Artificial Intelligence, Digital Tools and Forgiveness

Perspective on the Potential of AI in the Fields of Forgiveness, Reconciliation, and Peacebuilding

Forgiveness as we define it in Salam Institute is a voluntary process of transforming negative emotions, thoughts, and attitudes toward an offender into more constructive responses, without denying the harm or abandoning the pursuit of justice. It is distinct from forgetting, excusing, or condoning wrongdoing.

As an increasing share of human conflict, apology and reconciliation unfolds through digital platforms and as AI systems increasingly mediate interpersonal communication, an important question emerges: What happens to forgiveness when artificial intelligence enters the process? This briefing draws on insights from psychology, peacebuilding and technology ethics to examine current knowledge on this issue, identify emerging opportunities and challenges and highlight key questions that warrant further research.

What is Forgiveness?

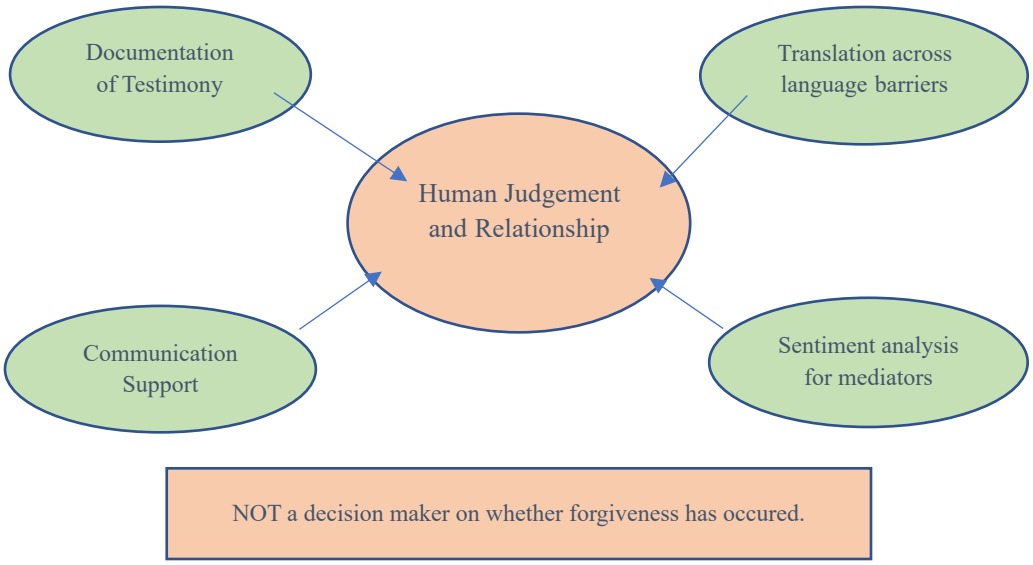
Is

A Voluntary shift in emotion and attitute toward an offender while still acknowledging harm and the pusuit of

Is not
Forgetting, Excusing,
Condoning the wrong.
Forgiveness is not being weak.

Forgiveness can coexist with accountability. It does not require the offender's presence, remorse or an admission of guilt.

AI's most tested role is in the mediated setting: crisis communication, customer service recovery and documentation of large-scale conflicts. Studies on AI generated apologies shows that tone, wording and small cues like emojis can shift how willing a person is to forgive. Machine-learning tools can also assist in helping scholars identify patterns across large amounts if conflict related text, which may be useful in recognising tensions. Lately, peacebuilding organizations are experimenting on how AI can be used to support the efforts for forgiveness and reconciliation. In educational settings, generative AI can also help adapt peace and forgiveness materials across different languages, cultural contexts and learning needs. Simultaneously, AI use in sensitive peacebuilding contexts raises concerns about privacy, consent, surveillance, algorithmic bias, and the potential misuse of personal data, particularly when data collected for humanitarian purposes can be repurposed in ways that expose vulnerable groups to additional risks. The role of AI can be best understood as a support for human judgment rather than a substitute for human decision-making in matters of forgiveness. AI is proving useful as an aid to human judgement in forgiveness-related processes not a s a decisionmaker that determines whether forgiveness has occurred.



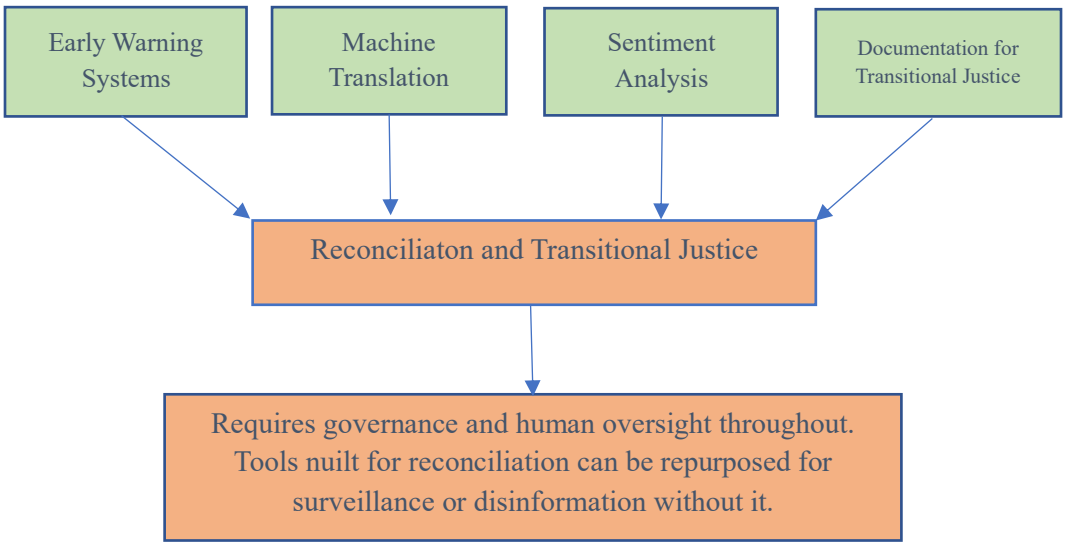
Digital technology is reshaping how we seek and grant forgiveness. AI is not changing the psychology of forgiveness; it is changing the context, signals, tools, and scale through which forgiveness and reconciliation can be pursued.

Researchers found that delivering apologies via texts or video calls rather than in person alters how people judge sincerity. Apologies lose nonverbal cues, voice intonation and facial expressions that traditionally signal genuine remorse, so the audiences lean on substitutes such as response speed, word choice and the use of emojis and typeface. Studies of public apology videos on platforms like YouTube have demonstrated that the audience's perception of sincerity and the willingness to forgive hinge heavily on how a message is framed rather than the actual act of apologizing. Digital channels haven't yet replaced the psychology of forgiveness, what they have changed is the signals people use to decide whether an apology is real.

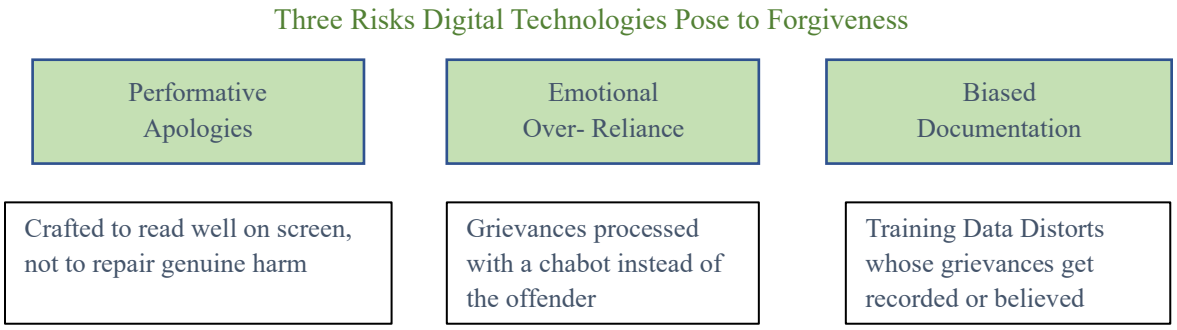
Addressing the evidence gap on AI supported forgiveness interventions requires first examining the broader landscape on digital mental health. There is substantial evidence on how AI supported mental health tools broadly, chatbots built on approaches like cognitive behavioural therapy, acceptance and commitment therapy have shown feasibility and relatively modest effectiveness for anxiety, stress and low mood, particularly in areas where access to human therapists is limited. However, there is not well-established evidence-based AI intervention that specifically targets interpersonal forgiveness in published research. The closest is AI tools that help people process grievances related distress, reflection on conflict and drafting a considered response. Whether these tools can responsibly supplement the weaknesses of traditional forgiveness initiatives is still an open question this is best explored through input from forgiveness researchers and clinicians rather than assumed from adjacent mental health results.

The expansion of AI into the forgiveness space has raised urgent ethical concerns, a study conducted at Brown University found that AI counselling chatbots systematically violated basic mental health ethical standards, offering as false sense of empathy and mishandling sensitive topics, risks that would apply directly to anything framed as aiding someone to forgive or be forgiven. Commentors in Nature Machine Intelligence have also warned that simulated empathy can foster emotional dependency. A 2026 AI safety report found measurable rates of users forming un unhealthy emotional attachment to AI chatbots. Growing body of scholarship emphasizes that forgiveness is inseparable from accountability, consent and authentic human relationships, AI should support the parties involved in a forgiveness process rather than stand in their place.

At the Societal level, AI plays a key role in forgiveness, peacebuilding and reconciliation. A major systemic review conducted in 2025 by the International Panel on the Information Environment drawing on over 600 studies, identified concrete uses of AI in peacebuilding: early warning systems that flag rising conflict risks, language translation, sentiment analysis tools that support mediators working in areas with language and cultural barriers and also AI assisted documentation for transitional justice and truth commissions). AI-assisted dialogue platforms can also help facilitators identify common ground, analyse participant input, and map areas of agreement and disagreement. They can also help enable participation across geographic, language, and identity divisions, supporting a wider range of perspectives to be included in dialogue processes. A 2026 report from the AI Negotiation Forum found that negotiators are already using AI for preparation, analysis, drafting and sense making. AI can therefore help support preparation prior to difficult negotiations and dialogue by helping enable practitioners to conduct background research, organise relevant information, explore potential scenarios and test their arguments in advance. The United States Institute of Peace has argued that peacebuilding organizations need a seat at the decision-making table to shape how these tools are built, rather than leaving it entirely to governments and technology companies. However, researcher across this field have cautioned that the same tools built for reconciliation can be repurposed for surveillance or disinformation, also governance and human oversight should be treated as essential and not optional.



Barriers created by Digital technologies and specially AI forgiveness. There are three core risks that stand out. First, performative digital apologies, tailored for how they read and appear on screen rather than for genuine repair, can create the appearance of forgiveness without the actual substance for it. Linguistic sincerity refers to whether an apology sounds sincere through its wording and tone. On the other hand, relational sincerity depends on the broader relationship, including accountability, trust, and whether the apology is succeeded by visible, meaningful action. Second, emotional over reliance of AI chatbots can lead people to work through their grievances with a program rather than with the person who wronged them, this can undermine the relational repair that forgiveness is meant to achieve. Third, AI systems that are used in justice, mediation or documentation inherit the biases of their training data, which can distort whose grievances are recorded, believed to be genuine or acted upon a pressing concern in transitional justice and human rights contexts. This is especially significant when the data used by AI systems does not sufficiently reflect local information, languages, or cultural contexts. In these situations, local perspectives of deep importance may be overlooked, demonstrating the importance of human oversight and contextual knowledge.



Prevailing research on forgiveness education suggests, there are 3 useful cases for AI. First, education should build literacy that is context specific to apology and conflict helping in reading sincerity cues in text-based communication rather than assuming that digital apologies carry more weight than face to face apologies. AI systems could also help support forgiveness education through simulated conflict scenarios, reflection exercises feedback on communication, and customised learning. These tools may assist learners in practicing how to interpret apologies, consider several different perspectives, and contemplate on how wording, context, and communication impact perceptions of sincerity. It could also help adapt educational materials across multiple languages and cultural contexts. Second, limits of AI emotional support should be explicit and clear, so people use these tools as a supplement for forgiveness initiatives rather than a replacement. Such AI systems should stay closely tied to counselors, mediators, educators, civil workers and faith community leaders who can identify ethical gaps and failures like false empathy, mismanaged crises and biased outputs that current AI systems are documented to make.

Practitioner's recommendations

"The interviewed practitioners identified several practical recommendations:

- Encourage closer collaboration between AI developers and forgiveness/peacebuilding researchers so that tools are built with a better understanding of what forgiveness requires
- Pilot AI-assisted reconciliation tools in low-stakes settings, such as community mediation programs, in collaboration with AI developers, peacebuilding organisations, and research institutions, before using them in high-stakes contexts like transitional justice
- Evaluate these pilots' using measures such as empathy, willingness to engage in dialogue, perceived sincerity, trust, willingness to consider forgiveness, and quality of interaction
- Develop clearer standards, guidelines, or certification for tools that claim to support emotional healing, so organisations and individuals can better distinguish well-designed tools from risky ones
- Keep human facilitators central to the process. AI may help prepare the ground, but forgiveness still requires human guidance and interaction

Future Directions

The literature review supports the consistent conclusion: AI has a growing role to play around forgiveness in documentation, translation, communication support and early warning but does not yet have a validated role in the forgiveness space. Salam defines it as a voluntary transformation of emotion and attitude that neither denies harm does not abandon justice; it is not a status that a technology can certify.

Looking ahead, three directions seem most promising to pursue. First, closer collaboration between AI developers and peacebuilding or forgiveness researchers, who currently work in isolation from one another, so tools are designed with an accurate understanding of what forgiveness requires. Second, piloting AI assisted reconciliation tool in a low-stakes setting, such as a community mediation program before

considering its use in a high-stakes transitional justice context. Third, supporting the development of clearer standards for tools that claim to support emotional healing, organisations like Salam and the communities we serve can distinguish between a well-designed tool from a risky one. In conclusion, human facilitators remain an essential part, AI can prepare the ground, translate, document and flag risks but the work of forgiveness itself still depends on human relationships and guidance.

References

1. Ashby, H. (2023, December). *A role for AI in peacebuilding*. United States Institute of Peace. https://www.usip.org/publications/2023/12/role-ai-peacebuilding
2. Bengio, Y., Clare, S., Prunkl, C., Andriushchenko, M., Bucknall, B., Murray, M., Bommasani, R., Casper, S., Davidson, T., Douglas, R., Duvenaud, D., Fox, P., Gohar, U., Hadshar, R., Ho, A., Hu, T., Jones, C., Kapoor, S., Kasirzadeh, A., Manning, S., ... Zhang, Y.-Q. (2026). *International AI safety report 2026*. arXiv. https://arxiv.org/abs/2602.21012
3. Bruderlein, C. (2026). *Exploring pathways for responsible use of AI in negotiation and diplomacy: Final report.* AI Negotiation Forum. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6925859
4. Croicu, M. (2024). *Deep active learning for data mining from conflict text corpora.* arXiv. https://arxiv.org/html/2402.01577v1?utm_source=perplexity
5. De Freitas, J., & Cohen, I. (2025). Emotional risks of AI companions demand attention. *Nature Machine Intelligence*. https://www.nature.com/articles/s42256-025-01093-9
6. Devidal, P. (2024). *Lost in digital translation? The humanitarian principles in the digital age.* International Review of the Red Cross, 106(925), 120–154. https://www.cambridge.org/core/journals/international-review-of-the-red-cross/article/abs/lost-in-digital-translation-the-humanitarian-principles-in-the-digital-age/A4E233798E3AFA2516FEAA750E249AA3
7. Fietta, V., Rizzi, S., Gios, L., Selmi, S., De Luca, C., Pederiva, L., Poggianella, S., Pavesi, M. C., Campregher, M., Lazzeri, S., Cantarelli, S., Purgato, M., Barbui, C., Navarin, N., Gabrielli, S., Monaro, M., Forti, S., & Ferro, A. (2025). Chatbot-based version of a World Health Organization–validated intervention for stress management in patients with breast cancer (Self-Help Plus): Protocol for a pilot feasibility study. *JMIR Research Protocols, 14*, Article e65837. https://doi.org/10.2196/65837
8. Hirblinger, A. T. (2026). *Digital peacebuilding? The socio-technicality of transforming armed conflict*. Oxford University Press. https://doi.org/10.1093/9780198986553.001.0001
9. Iftikhar, Z., et al. (2025, October 21). *AI chatbots systematically violate mental health ethics standards*. Brown University News. https://www.brown.edu/news/2025-10-21/ai-mental-health-ethics
10. Impunity Watch. (2025, January). *Artificial intelligence and transitional justice: Framing the connections* [Overview paper]. Overview-Paper-Report-Artificial-Intelligence-and-Transitional-Justice-Framing-the-Connections-Impunity-Watch-Jan-2025.pdf
11. International Panel on the Information Environment. (2025). *Artificial intelligence and peacebuilding: Opportunities and challenges* (C. Zelizer, F. Ogenga, L. Schirch, E. Tauchnitz, P. N. Howard, & S. Valenzuela, Eds.; Technical Paper TP2025.3). https://doi.org/10.61452/RNGW7145
12. PIRG Education Fund & Consumer Federation of America. (2026). *The risks of AI companion chatbots as mental health support*. https://pirg.org/edfund/resources/ai-chatbot-therapy
13. Sandlin, J. K., & Gracyalny, M. L. (2018). Seeking sincerity, finding forgiveness: YouTube apologies as image repair. *Public Relations Review, 44*(3), 393–406.
14. Sumulong, L. (2025, January 21). *My journey in using artificial intelligence framed in social ethics and peacebuilding*. PeaceBuilders Community. https://peacebuilderscommunity.org/2025/01/my-journey-in-using-artificial-intelligence-framed-in-social-ethics-and-peacebuilding/
15. Various authors. (2026). *Exploring the ethical concerns in user reviews of mental health apps using topic modeling and sentiment analysis*. arXiv. https://arxiv.org/abs/2602.18454
16. Xiang, R., & Park, H. J. (2026). The interplay of typeface and apology source in shaping sincerity and forgiveness. *Journal of Retailing and Consumer Services, 90*. https://www.sciencedirect.com/science/article/abs/pii/S096969892500476X
17. Yang, G., & Shao, B. (2025). Human-AI collaborative recovery: How recovery sequence and strategy order drive consumer forgiveness. *Journal of Retailing and Consumer Services, 87*, Article 104452. https://doi.org/10.1016/j.jretconser.2025.104452