton, P.; Deliege, A.; Rabhi, C.; and Ayari, R. 2023. From entertainment to extortion: Tunisian children's online experiences. *Child Abuse & Neglect*, 145: 106405.
- McDonald, N.; Schoenebeck, S.; and Forte, A. 2019. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for CSCW and HCI practice. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW): 1–23.
- Mian, T.; and Alatawi, E. 2023. Investigating How Parental Perceptions of Cybersecurity Influence Children's Safety in the Cyber World: A Case Study of Saudi Arabia. *Intelligent Information Management*, 15: 350–372.
- Moubarak, H. F. A.; and Afthanorhan, A. 2024. Risk assessments of virtual interactions on Saudi families. *Humanities and Social Sciences Communications*, 11: 1–9.
- Nemko Digital. 2025. AI Regulation in Saudi Arabia: Laws, Standards & Compliance. <https://digital.nemko.com/regulations/ai-regulation-in-saudi-arabia>. Accessed: 2026-05-21.
- Neugnot-Cerlioli, M.; and Laurenty, O. M. 2024. The future of child development in the AI era. Cross-disciplinary perspectives between AI and child development experts. *arXiv preprint arXiv:2405.19275*.
- OpenAI. 2025. Building more helpful ChatGPT experiences for everyone. OpenAI.
- Othman, B. E. M.; Abdelhai, S. M. Z.; Ibrahim, M. I. T.; Hamed, A. E. M.; Barakat, A. M.; and Elsayed, E. E. 2025. Unravelling cyberbullying among Egyptian adolescents: the protective influence of self-control and moral identity with gender and socioeconomic dynamics. *BMC Psychology*, 13.
- Panhans, D.; Hoteit, L.; Yousuf, S.; Breward, T.; AlFaadhel, A.; and AlShaalan, B. 2022. Why Children Are Unsafe in Cyberspace. BCG.
- PRAW. 2025. PRAW: The Python Reddit API Wrapper. <https://praw.readthedocs.io/en/stable/>. Accessed: 2025-04-15.
- Ravi, P.; Broski, A.; Stump, G.; Abelson, H.; Klopfer, E.; and Breazeal, C. 2023. Understanding Teacher Perspectives and Experiences after Deployment of AI Literacy Curriculum in Middle-school Classrooms. *arXiv preprint arXiv:2312.04839*.
- Reddit. 2025. Reddit User Agreement. <https://redditinc.com/policies/user-agreement>.
- Rehima, M. H.; Al-Tkaynehb, K. M.; and Al-Shamsc, W. K. 2020. The Factors of Cyberbullying among the Basic Stage Students in the City of Al-Ain: A study on a Sample of Parents in The City of Al-Ain. *International Journal of Innovation, Creativity and Change*.
- Saudi Data and Artificial Intelligence Authority (SDAIA). 2021. Personal Data Protection Law. Royal Decree No. M/19. As amended by Royal Decree No. M/148 (2023). Saudi Data and Artificial Intelligence Authority. Accessed: 2026-05-20.
- Saudi Gazette. 2019. Majority of parents worried about their child's online safety. <https://saudigazette.com.sa/article/578232>. Survey reported in Saudi Gazette with Kaspersky.

- Saudi Ministry of Human Resources and Social Development (MHRSD). 2014. Child Protection Law. Royal Decree No. M/14. Accessed: 2026-05-20.
- Saudi Ministry of Interior. 2007. Anti-Cyber Crime Law. Royal Decree No. M/17. Accessed: 2026-05-20.
- Sağlam, R. B.; Miller, V.; and Franqueira, V. N. L. 2023. A Systematic Literature Review on Cyber Security Education for Children. *IEEE Transactions on Education*, 66(3): 274–286.
- SDAIA. 2023. AI Ethics Principles. <https://sdaia.gov.sa/>. Saudi Data and Artificial Intelligence Authority. Accessed: 2026-05-20.
- SDAIA. 2024a. AI Adoption Framework. <https://sdaia.gov.sa/>. Saudi Data and Artificial Intelligence Authority. Accessed: 2026-05-20.
- SDAIA. 2024b. Generative AI Guidelines for Government Entities. <https://sdaia.gov.sa/>. Saudi Data and Artificial Intelligence Authority. Accessed: 2026-05-20.
- SDAIA. 2024c. Generative AI Guidelines for the Public. <https://sdaia.gov.sa/>. Saudi Data and Artificial Intelligence Authority. Accessed: 2026-05-20.
- Shakir, U. 2025. Google is going to let kids use its Gemini AI. The Verge.
- Sharevski, F.; and Vander Loop, J. 2024. Children, parents, and misinformation on social media. In *2024 IEEE Symposium on Security and Privacy (SP)*, 1536–1553. IEEE.
- Sharma, T.; Kyi, L.; Wang, Y.; and Biega, A. J. 2024. "I'm not convinced that they don't collect more than is necessary": {User-Controlled} Data Minimization Design in Search Engines. In *33rd USENIX Security Symposium (USENIX Security 24)*, 2797–2812.
- Sharma, T.; Zhou, Y.; and Berisha, V. 2025. PRAC3 (Privacy, Reputation, Accountability, Consent, Credit, Compensation): Voice Actors in AI Data-Economy. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, volume 8, 2360–2372.
- Slattey, P.; Saeri, A. K.; Grundy, E. A.; Graham, J.; Noetel, M.; Uuk, R.; Dao, J.; Pour, S.; Casper, S.; and Thompson, N. 2024. The AI risk repository: A comprehensive meta-review, database, and taxonomy of risks from artificial intelligence. *arXiv preprint arXiv:2408.12622*.
- Su, J.; Ng, D. T. K.; and Chu, S. K.-W. 2023. Artificial Intelligence (AI) Literacy in Early Childhood Education: The Challenges and Opportunities. *Computers and Education: Artificial Intelligence*, 4: 100124.
- Sun, K.; Zou, Y.; Radesky, J.; Brooks, C.; and Schaub, F. 2021. Child safety in the smart home: parents' perceptions, needs, and mitigation strategies. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2): 1–41.
- Teimouri, M.; Hassan, M. S.; Griffiths, M. D.; Benrazavi, S. R.; Bolong, J.; Daud, A.; and Adzharuddin, N. A. 2016. Assessing the Validity of Western Measurement of Online Risks to Children in an Asian Context. *Child Indicators Research*, 9: 407–428.
- Touretzky, D. S.; Gardner-McCune, C.; Breazeal, C.; Martin, F. G.; and Seehorn, D. W. 2019a. A Year in K-12 AI Education. *AI Magazine*, 40: 88–90.
- Touretzky, D. S.; Gardner-McCune, C.; Martin, F. G.; and Seehorn, D. W. 2019b. Envisioning AI for K-12: What Should Every Child Know about AI? In *AAAI Conference on Artificial Intelligence*.
- UNICEF. 2019. Digital Literacy for Children Exploring definitions and frameworks. UNICEF.
- Williams, O.; Choong, Y.-Y.; and Buchanan, K. 2023. Youth understandings of online privacy and security: A dyadic study of children and their parents. In *Nineteenth Symposium on Usable Privacy and Security (SOUPS 2023)*, 399–416.
- Williams, R.; Ali, S.; Alcantara, R.; Burghle, T.; Alghowinem, S.; and Breazeal, C. 2024. Doodlebot: An Educational Robot for Creativity and AI Literacy. In *2024 19th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 772–780.
- Williams, R.; Ali, S.; Devasia, N.; DiPaola, D.; Hong, J.; Kaputsos, S. P.; Jordan, B.; and Breazeal, C. 2022. AI + Ethics Curricula for Middle School Youth: Lessons Learned from Three Project-Based Curricula. *International Journal of Artificial Intelligence in Education*, 1–59.
- X. 2025. About Parental Consent on X. <https://help.x.com/en/using-x/parental-consent>.
- Young, J.; Jawara, L. M.; Nguyen, D. N.; Daly, B. P.; Huh-Yoo, J.; and Razi, A. 2024. The Role of AI in Peer Support for Young People: A Study of Preferences for Human- and AI-Generated Responses. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*.
- Yu, Y.; Liu, Y.; Zhang, J.; Huang, Y.; and Wang, Y. 2025. Understanding Generative AI Risks for Youth: A Taxonomy Based on Empirical Data. *arXiv preprint arXiv:2502.16383*.
- Yu, Y.; Sharma, T.; Hu, M.; Wang, J.; and Wang, Y. 2024. Exploring Parent-Child Perceptions on Safety in Generative AI: Concerns, Mitigation Strategies, and Design Implications. *arXiv preprint arXiv:2406.10461*.
- Zawya. 2018. Children in Saudi Arabia among Youngest in the World to Receive First Mobile Device: Norton Research.
- Zhang, S.; Cagiltay, B.; Li, J.; Sullivan, D.; Mutlu, B.; Kirkorian, H.; and Fawaz, K. 2025a. Exploring Families' Use and Mediation of Generative AI: A Multi-User Perspective. *arXiv preprint arXiv:2504.09004*.
- Zhang, Y.; Zhao, D.; Hancock, J. T.; Kraut, R.; and Yang, D. 2025b. The Rise of AI Companions: How Human-Chatbot Relationships Influence Well-Being. *arXiv:2506.12605*.
- Zhang, Z.; Jia, M.; Yao, B.; Das, S.; Lerner, A.; Wang, D.; and Li, T. 2023. It's a fair game, or is it? Examining how users navigate disclosure risks and benefits when using LLM-based conversational agents. *arXiv preprint arXiv:2309.11653*.