Machine Learning (ML)

Oseni et al. (2021) thoroughly investigated several aspects of security and privacy for AI and covered several aspects, including the development of secure AI applications, ML task categories, deep learning and federation learning, advanced attacks that would exploit AI applications and measures their threats, cyber defense methods for protecting AI systems against adversarial attacks, and challenges in the security and privacy domain of AI. Perino, Katevas et al. (2022) considered privacy preserving AI in future networks, in which the main interest of the paper is the vulnerability of AI models to new attack scenarios and vectors. The paper refers to the fact that AI models and tools introduce new vulnerabilities and attack vectors in telecommunication environments, placing privacy at. The methods mainly explored that handle these issues are privacy-preserving AI methods, such as federated learning, differential privacy, policy-based AI, and trusted execution environments. Zhang, Al Hamadi, Damiani, Yeun, & Taher (2022) surveyed AI methods, specifically explainable AI methods (XAI) for cyber security applications. In this survey, there is a taxonomy that is first based on explainable AI (XAI) frameworks and data for cybersecurity, and the AI applications for cybersecurity. This taxonomy is spread in the lower level of different fields such as malware, spam, bots, fraud, and phishing. Relevant current solutions are presented in order to address this. Khalid, Qayyum, Bilal, Al-Fuqaha, & Qadir (2023) surveyed the most updated and relevant approaches for preserving privacy in AI-based healthcare applications. The paper relates to privacy-preserving techniques such as federated learning and hybrid techniques. The paper also covers possible privacy attacks, and security problems in this field. A taxonomy of privacy attacks is developed with techniques that can be used to protect against such attacks involving healthcare datasets and AI models. For every technique the advantages and disadvantages are shown. The different privacy-preserving machine learning (PPML) techniques with their problematic privacy aspects are shown, and finally the open research issues that require further development are addressed. Meurisch, Bayrak, and Muhlhauser (2020) introduced a platform called PrivAI, which incorporates privacy-by-design principles to provide personalized AI services to users while safeguarding their privacy through data decentralization. This approach involves using personal data to establish control over it. This is done by implementing two main mechanisms: 'divisible' AI algorithms and internal/external confidential processing for pdata stores (PDS), which is the data system that is under a user's control—including the specification of the location. These mechanisms have been shown to have accurate classification results and reasonable performance overhead for AI-based services. This approach is shown to be better than the conventional data management approaches to privacy that mostly rely on protocols, access control mechanisms, and policies. The research demonstrates a good method Okprivacy-preserving AI services using data decentralization, with a unique quality of preserving the intellectual property of providers. Gupta et al. (2020) surveyed and investigated AI techniques and tools for smart contract (SC) privacy protection and analyzes open issues and challenges for AI-based SC. They present a case study of retail marketing that uses AI and SC to preserve security and privacy. The paper presents an overview of the state-of-the-art SC security vulnerabilities, and reviews different SC platforms. It then presents the security vulnerabilities in SC and the possible solutions. These solutions are of high communication and computation costs, so the integration of AI in SC is suggested to solve these issues. Finally, the survey discusses the open issues and research challenges, that occur due to SC and AI integration issues. Villegas-Ch & Garcia-Ortiz (2023) discuss the challenges faced by the rapid expansion of AI in terms of data security and privacy. It proposes a comprehensive framework to address these issues. The paper reviews previous research on security and privacy in AI, highlighting advancements and limitations. It also identifies open research areas and gaps that need attention for improving current frameworks. The framework development focuses on data protection in AI, with the importance of

safeguarding data used in AI models. It describes policies, practices, and approaches to ensure data security and integrity. The paper also examines the security of AI, analyzing vulnerabilities and risks in AI systems and providing examples of potential attacks and manipulations. It presents security frameworks to handle these risks. Additionally, the ethical and regulatory aspects relevant to security and privacy in AI are considered, offering an overview of existing regulations and guidelines.

Cheng et al. (2020) focused on federated learning (FL) for privacy preserving AI. The paper defines and categorizes FL, and explains its architecture, and then focuses on two main uses and their implementation. The first is FL application in finance, for example federated risk control (FedRiskCtrl) for small and micro enterprise (SME) loans. The second use is an FL application in edge computing. It is an example of federated computer vision (FedVision) for object detection. Both of these were deployed by WeBank, a digital bank launched in 2014 in China, aiming to provide more convenient financial services to micro, small, and medium-sized enterprises (MSMEs) as well as to the general public. The paper shows the benefits of using FL as a good solution for privacy preserving AI, and also presents the different challenges in this approach, including the communication links between the local data owners and the coordinator, and the possibility of the need to manage a very large number of local data owners. Zhu, Ye, Wang, Zhou, and Philip (2020) showed how differential privacy can help address emerging issues in major areas of AI, including machine learning, deep learning, and multi-agent systems. Differential privacy provides properties like privacy preservation, stability, fairness, and composition that can help improve various aspects of these areas. For example, in ML, differential privacy techniques enable private learning and achieving fairness and privacy simultaneously. In deep learning, differential private mechanisms have been applied to distributed and federated learning. In multi-agent systems, differential privacy guarantees have been used for tasks like reinforcement learning, auctions, and game theory. The paper analyzes different applications of differential privacy across these areas and outlines directions for future research. Rodriguez-Barroso et al. (2020) showed the integration of federated learning (FL) and differential privacy (DP) in the context of AI services that prioritize data privacy at the network edge. It reviews existing FL and DP software tools, identifies their limitations, and introduces a federated learning framework named Sherpa.ai, designed to offer a unified solution that incorporates both FL and DP. The framework facilitates the development of AI services that are both privacy-preserving and efficient. Through detailed analysis and comparison, the study highlights the need for a holistic approach to FL and DP, addressing the gap in current methodologies and tools. The paper also outlines methodological guidelines for implementing AI services with FL and DP, supported by the Sherpa.ai framework, and demonstrates its application through classification and regression. Jain and Ghanavati (2020) explored the privacy implications of the increasing use of AI and ML in various digital services and devices. They highlight the privacy risks associated with the collection of large datasets necessary for training AI models and discuss current research efforts aimed at developing privacy-preserving techniques. These efforts include differential privacy, federated learning, and user-focused tools for controlling personal data. The paper emphasizes that privacy-preserving research for AI is still in its infancy and calls for more robust solutions to address these challenges.

Mahmud et al. (2022) explored the use of AI and ML for the diagnosis and support of individuals with autism spectrum disorder (ASD). The authors highlight the challenges in ASD diagnosis and treatment, emphasizing the need for explainable and privacy-preserving AI models. The paper examines various AI-based methods, including explainable AI and federated learning, to enhance ASD care, with focus on developing personalized solutions while safeguarding privacy. Bai et al. (2021) focused on the development of a privacy-preserving federated learning framework to improve diagnostic accuracy while addressing data privacy concerns. The paper highlights the limitations of the standard RT-PCR tests for

COVID-19, noting their variable sensitivity and the risk of false negatives. It points out the potential of using chest computed tomography (CT) scans, which show distinct radiological features in COVID-19 patients, as a complementary diagnostic tool. To overcome data privacy and security issues that come with the collection of large-scale medical datasets necessary for training robust AI models, the authors introduce the Unified CT-COVID AI Diagnostic Initiative (UCADI). This initiative employs a federated learning approach where the AI model is trained across multiple institutions without exchanging data, thus preserving privacy. The study demonstrates the potential of federated learning in developing secure, efficient, and scalable AI solutions for diagnosing COVID-19 and possibly other diseases, paving the way for future advancements in digital health and AI-driven diagnostics. Kaissis, Makowski, Ruckert, and Braren (2020) discussed the critical issue of privacy and security in the use of AI within medical imaging. They highlight the challenge of limited dataset availability for AI training due to strict privacy regulations and the lack of standardized electronic medical records. The paper underscores the potential of federated learning, which allows local processing of data, maintaining privacy while enabling the collective improvement of algorithms across multiple institutions. Furthermore, the authors delve into various methods to enhance data security and privacy in AI applications, such as differential privacy, homomorphic encryption, and secure multi-party computation. These techniques help protect sensitive patient data during AI processing. The paper also addresses potential vulnerabilities that could lead to privacy breaches, emphasizing the need for robust security measures in medical AI applications.

Zhang, Wu, Tian, Zhang, and Lu (2021) surveyed AI from the perspective of ethics and privacy. The paper provides a bibliometric view of works conducted on the topic and extracts the data from these papers into several domains. In their paper, an analysis was performed on articles indexed in the Web of Science under the topic of investigations on ethical issues surrounding AI. The research profiled the AI ethical issues in both hierarchical and time dimensions by using bibliometric approaches, including topical hierarchical trees and scientific evolutionary pathways, which helped answer the questions regarding specific ethical issues of concern related to AI and how public interests concerning these issues have changed over time. Bendechache et al. (2021) described a project designed to engage 500 Dublin teenagers in workshops to explore AI, ethics, and privacy. The program, aimed at disadvantaged youth, is a 15-week interactive workshop series encouraging students to evaluate AI's impact on their lives and consider science, technology, engineering and mathematics (STEM) careers. The project uses co-creation workshops to shape the content and assesses its effectiveness through various metrics, emphasizing student understanding, data literacy, and interest in STEM. Bak et al. (2022) discuss the necessary ethical considerations as AI technologies become increasingly integrated into daily life. It covers foundational ethical principles like Asimov’s Three Laws of Robotics, issues concerning robot rights and moral agency, the opaqueness and biases within AI systems, privacy concerns, automation's impact on employment, and the necessity for international AI ethics policies.

The article emphasizes the importance of ethical AI development to address the potential risks and ensure that technology benefits society fairly and responsibly.

Dilmaghani et al. (2019) surveyed and discussed the impact of big data on privacy and security in the context of ML and AI systems. They focus on the potential risks and vulnerabilities associated with using sensitive data to train AI systems, while also acknowledging the increased demand for high-quality AI in various sectors. The paper provides an overview of existing threats posed by big data and proposes an adversarial model to investigate attacks. It further analyzes defense strategies and countermeasures to handle these risks. The paper also examines Standards Developing Organizations (SDOs) that provide guidelines to ensure privacy and security in big data and AI systems. The authors' ultimate goal was to bridge the gap between research and standardization to enhance the consistency, efficiency, and trustworthiness of AI system development. Finally, they suggest privacy preserving solutions

for data breach, data poisoning, model extraction, and bias in data and evasion. Gulmezoglu et al. (2019) investigated attacks on mobile devices that exploit the microarchitecture of modern processors and compromise user privacy. These attacks rely on the fact that applications leave traces in the processor, which can be used by malware to infer user activities. The research demonstrates that the effectiveness of these inference attacks can be significantly enhanced through the application of advanced AI techniques. Specifically, the researchers focus on profiling the activity in the last-level cache of ARM processors. They employed a monitoring technique named Prime+Probe to obtain cache traces and used deep learning methods such as convolutional neural networks, to classify these traces. The researchers successfully demonstrated their approach on a regular Android phone, by launching a successful attack within minutes, even from an unprivileged app without any special permissions. The app accurately detected running applications, opened websites, and streaming videos with up to 98% accuracy, and the profiling phase took at most 6 seconds. The researchers attributed this success to the deep learning ability which compensates for the inherent noise and unfavorable characteristics of the cache monitoring process. Overall, these findings highlight the increasing ease with which inference attacks can be executed, emphasizing the need for countermeasures to protect user privacy in mobile phone applications. The paper emphasizes the importance of protection against inference attacks and suggests that a comprehensive solution requires collaboration between hardware manufacturers, operating system designers, and application developers. Qiu, Liu, Zhou, and Wu (2019) surveyed recent progress in adversarial attacks and defense technologies mainly in deep learning. They introduce the causes and characteristics of adversarial samples, as well as adversarial capabilities and goals. Adversarial attacks in the training stage include modifying the training dataset, label manipulation, and input feature manipulation. Testing stage attacks are divided into white-box and black-box attacks. Applications include computer vision, natural language processing, cybersecurity, and the physical world. Defenses modify data, models, or use auxiliary tools. Data defenses include adversarial training and gradient hiding, while model defenses include defensive distillation, feature squeezing, and deep contractive networks. Their paper provides several examples of current research and defense solutions for adversarial attacks. Zhou et al. (2022) surveyed current research on adversarial attacks and defenses in DL. The paper focuses on poisoning attacks and during training, on other adversarial attacks. The authors claim that there is a lack of standard evaluation methods to assess the real threat of adversarial attacks and the robustness of DL models. The paper reviews the existing literature on adversarial attacks and proposes an analysis framework from a cybersecurity perspective to understand and evaluate these attacks systematically.

The framework maps the stages of an adversarial attack to the lifecycle of an advanced persistent threat (APT) and provides a similar framework for adversarial defenses. The survey concludes with suggestions for future research to improve robustness against adversarial attacks and defenses. The paper claims that despite the impressive performance of deep neural networks, there are security concerns regarding their vulnerability to adversarial samples, and the survey aims to provide a standardized evaluation process for assessing the security and robustness of deep learning models against adversarial attacks. Ma et al. (2023) surveyed privacy and security issues that can arise in distributed ML systems. It proposes a framework that divides distributed ML into four levels based on the type of information exchanged: data, models, knowledge, and results. For each level, it analyzes potential attacks and defenses. It also covers challenges like balancing utility and privacy and proposes areas for further research. Bae et al. (2018) examined the vulnerability of deep learning models to security and privacy threats, including evasion and poisoning attacks. They discuss various attack strategies and their impact on model integrity and data privacy. The paper also explores defense mechanisms against these attacks, emphasizing the need for secure and private artificial intelligence systems. Through a comprehensive analysis, the authors highlight the

challenges in protecting deep learning models from malicious threats and suggest directions for future research to enhance AI security and privacy. Thuraisingham (2020) discusses the potential of AI to benefit humanity while also acknowledging the challenges of cyber-attacks and privacy violations. The paper emphasizes the importance of AI in improving healthcare, finance, and various other sectors, and highlights initiatives like "AI for Good." However, it raises concerns about the security of AI systems and privacy breaches that can result from AI technologies. The paper uses protecting children and their rights as an example to illustrate how AI can both safeguard and potentially endanger vulnerable groups, underscoring the need for advancements in robustness against adversarial AI and in privacy-aware AI to address these challenges. Chen (2020) explored personal privacy protection techniques from the standpoint of social engineering. Chen reveals that by completely diversifying the personal account group, such as utilizing multiple mobile phone numbers and account bindings, and not using registration-free login features, AI can split personal information into multiple user profiles. Consequently, AI discards these profiles due to incomplete information, effectively ensuring the protection of personal information. Chen also briefly describes and analyzes the privacy protection problems that are connected to the field of AI. Al-Khassawneh (2023) provides a comprehensive overview of adversarial attacks against AI applications, covering topics such as the knowledge and capabilities of adversaries, methods for producing adversarial examples, and existing cyber defense models. He explores cyber countermeasures to protect AI applications from attacks and proposes a systematic approach for formulating strategies against ML and AI threats, highlighting the importance of understanding attackers' intentions and methods to safeguard AI applications effectively.

Amaral et al. (2021) discuss the use of AI to automate the examination of the completeness of privacy policies in compliance with the GDPR. First, a conceptual model and a set of completeness criteria based on the provisions of GDPR were developed. Then the researchers used NLP and supervised ML to identify GDPR-relevant content in privacy policies and checked them against the criteria. The approach was evaluated using 234 real privacy policies from the fund industry, with a precision of 92.9% and a recall of 89.8%. Compared to a baseline keyword search approach, the AI-based approach shows a significant improvement in precision and recall.

The research is expected to enhance the completeness criteria when considering the meaning of sentences containing metadata and assessing the generalizability of the approach beyond the fund industry. Humerick (2017) focused on the problematic aspects of the implementation of the GDPR together with the development and use of AI. These problems occur because AI relies heavily on data to expand its knowledge, but the GDPR emphasizes consumer control over personally identifiable information (PII) and creates stricter legal and operational obstacles for controlling and processing such data. The GDPR grants various rights to EU citizens regarding their personal data, such as the right to erasure stored information, the right to receive explanations about automated decisions, and data portability. The territorial scope of the GDPR also applies to the data of European citizens, regardless of their location. This raises compliance issues for business leaders and legislatures worldwide. However, to remain competitive in developing AI, the research claims that the EU needs to find a balance between data privacy and AI advancement. The EU's strict implementation of the GDPR regarding the use of personal data of any European citizen in AI systems may provide ways for the AI industry to circumvent the GDPR. The author claims that the EU must recognize and address the potential impact of the GDPR on the AI industry to avoid falling behind in AI development.

Murdoch (2021) discussed AI in the field of healthcare. In this field technologies are rapidly advancing and are expected to have a significant impact on the healthcare industry. AI is particularly useful in radiology for analyzing diagnostic imagery. However, there are concerns regarding the access, use, and control of patient health information in the implementation of AI commercial healthcare. Commercial partners and private entities often own and control AI

technologies, which can be closed (black boxed) and difficult to oversee. Public and private partnerships have resulted in privacy breaches and lack of patient confidentiality. Regulations and safeguards are needed to protect patient privacy and maintain public trust. The problem of re-identification poses additional privacy risks, even when data is anonymized. Carefully constructed contracts and the use of generative data could help mitigate privacy concerns. Overall, there is a need for robust regulation, oversight, and innovative data protection methods in the development and implementation of AI healthcare. Tom et al. (2020) addressed privacy concerns arising from the integration of AI in ophthalmology. It discusses the challenges in ensuring data privacy with the increasing use of large datasets and AI technologies in this field. The paper highlights the ethical and legal considerations, data protection methods, and consent models in the context of AI application in ophthalmology. It also explores case studies and new approaches to protect patient data while leveraging AI for better healthcare outcomes. Roy (2022) discussed the use of AI to enhance privacy protection of healthcare data. He outlines various techniques like data encryption, access restrictions, anomaly detection, and data anonymization that can be used. He also discusses the roles, benefits, challenges and future directions of applying AI in healthcare data privacy. Some key models like Random Forest, SVM and neural networks were evaluated for accuracy. In the paper Roy also examines important regulations such as HIPAA and GDPR as well as case studies. Bak, Fritzsche, Mayrhofer, and McLennan (2022) discuss the ethical and practical trade-offs between health data privacy and access within the context of AI in healthcare across the European Union. They emphasize the need for a balanced approach to data governance that allows the advancement of AI technologies while protecting individual privacy. The authors argue that significant public funds invested in AI development are wasted if data access is overly restricted by privacy measures. They propose a fair and inclusive engagement process to establish values underlying national data governance policies, aiming for a balance that fosters AI development without compromising privacy.

Fritchman et al. (2018) explored the application of secure multiparty computation (SMC) to implement privacy-preserving machine learning (PPML) for healthcare. They present a framework that allows health organizations the ability to securely compute machine learning predictions over encrypted data, thus maintaining the confidentiality of both the data and the model. In the paper the authors detail the cryptographic protocols developed for classification with tree ensembles, such as Random forests and boosted decision trees, and discuss their integration into a scalable cloud-based platform for clinical outcome prediction. This approach not only protects patient privacy but also complies with data governance policies. Kirienko et al. (2021) reviewed the effectiveness of distributed learning as a non-inferior approach compared to centralized and locally trained ML models in medical applications. Their findings show that distributed learning performs comparably to centralized training, where in most cases distributed learning outperformed locally trained models. Their approach offers a promising solution for AI-based research and practice, especially in scenarios requiring privacy preservation and large, diverse datasets. Puiu et al. (2021) addressed the challenges of data privacy and the need for explainability in AI applications in cardiovascular imaging. The authors discuss the implementation of homomorphic encryption to ensure privacy while processing medical data and emphasize the importance of developing explainable AI algorithms to enhance their acceptance in clinical settings. in the paper the authors highlight methods to preserve data privacy without compromising the utility of the data, aiming to balance ethical considerations with technological advancements in medical imaging.

Ishii (2019) conducted a comparative legal study to analyze challenges of privacy and personal data protection stemming from robots equipped with AI. The aim was to understand issues and offer policy suggestions. Some benefits of AI robots include use in transportation, healthcare, and education. However, risks to privacy arise from technologies like machine learning, sensors, and human-like appearances that psychologically influence people. Privacy

laws in the EU, USA, Canada, and Japan provide some protection but have limitations in addressing issues uniquely created by AI like discriminatory decisions, lack of transparency, and hindered consent. The unpredictability of AI makes statutory regulation difficult. Approaches such as PbD, which was pioneered in Canada, are seen as more flexible and preferable to detailed laws that may become outdated. PbD aims to protect privacy without specific efforts while allowing AI development. Issues like biased algorithm decisions, the "algorithmic black box" problem, and assigning responsibility for harm caused by algorithms are challenges without easy solutions. Determining the best legal framework while proceeding is important.

Willems et al. (2023) investigated how perceived usefulness, data sharing requirements, and privacy concerns influence citizens' willingness to use AI-driven public services. It aims to test the privacy calculus theory and privacy paradox. An experiment was conducted with Austrian citizens where they received questionnaires about a hypothetical AI Chatbot app for public services. The Chatbot's name and amount of personal data required to use it varied. Usefulness was the main factor increasing the willingness to use the app. General privacy concerns lowered willingness but did not interact with the data sharing factor. This supports the existence of the privacy paradox in the public service context. Zhdanov, Bhattacharjee, and Bragin (2022) focused on integrating principles of fairness, accountability, and transparency (FAT) with privacy considerations in AI systems used in business decision-making. The authors developed a FAT-based framework and applied it to a privacy-constrained dataset, demonstrating that FAT principles can coexist in a well-designed AI system without significantly impacting predictive performance.

Their approach aims to make AI decisions more ethical, less controversial, and thus more trustworthy, contributing to the emerging global discourse on AI policy. Wang et al. (2019) introduced a novel approach for integrating privacy protection into the of aggregation of public preferences process regarding moral dilemmas in AI, e.g., autonomous vehicles in accident scenarios. By adopting differential privacy, the research presents methods to safeguard individuals' sensitive moral preferences while achieving accurate aggregate moral decision models. In the study the authors evaluate these methods using both synthetic and real-world data, demonstrating their effectiveness in preserving privacy without significantly compromising the accuracy of the collective moral preference model.

Tucker et al. (2018) discussed the complex relationship between AI and privacy economics. They focus on three main themes: data persistence, data repurpose, and data spillovers. The paper examines how AI challenges traditional privacy models, particularly regarding data's the use of long-term and unforeseen consequences. It emphasizes the need for a deeper understanding of privacy in the context of AI, suggesting that future research should explore how AI may affect privacy preferences, the predictive value of the data over time, and the broader economic impact of data usage and repurposing. Toch and Birman (2018) discuss privacy concerns related to advances in ML and AI. The ability of AI systems to collect and analyze vast amounts of personal data in order to infer and predict human behavior poses a threat to values such as the freedom to be unpredictable and the right to change. For example, decisions like whether or not to have children are often influenced by fear and uncertainty. By diminishing these elements, AI could unintentionally have a negative impact on birth rates. While current privacy approaches focus on data collection and use, they do not address issues concerning the inference of personal attributes and behaviors from data. The authors propose a framework for behavioral privacy that considers both general model robustness and personal privacy when analyzing datasets. They analyze the shortcomings of existing privacy theories in addressing AI’s inferential abilities and suggest new frameworks to help understand AI’s potential privacy impact. They also introduce a technical privacy measure to bridge the gap between legal and technical perspectives on AI and privacy. The authors focus on the

challenges of maintaining privacy in the face of AI's increasing ability to predict and infer individual behavior based on collected data.

Rahman et al. (2020) present a privacy-preserving framework for AI-enabled service composition in edge networks. They address the challenges of maintaining privacy in AI tasks distributed across edge devices, while proposing a model that uses fully homomorphic encryption (FHE) to secure quality-of-service (QoS) data. The paper demonstrates how AI-based service composition can be achieved on encrypted QoS data, ensuring protection of privacy from potential attacks on edge nodes. The study includes experimental evaluations using synthetic QoS datasets to validate the framework's effectiveness. McStay (2020) examined the privacy implications of emotional AI technologies. He discusses the use of affective computing in various contexts like advertising, policy, and healthcare, with focus on non-identifying emotional AI practices. The paper draws insights from interviews, a UK survey, and a multi-stakeholder workshop, revealing a weak consensus on privacy concerns across different groups. McStay highlights the need for regulatory action on emotional AI and soft biometrics, emphasizing the importance of maintaining privacy and dignity in the face of emerging technologies. Roemmich et al. (2023) explored the consequences of emotional AI applications in workplace settings. They conducted interviews with U.S. workers to understand their perceptions and experiences related to emotion AI, which is increasingly used to monitor and manage workers' emotional states.

Their findings reveal significant privacy concerns, as workers view emotion AI as an intrusive technology that risks violating their emotional privacy and potentially influencing their emotional labor. The study emphasizes the need for robust policies and a reevaluation of the ethical implications of using such technologies, highlighting the potential for emotional AI to potentially impact workplace dynamics, privacy rights, and personal autonomy in detrimental ways. Saura et al. (2022) investigated privacy concerns related to the use of AI by governments. They focused on how governments utilize AI for collective behavior analysis and the implications for citizens' privacy. Through systematic literature review, in-depth interviews, and data-mining techniques, the authors identify the main uses of AI by governments and citizens' concerns about privacy. The findings reveal various AI strategies employed by governments and outline the risks to privacy, suggesting the need for regulations focused on ethical data collection and usage practices. Kelley et al. (2023) explored public perceptions across ten countries on the impact of AI on privacy. They revealed four main themes of concern: data vulnerability, personal data sensitivity, data collection without consent, and state surveillance. These concerns are well-reasoned and align with expert opinions. The study suggests the need to guide public priorities to mitigate AI's privacy impacts and foster informed civic participation. Kronemann et al. (2023) explored how artificial AI impacts consumer information disclosure. They investigated the effects of AI anthropomorphism, personalization, and privacy concerns on consumer attitudes towards sharing private information. Using the personalization-privacy paradox (PPP) and privacy calculus theory (PCT), they propose a conceptual model with research propositions to understand consumer behavior towards AI. Their findings suggest that while personalization and anthropomorphism encourage information disclosure, privacy concerns have a negative effect. Their research contributes to understanding the balance between AI benefits and privacy risks from a consumer perspective. Kerry et al. (2020) published a report that discusses the profound implications that AI technologies have on privacy, with emphasis on the exponential growth of data and the transformative effects of AI and ML. Highlighting the privacy challenges posed by the vast collection and analysis of data, they outline the current legislative efforts and policy considerations aimed at protecting privacy in the AI-driven world. In their report the authors stress the importance of crafting comprehensive privacy legislation that addresses these emerging concerns without stifling AI innovation. Hu et al. (2023) advocated for recognizing data infrastructure as a vital component of national security within the context of AI

governance. They discuss the risks of self-regulated AI industries, while highlighting the need for structured data privacy and integrity to enhance cybersecurity and counter information warfare. In their paper the authors argue that a federally coordinated data infrastructure can significantly support national security by integrating rigorous privacy standards into the design and operation of AI systems. Mazurek and Malagocka (2019) examined the evolving perspectives on privacy and data protection, especially in relation to AI. They highlight how the perception of privacy affects regulatory approaches and business strategies. They also discuss the impact of AI on customer-company relationships, emphasizing the balance between technological development and consumer protection. They underscore the importance of aligning AI advancements with ethical and legal frameworks to ensure consumer wellness. Holzinger et al. (2021) examined the implications of AI in promoting the United Nations' Sustainable Development Goals (SDGs). The examine whether AI has significant potential to advance these goals, while emphasizing that the deployment of AI raises substantial security, safety, and privacy concerns that must be rigorously addressed.

They underline the dual nature of AI technology, its ability to drive positive changes (e.g., in smart agriculture and health systems) alongside the risk of introducing new vulnerabilities and challenges. Particularly, the authors highlight how AI can impact various aspects of security and privacy, stressing the importance of developing AI systems that are not only effective but also secure, transparent, and aligned with ethical standards. Ahmadi Mehri and Tutschku (2017) explored the integration of privacy and trust into cloud-based AI marketplaces. They propose the concept of a "virtual premise" to implement PbD within cloud environments, thereby enhancing users' control over their data and increasing trust. The authors discuss the importance of privacy and trust in the adoption of cloud services, especially when handling sensitive data in AI applications.

References

Ahmadi Mehri, V., & Tutschku, K. (2017). Flexible privacy and high trust in the next generation internet: The use case of a cloud-based marketplace for AI. *SNCNW-Swedish National Computer Networking Workshop, Halmstad.* Halmstad university. Retrieved from https://www.diva-portal.org/smash/get/diva2:1128475/FULLTEXT01.pdf

Al-Khassawneh, Y. A. (2023). A review of artificial intelligence in security and privacy: Research advances, applications, opportunities, and challenges. *Indonesian Journal of Science and Technology, 8*(1), 79--96. Retrieved from https://ejournal.kjpupi.id/index.php/ijost/article/view/9

Amaral, O., Abualhaija, S., Torre, D., Sabetzadeh, M., & Briand, L. C. (2021). AI-enabled automation for completeness checking of privacy policies. *IEEE Transactions on Software Engineering, 48*(11), 4647--4674. doi:https://doi.org/10.1109/TSE.2021.3124332

Bae, H., Jang, J., Jung, D., Jang, H., Ha, H., Lee, H., & Yoon, S. (2018). Security and privacy issues in deep learning. *arXiv preprint arXiv:1807.11655*. doi:https://doi.org/10.48550/arXiv.1807.11655

Bai, X., Wang, H., Ma, L., Xu, Y., Gan, J., Fan, Z., . . . others. (2021). Advancing COVID-19 diagnosis with privacy-preserving collaboration in artificial intelligence. *Nature Machine Intelligence, 3*(12), 1081--1089. doi:https://doi.org/10.1038/s42256-021-00421-z

Bak, M. a., Fritzsche, M.-C., Mayrhofer, M. T., & McLennan, S. (2022). You can't have AI both ways: balancing health data privacy and access fairly. *Frontiers in genetics, 13*, 929453. doi:https://doi.org/10.3389/fgene.2022.929453

Bendechache, M., Tal, I., Wall, P., Grehan, L., Clarke, E., Odriscoll, A., . . . Brennan, R. (2021). AI in my life: AI, ethics & privacy workshops for 15-16-year-olds. *Companion Publication of the 13th ACM Web Science Conference 2021*, (pp. 34--39). doi:https://doi.org/10.1145/3462741.3466664

Chen, Z. (2020). Privacy Protection Technology in the Age of A.I. *IOP Conference Series: Materials Science and Engineering, 750*(1), 012103. doi:10.1088/1757-899X/750/1/012103

Cheng, Y., Liu, Y., Chen, T., & Yang, Q. (2020). Federated learning for privacy-preserving AI. *Communications of the ACM, 63*(12), 33-36. doi:http://dx.doi.org/10.1145/3387107

Dilmaghani, S. a., Danoy, G., Cassagnes, N., Pecero, J., & Bouvry, P. (2019). Privacy and security of big data in AI systems: A research and standards perspective. *2019 IEEE international conference on big data (big data)* (pp. 5737--5743). IEEE. doi:https://doi.org/10.1109/BigData47090.2019.9006283

Fritchman, K., Saminathan, K., Dowsley, R., Hughes, T., De Cock, M., Nascimento, A., & Teredesai, A. (2018). Privacy-preserving scoring of tree ensembles: A novel framework for AI in healthcare. *2018 IEEE international conference on big data (Big Data)* (pp. 2413--2422). IEEE. doi:https://doi.org/10.1109/BigData.2018.8622627

Gulmezoglu, B., Zankl, A., Tol, M. C., Islam, S., Eisenbarth, T., & Sunar, B. (2019). Undermining user privacy on mobile devices using AI. *Proceedings of the 2019 acm asia conference on computer and communications security*, (pp. 214--227). doi:https://doi.org/10.1145/3321705.3329804

Gupta, R., Tanwar, S., Al-Turjman, F., Italiya, P., Nauman, A., & Kim, S. W. (2020). Smart contract privacy protection using AI in cyber-physical systems: tools, techniques and challenges. *IEEE access, 8*, 24746--24772. doi: 10.1109/ACCESS.2020.2970576

Holzinger, A., Weippl, E., Tjoa, A. M., & Kieseberg, P. (2021). Digital transformation for sustainable development goals (sdgs)-a security, safety and privacy perspective on ai. *International cross-domain conference for machine learning and knowledge extraction* (pp. 1--20). Springer. doi:https://doi.org/10.1007/978-3-030-84060-0_1

Hu, M., Behar, E., & Ottenheimer, D. (2023). National Security and Federalizing Data Privacy Infrastructure for AI Governance. *Fordham L. Rev., 92*, 1829. Retrieved from https://ssrn.com/abstract=4775242

Humerick, M. (2017). Taking AI personally: how the EU must learn to balance the interests of personal data privacy \& artificial intelligence. *Santa Clara High Tech. LJ, 34*, 393. Retrieved from https://heinonline.org/HOL/LandingPage?handle=hein.journals/sccj34&div=19&id=&page=

Ishii, K. (2019). Comparative legal study on privacy and personal data protection for robots equipped with artificial intelligence: looking at functional and technological aspects. *AI & society*, 509--533. doi:https://doi.org/10.1007/s00146-017-0758-8

Jain, V., & Ghanavati, S. (2020). Is it possible to preserve privacy in the age of ai? *PrivateNLP@ WSDM*, (pp. 32--36). doi:https://doi.org/10.1145/1122445.1122456

Kaissis, G. A., Makowski, M. R., Ruckert, D., & Braren, R. F. (2020). Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence, 2*(6), 305--311. doi:https://doi.org/10.1038/s42256-020-0186-1

Kelley, P. G., Cornejo, C., Hayes, L., Jin, E. S., Sedley, A., Thomas, K., . . . Woodruff, A. (2023). There will be less privacy, of course": How and why people in 10 countries expect AI will affect privacy in the future. *Nineteenth Symposium on Usable Privacy and Security (SOUPS 2023)*, (pp. 579--603). Retrieved from https://www.usenix.org/conference/soups2023/presentation/kelley

Kerry, C. F., & others. (2020). Protecting privacy in an AI-driven world. *Brookings Institution*. doi:https://coilink.org/20.500.12592/72dj71

Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A., & Qadir, J. (2023). Privacy-preserving artificial intelligence in healthcare: Techniques and applications. *Computers in Biology and Medicine, 158*. doi:https://doi.org/10.1016/j.compbiomed.2023.106848

Kirienko, M., Sollini, M., Ninatti, G., Loiacono, D., Giacomello, E., Gozzi, N., . . . Chiti, A. (2021). Distributed learning: a reliable privacy-preserving strategy to change multicenter collaborations using AI. *European Journal of Nuclear Medicine and Molecular Imaging, 48*, 3791--3804. doi:https://doi.org/10.1007/s00259-021-05339-7

Kronemann, B., Kizgin, H., Rana, N., & K. Dwivedi, Y. (2023). How AI encourages consumers to share their secrets? The role of anthropomorphism, personalisation, and privacy concerns and avenues for future research. *Spanish Journal of Marketing-ESIC, 27*(1), 3--19. doi:https://doi.org/10.1108/SJME-10-2022-0213

Ma, C., Li, J., Wei, K., Liu, B., Ding, M., Yuan, L., . . . Poor, H. V. (2023). Trusted AI in multiagent systems: An overview of privacy and security for distributed learning. *Proceedings of the IEEE, 111*(9), 1097--1132. doi:https://doi.org/10.1109/JPROC.2023.3306773

Mahmud, M., Kaiser, M. S., Rahman, M. A., Wadhera, T., Brown, D. J., Shopland, N., . . . others. (2022). Towards explainable and privacy-preserving artificial intelligence for personalisation in autism spectrum disorder. *International Conference on Human-Computer Interaction* (pp. 356--370). Springer. doi:https://doi.org/10.1007/978-3-031-05039-8_26

Mazurek, G., & Malagocka, K. (2019). Perception of privacy and data protection in the context of the development of artificial intelligence. *Journal of Management Analytics, 6*(4), 344--364. doi:https://doi.org/10.1080/23270012.2019.1671243

McStay, A. (2020). Emotional AI, soft biometrics and the surveillance of emotional life: An unusual consensus on privacy. *Big Data & Society, 7*(1), 2053951720904386. doi:https://doi.org/10.1177/2053951720904386

Meurisch, C., Bayrak, B., & Muhlhauser, M. (2020). Privacy-preserving AI services through data decentralization. *Proceedings of The Web Conference 2020*, (pp. 190--200). doi:https://doi.org/10.1145/3366423.3380106

Murdoch, B. (2021). Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Medical Ethics, 22*, 1--5. doi:https://doi.org/10.1186/s12910-021-00687-3

Oseni, A., Moustafa, N., Janicke, H., Liu, P., Tari, Z., & Vasilakos, A. (2021). Security and privacy for artificial intelligence: Opportunities and challenges. *arXiv preprint arXiv:2102.04661*. doi:https://doi.org/10.48550/arXiv.2102.04661

Perino, D., Katevas, K., Lutu, A., Marin, E., & Kourtellis, N. (2022). Privacy-preserving AI for future networks. *Communications of the ACM, 65*(4), 52--53. doi:https://doi.org/10.1145/3512343

Puiu, A., Vizitiu, A., Nita, C., Itu, L., Sharma, P., & Comaniciu, D. (2021). Privacy-preserving and explainable AI for cardiovascular imaging. *Studies in Informatics and Control, 30*(2), 21--32. Retrieved from https://sic.ici.ro/vol-30-no-2-2021/privacy-preserving-and-explainable-ai-for-cardiovascular-imaging/

Qiu, S., Liu, Q., Zhou, S., & Wu, C. (2019). Review of Artificial Intelligence Adversarial Attack and Defense Technologies. *Applied Sciences, 9*(5), 909. doi:https://doi.org/10.3390/app9050909

Rahman, M. S., Khalil, I., Atiquzzaman, M., & Yi, X. (2020). Towards privacy preserving AI based composition framework in edge networks using fully homomorphic encryption. *Engineering Applications of Artificial Intelligence, 94*, 103737. doi:https://doi.org/10.1016/j.engappai.2020.103737

Rodriguez-Barroso, N., Stipcich, G., Jimenez-Lopez, D., Ruiz-Millan, J. A., Martinez-Camara, E. a., Gonzalez-Seco, G., . . . Herrera, F. (2020). Federated Learning and Differential Privacy: Software tools analysis, the Sherpa. ai FL framework and methodological guidelines for preserving data privacy. *Information Fusion, 64*, 270--292. doi:https://doi.org/10.1016/j.inffus.2020.07.009

Roemmich, K., Schaub, F., & Andalibi, N. (2023). Emotion AI at work: Implications for workplace surveillance, emotional labor, and emotional privacy. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, (pp. 1--20). doi:https://doi.org/10.1145/3544548.3580950

Roy, S. (2022). Privacy prevention of health care data using AI. *Journal of Data Acquisition and Processing, 37*(3), 769. doi:http://dx.doi.org/10.5281/zenodo.7699408

Saura, J. R., Ribeiro-Soriano, D., & Palacios-Marques, D. (2022). Assessing behavioral data science privacy issues in government artificial intelligence deployment. *Government Information Quarterly, 39*(4), 101679. doi:https://doi.org/10.1016/j.giq.2022.101679

Thuraisingham, B. M. (2020). Can ai be for good in the midst of cyber attacks and privacy violations? a position paper. *Proceedings of the Tenth ACM Conference on Data and Application Security and Privacy*, (pp. 1--4). doi:https://doi.org/10.1145/3374664.3379334

Toch, E., & Birman, Y. (2018). Towards Behavioral Privacy: How to Understand AI's Privacy Threats in Ubiquitous Computing. *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, 931--936. doi:https://doi.org/10.1145/3267305.3274155

Tom, E., Keane, P. A., Blazes, M., Pasquale, L. R., Chiang, M. F., Lee, A. Y., . . . Force, A. A. (2020). Protecting data privacy in the age of AI-enabled ophthalmology. *Translational vision science & technology, 9*(2), 36--36. doi:https://doi.org/10.1167/tvst.9.2.36

Tucker, C., Agrawal, A., Gans, J., & Goldfarb, A. (2018). Privacy, algorithms, and artificial intelligence. *The economics of artificial intelligence: An agenda*, 423--437. Retrieved from https://www.nber.org/system/files/chapters/c14011/c14011.pdf

Villegas-Ch, W., & Garcia-Ortiz, J. (2023). Toward a comprehensive framework for ensuring security and privacy in artificial intelligence. *Electronics, 12*(18), 3786. doi:https://doi.org/10.3390/electronics12183786

Wang, T., Zhao, J., Yu, H., Liu, J., Yang, X., Ren, X., & Shi, S. (2019). Privacy-preserving crowd-guided AI decision-making in ethical dilemmas. *Proceedings of the 28th ACM international conference on information and knowledge management*, (pp. 1311--1320). doi:https://doi.org/10.1145/3357384.3357954

Willems, J., Schmid, M. J., Vanderelst, D., Vogel, D., & Ebinger, F. (2023). AI-driven public services and the privacy paradox: do citizens really care about their privacy? *Public Management Review, 25*(11), 2116--2134. doi:https://doi.org/10.1080/14719037.2022.2063934

Zhang, Y., Wu, M., Tian, G. Y., Zhang, G., & Lu, J. (2021). Ethics and privacy of artificial intelligence: Understandings from bibliometrics. *Knowledge-Based Systems, 222*, 106994. doi:https://doi.org/10.1016/j.knosys.2021.106994

Zhang, Z., Al Hamadi, H., Damiani, E., Yeun, C. Y., & Taher, F. (2022). Explainable artificial intelligence applications in cyber security: State-of-the-art in research. *IEEE Access, 10*, 93104--93139. doi:10.1109/ACCESS.2022.3204051

Zhdanov, D., Bhattacharjee, S., & Bragin, M. A. (2022). Incorporating FAT and privacy aware AI modeling approaches into business decision making frameworks. *Decision Support Systems, 155*, 113715. doi:https://doi.org/10.1016/j.dss.2021.113715

Zhou, S., Liu, C., Ye, D., Zhu, T., Zhou, W., & Yu, P. S. (2022). Adversarial attacks and defenses in deep learning: From a perspective of cybersecurity. *ACM Computing Surveys, 55*(8), 1--39. doi:https://doi.org/10.1145/3547330

Zhu, T., Ye, D., Wang, W., Zhou, W., & Philip, S. Y. (2020). More than privacy: Applying differential privacy in key areas of artificial intelligence. *IEEE Transactions on Knowledge and Data Engineering, 34*(6), 2824--2843. doi:https://doi.org/10.1109/TKDE.2020.3014246