Jiang, B. et al. (Eds.) (2025). Proceedings of the 33rd International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Privacy in the Age of Robotics: Protecting Personal Data in Classrooms

Antun DROBNJAK[a*], Ivan TERZIC[a], Branko KIRIN[b] & Ivica BOTICKI[a]
[a]*Faculty of electrical engineering and computing, University of Zagreb, Croatia*
[b]*LAFCO TECH, Zagreb, Croatia*
*antun.drobnjak@fer.unizg.hr

Abstract: This article explores the privacy challenges posed by artificial intelligence in embodied robotic systems and proposes technical, design, and governance responses. Robots generate raw sensor data and derived inferences, creating distinctive risks in human–robot interaction such as incidental capture, inferential leakage, algorithmic bias, and third-party exposure. Existing consent models and legal frameworks (GDPR, CCPA, EU AI Act) provide only partial protection, especially in contexts where robots operate persistently and without meaningful choice for users or bystanders. A classroom case study illustrates these concerns, showing how educational robots can expose children and teachers to privacy harms while also pointing to mitigation strategies, including privacy-by-design, transparency indicators, configurable controls, and privacy-enhancing technologies like edge AI and federated learning. The discussion emphasizes interdisciplinary collaboration, participatory deployment, and privacy impact assessment. The article concludes that embedding dignity and digital self-determination into system design and governance is essential for aligning innovation with accountability and trust.

Keywords: Privacy in robotics, Human-Robot Interaction (HRI), Educational robots, Artificial Intelligence in Education

1. Introduction – Embodied Privacy Risks

The rapid development of robotics is creating new opportunities across various fields, including education, healthcare, logistics, and domestic life (Henschel et al., 2021; Morgan et al., 2022; Ouyang & Xu, 2024). Unlike traditional digital technologies such as laptops and smartphones, which require deliberate user interaction, robots are embodied agents comprised of sensors and actuators that are capable of sensing, moving, and acting within the physical environment to accomplish a set of tasks. This embodiment introduces distinct privacy challenges, as robots collect and process data about individuals and their surroundings, often without explicit or conscious consent (Lin et al., 2011).

Artificial intelligence (AI) is a branch of computer science focused on creating systems that mimic human abilities such as learning, reasoning, and decision-making (Norvig & Intelligence, 2002). Unlike rule-based software, AI is data-driven and adaptive, making it effective in dynamic settings but also raising concerns about transparency, fairness, and accountability (Radanliev, 2025; Samana, 2023). The integration of AI into robotics amplifies both opportunity and risk, with privacy emerging as a central concern. The learning and adaptive capacities of AI-enabled robots magnify privacy risks. For example, model training may expose training membership through inference attacks, such as membership inference (Shokri et al., 2017), while model architectures themselves may leak sensitive attributes even under trusted conditions (Mireshghallah et al., 2020).

Contemporary data-protection regimes, notably the EU General Data Protection Regulation (GDPR, 2016) and the California Consumer Privacy Act (CCPA, 2020), provide important legal baselines but frequently struggle to keep pace with rapidly evolving AI-robotics capabilities. These regulatory gaps blur liability, accountability, and consent, especially in high-stakes contexts like remote surgery, autonomous decisions, or safety-critical interventions where tracing errors or data misuse is complex. Addressing them requires coordinated technical, legal, and ethical measures.

2. Data Collection in Robotics

Robotic platforms employ diverse sensors—cameras, microphones, LiDAR, infrared, and inertial units—that produce raw data (video, audio, depth maps, motion traces) and derived artefacts (maps, trajectories, behavioral profiles, biometric attributes) (Cadena et al., 2016; Siciliano & Khatib, 2016). These are essential for localization, mapping, perception, manipulation, and HRI (Lin et al., 2014; Thrun, 2002), but they raise privacy concerns beyond conventional information systems (Koops & Leenes, 2014).

First, the boundary between minimal sensing (e.g., obstacle detection) and intrusive inference (e.g., facial recognition, gait or emotion analysis) is opaque to non-experts (Lee et al., 2011). Second, robots in public or multi-user settings capture data about bystanders without consent, implicating third-party rights (Dietrich et al., 2023). Third, many systems lack transparency about collection, processing, retention, and redress, weakening consent and fairness (Koops & Leenes, 2014; Richards & King, 2014).

Taken together, robotic sensing expands both the scope of data and the capacity to infer sensitive attributes, amplifying ethical and regulatory challenges. Addressing these requires sensor-level design choices, data minimization, explainable pipelines, and governance mechanisms attuned to embodied sensing and mobility.

2.1 User Control, Consent, and Transparency

Traditional consent mechanisms are frequently ill-suited to robotic contexts because robots are mobile, autonomous, and embedded within socio-technical environments inhabited by multiple stakeholders. Robots can record people without their awareness, challenging informed consent and especially in classrooms, workplaces, and clinics where interaction may be unavoidable (Sharkey, 2016; van Wynsberghe et al., 2022). The asymmetry of knowledge and control between system operators, designers, and incidental subjects complicates attribution of consent and responsibility.

Proposed mitigations emphasize both technical and procedural measures. At the technical level, persistent and unambiguous transparency cues such as visual, auditory, or haptic indicators that signal when cameras, microphones, or other sensors are active can improve bystander awareness and support situational trust. Complementary user controls that permit teachers, caregivers, or administrators to disable, limit, or configure sensing modalities provide practical mechanisms for restricting data collection in sensitive contexts. Crucially, transparency must encompass more than perceptual indicators: meaningful consent requires accessible disclosures about data flows, retention schedules, purposes of processing, and mechanisms for access, correction, and deletion. Without such explanatory and governance elements, passive notification is unlikely to translate into genuine agency for affected individuals (Richards & King, 2014; Westin, 2003).

2.2 Privacy by Design and Negative Design

Privacy-by-Design has gained broad endorsement in policy and scholarly discourse as a foundational principle for protecting personal data (Cavoukian, 2009). Applied to robotic systems, this principle requires that privacy safeguards be integrated into system architectures and development lifecycles rather than appended as post-hoc fixes. Practical measures include privacy-preserving default settings (e.g., conservative logging policies, sensor duty-cycling), edge processing to keep sensitive data local, encrypted storage, and fine-grained access controls that limit data exposure to only those modules or actors with a demonstrated need.

Complementing affirmative design is the notion of negative design, which purposefully constrains robotic capabilities to prevent privacy harm. Negative-design strategies for embodied agents include context-aware suspension of sensing (e.g., geofenced “no-capture” zones and behaviors like turning away in bathrooms or clearly private interactions), aggressive data-minimization and storage-limitation (e.g., deletion of non-essential recordings), and hard defaults against sensitive inferences (e.g., facial recognition or affect-state inference disabled by default). These strategies are consistent with privacy-by-design/default obligations and with emerging legal limits on certain biometric/affective inferences, while aligning with contextual-integrity's emphasis on protecting dignity and autonomy in specific settings (GDPR, 2016; EDPB, Guidelines 4/2019, 2020; Schiff et al., 2007).

2.3 Technical Safeguards: Privacy-Enhancing Technologies

Building on privacy-centric design and negative-design constraints, privacy-enhancing technologies (PETs) provide concrete mechanisms to operationalize those principles in robotic systems. Edge AI (on-device processing) aligns with privacy-by-design by keeping raw sensor data local—reducing cloud dependence and exposure during transmission or centralized storage (Zhou et al., 2019). Federated learning allows multiple robots to improve a model collaboratively without sharing raw data, preserving individual privacy while supporting collective learning (McMahan et al., 2017). Differential privacy adds quantifiable protection by injecting calibrated noise into outputs or model updates so that contributions of any individual subject cannot be reverse-engineered (Abadi et al., 2016). Homomorphic encryption and secure aggregation further protect collaboration by enabling computation on encrypted data and privacy-preserving combination of model updates, respectively (Bonawitz et al., 2017; Gentry, 2009). In robotics, these PETs increasingly appear in privacy-preserving perception and mapping, e.g., SLAM/localization pipelines that restrict what is revealed or computed (Geppert et al., 2022; Shibuya et al., 2020). Taken together, PETs complement architectural measures (local processing, conservative defaults, negative design) by providing technical guarantees that limit data exposure, constrain inferential risks, and support compliance with data-minimization and privacy-by-design obligations.

3. Case Study: Educational Robots in Classroom Settings

The introduction of educational robots into classrooms provides a concrete example of how privacy concerns manifest in practice. Consider a mobile robot equipped with cameras and object detection algorithms, designed to teach students about robotics and machine learning, the platform simultaneously generates primary sensor outputs (video, audio, depth) and secondary inferences (face images, trajectories, interaction logs), each of which may implicate student privacy.

3.1 Privacy risks and harms

Privacy risks in classroom deployments arise along several dimensions. First, there is a well-documented opacity for non-expert users between "minimally invasive" functions (such as proximity sensing for obstacle avoidance) and more intrusive inferences (e.g., facial recognition or affect estimation). This obscurity complicates contextual privacy judgments and informed assessment of risk in HRI (Lee et al., 2011; Serholt et al., 2017). Inference risks also include downstream repurposing of logs (movement, participation, social interaction) for profiling, a phenomenon emphasized in the data-ethics literature on big-data inference (Richards & King, 2014; Wachter & Mittelstadt, 2019).

Second, embodied sensing routinely captures bystanders and other third parties who have not provided meaningful consent; especially acute in classrooms where participation may be mandatory and power asymmetries are high (Serholt et al., 2017; Sharkey, 2016). This places a premium on transparency, default minimization, and child-specific safeguards recognized in EU data-protection doctrine ("GDPR Arts. 5 & 25," n.d.).

Third, fairness and accuracy risks accompany computer-vision pipelines commonly used in educational robots. Large-scale evaluations show systematic performance disparities by demographic attributes in face analysis and detection (Buolamwini & Gebru, 2018; Wilson et al., 2019), raising the prospect of disproportionate harm to minoritized students via misclassification. Finally, security remains a persistent concern: robotics stacks and connected peripherals have exhibited exploitable vulnerabilities that enable eavesdropping or data exfiltration (Yaacoub et al., 2022). Recent cases with consumer "toy robots" show risks like unauthorized video chat, account takeovers, and children's data leaks; threats that could extend to schools if similar components are used (Kaspersky, 2024).

3.2 Technical and procedural mitigations

A defensible posture combines privacy-by-design defaults with PETs. Edge processing keeps raw sensor data local, reducing exposure from transmission or cloud aggregation and aligning with GDPR's data-minimization and "data protection by design and by default" obligations ("GDPR Arts. 5 & 25," n.d.; EDPB, Guidelines 4/2019, 2020; Zhou et al., 2019). Where aggregate learning is needed across devices, federated learning avoids sharing raw data and has mature threat models and deployment patterns (Kairouz et al., 2021). For analytics that produce counts or models, differential privacy provides quantifiable protection against singling-out (Dwork & Roth, 2014).

When cameras are necessary, anonymization/obfuscation can lower identifiability: classical k-anonymity for tabular logs (Sweeney, 2002), automatic face blurring for video, and emerging privacy-preserving visual-SLAM/localization methods that conceal scene appearance by operating on line-cloud or transformed features instead of raw images (Geppert et al., 2022; Shibuya et al., 2020). Systems should expose configurable "privacy modes" to disable sensing or reduce sampling during sensitive activities, alongside persistent recording indicators and child-appropriate notices (*Age Appropriate Design*, 2025).

Retention and accountability measures matter as much as front-end minimization: default short retention with automatic deletion, purpose limitation, and auditable access-logging are expected under GDPR principles and design-by-default guidance ("GDPR Art. 5," n.d.; EDPB, Guidelines 4/2019, 2020).

3.3 Legal and regulatory considerations

Deployments in schools implicate heightened duties because minors' data merit "specific protection" ("Recital 38 - GDPR," n.d.). Controllers must identify a lawful basis, respect purpose limitation and minimization ("GDPR Art. 5," n.d.), implement data protection by design and by default ("GDPR Art. 25," n.d.), and conduct a Data Protection Impact Assessment where processing is likely high-risk—for example, systematic monitoring of publicly accessible areas. The EU AI Act also bans emotion-recognition systems in educational settings and tightens controls on certain biometric uses, clarifying boundaries for classroom AI. National practice varies: for example, New York State prohibits K-12 schools from using facial recognition, citing accuracy and civil-rights concerns—illustrating how local regulators may go beyond EU-level baselines in school contexts (*NYSED*, 2023). Beyond EU/US contexts, regulatory approaches can differ significantly—for instance, some Asia-Pacific jurisdictions emphasize data localization and state oversight, while others lack comprehensive protections, creating divergent expectations for educational deployments.

4. Reflection & Conclusion

Privacy in robotics extends beyond legal compliance to encompass dignity, autonomy, and digital self-determination. Robots operating in homes, classrooms, and clinics continuously sense, infer, and act, raising questions about how to protect vulnerable groups, whether robot-specific rules on consent and transparency are needed, and how social norms must adapt to persistent observation. Addressing these challenges demands interdisciplinary collaboration and participatory engagement, supported by technical safeguards (privacy-by-design, PETs, negative design) and organizational measures such as procurement standards and training. Privacy should be evaluated not only in terms of compliance but also by its impact on autonomy. Viewing it as protection of the digital self reorients development toward human well-being, grounded in proportionality, accountability, and respect. Embedding these values increases the likelihood that social and educational robots empower rather than surveil.

While robotic systems promise significant benefits, their embodied sensing and inferential capacities generate privacy risks that current regulations and design practices cannot fully address. Mitigation requires integrating privacy into system architectures, deploying safeguards such as edge processing and encryption, and strengthening governance through transparency, user control, and enforceable oversight. In education, privacy protections foster trust between students and institutions. They also support inclusivity and equity by ensuring all learners benefit without disproportionate risks.

References

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep Learning with Differential Privacy. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 308–318. https://doi.org/10.1145/2976749.2978318

Age appropriate design: A code of practice for online services. (2025). ICO. https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/age-appropriate-design-a-code-of-practice-for-online-services/

Art. 5 GDPR – Principles relating to processing of personal data. (n.d.). General Data Protection Regulation (GDPR). Retrieved August 23, 2025, from https://gdpr-info.eu/art-5-gdpr/

Art. 25 GDPR – Data protection by design and by default. (n.d.). General Data Protection Regulation (GDPR). Retrieved August 23, 2025, from https://gdpr-info.eu/art-25-gdpr/

Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., & Seth, K. (2017). Practical Secure Aggregation for Privacy-Preserving Machine Learning. Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 1175–1191. https://doi.org/10.1145/3133956.3133982

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency, 77–91. https://proceedings.mlr.press/v81/buolamwini18a.html

Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I., & Leonard, J. J. (2016). Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age. IEEE Transactions on Robotics, 32(6), 1309–1332.

California Consumer Privacy Act, California Civil Code § 1798.100 et seq. (2020). https://oag.ca.gov/privacy/ccpa

Calo, M. R. (2020). 12 robots and privacy. In Machine Ethics and Robot Ethics (pp. 491–505). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781003074991-39/12-robots-privacy-ryan-calo

Cavoukian, A. (2009). Privacy by design: The 7 foundational principles. Information and Privacy Commissioner of Ontario, Canada, 5(2009), 12.

Dietrich, M., Krüger, M., & Weisswange, T. H. (2023). What should a robot disclose about me? A study about privacy-appropriate behaviors for social robots. Frontiers in Robotics and AI, 10, 1236733.

Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. Foundations and Trends® in Theoretical Computer Science, 9(3–4), 211–407.

GDPR Arts. 5 & 25. (n.d.). General Data Protection Regulation (GDPR). Retrieved August 23, 2025, from https://gdpr-info.eu/art-5-gdpr/

General Data Protection Regulation, Regulation (EU) 2016/679 (2016). https://gdpr-info.eu

Gentry, C. (2009). Fully homomorphic encryption using ideal lattices. Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing, 169–178.

Geppert, M., Larsson, V., Schonberger, J. L., & Pollefeys, M. (2022). Privacy Preserving Partial Localization. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 17316–17326

Guidelines 4/2019 on Data Protection by Design and by Default (2020).

Henschel, A., Laban, G., & Cross, E. S. (2021). What Makes a Robot Social? A Review of Social Robots from Science Fiction to a Home or Hospital Near You. Current Robotics Reports, 2(1), 9–19.

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., & Cummings, R. (2021). Advances and open problems in federated learning. Foundations and Trends® in Machine Learning, 14(1–2), 1–210.

Kaspersky. (2024, February 27). Smart toy vulnerabilities could let cybercriminals video-chat with kids. https://www.kaspersky.com/about/press-releases/smart-toy-vulnerabilities-could-let-cybercriminals-video-chat-with-kids

Koops, B.-J., & Leenes, R. (2014). Privacy regulation cannot be hardcoded. A critical comment on the 'privacy by design' provision in data-protection law. International Review of Law, Computers & Technology, 28(2), 159–171. https://doi.org/10.1080/13600869.2013.801589

Lee, M. K., Tang, K. P., Forlizzi, J., & Kiesler, S. (2011). Understanding users' perception of privacy in human-robot interaction. Proceedings of the 6th International Conference on Human-Robot Interaction, 181–182. https://doi.org/10.1145/1957656.1957721

Lin, P., Abney, K., & Bekey, G. (2011). Robot ethics: Mapping the issues for a mechanized world. Artificial Intelligence, 175(5–6), 942–949.

Lin, P., Abney, K., & Bekey, G. A. (2014). Robot ethics: The ethical and social implications of robotics. MIT press.

McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. Artificial Intelligence and Statistics, 1273–1282.

Mireshghallah, F., Taram, M., Vepakomma, P., Singh, A., Raskar, R., & Esmaeilzadeh, H. (2020). Privacy in Deep Learning: A Survey (No. arXiv:2004.12254). arXiv.

Morgan, A. A., Abdi, J., Syed, M. A. Q., Kohen, G. E., Barlow, P., & Vizcaychipi, M. P. (2022). Robots in Healthcare: A Scoping Review. Current Robotics Reports, 3(4), 271–280.

Norvig, P. R., & Intelligence, S. A. (2002). A modern approach. Prentice Hall Upper Saddle River, NJ, USA

Rani, M., Nayak, R., & Vyas, OP (2015). An Ontology-Based Adaptive Personalized e-Learning System, Assisted by Software Agents on Cloud Storage. Knowledge-Based Systems, 90, 33–48.

Ouyang, F., & Xu, W. (2024). The effects of educational robotics in STEM education: A multilevel meta-analysis. International Journal of STEM Education, 11(1), 7.

Radanliev, P. (2025). AI Ethics: Integrating Transparency, Fairness, and Privacy in AI Development. Applied Artificial Intelligence, 39(1), 2463722. https://doi.org/10.1080/08839514.2025.2463722

Recital 38—Special Protection of Children's Personal Data. (n.d.). General Data Protection Regulation (GDPR). Retrieved August 23, 2025, from https://gdpr-info.eu/recitals/no-38/

Richards, N. M., & King, J. H. (2014). Big data ethics. Wake Forest L. Rev., 49, 393.

Samana, S. (2023). Rule-Based Vs. Machine Learning AI: Which Produces Better Results? Pecan AI. https://www.pecan.ai/blog/rule-based-vs-machine-learning-ai-which-produces-better-results/

Schiff, J., Meingast, M., Mulligan, D. K., Sastry, S., & Goldberg, K. (2007). Respectful cameras: Detecting visual markers in real-time to address privacy concerns. 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, 971–978.

Serholt, S., Barendregt, W., Vasalou, A., Alves-Oliveira, P., Jones, A., Petisca, S., & Paiva, A. (2017). The case of classroom robots: Teachers' deliberations on the ethical tensions. AI & SOCIETY, 32(4), 613–631. https://doi.org/10.1007/s00146-016-0667-2

Sharkey, A. J. C. (2016). Should we welcome robot teachers? Ethics and Information Technology, 18(4), 283–297. https://doi.org/10.1007/s10676-016-9387-z

Shibuya, M., Sumikura, S., & Sakurada, K. (2020). Privacy Preserving Visual SLAM. In A. Vedaldi, H. Bischof, T. Brox, & J.-M. Frahm (Eds.), Computer Vision – ECCV 2020 (Vol. 12367, pp. 102–118). Springer International Publishing. https://doi.org/10.1007/978-3-030-58542-6_7

Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). Membership Inference Attacks against Machine Learning Models (No. arXiv:1610.05820). arXiv.

Siciliano, B., & Khatib, O. (Eds.). (2016). Springer Handbook of Robotics. Springer International Publishing. https://doi.org/10.1007/978-3-319-32552-1

State Education Department Issues Determination on Biometric Identifying Technology in Schools. (2023). New York State Education Department. https://www.nysed.gov/news/2023/state-education-department-issues-determination-biometric-identifying-technology-schools

Sweeney, L. (2002). k-ANONYMITY: A MODEL FOR PROTECTING PRIVACY. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(05), 557–570.

Thrun, S. (2002). Probabilistic robotics. Communications of the ACM, 45(3), 52–57.

van Wynsberghe, A., Ley, M., & Roeser, S. (2022). Ethical Aspects of Human–Robot Collaboration in Industrial Work Settings. In M. I. Aldinhas Ferreira & S. R. Fletcher (Eds.), The 21st Century Industrial Robot: When Tools Become Collaborators (pp. 255–266).

Wachter, S., & Mittelstadt, B. (2019). A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI. Columbia Business Law Review, 2019(2), 494–620.

Westin, A. F. (2003). Social and Political Dimensions of Privacy. Journal of Social Issues, 59(2), 431–453. https://doi.org/10.1111/1540-4560.00072

Wilson, B., Hoffman, J., & Morgenstern, J. (2019). Predictive Inequity in Object Detection (No. arXiv:1902.11097). arXiv. https://doi.org/10.48550/arXiv.1902.11097

Yaacoub, J.-P. A., Noura, H. N., Salman, O., & Chehab, A. (2022). Robotics cyber security: Vulnerabilities, attacks, countermeasures, and recommendations. International Journal of Information Security, 21(1), 115–158. https://doi.org/10.1007/s10207-021-00545-8

Zhou, Z., Chen, X., Li, E., Zeng, L., Luo, K., & Zhang, J. (2019). Edge intelligence: Paving the last mile of artificial intelligence with edge computing. Proceedings of the IEEE, 107(8), 1738–1762.