Privacy Risks for Underaged User Groups in LLM-based Conversational Agent Interactions

ANONYMOUS AUTHOR(S)

Generative AI, particularly Large Language Model (LLM)-based conversational agents (CAs), presents emerging privacy risks, especially for younger users. Children and adolescents, who are more vulnerable to online risk-taking, may struggle to recognize and mitigate these risks, leading to excessive personal data disclosures. This workshop paper explores privacy issues in LLM-based CAs, drawing on existing research to highlight the unique vulnerabilities of younger users. Given the challenges younger users face in assessing these risks, we emphasize the need for AI systems that balance privacy protection with user autonomy. We discuss design considerations, including clearer privacy policies, adaptive content filters, and educational safeguards, to ensure safer interactions without overly restricting user engagement.

CCS Concepts: • **Human-centered computing → Human computer interaction (HCI)**.

Additional Key Words and Phrases: Open-ended AI, Anthropomorphization

ACM Reference Format:

Anonymous Author(s). 2018. Privacy Risks for Underaged User Groups in LLM-based Conversational Agent Interactions. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 7 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Generative Artificial Intelligence (AI) encompasses machine learning models capable of producing original content—including text, images, music, video, and code—by identifying and replicating patterns from previously analyzed data [2]. The rapid progress in this field has led to its integration into mainstream consumer applications, particularly in conversational agents (CAs) powered by Large Language Models (LLMs) (e.g., [24, 36, 37, 43]). These systems enable functionalities such as text-to-speech, as seen in CAs like OpenAI's ChatGPT [5, 34, 38]. Thereby, disclosing private and sensitive information in LLM-based CAs can expose users to a range of emerging privacy and security risks [12, 42, 54, 62] which are partly novel and partly amplify existing online privacy risk [30].

Among all user groups, children, and adolescents are more prone to taking risks online [22] and may struggle to assess dangerous situations including the ones emerging from AI use. This makes them more susceptible to being misled and to underestimate the long-term consequences of their actions in the digital world compared to adults [9, 48].

In this workshop paper, we explore relevant research on privacy issues in LLM-based CAs and outline their impact on younger user groups. We emphasize the importance of designing CAs that mitigate privacy risks while ensuring that even less risk-aware users can engage autonomously. Our discussion aims to promote a balanced approach to CA design — one that protects user privacy without being overly restrictive or paternalistic. Rather than imposing strict controls that limit user autonomy, the goal should be to develop systems that provide meaningful privacy protections while allowing users, especially those less aware of privacy risks, to engage freely and make informed decisions.

Manuscript submitted to ACM

2 Background on User Privacy Risks in LLM-Based Conversational Agents

Based on Solove´s established privacy taxonomy [46], we outline privacy risks for data collection, processing as well as disseminating and provide background on current technical privacy measures.

Privacy issues in data collection. In digital environments, personal data collection through cookies and metadata is often viewed negatively due to inadequate privacy notices [1, 21]. The rise of AI further obscures data collection, reducing transparency around user consent [21]. Specifically, LLMs rely on vast amounts of personal data for training, further intensifying surveillance risks by expanding data collection [31]. Moreover, the anthropomorphizing of dialogue-based LLM applications poses particular risks of privacy violations due to users' tendency to overshare [54]. Studies show that users may overestimate LLMs' capabilities, perceiving them as human-like and disclosing more information than intended [31, 39, 54]. Users are also more likely to share personal details with helpful, human-like chatbots and accept privacy intrusions [23, 40, 52]. Even when aware the chatbot is not human, its human-like traits and lack of social judgment encourage oversharing, which can be exploited for harmful purposes like promoting addictive content [40, 54].

Privacy issues in data processing. This vast amount of data collected is often not only stored on the service providers' databases but processed and reused for machine learning purposes and shared with third parties [31, 53]. In the context of AI, including LLMs, privacy violations can occur during data processing [31, 54]. Based on the data captured, specific data points may be linked and combined to identify individuals, leading to inferences about their personality, social characteristics, and emotional attributes. LLMs, for example, can enhance the accuracy of predictions about sensitive traits such as sexual orientation, gender, or religious beliefs [54]. This capability may lead to the creation of detailed profiles that include true and sensitive information without the individual's knowledge or consent. Without users' consent, their data risks being repurposed for objectives beyond the original intent [31]. End-users are oftentimes neither informed nor given control over how their data is utilized, including its inclusion in training datasets. This way of handling personal information leaves it exposed to potential leaks and unauthorized access.

Privacy issues in data dissemination. The dissemination of sensitive information from LLM-trained data poses significant risks [32, 54]. For instance, GPT-2 has unintentionally disclosed personal information, like phone numbers and email addresses, from its training data [13]. Similarly, GPT-3-based Co-pilot exposed sensitive API keys, potentially allowing unauthorized database access [27]. Beyond sharing genuine sensitive data, LLM also risks spreading false or misleading content [54]. Additionally, LLMs and text-to-image (TTI) models present copyright and cybersecurity concerns [10, 54]. LLMs may generate content that, while not directly violating copyright, exploits original creators' ideas [10]. This is particularly evident in TTI models, like DALL-E, which can mimic specific artists' styles, and potentially be engineered to exploit certain images for financial benefit [10, 47]. Both technologies also pose cybersecurity threats, such as generating personalized phishing emails or convincing visuals for scams [10, 54].

Privacy safeguards in LLM. Much research has been dedicated to protecting user privacy in LLM usage, leading to the development of various frameworks and methods. For example, pre-processing techniques filter or replace sensitive information, such as personal identifiers (e.g., names, addresses), before it reaches the model [25]. Differential privacy adds noise to data or model updates, preventing the model from memorizing and leaking specific user data while maintaining overall performance [45]. Federated learning further enhances privacy by training models in a decentralized manner, minimizing data exposure by keeping information local [65]. Beyond LLM models themselves, limited research has explored how users, and in particular underage, interact and respond to privacy issues with CA based on LLM [61, 63].

3 Children and Adolescents as Vulnerable User Groups for LLM-based Conversational Agents

Existing research on children and adolescents' interactions with generative AI has predominantly focused on educational applications, often aiming to foster creativity and understanding of machine learning processes (e.g., [3, 4, 35]). However, ethical considerations and potential risks associated with generative AI remain significantly underexplored, particularly for underage user groups (e.g., [26, 60, 61]). While studies on children's engagement with AI-driven technologies are scarce, research on related digital interactions suggests that children struggle to understand the full extent of privacy risks. For example, Zhao et al. [64] examined children's privacy reactions to online in-app games and found that while younger children (ages 6–10) could identify overt threats like inappropriate content or oversharing of personal information, they had difficulty recognizing more subtle risks such as implicit data collection through tracking and in-app recommendations. When using state-of-the-art LLM-based CAs, it is reasonable to expect that children may face similar, if not heightened, challenges in identifying and mitigating privacy risks.

What supports concerns of heightened privacy risks for minors is research on chatbot interactions demonstrating that children and adolescents's perceptions of CAs can significantly influence their disclosure behaviour. Indeed, Pérez-Marín and Pascual-Nieto [41] found that minors aged 11–14 responded to chatbots differently based on their perceived personalities and moods, often exhibiting increased self-disclosure when the chatbot displayed emotions. Importantly, children tended to anthropomorphize these conversational agents, treating them as human-like friends rather than digital tools. This perceived humanness not only strengthened prosocial behaviours but also led children to share personal information more readily. Although this study did not examine AI-powered chatbots, it underscores the tendency of children to form social connections with CAs, which may amplify privacy risks when interacting with more advanced LLM-driven systems [41]. LLM-based CAs exhibit even more human-like interactions and encourage free-form interactions that can elicit personal disclosures [23, 40, 52, 54] leading to an increase in both traditional privacy risks and novel concerns related to AI's data memorization capabilities.

Personal information can be collected not only explicitly but also implicitly through subtle conversational cues [29]. This presents another significant challenge to user awareness, particularly for minors, who may not fully comprehend how their data is being captured and processed [20]. Indeed, a 2021 international study by [United Nations Children's Fund (UNICEF)] found that adolescents generally had low awareness of AI-related risks. While only a few study participants demonstrated a clear awareness of the dangers, the majority either had naïve mental models of (potentially risky) data flows or lacked awareness altogether [51].

From a legal perspective, persons are considered children up to the age of 16 (e.g., [11, 17, 18]). Taking the GDPR as an example [17], under Article 8, online services including LLM-based CAs AI must obtain verifiable parental consent before collecting personal data from children under 16 (or 13 in Norway, Spain, Sweden, and the United Kingdom). This ensures that minors do not unknowingly share sensitive information without oversight. Furthermore, transparency is another key aspect of GDPR compliance. Article 12 requires AI chatbot providers to present clear, child-friendly explanations regarding data collection and processing. While these are the legal rules, practically this does not entirely prevent the usage of CAs by teenagers as many AI-powered chatbot providers (e.g., [14, 19, 38]) can be subscribed to through an email account. The service provider asks for the birthdate but one may argue that this can easily be faked. Furthermore, the majority of these popular chat services provide an untailored privacy policy for all service users which is further characterized by long and partly complicated texts, not prominently displayed in the user interface.

While the legal protection spans up until 16, from a developmental perspective persons from age 10 to 19 can already be considered adolescents [55] and people in this age group are characterized by being highly social and more prone to

risk-taking than younger children [49]. This means, that teenagers often underestimate their ability to avoid risk while simultaneously engaging in more risky behaviours - also online [8, 58] - due to their tendency to underestimate the risks themselves [15]. This further highlights that children and especially adolescents are very vulnerable to privacy risks when interacting with LLM-based chatbots.

4 Call to Action for Potential Mitigation Approaches

While adolescents seek independence as part of the individuation process from their parents, they are still less capable than adults to manage online - including AI- risks without some level of guidance [8, 58]. However, some degree of risk-taking and autonomy-seeking is a natural and essential aspect of adolescence [7]. Thus, restricting these experiences may hinder developmental growth, as teens rely on them to gradually separate from their parents and develop into well-adjusted, independent adults [8, 49, 50]. There is therefore a need for a "safe space" for adolescents enabling both privacy self-regulation and sufficient parental control [56]. This aligns with Value Sensitive Design (VSD) [6], which suggests shifting from restrictive parental control towards collaborative safety mechanisms that balance parental oversight with adolescent autonomy. Currently, a wide array of parental safety controls exists, but they prioritize authoritarian restriction and privacy-invasive monitoring by parents [56, 57]. These approaches not only risk damaging parent-teen trust but also fail to account for the dynamic and unpredictable nature of generative AI [60, 61], which make output monitoring difficult. Moreover, defining appropriate content for teenagers remains a challenge, as they are not a homogeneous group [61]. Even more so, there seems to be a disparity between parents' assumptions about AI usage and the reality of adolescents' engagement with these technologies [61]. While parents tend to focus on mainstream tools like ChatGPT, teens are increasingly drawn to character-based chatbots on social media platforms, which may blur the line between AI companionship and human-like relationships [14, 43, 61].

This raises an important question: what are teens using chatbots for, and is this usage beneficial or problematic? While some may use them for harmless exploration or self-expression, research suggests that excessive chatbot interaction can lead to problematic usage patterns, including increased privacy disclosure and even emotional dependence [33, 59–61]. A particularly alarming case involved a teenager who tragically took his own life after developing an intense relationship with a chatbot [44]. While preventing such incidents is paramount, it is crucial to consider whether privacy-preserving designs and structured emotional regulation tools [6] could have helped mitigate such risks. More effective age-restricting technologies, such as platforms with strict age verification and promotion of shared safety settings such as parental alerts and risk ratings, offer one potential solution [6]. However, this leads to challenges of their own, including the sufficient preservation of the minor users' privacy with respect to the legal guardian in control of the safety settings and potentially the provider of the Generative AI system too. The ideal approach would allow teenagers to explore independently while enabling parental intervention solely when needed. With these features, parents could rely on alerts or risk-ranking rather than direct content access, while in more dangerous cases AI-generated outputs remain reviewable for refining risk detection [61]. However, such approaches and details of potential solutions for privacy in LLM-based chatbots need further exploration and research.

Similarly, age-appropriate privacy policies must be provided - an issue not only related to LLM-based CAs. Lengthy and complex traditional legal warnings are often unsuitable for younger audiences; instead, risk communication strategies should incorporate visual storytelling, interactive guidance, and contextual nudges to promote safer online behaviors [16, 28]. Finally, research [6] also suggests that chatbots themselves could integrate real-time feedback mechanisms that encourage responsible disclosure behaviors while preserving a user's sense of autonomy. How to design those mechanisms is another very interesting avenue for future research.

References

[1] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. 2015. Privacy and human behavior in the age of information. *Science* 347, 6221 (2015), 509–514. doi:10.1126/science.aaa1465 arXiv:https://www.science.org/doi/pdf/10.1126/science.aaa1465

[2] A. Aggarwal, M. Mittal, and G. Battineni. 2021. Generative adversarial network: An overview of theory and applications. *International Journal of Information Management Data Insights* 1, 1 (2021), 100004.

[3] Safinah Ali, Daniella DiPaola, and Cynthia Breazeal. 2021. What are GANs?: introducing generative adversarial networks to middle school students. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 15472–15479.

[4] Safinah Ali, Prerna Ravi, Randi Williams, Daniella DiPaola, and Cynthia Breazeal. 2024. Constructing dreams using generative AI. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 23268–23275.

[5] NAIF J ALOTAIBI. 2024. ChatGPT - image generator. https://chat.openai.com/g/g-pmuQfob8d-image-generator. (Accessed on 03/27/2024).

[6] Karla Badillo-Urquiola, Chhaya Chouhan, Stevie Chancellor, Munmun De Choudhary, and Pamela Wisniewski. 2020. Beyond parental control: designing adolescent online safety apps using value sensitive design. *Journal of adolescent research* 35, 1 (2020), 147–175.

[7] Diana Baumrind. 1987. A developmental perspective on adolescent risk taking in contemporary America. *New directions for child and adolescent development* 1987, 37 (1987), 93–125.

[8] D. Baumrind. 2005. Patterns of parental authority and adolescent autonomy. *New Directions for Child and Adolescent Development* 2005, 108 (2005), 61–69.

[9] J. Bessant. 2008. Hard wired for risk: Neurological science, 'the adolescent brain' and developmental theory. *Journal of Youth Studies* 11, 3 (2008), 347–360.

[10] Charlotte Bird, Eddie Ungless, and Atoosa Kasirzadeh. 2023. Typology of Risks of Generative Text-to-Image Models. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (Montréal, QC, Canada) *(AIES '23)*. Association for Computing Machinery, New York, NY, USA, 396–410. doi:10.1145/3600211.3604722

[11] California Attorney General's Office. 2024. California Consumer Privacy Act (CCPA). https://oag.ca.gov/privacy/ccpa Accessed: 2024-02-20.

[12] Nicholas Carlini, Daphne Ippolito, Matthew Jagielski, Katherine Lee, Florian Tramer, and Chiyuan Zhang. 2022. Quantifying memorization across neural language models. *arXiv preprint* arXiv:2202.07646 (2022).

[13] Nicholas Carlini, Florian Tramèr, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, Tom Brown, Dawn Song, Úlfar Erlingsson, Alina Oprea, and Colin Raffel. 2021. Extracting Training Data from Large Language Models. In *30th USENIX Security Symposium (USENIX Security 21)*. USENIX Association, Virtual Event, 2633–2650. https://www.usenix.org/conference/usenixsecurity21/presentation/carlini-extracting

[14] Character.AI. 2024. Character.AI - Conversational AI Platform. https://character.ai/ Accessed: 2024-02-20.

[15] Lawrence D Cohn, Susan Macfarlane, Claudia Yanez, and Walter K Imai. 1995. Risk-perception: differences between adolescents and adults. *Health psychology* 14, 3 (1995), 217.

[16] John Dempsey, Gavin Sim, Brendan Cassidy, and Vinh-Thong Ta. 2022. Children designing privacy warnings: Informing a set of design guidelines. *International Journal of Child-Computer Interaction* 31 (2022), 100446.

[17] European Parliament and Council. 2016. Regulation (EU) 2016/679 - General Data Protection Regulation (GDPR). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679 Accessed: 2024-02-20.

[18] Federal Trade Commission (FTC). 2024. Children's Online Privacy Protection Rule (COPPA). https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa Accessed: 2024-02-20.

[19] Google. 2024. Google Gemini. https://gemini.google.com/?hl=es-ES Accessed: 2024-02-20.

[20] Ece Gumusel. 2025. A literature review of user privacy concerns in conversational chatbots: A social informatics approach: An Annual Review of Information Science and Technology (ARIST) paper. *Journal of the Association for Information Science and Technology* 76, 1 (2025), 121–154.

[21] Joanne Hinds and Adam N Joinson. 2018. What demographic attributes do our digital footprints reveal? A systematic review. *PloS one* 13, 11 (2018), e0207112. https://doi.org/10.1371/journal.pone.0207112

[22] A. Hope. 2007. Risk taking, boundary performance and intentional school internet "misuse". *Discourse: studies in the cultural politics of education* 28, 1 (2007), 87–99.

[23] Carolin Ischen, Theo Araujo, Hilde Voorveld, Guda van Noort, and Edith Smit. 2020. Privacy concerns in chatbot interactions. In *Chatbot Research and Design: Third International Workshop (Lecture Notes in Computer Science)*. Springer, Amsterdam, Netherlands, 34–48. https://doi.org/10.1007/978-3-030-39540-7_3

[24] John Snow Labs. 2024. Medical Chatbot - John Snow Labs. https://www.johnsnowlabs.com/medical-chatbot/ Accessed: 2024-02-20.

[25] Zhigang Kan, Linbo Qiao, Hao Yu, Liwen Peng, Yifu Gao, and Dongsheng Li. 2023. Protecting User Privacy in Remote Conversational Systems: A Privacy-Preserving Framework Based on Text Sanitization. arXiv:2306.08223

[26] S. Khan, M. Iqbal, O. Osho, K. Singh, K. Derrick, P. Nelson, and B. Knijnenburg. 2024. Teaching Middle Schoolers about the Privacy Threats of Tracking and Pervasive Personalization: A Classroom Intervention Using Design-Based Research. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–26.

[27] Amit Kulkarni. 2021. *GitHub Copilot AI is Leaking Functional API Keys.* Analytics Drift. https://analyticsdrift.com/github-copilot-ai-is-leaking-functional-api-keys/ (Accessed on 08/26/2024).

[28] Priya Kumar, Shalmali Milind Naik, Utkarsha Ramesh Devkar, Marshini Chetty, Tamara L Clegg, and Jessica Vitak. 2017. 'No Telling Passcodes Out Because They're Private' Understanding Children's Mental Models of Privacy and Security Online. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–21.

[29] Martha Larson, Nelleke Oostdijk, and Frederik Zuiderveen Borgesius. 2021. Not directly stated, not explicitly stored: Conversational agents and the privacy threat of implicit information. In *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*. 388–391.

[30] Hao-Ping Lee, Yu-Ju Yang, Thomas Serban Von Davier, Jodi Forlizzi, and Sauvik Das. 2024. Deepfakes, Phrenology, Surveillance, and More! A Taxonomy of AI Privacy Risks. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–19.

[31] Hao-Ping (Hank) Lee, Yu-Ju Yang, Thomas Serban Von Davier, Jodi Forlizzi, and Sauvik Das. 2024. Deepfakes, Phrenology, Surveillance, and More! A Taxonomy of AI Privacy Risks. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 775, 19 pages. doi:10.1145/3613904.3642116

[32] Sangmin Lee, Edmund L. Wong, Deepak Goel, Mike Dahlin, and Vitaly Shmatikov. 2013. πBox: A Platform for Privacy-Preserving Apps. In *10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13)*. USENIX Association, Lombard, IL, 501–514. https://www.usenix.org/conference/nsdi13/technical-sessions/presentation/lee_sangmin

[33] Robert Mahari and Pat Pataranutaporn. 2024. We need to prepare for 'addictive intelligence. *MIT Technology Review. https://www. technologyreview. com/2024/08/0* 5 (2024), 1095600.

[34] D. Milmo. 2023. ChatGPT hits 100 million users in two months, making it the fastest-growing app ever. *The Guardian* (2023). Retrieved October 4, 2024, from https://www.theguardian.com/technology/2023/feb/02/chatgpt-100-million-users-open-ai-fastest-growing-app.

[35] Michele Newman, Kaiwen Sun, Ilena B Dalla Gasperina, Grace Y Shin, Matthew Kyle Pedraja, Ritesh Kanchi, Maia B Song, Rannie Li, Jin Ha Lee, and Jason Yip. 2024. " I want it to talk like Darth Vader": Helping Children Construct Creative Self-Efficacy with Generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–18.

[36] Emmanuel Opara, Adalikwu Mfon-Ette Theresa, and Tolorunleke Caroline Aduke. 2023. ChatGPT for teaching, learning and research: Prospects and challenges. *Opara Emmanuel Chinonso, Adalikwu Mfon-Ette Theresa, Tolorunleke Caroline Aduke (2023). ChatGPT for Teaching, Learning and Research: Prospects and Challenges. Glob Acad J Humanit Soc Sci* 5 (2023).

[37] OpenAI. 2024. Introducing the GPT Store. https://openai.com/blog/introducing-the-gpt-store. (Accessed on 03/27/2024).

[38] OpenAI. 2024. OpenAI - Advancing Artificial Intelligence. https://openai.com/ Accessed: 2024-02-20.

[39] Google PAIR. 2019. *People + AI Guidebook*. Google Research. https://design.google/aiguidebook

[40] Arielle Pardes. 2018. *The Emotional Chatbots Are Here to Probe Our Feelings*. Condé Nast. https://www.wired.com/story/replika-open-source/

[41] D. Pérez-Marín and I. Pascual-Nieto. 2013. An exploratory study on how children interact with pedagogic conversational agents. *Behaviour & Information Technology* 32, 9 (2013), 955–964. doi:10.1080/0144929X.2012.687774

[42] Charith Peris, Christophe Dupuy, Jimit Majmudar, Rahil Parikh, Sami Smaili, Richard Zemel, and Rahul Gupta. 2023. Privacy in the Time of Language Models. doi:10.1145/3539597.3575792

[43] Replika. 2024. Replika - The AI Companion Who Cares. https://replika.com/ Accessed: 2024-02-20.

[44] Kevin Roose. 2024. Character.AI Faces Lawsuit After Teen's Tragic Death. *The New York Times* (23 October 2024). https://www.nytimes.com/2024/10/23/technology/characterai-lawsuit-teen-suicide.html Accessed: 2025-02-21.

[45] Tanmay Singh, Harshvardhan Aditya, Vijay K Madisetti, and Arshdeep Bahga. 2024. Whispered tuning: Data privacy preservation in fine-tuning llms through differential privacy. *Journal of Software Engineering and Applications* 17, 1 (2024), 1–22. doi:10.4236/jsea.2024.171001.

[46] Daniel J Solove. 2005. A taxonomy of privacy. *U. Pa. l. Rev.* 154 (2005), 477.

[47] Gowthami Somepalli, Vasu Singla, Micah Goldblum, Jonas Geiping, and Tom Goldstein. 2023. Diffusion art or digital forgery? investigating data replication in diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Vancouver, Canada, 6048–6058. https://doi.org/10.48550/arXiv.2212.03860

[48] L. Steinberg. [n. d.]. A social neuroscience perspective on adolescent risk-taking. In *Biosocial theories of crime*.

[49] L. Steinberg. 2004. Risk taking in adolescence: what changes, and why? *Annals of the New York Academy of Sciences* 1021, 1 (2004), 51–58.

[50] L. Steinberg, S. D. Lamborn, N. Darling, N. S. Mounts, and S. M. Dornbusch. 1994. Over-time changes in adjustment and competence among adolescents from authoritative, authoritarian, indulgent, and neglectful families. *Child Development* 65, 3 (1994), 754–770.

[51] United Nations Children's Fund (UNICEF). 2021. Adolescent perspectives on artificial intelligence. UNICEF Office of Research - Innocenti. Retrieved from https://www.unicef.org/innocenti/stories/adolescent-perspectives-artificial-intelligence.

[52] Evert Van den Broeck, Brahim Zarouali, and Karolien Poels. 2019. Chatbot Advertising Effectiveness: When Does the Message Get Through? *Computers in Human Behavior* 98 (September 2019), 150–157. doi:10.1016/j.chb.2019.04.009

[53] Michael Veale, Midas Nouwens, and Cristiana Santos. 2022. Impossible asks: can the transparency and consent framework ever authorise real-time bidding after the Belgian DPA decision? *Technology and Regulation* 2022 (2022), 12–22. https://doi.org/10.26116/

[54] Laura Weidinger, Jonathan Uesato, Maribeth Rauh, Conor Griffin, Po-Sen Huang, John Mellor, Amelia Glaese, Myra Cheng, Borja Balle, Atoosa Kasirzadeh, et al. 2022. Taxonomy of risks posed by language models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 214–229.

[55] Canadian Paediatric Society (WHO). 2003. Age limits and adolescents. *Paediatrics Child Health* 8, 9 (2003), 577–578. doi:10.1093/pch/8.9.577

[56] Pamela Wisniewski, Arup Kumar Ghosh, Heng Xu, Mary Beth Rosson, and John M Carroll. 2017. Parental control vs. teen self-regulation: Is there a middle ground for mobile online safety?. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*.

51–69.

[57] Pamela J Wisniewski, Jessica Vitak, and Heidi Hartikainen. 2022. Privacy in adolescence. In *Modern socio-technical perspectives on privacy*. Springer International Publishing Cham, 315–336.

[58] J. Youniss. 1985. *Adolescent relations with mothers, fathers, and friends*. University of Chicago Press.

[59] Sen-Chi Yu, Hong-Ren Chen, and Yu-Wen Yang. 2024. Development and validation the Problematic ChatGPT Use Scale: a preliminary report. *Current Psychology* 43, 31 (2024), 26080–26092.

[60] Y. Yu. 2025. Safeguarding Children in Generative AI: Risk Frameworks and Parental Control Tools. In *Companion Proceedings of the 2025 ACM International Conference on Supporting Group Work*. 121–123.

[61] Yaman Yu, Tanusree Sharma, Melinda Hu, Justin Wang, and Yang Wang. 2024. Exploring Parent-Child Perceptions on Safety in Generative AI: Concerns, Mitigation Strategies, and Design Implications. *arXiv preprint arXiv:2406.10461* (2024).

[62] Chiyuan Zhang, Daphne Ippolito, Katherine Lee, Matthew Jagielski, Florian Tramèr, and Nicholas Carlini. 2021. Counterfactual memorization in neural language models. *arXiv preprint* arXiv:2112.12938 (2021).

[63] Zhiping Zhang, Michelle Jia, Hao-Ping Lee, Bingsheng Yao, Sauvik Das, Ada Lerner, Dakuo Wang, and Tianshi Li. 2024. "It's a Fair Game", or Is It? Examining How Users Navigate Disclosure Risks and Benefits When Using LLM-Based Conversational Agents. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Honolulu, HI, USA, 1–26. doi:10.1145/3613904.3642385

[64] J. Zhao, G. Wang, C. Dally, P. Slovak, J. Edbrooke-Childs, M. Van Kleek, and N. Shadbolt. 2019. "I Make Up a Silly Name": Understanding Children's Perception of Privacy Risks Online. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 1–13.

[65] JiaYing Zheng, HaiNan Zhang, LingXiang Wang, WangJie Qiu, HongWei Zheng, and ZhiMing Zheng. 2024. Safely Learning with Private Data: A Federated Learning Framework for Large Language Model. arXiv:2406.14898 Available at https://arxiv.org/abs/2406.14898.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009