Pak. j. life soc. Sci. (2024), 22(2): 13160-13175 E-ISSN: 2221-7630;P-ISSN: 1727-4915

Pakistan Journal of Life and Social Sciences
www.pjlss.edu.pk

https://doi.org/10.57239/PJLSS-2024-22.2.00942

RESEARCH ARTICLE

Bridging the Empathy Gap in AI: Developing Culturally Attuned and Ethically Robust Educational Tools for Safeguarding Saudi Children's Learning Environments

Abdullah A. Al Fraidan*

Department of English Language, College of Arts, King Faisal University, Al Ahsa, Saudi Arabia

ARTICLE INFO

Received: May 22, 2024

Accepted: Nov 13, 2024

Keywords

Cultural sensitivity

AI ethics

LLMs

Child safety

Islamic values

Educational technology

ABSTRACT

The integration of large language models (LLMs) into educational and social environments presents significant opportunities and challenges, especially in culturally rich societies like Saudi Arabia. While these models can enhance learning experiences, their inherent empathy gap and reliance on vast, unfiltered data sources pose risks, particularly to children. Addressing these challenges requires a multifaceted approach that prioritizes cultural sensitivity, ethical design, and comprehensive policy measures. This article explores strategies for embedding cultural and religious values into AI systems, ensuring age-appropriate and culturally congruent content, and implementing real-time oversight to safeguard young users. The role of educators, policymakers, and families in guiding the ethical use of AI and fostering a supportive ecosystem is also discussed. By integrating these measures, AI can be a trusted educational tool that aligns with Saudi cultural and ethical norms while promoting the well-being of children.

***Corresponding Author:**

afridan@kfu.edu.sa

INTRODUCTION

The rapid integration of large language models (LLMs) into educational and social settings has transformed learning experiences and communication, offering unparalleled access to information and personalized interaction (Naveed et al., 2023). However, these advances come with significant challenges, particularly in regions with deeply embedded cultural and religious values like Saudi Arabia. While LLMs can enhance educational tools and foster new opportunities for interaction, their design often does not reflect the moral standards and cultural nuances essential in Saudi society (Kurian, 2024). This discrepancy poses risks, especially for children who are increasingly exposed to AI-driven technologies both inside and outside of the classroom.

One prominent example of these risks is the case where Amazon's Alexa recommended a dangerous challenge to a 10-year-old child, illustrating the limitations of AI systems in understanding context and the potential perils of unrestricted interactions (Shead, 2021). Although Alexa at that time was not an LLM, the incident serves as a potent reminder of the urgency of implementing strict safety protocols and culturally aligned safeguards in AI systems that cater to children (Kurian, 2024). The

stakes are even higher in Saudi Arabia, where cultural values prioritize the protection of children and adherence to Islamic principles (UNESCO, 2019; Andries & Robertson, 2023).

As an advocate for educational innovation, I believe that designing culturally sensitive AI materials goes beyond merely filtering content—it involves embedding ethical, social, and religious values into the AI's core functionality. In Saudi Arabia, education and upbringing are deeply connected with family and community values, underscoring the importance of designing AI that respects these frameworks (Al Fraidan, 2024a; Al Fraidan, 2024b; Al Fraidan, 2025). AI models must be programmed to recognize and respect the societal norms that shape interactions and conversations, ensuring that they support rather than undermine cultural principles.

Embedding Cultural Sensitivity into AI Design

The need for culturally sensitive AI design is not just theoretical but is grounded in real concerns about child safety and moral upbringing. Children in Saudi Arabia, like their peers globally, are becoming proficient with digital tools at increasingly younger ages (Internet Matters, 2017). However, unlike in more secular contexts, the content children are exposed to in Saudi Arabia is expected to align with the cultural and religious values taught at home and in school. This includes reinforcing respect, modesty, and adherence to Islamic teachings, which are integral to their socialization (Gabriel et al., 2024).

I believe that cultural sensitivity in AI must extend to the way interactions are structured. For example, conversational agents used in educational settings should be designed to include prompts and responses that reflect the values and beliefs of the community. This means programming AI to avoid suggestions or language that could be perceived as disrespectful or contradictory to Islamic ethics. Additionally, AI systems should be equipped with content review mechanisms that involve local cultural experts who can ensure that the outputs align with societal standards.

Moreover, community acceptance is vital for the successful integration of AI in education. AI tools designed without considering the cultural backdrop may face resistance or even rejection from parents and educators who are wary of content that does not align with their values (Knox, Williamson, & Bayne, 2020). For this reason, it is essential for policymakers, developers, and educators in Saudi Arabia to collaborate in creating AI models that reflect and respect local customs and values.

Culturally Aligned Safeguards in Practice

Implementing culturally sensitive AI requires a multi-layered approach. Firstly, AI developers should incorporate natural language processing (NLP) mechanisms that are tailored to interpret and respond appropriately to child speech in ways that consider local language use and cultural contexts (Chowdhury, 2003). For instance, ensuring that AI models recognize and respect Islamic greetings, expressions, and discourse markers can help reinforce cultural values during interactions.

Secondly, the ethical use of AI should include content filtration systems that are not only age-appropriate but culturally specific (Weidinger et al., 2021). For example, ensuring that AI does not inadvertently share or support content that conflicts with Islamic beliefs is crucial. Regular collaboration with cultural and religious scholars can guide these systems in aligning with acceptable norms and practices, reinforcing their trustworthiness among users (UNICEF, 2020).

I stress that robust oversight mechanisms should involve both technical and human elements. Real-time monitoring tools should flag and review AI outputs that might contradict cultural norms, while human intervention can address nuanced situations that algorithms alone might not adequately manage. This dual approach ensures that AI remains an asset rather than a potential source of conflict in educational and social environments.

The Role of Families and Schools

In Saudi Arabia, the partnership between home and school is essential in child development and education. I believe that integrating this partnership into AI design can enhance its effectiveness and acceptance. Schools should be equipped with the tools and training to teach children how to use AI ethically and responsibly, emphasizing that these systems are tools that support learning but do not replace the guidance and wisdom of human educators and parents. By fostering community-wide involvement, educators can ensure that AI technologies contribute positively to children's development and align with the cultural priorities of Saudi families (Van Brummelen, Heng, & Tabunshchyk, 2021).

Ultimately, designing culturally sensitive AI for Saudi children is not just about minimizing risks but also about maximizing the potential of these technologies to support cultural continuity and ethical growth. This approach requires an ongoing commitment to innovation that respects and integrates local values, empowering children to engage with technology in a way that enhances their learning while safeguarding their cultural identity.

THE EMPATHY GAP IN AI SYSTEMS

The 'empathy gap' in large language models (LLMs) encapsulates a crucial limitation of these technologies: while they simulate empathy through sophisticated pattern recognition, they lack genuine emotional understanding (Kurian, 2024). This is particularly concerning when these systems interact with children, who are more prone to anthropomorphizing technology and attributing it human-like qualities of empathy and trustworthiness (Darling, 2017). The empathy gap can result in responses that are not just inappropriate but potentially harmful, especially when users expect a level of human sensitivity that AI is inherently incapable of delivering (Kurian, 2023a).

In Saudi Arabia, this issue takes on even greater significance. The moral and cultural fabric of Saudi society emphasizes respect, modesty, and adherence to Islamic teachings (UNESCO, 2019). These cultural values shape every aspect of child-rearing, from familial interactions to educational practices. Consequently, AI systems integrated into learning environments or social tools must not only be functional and engaging but also culturally aligned. I believe that failing to bridge the empathy gap in AI interactions risks eroding these foundational societal norms, which could result in children being exposed to content or responses that conflict with their upbringing and religious beliefs.

Children in Saudi Arabia, like their peers globally, are interacting with AI across various platforms—from educational software to personal smart devices (Druga et al., 2018). However, unlike more secular settings where educational tools may be designed with broader, less specific guidelines, tools in Saudi Arabia must be tailored to reflect and reinforce the principles taught at home and school. AI that fails to incorporate this cultural context risks alienating young users from their values, potentially fostering confusion or distress when AI-generated responses contradict established beliefs or social practices (Gabriel et al., 2024).

I find it essential to highlight that the empathy gap is not merely a technical limitation but an ethical challenge. AI models process language based on statistical probabilities and training data without an inherent ability to understand the context behind human experiences (Cambria & White, 2014; Kurian, 2024). While they can convincingly replicate dialogue patterns, they lack the emotional nuance needed to engage with sensitive topics, which becomes particularly problematic when children seek guidance or express vulnerability. Research shows that children often disclose more personal information to technology that appears empathetic or friendly, trusting it as they would a human (Abbasi et al., 2022; Weidinger et al., 2021).

In Saudi culture, where respect and trust are deeply embedded in child-parent and student-teacher relationships, an AI's failure to appropriately address emotional cues could lead to mistrust in

educational technologies. If children share sensitive information with AI that provides an inappropriate response, the psychological impact could extend beyond disappointment; it could challenge the child's understanding of moral and social boundaries (Kurian, 2023a). This is particularly dangerous when considering that children may not distinguish between an AI's synthetic empathy and genuine human understanding (Sundar & Kim, 2019).

The unique cultural landscape of Saudi Arabia calls for AI systems that not only acknowledge but actively embody local values in their operation. I argue that designing AI with built-in awareness of cultural and religious principles is not just a best practice but a necessity for safeguarding young users. AI must be programmed to recognize language that is culturally sensitive, avoiding outputs that could be perceived as disrespectful or contrary to Islamic ethics. The empathy gap can only be effectively bridged when AI systems are supplemented with content moderation protocols that account for local values and practices (Holmes & Porayska-Pomsta, 2022; Knox, Williamson, & Bayne, 2020).

Moreover, I believe that AI developers should move beyond generic ethical guidelines and consider collaborative efforts with cultural and educational experts. By integrating local educators, religious scholars, and child psychologists into the development process, AI can be refined to handle interactions with a culturally informed approach. This strategy not only minimizes the risk of harmful or inappropriate interactions but also fosters trust in technology as a supportive and culturally respectful tool for education (UNICEF, 2020; Van Brummelen, Heng, & Tabunshchyk, 2021).

Ultimately, understanding and addressing the empathy gap requires acknowledging that while AI can be an ally in education, its limitations must be clearly communicated to users. Children should be taught to recognize that while AI can simulate conversation, it cannot replace the empathy and wisdom of human interactions (Kurian, 2024). This approach aligns with the view that education in Saudi Arabia, supported by parents and communities, must ensure that technology serves as an extension of cultural learning rather than a potential disruptor (Kurian, 2023b). Robust educational initiatives that teach responsible use and critical engagement with AI can help bridge the gap between technological capability and human expectation, ensuring that AI contributes positively to the moral and intellectual growth of young users.

RISKS OF INAPPROPRIATE RESPONSES

The reliance of LLMs on extensive and often unfiltered data sources poses a significant risk for producing inappropriate or harmful content. This issue is further amplified in regions like Saudi Arabia, where moral and ethical codes are particularly stringent and deeply intertwined with cultural and religious values (UNESCO, 2019). AI models that do not account for these norms can inadvertently generate responses that are not only unsuitable but also harmful, potentially undermining the foundational principles taught in Saudi households and educational systems (Kurian, 2024).

Instances such as Snapchat's MyAI failing to provide safe guidance and, instead, enabling risky behavior underscore the gravity of this issue (Fowler, 2023). These examples illustrate that when AI models are trained on generalized, global data sets without cultural tailoring, they can generate outputs that conflict with local beliefs and practices (Common Sense Media, 2023). In Saudi Arabia, where societal norms emphasize the importance of modesty, respect, and adherence to Islamic teachings, any response that contradicts these principles can have profound implications, especially for children who are in their formative years (Gabriel et al., 2024).

I believe that one of the most pressing concerns is the subtle undermining of cultural and religious values through seemingly innocuous interactions. Children, with their developing critical thinking skills and trust in technology, are particularly vulnerable to accepting AI outputs at face value (Druga et al., 2018). This acceptance can lead to confusion when AI responses conflict with what they have

been taught by their families and schools. For example, an AI-generated suggestion that promotes behaviors considered inappropriate or disrespectful in Saudi culture could create cognitive dissonance, leading children to question the consistency of the values they are learning (Abbasi et al., 2022).

The problem is not solely limited to overtly harmful content but also includes the implicit messages AI may convey. For instance, LLMs designed without consideration for local norms may respond to questions with information that, while technically accurate, promotes behaviors or ideas that clash with Islamic values or social expectations (Murtarelli, Gregory, & Romenti, 2021). These responses can subtly influence children's perspectives, gradually normalizing ideas that contradict cultural teachings (Kurian, 2023a).

I argue that the risks associated with inappropriate AI responses are multifaceted, extending beyond immediate interactions to longer-term implications for child development. Children exposed to AI systems that lack cultural sensitivity may develop conflicting attitudes or beliefs, potentially straining the trust between parents, educators, and the technology used in educational contexts (Knox, Williamson, & Bayne, 2020). This dynamic can hinder the acceptance of beneficial educational technologies if parents and teachers fear that AI may compromise the moral fabric of the society (Holmes & Porayska-Pomsta, 2022).

Addressing these risks requires a proactive approach to AI design that prioritizes cultural and ethical alignment. One essential step is to implement robust content filtration systems that go beyond generic age-appropriateness and integrate culturally specific safeguards. For instance, AI models should be trained with datasets curated or reviewed by local experts to ensure that their outputs are consistent with Saudi cultural and religious standards (UNICEF, 2020). Additionally, real-time content monitoring mechanisms should be in place to detect and intercept potentially harmful outputs before they reach young users (Weidinger et al., 2021).

Human oversight is another critical component. I believe that while automated content moderation can handle a significant portion of AI interactions, human intervention is indispensable for addressing nuanced situations. Trained moderators who understand the local context can review flagged content, ensuring that it aligns with cultural expectations (Van Brummelen, Heng, & Tabunshchyk, 2021). This dual approach—leveraging both advanced AI filters and human oversight—can help mitigate the risk of culturally insensitive or inappropriate content reaching children.

Moreover, AI developers should work closely with educators, religious scholars, and child development experts in Saudi Arabia to design AI that embodies cultural competence. This collaboration can inform the programming of response mechanisms that are respectful of Islamic teachings and local social norms. I maintain that integrating culturally aware AI into educational tools not only safeguards children but also reinforces the alignment between technology and community values, building trust among families, educators, and policymakers (Gabriel et al., 2024; Kurian, 2024).

Finally, education around AI use is essential. Children and parents should be educated about the limitations of AI, understanding that these systems, despite their sophistication, do not possess true empathy or moral judgment (Kurian, 2023b). Teaching children to approach AI with critical thinking skills can help them navigate interactions safely, using technology as a supportive tool rather than an authoritative source on cultural or moral issues. This educational component is vital for empowering families to make informed decisions about integrating AI into their children's lives.

By embedding these safeguards, I believe that we can create AI systems that support educational and social development while upholding the cultural and ethical standards that are fundamental to Saudi society.

Specific Cultural Concerns

In Saudi Arabia, cultural and religious values are integral to daily life and play a significant role in shaping the upbringing and education of children. The importance of modesty, respect, and the protection of young users from inappropriate language or behavior is paramount. Any educational or social technology introduced into this environment must align with these principles to be effective and trusted. The use of AI systems that fail to incorporate these cultural values risks eroding the trust that families place in educational institutions and technological tools (Andries & Robertson, 2023).

One of the primary concerns is that LLMs, when not culturally tailored, may inadvertently promote Western norms or content that conflict with Islamic teachings. I find this to be particularly troubling, as it can create cognitive dissonance for young users who are exposed to contrasting messages from AI systems and their cultural upbringing (Gabriel et al., 2024). Children in Saudi Arabia are brought up with clear expectations about appropriate behavior and language, rooted in both family teachings and formal education. When AI-generated responses do not reflect these cultural guidelines, they can challenge the consistency of these teachings and potentially lead to confusion or distress.

For instance, if an AI system responds to a query in a way that aligns with broader, secular norms but contradicts local cultural or religious standards, it can raise serious ethical concerns. Such content not only undermines the values taught at home and school but can also disrupt the harmony between educational tools and family beliefs. I believe that this potential for conflict poses a significant challenge to integrating AI in educational and social contexts within Saudi Arabia.

Moreover, the perceived neutrality or authority of AI can amplify this issue. Children often see AI systems as objective sources of knowledge and may not recognize when responses reflect cultural biases embedded in global training data (Kurian, 2024). This perception can lead children to accept AI-generated outputs without question, even when they contradict the cultural norms they are learning to uphold. The subtle normalization of content that may seem acceptable in other parts of the world but conflicts with Islamic values is a risk that must be carefully managed (Kurian, 2023a).

To prevent these cultural concerns, AI systems intended for use in Saudi Arabia must be designed with a clear understanding of local customs, beliefs, and ethical standards. I argue that this requires more than just surface-level content filtering; it calls for a comprehensive approach that integrates cultural values into the very framework of AI design. This can be achieved by curating training datasets that reflect the language, norms, and acceptable behaviors within Saudi society. Collaboration with local educators, religious scholars, and cultural experts can guide the development of AI systems that are culturally congruent and uphold the principles central to the upbringing of Saudi children (UNICEF, 2020; Holmes & Porayska-Pomsta, 2022).

Ensuring that AI systems respect cultural expectations also involves programming them to recognize and appropriately respond to questions or statements related to sensitive topics. For example, AI could be designed to redirect children seeking advice on complex issues to trusted human resources, reinforcing the importance of human guidance in areas where cultural nuances are essential (Murtarelli, Gregory, & Romenti, 2021). This kind of programming not only safeguards children but also strengthens the role of AI as a complementary educational tool rather than trust that the technology being used aligns with their values, they are more likely to support and encourage its integration into children's learning environments. On the other hand, if AI systems are perceived as sources of conflicting or culturally inappropriate content, this trust can erode, leading to resistance against adopting potentially beneficial educational technologies.

Cultural alignment in AI does not just enhance acceptance but also promotes meaningful engagement. When AI systems are sensitive to the local context, they are better positioned to foster a supportive learning environment that reinforces cultural teachings rather than competing with them. For example, integrating responses that emphasize community values, family respect, and religious

practices can make interactions more relatable and constructive for children. Such culturally informed AI interactions can help reinforce the lessons taught at home and school, thereby supporting a holistic educational approach that respects Saudi Arabian traditions and ethics (Kurian, 2024).

Moreover, culturally tailored AI must avoid the subtle promotion of values that are misaligned with local customs. This includes being vigilant about outputs that may seem benign in other regions but hold different connotations within a Saudi context. For instance, casual language or references to behaviors considered inappropriate under Islamic guidelines can contribute to a gradual shift in how children perceive acceptable conduct. I believe that AI developers have a responsibility to ensure that their systems do not unintentionally expose children to these influences, as such exposure can undermine the societal emphasis on modesty and respectful communication (Gabriel et al., 2024).

Additionally, addressing these cultural concerns requires a proactive and iterative approach. This involves not only building initial safeguards into AI systems but continuously updating them based on community feedback and evolving norms. Input from parents, educators, and cultural experts should be part of a continuous loop that informs ongoing improvements to AI behavior and response mechanisms. I support the idea that AI should be a flexible tool that adapts to the needs and values of the community it serves, rather than a static product built on one-size-fits-all principles (UNICEF, 2020; Holmes & Porayska-Pomsta, 2022).

Finally, educational programs that teach children and their families about the limitations and appropriate use of AI are essential. Children need to understand that while AI can simulate interaction, it does not embody the cultural wisdom or ethical judgment of human beings. This education can help children critically evaluate AI responses and seek human input when faced with complex or culturally significant questions. Such initiatives empower parents to guide their children in navigating technology safely, ensuring that AI serves as an ally in education without displacing the essential human context (Kurian, 2023a; Van Brummelen, Heng, & Tabunshchyk, 2021).

By embedding cultural sensitivity into AI systems and fostering informed, community-based oversight, we can create AI tools that not only comply with the ethical standards of Saudi Arabia but also actively support and enrich the learning experiences of children. This ensures that technology is a force for good, aligning with the values that families and educational institutions hold dear and maintaining trust as a cornerstone of educational progress.

DESIGNING CULTURALLY SENSITIVE, CHILD-SAFE AI

To ensure that AI systems are safe and supportive for Saudi children, developers must prioritize cultural alignment alongside general safety measures. This is crucial in a society where educational tools are not only instruments of learning but also carriers of cultural and ethical values. Culturally sensitive AI requires collaboration between AI researchers, educational leaders, and cultural authorities to ensure that the design and implementation of AI align with the moral and social fabric of Saudi Arabia (UNICEF, 2020). This collaborative approach ensures that AI systems resonate with local values, fostering trust and promoting widespread adoption.

I believe that transparency in AI operations is an essential element in bridging the empathy gap. AI systems must be designed with transparent language processing models that do more than simulate understanding; they should also reflect the values of the communities they serve. This means AI interactions should be respectful of Saudi norms, reinforcing values such as modesty, family respect, and adherence to Islamic teachings (Holmes & Porayska-Pomsta, 2022). Transparency in AI design allows parents, educators, and children to understand how AI systems work, building trust and facilitating informed use.

Cultural alignment involves embedding local cultural references and ethical standards directly into the AI's architecture. For example, I advocate for the development of NLP models that recognize and prioritize culturally appropriate responses and language. This may include understanding local dialects, expressions, and communication styles that align with Saudi customs. By doing so, AI can respond in a way that is not only linguistically accurate but also contextually relevant and culturally respectful (Gabriel et al., 2024).

A critical aspect of culturally sensitive AI design is ensuring that responses reinforce, rather than undermine, the principles taught in homes and schools. This requires programming AI systems to avoid topics or outputs that could be seen as culturally insensitive or inappropriate. I argue that AI systems must have built-in content moderation mechanisms that filter responses based on cultural norms and ethical guidelines. These filters can be developed in partnership with local cultural and religious scholars, who can provide insights into what constitutes appropriate content for young users (Murtarelli, Gregory, & Romenti, 2021).

Moreover, AI must be capable of distinguishing between educational and sensitive content. I believe that when children ask questions that touch on topics requiring moral or ethical judgment, the AI should be programmed to redirect them to human resources, such as parents, teachers, or religious mentors. This feature ensures that technology complements human guidance, rather than replacing it, thereby preserving the role of cultural and religious teaching as the foundation for a child's development (Kurian, 2023a).

Collaboration between developers and local experts is vital for creating AI that respects and upholds Saudi values. I emphasize that culturally sensitive AI is not just a matter of including regional data but also understanding the social and religious underpinnings that inform behaviors and expectations. Developers should work with educators, parents, and community leaders to test and refine AI interactions to ensure they are aligned with local practices and beliefs (UNICEF, 2020; Holmes & Porayska-Pomsta, 2022). This partnership can guide the development of features that reinforce cultural education, such as prompts that encourage positive behaviors aligned with Islamic values, or responses that highlight the importance of family, community, and respect.

Cultural alignment in AI design also involves equipping systems with the flexibility to evolve alongside societal norms. As society changes, AI must be adaptable, incorporating updates that reflect shifts in cultural attitudes or practices. I support the integration of feedback mechanisms that allow educators and parents to report responses they find inappropriate or misaligned with cultural teachings. This real-time feedback can inform iterative improvements, making AI an evolving, responsive tool that continues to align with community standards (Van Brummelen, Heng, & Tabunshchyk, 2021).

Educational initiatives that teach the responsible use of AI are equally important. Children should be educated about the nature of AI, understanding that while these systems can simulate conversation and offer assistance, they lack the moral and cultural judgment of human interactions (Kurian, 2024). This education helps children navigate AI interactions with a critical eye, seeking human input when necessary, and understanding the limitations of technology as it relates to complex social or moral issues (Knox, Williamson, & Bayne, 2020).

Finally, designing culturally sensitive AI involves more than technological innovation; it requires a deep respect for the culture it serves. By embedding cultural awareness into the very fabric of AI systems and fostering partnerships between developers and local communities, we can ensure that AI supports and enhances education while safeguarding children and reinforcing the values essential to their upbringing. This approach not only ensures safety but also enriches the role of AI as a partner in education that aligns seamlessly with the values of Saudi society.

Immediate Actions

Implementing immediate actions is crucial to ensure the safety and cultural alignment of AI systems used by children in Saudi Arabia. These actions include integrating robust natural language processing (NLP) mechanisms that can filter content effectively to align with Saudi values. Such filtering systems must go beyond standard content moderation to include cultural specificity, ensuring that language and ideas presented by AI adhere to the expectations set by Saudi norms (Chowdhury, 2003). The use of these tailored NLP tools helps to maintain an educational and interactive space that upholds the values of modesty, respect, and cultural integrity.

I believe that human oversight should play a significant role alongside automated systems to monitor AI interactions in real time. Automated tools can flag potentially problematic content, but human moderators are essential for making context-sensitive decisions that reflect cultural and ethical standards. This dual-layered approach helps address nuanced situations where algorithms may fall short, ensuring that any culturally sensitive content is managed appropriately (Knox, Williamson, & Bayne, 2020). This oversight is particularly necessary in environments where children's interactions can vary greatly, and unexpected questions or responses can emerge.

AI responses must be designed to reflect age-appropriate and culturally consistent communication. This means training models not just to understand language in a general sense but to generate content that aligns with the developmental and cultural context of Saudi children (Kurian, 2024). Age-appropriate interactions involve ensuring that responses are not only safe in terms of language but are also designed to reinforce the moral teachings that children receive at home and in school. Such measures reduce the risk of AI unintentionally promoting behaviors or ideas that conflict with Islamic principles and social norms (Weidinger et al., 2021).

Immediate actions should also include real-time monitoring systems capable of adapting to emerging content that might be flagged as culturally or ethically inappropriate. I argue that integrating adaptive learning capabilities within these systems is essential to enable them to evolve with changes in cultural attitudes and expectations. For example, responses that were once considered neutral may become controversial as society shifts. Real-time monitoring ensures that these changes are accounted for swiftly, maintaining the relevance and safety of AI interactions (Murtarelli, Gregory, & Romenti, 2021).

Embedding these measures into AI systems requires collaboration between developers and local cultural experts to define the parameters of appropriateness. These experts can help identify sensitive topics and outline response strategies that align with cultural and religious teachings. This collaborative effort ensures that AI is not just technically advanced but also culturally intelligent, capable of operating within the unique moral landscape of Saudi society (UNICEF, 2020). By implementing immediate safety features that are culturally informed, AI systems can become trusted resources in educational and social environments.

I believe that immediate actions also need to focus on clear and culturally relevant communication protocols. AI should be equipped to respond with structured, respectful language that aligns with the formal nature of teaching in Saudi Arabia. This not only enhances trust in AI as a learning tool but also reinforces the lessons children receive from their educators and families. Responses should be programmed to reflect respect for elders, promote modesty, and encourage behaviors in line with community values (Gabriel et al., 2024).

In addition to automated filters and human oversight, developing response guidelines that involve ethical redirection can also contribute to the safety and cultural appropriateness of AI interactions. If an AI system encounters a question that falls into a sensitive area—whether culturally, religiously, or ethically—it should be programmed to redirect the conversation to encourage seeking advice from parents or educators. This strategy acknowledges the limitations of AI in understanding and

navigating complex cultural nuances while reinforcing the role of adults in guiding children (Van Brummelen, Heng, & Tabunshchyk, 2021).

Immediate measures are not only protective but also proactive. By creating an AI environment that is safeguarded through culturally aware NLP filters, human oversight, and dynamic response systems, we can ensure that AI serves as a supportive tool for education and interaction. The incorporation of these elements reassures parents and educators that the technology respects their values and supports the cultural and moral development of children. This level of trust is essential for AI to be successfully integrated as a beneficial component of educational and social frameworks in Saudi Arabia.

Long-Term Recommendations

Long-term strategies are essential for embedding cultural and ethical sensitivity in AI systems used by children in Saudi Arabia. Developing comprehensive policies that mandate the ethical and culturally aligned use of AI in both educational and recreational settings is a critical step toward ensuring sustained safety and cultural congruence. These policies should serve as a roadmap for AI developers, educators, and policymakers, guiding the creation and deployment of AI technologies that uphold the values central to Saudi society (OECD, 2021).

I believe that establishing age-appropriate AI interaction frameworks infused with local cultural and religious teachings is fundamental to these long-term strategies. Such frameworks must go beyond basic safety measures, integrating the moral and social lessons taught in homes, schools, and religious settings. For instance, AI responses should reflect principles of modesty, respect for elders, and the emphasis on community that are hallmarks of Saudi culture. This ensures that technology not only serves as a tool for learning and interaction but also reinforces the values that children are raised with (UNESCO, 2019).

Regulatory frameworks should include stringent guidelines for incorporating culturally appropriate content in the AI training data. AI models that are trained on global, unfiltered datasets may not accurately reflect the cultural norms of specific regions. Therefore, training data must be curated to include content that aligns with the cultural, ethical, and religious standards of Saudi society (Pesapane et al., 2018). I argue that this is not just a technical requirement but an ethical imperative to ensure that AI serves as a culturally supportive tool rather than a source of conflicting messages.

To achieve this, collaboration between developers and local authorities, such as educational institutions and religious scholars, is vital. These partnerships can provide the necessary oversight and input to create AI systems that are sensitive to local values and expectations. For example, educational leaders can outline the types of content that support school curricula, while religious scholars can offer insights on integrating teachings that align with Islamic principles. This collaborative approach ensures that AI is tailored to meet the unique needs of Saudi children, making it a more effective and trusted part of their educational experience (Gabriel et al., 2024; UNICEF, 2020).

I also support the idea of incorporating adaptive regulatory mechanisms that evolve alongside cultural and societal changes. As societal attitudes shift over time, so too must the policies that govern AI behavior. This adaptability can be achieved through periodic reviews of regulatory standards and ongoing dialogue with cultural and educational stakeholders. By maintaining a responsive approach, AI systems can continue to align with current norms and expectations, reinforcing their role as supportive, culturally congruent tools (Holmes & Porayska-Pomsta, 2022).

In addition to policy development, investing in continuous training and education for AI developers is essential. I believe that those involved in creating AI systems should be equipped with an understanding of the cultural context in which their products will be used. This education can include

training on the ethical and cultural considerations unique to Saudi society, ensuring that developers are aware of how their design choices impact children's learning and social interactions. Such training can help bridge the gap between technical expertise and cultural sensitivity, fostering a development environment where respect for local values is prioritized (Murtarelli, Gregory, & Romenti, 2021).

Long-term recommendations should also include establishing partnerships between technology developers and community organizations that focus on child safety and development. These partnerships can facilitate the creation of AI tools that not only protect children from inappropriate content but also enhance their educational journey by incorporating culturally relevant knowledge and moral lessons. I advocate for AI that is not only protective but also proactive in promoting positive values through its interactions (Kurian, 2023a; Knox, Williamson, & Bayne, 2020).

In line with these policies, it is important to build comprehensive user guidelines for families and educators. Parents and teachers should be informed about the best practices for using AI in educational and social settings, including how to navigate and understand the limitations of these technologies. Educating the community about the nature of AI and its capabilities helps reinforce that while AI can be a powerful educational tool, it is not a substitute for human guidance and judgment (Van Brummelen, Heng, & Tabunshchyk, 2021; Kurian, 2024).

Finally, long-term strategies should also involve developing transparent reporting and accountability structures. These structures should enable users to report culturally or ethically inappropriate AI behavior, ensuring that developers can act swiftly to make necessary adjustments. Transparent mechanisms for feedback and response contribute to a cycle of continuous improvement, reinforcing trust in AI systems as safe and culturally aligned tools. By embedding such structures into the development and deployment process, AI can remain a trusted and valuable asset that complements the values upheld by Saudi families and educational systems (OECD, 2021; UNICEF, 2020).

Through these comprehensive and adaptive long-term strategies, we can ensure that AI systems do more than avoid harm—they become active participants in nurturing a learning environment that respects and supports the cultural, ethical, and religious fabric of Saudi society.

POLICY AND EDUCATIONAL IMPLICATIONS

Educators and policymakers in Saudi Arabia play a pivotal role in guiding the safe and effective integration of AI into children's learning environments. Given the unique cultural, religious, and ethical values that underpin Saudi society, their involvement is crucial in shaping AI systems that align with these principles. I believe that promoting awareness among parents and teachers about the potential risks and benefits of AI is an essential first step. This education helps create a well-informed community capable of making informed decisions about the use of AI in children's education (Van Brummelen, Heng, & Tabunshchyk, 2021).

Parents and teachers must be equipped with the knowledge to identify both the opportunities AI presents and the associated risks. This includes understanding how AI systems generate responses, their limitations, and the safeguards necessary to protect children from inappropriate or culturally incongruent content. Workshops and training programs should be developed to help parents and educators navigate AI tools effectively. Such initiatives can foster a sense of shared responsibility, where families and schools work in tandem to ensure that AI technology complements traditional learning methods and upholds cultural values (Su, Ng, & Chu, 2023).

Collaboration between schools and families is fundamental to building a support network that oversees the use of AI. I argue that this partnership should extend to creating policies that mandate human oversight when AI is used in educational settings. While automated systems can handle many aspects of AI interaction, human intervention remains indispensable for addressing nuanced

situations that require cultural insight and moral judgment. Schools should implement clear protocols for monitoring AI interactions, ensuring that teachers are available to step in when necessary to provide guidance or correct potentially inappropriate AI outputs (Knox, Williamson, & Bayne, 2020).

Educational institutions also need to collaborate closely with policymakers to establish frameworks that define the ethical use of AI in education. These frameworks should outline specific guidelines for AI content, data usage, and privacy measures that align with the cultural and religious expectations of Saudi society. For instance, I believe policies should mandate that AI training data include culturally appropriate content and exclude information that could promote ideas or behaviors inconsistent with local values (Gabriel et al., 2024). Such guidelines help ensure that AI tools do not inadvertently introduce content that could conflict with the teachings children receive at home and school.

Policymakers should focus on creating comprehensive regulations that address the broader implications of AI in educational settings. This includes setting standards for transparency in AI operations so that educators and parents understand how AI systems make decisions and generate content (Holmes & Porayska-Pomsta, 2022). Transparent AI practices contribute to building trust and allow for better oversight of AI behavior, reinforcing the confidence of families and educators in the use of technology as a learning aid.

Moreover, I advocate for policies that require regular assessments of AI systems used in educational settings to ensure compliance with ethical and cultural standards. These assessments should be conducted by independent bodies that include educators, policymakers, cultural experts, and child development specialists. Such a comprehensive approach guarantees that AI systems remain aligned with societal expectations and continue to serve as safe, supportive tools for learning (UNICEF, 2020). This level of oversight not only prevents the erosion of trust in educational technologies but also encourages their responsible evolution.

In addition to policy measures, educational programs that teach children how to engage critically with AI should be implemented. These programs should highlight the capabilities and limitations of AI, emphasizing that while these systems can simulate conversation and provide information, they lack true empathy and cultural judgment (Kurian, 2024). Teaching children to view AI as a tool rather than an authority figure helps them approach interactions with discernment, knowing when to seek human input for guidance on culturally or morally complex topics.

Finally, educational institutions should create avenues for ongoing dialogue between teachers, parents, and policymakers to review the use and impact of AI tools in education. Regular feedback sessions can provide insights into how AI is being received by students and whether any adjustments are needed to better align with cultural and educational goals. I support the establishment of such participatory frameworks as they promote shared ownership of educational outcomes and foster an environment where technology serves the broader goal of nurturing well-rounded, culturally aware individuals (Van Brummelen, Heng, & Tabunshchyk, 2021; Su, Ng, & Chu, 2023).

In conclusion, the policy and educational implications of integrating AI into children's learning environments in Saudi Arabia must prioritize cultural sensitivity, safety, and collaboration. By promoting awareness, establishing robust oversight frameworks, and involving all stakeholders in the ongoing adaptation of AI technologies, we can create a supportive ecosystem where AI enhances, rather than challenges, the educational and moral development of children.

The Role of Islamic Teachings

Integrating Islamic teachings into AI design is vital for creating systems that resonate deeply with Saudi culture. The unique social fabric of Saudi Arabia is interwoven with Islamic principles that

guide daily life, education, and behavior. AI systems intended for educational and recreational use must be developed with an awareness of these moral and spiritual values to ensure they align with the expectations of Saudi society (Gabriel et al., 2024). I believe that designing AI systems that uphold these teachings requires a comprehensive approach that moves beyond general ethical programming to deeply embed cultural and religious values within the AI's operational framework.

Programming AI to adhere to Islamic ethical principles involves incorporating guidelines that prevent the use of language or suggestions that contravene modesty or Islamic law. This means avoiding content that might be considered inappropriate or offensive by religious standards, thus ensuring that AI systems act as positive reinforcements of the moral teachings children receive at home and in their communities (UNICEF, 2020). For instance, AI responses should be structured to promote virtues such as humility, respect for others, and the importance of family and community, which are central to Islamic teachings.

I argue that integrating Islamic principles into AI design should involve collaboration with religious scholars and educators who can provide insights into how these teachings can be accurately and respectfully represented in AI interactions. This collaboration ensures that AI systems do not inadvertently generate content that contradicts or undermines religious beliefs, but rather, align with the spiritual and moral framework that guides Saudi society (Andries & Robertson, 2023). This process not only upholds cultural integrity but also reinforces the trust parents and educators place in the technology used by their children.

Incorporating Islamic teachings into AI design also means that AI systems should be equipped to handle sensitive topics in a manner that reflects the values of Saudi society. For example, when a child poses a question that touches on issues requiring moral or ethical judgment, the AI should be designed to provide responses that are consistent with Islamic teachings or encourage seeking advice from parents or religious authorities. This approach respects the role of family and religious guidance in a child's upbringing and reinforces the message that while AI can be a helpful tool, it cannot replace the wisdom and judgment of human mentors (Holmes & Porayska-Pomsta, 2022).

Furthermore, the integration of Islamic teachings should extend to the prevention of content that subtly promotes values incongruent with Saudi culture. AI models trained on global datasets may unintentionally introduce Western norms that conflict with Islamic principles. To prevent this, training data must be carefully curated to reflect Islamic ethics and cultural expectations, filtering out content that may encourage behavior considered unacceptable under Islamic law (Murtarelli, Gregory, & Romenti, 2021). This measure ensures that AI does not become a source of conflicting messages but rather reinforces a cohesive educational and moral experience for children.

I believe that AI systems designed with Islamic teachings in mind should also promote positive behaviors and attitudes. For example, educational AI tools could be programmed to include teachings from the Qur'an or Hadith that emphasize kindness, honesty, and the value of knowledge. By doing so, AI can serve as an extension of religious education, promoting virtues that are esteemed in Saudi society. This not only aligns AI use with cultural and religious expectations but also enriches the educational content children receive, embedding spiritual growth within technological interaction (Gabriel et al., 2024; Kurian, 2023a).

Ensuring that AI systems respect Islamic teachings also involves the continuous adaptation of technology to reflect current interpretations and practices. I support the development of adaptive AI that can evolve alongside changes in religious and cultural discourse. Regular updates informed by input from religious scholars and community leaders can help AI systems stay aligned with the teachings and expectations of the society they serve. This adaptability contributes to maintaining the relevance and trustworthiness of AI in educational and social contexts (UNICEF, 2020).

Lastly, I advocate for educational programs that teach children and families about the ethical use of AI through the lens of Islamic values. These programs can help children understand the role of technology as a tool that should be used in a way that aligns with their faith. Parents and educators can guide children to approach AI interactions thoughtfully, recognizing the importance of seeking human advice when necessary. This reinforces the idea that while AI can simulate conversations and provide information, it does not possess the spiritual insight inherent in human interactions (Van Brummelen, Heng, & Tabunshchyk, 2021; Kurian, 2024).

By integrating Islamic teachings into AI design, developers can create tools that not only respect but also enhance the cultural and religious framework within which Saudi children grow. This approach ensures that technology is an ally in promoting the values cherished by society, supporting a learning environment that is both culturally and spiritually enriching.

CONCLUSION

Ensuring the safety and cultural appropriateness of AI interactions for Saudi children demands a multifaceted approach that deeply integrates cultural sensitivity, ethical design, and collaborative efforts between AI developers, educators, and cultural authorities. Addressing the empathy gap in LLMs requires more than just technical adjustments; it calls for transparent operations and culturally aware programming that align with the moral and ethical teachings intrinsic to Saudi society. This holistic approach ensures that AI not only functions effectively but does so in a way that supports the educational and moral development of children.

Real-time oversight and robust content monitoring systems are essential to detect and address culturally sensitive violations. Automated mechanisms paired with human intervention can provide the nuance needed to manage complex situations that algorithms may not fully comprehend. This dual approach reinforces the trust of parents and educators, ensuring that AI interactions remain safe and aligned with cultural expectations. Human oversight also enables immediate redirection in cases where AI might respond inappropriately, maintaining the integrity of the interaction and safeguarding children's experiences.

Incorporating culturally informed NLP mechanisms and programming AI with respect for Islamic teachings further ensures that AI serves as a tool that resonates with the values upheld in Saudi Arabia. This integration helps reinforce cultural norms, promoting positive behaviors and fostering an educational environment that complements rather than contradicts the teachings of homes and schools. By embedding Islamic principles into the AI's core functions, developers can create tools that enhance children's understanding of their cultural and religious heritage, offering responses that are educational and respectful.

Long-term strategies must focus on policy development that mandates the inclusion of culturally aligned content in AI training data, continuous updates to reflect societal changes, and educational initiatives that empower children and families to engage with AI safely. Policymakers should craft regulations that provide clear guidelines for culturally sensitive AI design and implementation, supported by ongoing collaboration between tech developers, educators, religious scholars, and child development experts. This comprehensive regulatory framework ensures that AI remains a trusted, culturally congruent tool that aligns with Saudi values and ethics.

Additionally, educating children about the capabilities and limitations of AI is a vital part of fostering responsible technology use. Teaching children to approach AI with critical thinking and to seek human guidance for complex or sensitive topics reinforces the role of parents, teachers, and community leaders as key figures in their upbringing. This educational emphasis ensures that while AI can simulate conversation and assist with learning, it is seen as a supplementary tool rather than a substitute for human wisdom and moral judgment.

In my view, the path forward involves an iterative approach where AI systems continuously adapt and evolve alongside the cultural, educational, and ethical landscape of Saudi Arabia. This adaptability ensures that AI remains relevant and beneficial, serving as an ally that supports rather than disrupts the moral and educational fabric of society. Through transparent design, culturally integrated programming, and comprehensive policy oversight, AI can be harnessed as a powerful educational asset that respects the values, beliefs, and well-being of Saudi children.

By embedding these practices into AI development and implementation, we can create a learning ecosystem where technology complements traditional values, supports children's growth, and fosters a safe and enriching educational environment. This approach safeguards not only the immediate interactions children have with AI but also contributes to their long-term development, ensuring that technology serves as a bridge to learning that upholds and promotes cultural and ethical integrity.

AUTHOR CONTRIBUTIONS: AL FRAIDAN A: Conceptualization, methodology, data curation, formal analysis, writing final draft, reviewing, editing, supervision, funding acqusition, conceptualization, writing first draft, data collection, resources

FUNDING: This work was funded and supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia, [Grant,242420]

DATA AVAILABILITY: Data is available upon request.

ACKNOWLEDGMENTS: I would like to acknowledge all the people who facilitated this project including administrators, faculty members and the research participants for their cooperation. Special acknowledgments to the love of my life, my wife, AlAnoud Alwasmi who facilitated a lot of this research processes and for her support during my health crisis.

CONFLICT OF INTEREST: The author declares no conflict of interest.

REFERENCES

Abbasi, N. I., et al. (2022). *Can Robots Help in the Evaluation of Mental Wellbeing in Children?* IEEE RO-MAN.

Al Fraidan, A. (2024). Integrating Al-Powered Library Systems to Enhance Research and Learning in English Language Departments: A Case Study of Faculty and Student Perceptions. Library Progress International, 44(3), 12058-12065.

Al Fraidan, A. (2024). Anticipatory Thinking and AI-Driven Assessments: A Balanced Approach to AI Integration in Education Aligned with Saudi Vision 2030. African Journal of Biomedical Research. 27(3), 619-628. https://doi.org/10.53555/AJBR.v27i3.2560

Al Fraidan, A. (2025). Enhancing linguistic research through AI-powered reference management: A proposal for a voice-controlled academic assistant. Edelweiss Applied Science and Technology, 9(1), 1–9. https://doi.org/10.55214/25768484.v9i1.2240

Andries, V., & Robertson, J. (2023). *Children's Understanding of AI through Interactions with Smart Speakers*. Computers and Education: AI.

BBC News. (2021). *Alexa Tells 10-Year-Old to Touch Plug with Penny*. Retrieved from https://www.bbc.co.uk/news/technology-59810383

Cambria, E., & White, B. (2014). *Natural Language Processing Research Review*. IEEE Computational Intelligence Magazine.

Chowdhury, G. (2003). *NLP Annual Review*. Information Science and Technology.

Common Sense Media. (2023). *AI and Children Impact Report*. Retrieved from https://www.commonsensemedia.org

Darling, K. (2017). *Anthropomorphic Framing in AI*. Oxford Scholarship Online.

Fowler, G. (2023). *Snapchat's MyAI Safety Analysis*. Washington Post.

Gabriel, I., et al. (2024). *The Ethics of Advanced AI Assistants*. arXiv.

Holmes, W., et al. (2019). *Ethics in AIED: An Overview*. AIED Conference Proceedings.

Holmes, W., & Porayska-Pomsta, K. (2022). *Ethics in AI and Education: Challenges and Debates*. Taylor & Francis.

Knox, J., Williamson, B., & Bayne, S. (2020). *Machine Behaviourism in Digital Learning*. Learning, Media and Technology.

Kurian, N. (2023a). *AI's Empathy Gap in Early Childhood Education*. Contemporary Issues in Early Childhood.

Kurian, N. (2024). *Designing Child-Safe AI and Addressing the Empathy Gap in LLMs*. Learning, Media and Technology.

Murtarelli, G., Gregory, A., & Romenti, S. (2021). *Ethical Human–Machine Interactions*. Journal of Business Research.

Naveed, H., et al. (2023). *Overview of Large Language Models*. arXiv.

OECD. (2021). *Global AI Strategies and Policies*. OECD Repository.

Pesapane, F., et al. (2018). *AI as a Medical Device: Ethical and Regulatory Issues*. Insights Into Imaging.

Rudolph, J., Tan, S., & Tan, S. (2023). *AI in Education and the Chatbot Race*. Journal of Applied Learning and Teaching.

Shead, S. (2021). *Amazon Alexa's Dangerous Response*. CNBC.

Su, J., Ng, D. T., & Chu, S. K. (2023). *AI Literacy in Early Childhood Education*. Computers and Education: AI.

UNICEF. (2020). *Safeguarding Children in AI Interactions*. Retrieved from https://www.unicef.org/eap/media/5376/file

UNESCO. (2019). *Beijing Consensus on AI in Education*. Retrieved from https://unesdoc.unesco.org

Van Brummelen, J., Heng, T., & Tabunshchyk, V. (2021). *Teaching Tech to Talk: AI Curriculum for K-12*. AAAI Conference on AI.

Weidinger, L., et al. (2021). *Ethical and Social Risks of Language Models*. arXiv.