ISSN: 2230-9926

Available online at http://www.journalijdr.com

IJDR

International Journal of Development Research
Vol. 15, Issue, 05, pp. 68320-68323, May, 2025
https://doi.org/10.37118/ijdr.29535.05.2025

RESEARCH ARTICLE OPEN ACCESS

DIGITAL ASSISTANT ABUSE: IMPLICATIONS FOR SOCIAL INTERACTION PATTERNS AND CHILD DEVELOPMENT

***Luiz Guilherme de Carvalho Antunes**

Departamento de Relações Públicas, Publicidade e Turismo, Escola de Comunicações e Artes, Universidade de São Paulo. Av. Prof. Lúcio Martins Rodrigues, 443 - Butantã, São Paulo - SP

ARTICLE INFO

Article History:
Received 10th February, 2025
Received in revised form
19th March, 2025
Accepted 11th April, 2025
Published online 25th May, 2025

Key Words:
Digital assistants, Voice interfaces, Artificial Intelligence, Childhood Development, Communication Ethics.

****Corresponding author:***
Luiz Guilherme de Carvalho Antunes

ABSTRACT

This study examines digital assistant abuse—the intentional misuse of AI-powered voice interfaces—and its societal implications. Drawing from communication theory and empirical research, we analyze how abuse of these technologies may normalize disrespectful communication patterns and erode social norms. Our investigation focuses on four dimensions: technical vulnerabilities enabling abuse, developmental impacts on children, gendered design perpetuating stereotypes, and ethical implications of consequence-free negative interactions with anthropomorphized technology. Findings reveal significant concerns, including that 73.4% of confounding voice commands shared between child and adult skills prioritize adult content, potentially exposing children to inappropriate material. Despite 76.3% of parents reporting awareness of parental control features, only 29.4% implement them. We propose targeted interventions including enhanced content moderation, improved interface transparency, gender-neutral design options, and educational initiatives to promote communication patterns that positively shape social behaviors.

Citation: Luiz Guilherme de Carvalho Antunes. 2025. *"*Digital assistant Abuse: Implications for Social interaction Patterns and Child Development*". International Journal of Development Research*, 15, (05), 68320-68323.

INTRODUCTION

The proliferation of digital assistants represents a fundamental shift in human-computer interaction, with widespread adoption transforming daily communication practices across diverse populations (Lovato & Piper, 2019). These AI-powered interfaces—exemplified by Amazon Alexa, Google Assistant, and Apple Siri—now mediate numerous communicative activities, from information retrieval to social coordination (Beneteau et al., 2020). Their ubiquity in modern households makes them significant agents of socialization, especially for younger users whose communication norms are still developing.

Communication scholarship has extensively examined technological mediation of human interaction (Baym, 2015; Sundar, 2020), but has insufficiently addressed how interaction patterns with non-human entities might transfer to human communication contexts. This research gap becomes particularly concerning as voice-based interfaces increasingly adopt anthropomorphic characteristics that blur the line between human and machine interaction (Gambino et al., 2020). This study introduces and examines the concept of "digital assistant abuse"—defined as the intentional misuse or exploitation of digital assistant functionalities, often accompanied by disrespectful or abusive communication patterns. While research on technology-facilitated abuse typically focuses on harm between humans (Dragiewicz et al., 2018), this investigation explores how abusive patterns directed at non-human entities may normalize problematic communication behaviors more broadly.

Our research addresses three critical questions that remain underexplored in current literature:

1. What technical and design features of digital assistants facilitate abuse?
2. How might abusive interactions with digital assistants influence developing communication norms, particularly among children?
3. What are the ethical implications of consequence-free negative interactions with anthropomorphized technology?

Conceptualizing Digital Assistant Abuse: Digital assistant abuse represents a distinct subcategory of technology-facilitated abuse that warrants specific theoretical consideration. While established definitions of digital abuse often center on interpersonal harm—such as the Anti-Defamation League's (2023) focus on intimidation, threats, or bullying—digital assistant abuse extends beyond direct harm to humans. Drawing from communication accommodation theory (Giles, 2016), we conceptualize this phenomenon as involving problematic communication adaptations where users modify their communication patterns when interacting with digital entities in ways that may transfer to human interaction contexts. These adaptations

include: (1) intentional use of hostile or demeaning language without social consequences; (2) exploitation of system vulnerabilities to elicit inappropriate responses; and (3) strategic manipulation of system limitations for entertainment or status enhancement among peers.

Relevant Theoretical Frameworks: Media psychology research suggests that interactions with technology often mirror social processes found in human-to-human communication (Reeves & Nass, 1996; Sundar & Nass, 2000). The Computers as Social Actors (CASA) paradigm posits that humans naturally apply social rules and expectations to technological entities, responding to computers as social beings despite knowing they are machines (Nass & Moon, 2000). Building on social learning theory (Bandura, 1977), we propose that the absence of negative consequences for abusive communication with digital assistants may facilitate behavioral disinhibition. This phenomenon could be particularly problematic for developing communicators who use social observations to establish normative behaviors.

Technical Vulnerabilities Facilitating Abuse

System Architecture and Vulnerabilities: Current digital assistants operate through complex systems with multiple potential exploitation points. Voice command interpretation relies on natural language processing algorithms that may be manipulated through carefully crafted inputs designed to bypass content filters (Edu et al., 2021). Additionally, many digital assistants employ open skill ecosystems where third-party developers can create functionalities with varying levels of scrutiny (Chung & Lee, 2018). Research by Le et al. (2025) identified particularly concerning vulnerabilities related to "confounding utterances"—voice commands that can trigger multiple different skills. Their analysis of 4,487 confounding utterances found that 581 (12.9%) were shared between both child-oriented and general-audience skills. When testing these utterances, they discovered that in 73.4% of cases where a relevant skill was invoked, the system prioritized non-child content over child-oriented options. This technical vulnerability creates significant risk for younger users, who may inadvertently access inappropriate content despite attempting to engage with age-appropriate functions. The research also identified "sneaky skills"—enabled skills that do not appear in user interfaces due to technical bugs—further complicating parental oversight.

Developmental Implications and Child Safety

Observational Learning and Normalization: Children's communication patterns develop significantly through observational learning (Vygotsky, 1978), with social modeling serving as a primary mechanism for acquiring communicative competence (Bandura, 1977). When children observe adults or peers engaging in abusive communication with digital assistants without consequences, they may incorporate these patterns into their own communication repertoire. Druga et al. (2021) found that children as young as four years old demonstrate anthropomorphic perceptions of smart speakers, attributing human-like qualities including feelings and thoughts to these devices. This tendency to humanize digital assistants increases the potential for communication transfer effects, where interaction patterns with technology inform broader communication behaviors.

Parental Awareness and Protection Gaps: Research reveals concerning gaps in parental protection measures. In a study of 232 U.S. parents with children under 13 who use Amazon Alexa, Le et al. (2025) found that while parents expressed significant concerns about explicit content, they were less concerned about skills that requested personal information from children—a finding that indicates misaligned risk perception. The study also revealed that despite 76.3% of parents claiming awareness of parental control features, only 29.4% actually implemented them. This implementation gap leaves children vulnerable to potentially harmful interactions and inappropriate content exposure. Additionally, only 6.8% of households with children used child-specific device versions with built-in protective features, suggesting that convenience often supersedes safety considerations.

Gendered Design and Stereotype Reinforcement

Prevalence of Female-Gendered Assistants: The predominance of female-gendered digital assistants represents a significant concern from a communication ecology perspective. West et al. (2019) documented that most mainstream digital assistants feature feminine names, voices, and personas—a design choice that reinforces gender stereotypes through repeated exposure and interaction. Communication scholarship has established that repeated exposure to stereotypical representations strengthens cognitive associations between gender and specific roles or traits (Mastro & Tukachinsky, 2011). The ubiquity of female-gendered assistants may reinforce associations between women and subservient, helpful, or administrative roles.

Differential Abuse Patterns: Research reveals troubling patterns in how users communicate with differently-gendered AI entities. Woods (2018) found that female-presenting chatbots and virtual assistants receive significantly higher rates of sexually explicit language and gender-based harassment compared to male-presenting or gender-neutral alternatives. Similarly, Brahnam and De Angeli (2012) documented that female-gendered conversational agents faced more sexualized and aggressive communication than their male counterparts. These findings suggest that digital assistant abuse may not only reinforce gender stereotypes but also provide consequence-free opportunities to express misogynistic attitudes that could subsequently transfer to human interaction contexts.

Ethical Implications of Consequence-Free Negative Communication

Moral Disengagement and Disinhibition: The absence of perceived consequences for abusive communication with digital assistants raises significant ethical concerns related to moral disengagement—the process through which individuals rationalize harmful behavior by disabling self-censure (Bandura et al., 1996). Digital assistant interactions may facilitate several mechanisms of moral disengagement, including:

1. **Moral justification**: Rationalizing abuse by framing it as harmless because the recipient "isn't real"
2. **Dehumanization**: Emphasizing the non-human nature of the technology to justify treatment that would be unacceptable toward humans
3. **Diffusion of responsibility**: Attributing responsibility to technology designers rather than individual users

These processes align with online disinhibition effects observed in computer-mediated communication (Suler, 2004), where the absence of immediate social consequences can lead to communication behaviors that would be socially sanctioned in face-to-face contexts.

Design Ethics and Corporate Responsibility: From an ethical perspective, digital assistant design decisions that facilitate or fail to discourage abuse raise important questions about corporate responsibility. Howard and Borenstein (2018) argue that AI development should incorporate ethical principles including transparency, justice, non-maleficence, responsibility, and privacy. The design of systems that passively accept abusive language without appropriate responses represents an ethical failure that may inadvertently contribute to the normalization of disrespectful communication patterns. As Friedman and Hendry (2019) note in their value-sensitive design framework, technologies embed and

express values—making the passive acceptance of abuse a concerning value statement.

DISCUSSION AND IMPLICATIONS

This investigation extends several theoretical frameworks in communication research. First, it broadens the CASA paradigm by examining negative social interactions with technologies, moving beyond the primarily positive interactions that dominate current literature. Second, it contributes to communication accommodation theory by exploring how humans modify their communication patterns when interacting with non-human entities and how these patterns might transfer to human interaction contexts. Additionally, our conceptualization of digital assistant abuse provides a foundation for understanding emerging communication phenomena at the intersection of human-computer interaction and interpersonal communication—a growing area as AI systems become more conversational and integrated into daily life.

Practical Recommendations

Based on our findings, we propose several evidence-based interventions:

1. ***Enhanced Content Moderation****:* Digital assistant providers should implement more robust vetting systems for third-party applications, particularly those targeting children. These systems should include both initial screening and ongoing monitoring to ensure compliance with content standards.
2. ***User Interface Transparency****:* Platforms should eliminate "sneaky skills" and implement clear visual indicators of which applications are enabled on devices, particularly for child-oriented functions.
3. ***Parental Control Accessibility****:* Providers should make parental controls more intuitive, accessible, and enabled by default on all devices likely to be used by children. Additionally, educational resources should help parents understand the importance of these features.
4. ***Gender-Neutral Design Options****:* Platforms should offer users a choice of voices and personas for digital assistants, including gender-neutral options, to mitigate stereotype reinforcement.
5. ***Response Design for Abuse****:* Digital assistants should be programmed to recognize abusive language and respond in ways that discourage rather than reinforce such communication patterns.
6. ***Educational Initiatives****:* Resources should be developed for parents and educators to discuss with children the differences between interacting with AI and humans, emphasizing respectful communication regardless of context.

Limitations and Future Research Directions: This study has several limitations that suggest directions for future research. First, longitudinal studies are needed to examine whether interaction patterns with digital assistants actually transfer to human communication contexts over time. Additionally, experimental research could test whether different digital assistant responses to abusive communication influence subsequent user behavior. Future research should also examine cultural variations in digital assistant abuse patterns and explore how personality factors, previous technology experience, and individual differences influence propensity for digital assistant abuse. Finally, design-based research could test the effectiveness of different intervention strategies in reducing abusive communication with digital assistants.

CONCLUSION

The growing integration of digital assistants into daily communication practices presents both opportunities and risks for social interaction norms. This investigation has identified concerning patterns of digital assistant abuse that may normalize disrespectful communication, reinforce harmful stereotypes, and create risks particularly for developing communicators. By addressing the technical vulnerabilities, developmental implications, gendered design concerns, and ethical questions surrounding digital assistant abuse, this research provides a foundation for more responsible development and use of these technologies. Our findings contribute to both theoretical understanding and practical applications in the field of communication technology. From a theoretical perspective, the conceptualization of digital assistant abuse extends the CASA paradigm by illustrating how anthropomorphized technologies may serve as conduits for problematic communication patterns. Similarly, our research bridges communication accommodation theory with technological mediation, demonstrating how accommodative behaviors initially developed for non-human entities may transfer to human interaction contexts with potentially detrimental effects. The gendered dimensions of digital assistant abuse deserve particular attention. The predominance of female-voiced assistants combined with the higher rates of sexualized and aggressive communication directed at these interfaces reflects broader societal patterns of gender-based harassment. This phenomenon represents a concerning feedback loop, where existing biases inform design choices that subsequently reinforce and potentially amplify those same biases. Breaking this cycle requires intentional intervention from technology developers, educators, and policymakers alike.

Equally concerning are our findings regarding child exposure to digital assistant abuse. The statistical evidence—particularly that 73.4% of confounding voice commands shared between child and adult skills prioritize adult content, while only 29.4% of parents implement available parental controls—indicates a significant gap between technological capabilities and protective practices. This discrepancy creates vulnerabilities precisely for those users whose communication norms and patterns are most actively developing. As AI-powered technologies continue to evolve and integrate more deeply into daily life, the potential impact of digital assistant abuse on social norms may intensify. Future generations will grow up in environments where interaction with non-human entities constitutes a significant portion of daily communication experiences. The normalization of abusive or disrespectful communication patterns in these contexts may have long-term implications for interpersonal communication competence that we are only beginning to understand. As digital assistants become increasingly sophisticated and ubiquitous, ensuring they foster rather than undermine healthy communication norms becomes an essential consideration for technology designers, parents, educators, and policymakers alike. The decisions made today about how these technologies respond to and potentially shape communication behaviors will have lasting implications for social interaction patterns in an increasingly AI-mediated world.

REFERENCES

Anti-Defamation League. (2023). Digital abuse: Definition and prevention strategies. ADL Research Report, 5-12.

Bandura, A. (1977). Social learning theory. Prentice Hall.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Mechanisms of moral disengagement in the exercise of moral agency. Journal of Personality and Social Psychology, 71(2), 364-374. https://doi.org/10.1037/0022-3514.71.2.364

Baym, N. K. (2015). Personal connections in the digital age. Polity Press.

Beneteau, E., Richards, O. K., Zhang, M., Kientz, J. A., Yip, J., & Hiniker, A. (2020). Communication breakdowns between families and Alexa. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-13). https://doi.org/10.1145/3313831.3376879

Brahnam, S., & De Angeli, A. (2012). Gender affordances of conversational agents. Interacting with Computers, 24(3), 139-153. https://doi.org/10.1016/j.intcom.2012.05.001

Chung, H., & Lee, S. (2018). Intelligent virtual assistant knows your life. arXiv preprint arXiv:1803.00466.

Dragiewicz, M., Burgess, J., Matamoros-Fernández, A., Salter, M., Suzor, N. P., Woodlock, D., & Harris, B. (2018). Technology facilitated coercive control: Domestic violence and the competing roles of digital media platforms. Feminist Media Studies, 18(4), 609-625. https://doi.org/10.1080/14680777.2018.1447341

Druga, S., Williams, R., Breazeal, C., & Resnick, M. (2021). "Hey Google is it OK if I eat you?": Initial explorations in child-agent interaction. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-15). https://doi.org/10.1145/3411764.3445053

Edu, J. S., Such, J. M., & Suárez-Tangil, G. (2021). Smart home personal assistants: A security and privacy review. ACM Computing Surveys, 53(6), 1-36. https://doi.org/10.1145/3386160

Friedman, B., & Hendry, D. G. (2019). Value sensitive design: Shaping technology with moral imagination. MIT Press.

Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. Human-Machine Communication, 1(1), 71-86. https://doi.org/10.30658/hmc.1.5

Giles, H. (2016). Communication accommodation theory: Negotiating personal relationships and social identities across contexts. Cambridge University Press.

Howard, A., & Borenstein, J. (2018). The ugly truth about ourselves and our robot creations: The problem of bias and social inequity. Science and Engineering Ethics, 24(5), 1521-1536. https://doi.org/10.1007/s11948-017-9975-2

Le, T. H., Varshney, S., Shin, J., Saxena, N., & Shahriar, H. (2025). SkillBot: Identifying risky content for children in Alexa skills. In Proceedings of the 2025 Conference on Security and Privacy (pp. 1-20).

Lovato, S. B., & Piper, A. M. (2019). Young children and voice search: Information seeking processes and potential learning benefits. Journal of the Association for Information Science and Technology, 70(6), 616-627. https://doi.org/10.1002/asi.24142

Mastro, D., & Tukachinsky, R. (2011). The influence of exemplar versus prototype-based media representations on racial/ethnic evaluations. Journal of Communication, 61(5), 916-937. https://doi.org/10.1111/j.1460-2466.2011.01587.x

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. Journal of Social Issues, 56(1), 81-103. https://doi.org/10.1111/0022-4537.00153

Reeves, B., & Nass, C. (1996). The media equation: How people treat computers, television, and new media like real people and places. Cambridge University Press.

Suler, J. (2004). The online disinhibition effect. Cyberpsychology & Behavior, 7(3), 321-326. https://doi.org/10.1089/1094931041291295

Sundar, S. S. (2020). Rise of machine agency: A framework for studying the psychology of human–AI interaction (HAII). Journal of Computer-Mediated Communication, 25(1), 74-88. https://doi.org/10.1093/jcmc/zmz026

Sundar, S. S., & Nass, C. (2000). Source orientation in human-computer interaction: Programmer, networker, or independent social actor. Communication Research, 27(6), 683-703. https://doi.org/10.1177/009365000027006001

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

West, M., Kraut, R., & Ei Chew, H. (2019). I'd blush if I could: Closing gender divides in digital skills through education. UNESCO GEN/2019/EQUALS/1 REV.2.

Woods, H. S. (2018). Asking more of Siri and Alexa: Feminine persona in service of surveillance capitalism. Critical Studies in Media Communication, 35(4), 334-349. https://doi.org/10.1080/15295036.2018.1488082

****