DOI 10.20535/2411-1031.2025.13.2.344715
UDC 004.8:004.056.5

OLEKSII SHAMOV

ADAPTIVE AI ARCHITECTURE FOR IMPLEMENTING PRIVACY-BY-DESIGN IN ACCORDANCE WITH GDPR

This article addresses one of the key challenges in modern intelligent systems engineering: the practical implementation of the Privacy-by-Design principle, enshrined in the General Data Protection Regulation (GDPR), within artificial intelligence architectures. Existing approaches, such as federated learning, differential privacy, and homomorphic encryption, while effective tools, create a rigid trade-off between the level of personal data protection, model utility (accuracy), and computational efficiency when applied statically. Such a unified "one-size-fits-all" approach is inefficient, as it leads to either excessive protection of non-sensitive data, which unjustifiably degrades performance, or insufficient protection for the most vulnerable categories of information. The objective of this research is to develop a conceptual framework for a novel artificial intelligence architecture that resolves this issue through dynamic, risk-oriented management of privacy mechanisms. The result of this study is a proposed adaptive hybrid architecture. The scientific novelty of this work lies in shifting from a static model of applying Privacy-Enhancing Technologies (PETs) to a flexible, multi-layered system. This system classifies data and model components in real-time based on their sensitivity level and associated risks. Depending on the risk level, the architecture dynamically applies an optimal set of protection tools: from basic federated learning with light differential privacy guarantees for low-risk data to the application of homomorphic encryption for the most critical computations. At the core of the architecture is an optimization model that aims to maximize model utility while minimizing computational costs, ensuring compliance with predefined privacy thresholds for each data category as required by GDPR. This approach enables the creation of more efficient, secure, and productive intelligent systems that meet modern regulatory demands.

Key words: artificial intelligence, GDPR, privacy-by-design, federated learning, differential privacy, homomorphic encryption, adaptive architecture, privacy-enhancing technologies.

1. Introduction. The era of artificial intelligence (AI) and big data has fundamentally and irrevocably transformed the technological landscape, becoming the driving force of the fourth industrial revolution. We are witnessing unprecedented innovations in all spheres of human activity: from personalized medicine, where algorithms predict disease development based on genomic data, and autonomous transport, which promises to enhance road safety, to intelligent financial systems capable of detecting complex fraudulent schemes in real-time. At the heart of this revolution lies the ability of complex algorithms, particularly machine and deep learning models, to identify non-obvious patterns in vast, multidimensional data arrays and make predictions with superhuman accuracy. However, this technological progress has a flip side. Its fuel is data, and this analytical capability is inextricably linked to the processing of personal data on an unprecedented scale. AI systems today operate not just with names and addresses, but with the most sensitive information about an individual: biometric indicators, medical diagnoses, behavioral patterns, financial transactions, and even emotional states. The uncontrolled use of such data creates significant and real risks to fundamental human rights and freedoms, including the right to privacy, protection against algorithmic discrimination, preservation of personal autonomy, and even freedom of thought. The awareness of these risks and the public demand for establishing digital sovereignty of the individual have led to the emergence of a new generation of data regulation legislation. The most significant and influential act in this area became the General Data Protection Regulation (GDPR), adopted by the

European Union in 2016. Thanks to its extraterritorial principle of application, the GDPR quickly transformed from a regional standard into a de facto global one, forcing organizations worldwide to reconsider their approaches to working with data. One of the cornerstones and most innovative provisions of the Regulation is Article 25, which enshrines the principles of "Data Protection by Design" and "Data Protection by Default". Together, they form the holistic concept of Privacy-by-Design, which requires a fundamental paradigm shift. It is a transition from an outdated, reactive security model, where protection measures were added to an already finished system ("bolted-on security"), to a proactive approach, where privacy is considered an integral part of the architecture, embedded in its foundation from the earliest stages of design. This principle requires engineers and developers not just to check boxes on a list of legal requirements, but to think in terms of privacy as one of the key functional characteristics of the system. For the engineering and technical community working in the field of AI, this requirement has become a serious practical and conceptual challenge. Traditional machine learning architectures, especially centralized models where all data is collected and processed in a single repository, are inherently contrary to key GDPR principles such as data minimization. Such centralized repositories not only create a single point of failure but also become attractive targets for attackers ("honey pots").

Therefore, the direct implementation of Privacy-by-Design requires a fundamental rethinking of the entire data lifecycle in AI systems: how data is collected, transmitted, stored, and used for training models. This task is complicated by the fact that any privacy protection measures inevitably conflict with other key non-functional requirements of the system. This gives rise to the so-called triumvirate of compromises, or the "privacy-accuracy-efficiency" trilemma. The implementation of privacy mechanisms (e.g., adding noise) often reduces the accuracy of predictions (model utility), while the use of cryptographic methods can increase computational complexity and training time by orders of magnitude. Finding an optimal, dynamic, and context-dependent balance within this triangle is one of the most important and complex scientific and practical tasks of modern intelligent systems engineering.

This article is dedicated to a thorough investigation of this problem and proposes ways to solve it. We aim to go beyond a simple review of existing privacy-enhancing technologies and their static combinations. Our goal is to propose a new architectural approach that will allow the creation of AI systems that not only formally comply with the letter of the law but do so in an efficient, flexible, and intelligent way. We present a concept that treats privacy not as a constraint, but as a manageable parameter in a multidimensional optimization problem. This will enable developers to build more trusted, resilient, and socially responsible intelligent systems, which is a necessary condition for their sustainable development and societal acceptance.

2. State-of-the-art. An analysis of current scientific research and engineering practices allows us to identify three main technological pillars on which AI architectures that implement Privacy-by-Design principles are built: Federated Learning (FL), Differential Privacy (DP), and Homomorphic Encryption (HE). Each of these approaches has unique advantages and disadvantages, and their combination opens the way to creating complex hybrid systems.

2.1. Federated Learning as a Decentralized Paradigm

Federated Learning, first proposed by Google researchers, has become a fundamental paradigm shift from centralized to decentralized machine learning. The main idea of FL is to train a global AI model on distributed data without the need to transfer this data to a central server. Instead, the global model is sent to client devices (e.g., mobile phones or hospital servers), where it is locally trained on local data. After that, only the model updates (e.g., gradients) are returned to the central server, not the data itself. These updates are aggregated to improve the global model, and the cycle repeats. As noted in the report of the European Data Protection Supervisor, FL inherently implements the principle of data minimization, which is a key requirement of the GDPR [1].

However, as K. Bonawitz et al, rightly point out in their survey [2], federated learning by itself is not a panacea. Although the raw data remains local, the transmitted model updates may contain enough information to carry out attacks, such as data reconstruction attacks or membership inference attacks, which allow an attacker to find out whose data was used for training. Thus, FL should be considered not as a standalone solution, but as a basic architecture that needs to be supplemented with other Privacy-Enhancing Technologies (PETs).

2.2. Differential Privacy as a Standard for Privacy Guarantees

Differential Privacy has become the gold standard for providing rigorous, mathematically provable privacy guarantees. The concept, formalized by C. Dwork [3], is that the result of any analysis (including the training of an Al model) should not significantly change if the data of one individual is removed from the initial dataset. In practice, this is achieved by adding carefully calibrated statistical noise to the data, intermediate results (e.g., gradients in FL), or final results. The level of privacy is quantitatively measured by the parameter epsilon (ε), where a smaller value of ε corresponds to a higher level of protection.

The combination of FL and DP (FL+DP) is the most common approach to building private AI systems. In such an architecture, clients add differentially private noise to their updates before sending them to the server. This approach was described in detail and implemented in the work of R. Shokri and V. Shmatikov [4], where they proposed a collaborative deep learning system that allows participants to train models while disclosing only a limited part of their updates. However, the main problem with DP, as convincingly shown by B. Jayaraman et al. in their practical study [5], is the fundamental trade-off between privacy and model utility. To achieve strong privacy guarantees (low ε), it is necessary to add a significant amount of noise, which can significantly degrade the model's accuracy. This challenge has spurred the development of more advanced architectures, such as the PATE (Private Aggregation of Teacher Ensembles) approach proposed by N. Papernot et al. [11], where privacy is achieved by aggregating the predictions of an ensemble of “teacher models” trained on disjoint data subsets. Nevertheless, the trade-off dilemma remains a central challenge for engineers working with DP.

2.3. Homomorphic Encryption as a Tool for Computations on Encrypted Data

Homomorphic Encryption is a cryptographic technique that allows computations to be performed directly on encrypted data without decrypting it. The result of such computations remains encrypted, and after decryption, it will be identical to the result that would have been obtained from computations on plaintext data. From a privacy perspective, this is an ideal solution, as the server performing the computations (e.g., aggregating gradients in FL) never has access to unencrypted data or updates. However, as noted in the survey by R. Podschwadt et al. [6], the main obstacle to the widespread adoption of HE is the extremely high computational costs and significant resource consumption. Operations on encrypted data are orders of magnitude slower than on plaintext. Despite significant progress in optimization, such as in the work of Y. Joo et al., where a speedup of up to 2.55 times was achieved for neural networks [7], the practical use of fully homomorphic encryption for training complex deep learning models remains extremely difficult. Because of this, HE is more often used for specific, less computationally intensive tasks, such as secure aggregation in FL.

2.4. Hybrid Approaches and Unresolved Problems

Understanding the limitations of each approach has prompted researchers to create hybrid architectures. One of the indicative works in this direction is the study by S. Truex et al. [8], which proposes a system that combines FL, DP, and Secure Multiparty Computation (SMC), in particular, using additive homomorphic encryption. In their system, clients add a smaller amount of differentially private noise, and secure aggregation using cryptography ensures that the server cannot see individual updates, which allows for increased model accuracy at the same level of privacy. Similar hybrid models that combine different PETs are considered in the comparative study by E. Shalabi et al. [9].

Despite significant progress, the literature analysis reveals a key unresolved problem: existing approaches, even hybrid ones, predominantly apply a static, unified level of protection to all data and computational stages. For example, the same privacy budget ε is applied to all gradient updates, regardless of how sensitive the data used to calculate them was. Such a "one-size-fits-all" approach is inefficient.

It either provides excessive protection for non-sensitive data, leading to an unjustified loss of accuracy and performance, or proves insufficient for protecting the most vulnerable information, such as data belonging to special categories according to Article 9 of the GDPR. This lack of flexibility and adaptability is the main gap that this article aims to address.

3. Formulation of the Article's Goals. Based on the analysis of unresolved problems in the implementation of Privacy-by-Design, the main goal of this article is to develop a conceptual framework for an adaptive, risk-oriented hybrid AI architecture. This architecture is intended to overcome the limitations of static approaches by dynamically managing privacy-enhancing technologies to optimize the balance in the "privacy-accuracy-efficiency" triad.

To achieve this goal, the following tasks are set:

1. To design a multi-level model for classifying data and computations by privacy risk level.
2. To develop the logic for dynamic selection and configuration of PETs (FL, DP, HE) according to the identified risk level.
3. To formulate the concept of an optimization model that underlies the architecture and manages the trade-offs between key system metrics.

The *scientific novelty* of this work lies in the transition from the paradigm of static, universal protection to dynamic, context-dependent privacy management. Unlike existing works that focus on combining PETs in a fixed configuration, we propose an intelligent system that treats the choice of protection mechanism as a real-time optimization task. The proposed architecture can adapt the level of protection to the sensitivity of specific data and the vulnerability of individual components of the AI model. This allows for a more granular and effective fulfillment of GDPR requirements while minimizing the negative impact on model utility and computational resources. Such a proposal is a new engineering approach to building trusted intelligent systems.

4. Adaptive, Risk-Oriented Hybrid Architecture. To address the stated challenges and overcome the fundamental trade-offs inherent in static approaches, we propose the concept of an Adaptive, Risk-Oriented Hybrid Architecture (AROHA). This architecture is not just a combination of existing technologies, but a new paradigm for designing intelligent systems, based on flexibility and contextual awareness. The term "adaptive" signifies the system's ability to dynamically change its behavior and protection configuration in response to changing data characteristics and privacy requirements. "Risk-oriented" emphasizes that the driving force for decision-making is not hard-coded rules, but a continuous assessment of potential risks to the rights and freedoms of data subjects, as required by the spirit of the GDPR. Finally, "hybrid" indicates that the architecture intelligently orchestrates the application of a whole spectrum of Privacy-Enhancing Technologies (PETs), using each where it is most appropriate.

The fundamental principle of AROHA lies in the recognition and practical application of the fact that not all data and not all computations are equally sensitive. For example, in a medical AI system, a patient's diagnosis carries significantly higher risks than general demographic statistics, and the gradients of the last layers of a neural network can reveal more information about specific training examples than the gradients of the initial layers. Applying a single, most stringent level of protection to the entire system for example, using homomorphic encryption for all operations without exception is extremely inefficient and practically infeasible. Such a "one-size-fits-all" approach creates a false dichotomy: either we choose maximum protection, which leads to exorbitant computational costs and significant degradation of model accuracy, or we settle for a weaker but more productive level of protection, which may be insufficient for special categories of data. AROHA moves beyond this

limited choice by differentiating protection approaches based on a continuous, granular risk assessment. Instead of building an impregnable fortress around the entire data array, it creates a multi-layered security system where the most valuable assets are protected by the strongest mechanisms, and less critical ones by more efficient and lightweight means. To implement this principle, the proposed architecture consists of three key interconnected modules that together form a closed loop of privacy management: a sensitivity assessment module, a dynamic PET application module, and an optimization model.

4.1. Multi-level Sensitivity Assessment Module

The first and primary component of the architecture is the Multi-level Sensitivity Assessment Module. Its task is to classify both input data and internal components of the AI model by their level of privacy risk. This classification is dynamic and can take into account various factors, including the type of data, the context of its use, and the vulnerability of a specific computational stage. A three-level classification model is proposed:

– *Level 1: Low Risk*. This level includes data that is not personal or has been effectively anonymized, as well as computations on those layers of the neural network that are known to be less prone to information leakage (e.g., initial convolutional layers in computer vision models). Parameter updates related to this level have a low potential for data reconstruction attacks.

– *Level 2: Medium Risk*. This level covers ordinary personal data not belonging to special categories (e.g., age, gender, geolocation data), as well as computations in the intermediate layers of the model that may indirectly disclose information about the input data. This level requires a balanced approach to protection.

– *Level 3: High Risk*. This level includes special categories of data as defined in Article 9 of the GDPR (e.g., medical diagnoses, biometric data, political opinions), financial information, as well as computations directly related to the model's input and output data. Gradients calculated for these layers are the most informative and pose the greatest threat in case of a leak. In addition, data belonging to vulnerable population groups may be assigned to this level.

Classification can be carried out based on metadata, automatic analysis of data semantics (e.g., using natural language processing models), and predefined security policies that comply with GDPR requirements and organizational policies.

4.2. Dynamic PET Application Module

The second component is the Dynamic PET Application Module. This module receives information from the Sensitivity Assessment Module and, depending on the assigned risk level, applies the appropriate protection mechanism. Its logic is as follows:

– *For Level 1 (Low Risk)*: The basic Federated Learning (FL) architecture is used. To provide a basic level of protection and comply with general principles, Differential Privacy (DP) with a high privacy budget (large ε value) can be applied to the updates. This adds a minimal amount of noise, which has almost no effect on the model's accuracy and convergence speed, but creates a formal privacy guarantee.

– *For Level 2 (Medium Risk)*: A combination of FL + DP with strict privacy guarantees (small ε value) is applied. To reduce the amount of noise needed to achieve the target ε at the server level, the aggregation of updates can be protected using cryptographic methods such as Secure Multiparty Computation (SMC) based on additive homomorphic encryption. This allows the server to compute the sum of the updates without seeing them individually, which reduces overall risks.

– *For Level 3 (High Risk)*: To process the most sensitive data and perform critical computations, the strongest tool is applied Homomorphic Encryption (HE). For example, if a model processes financial transactions to detect fraud, computations directly related to transaction amounts can be performed in an encrypted form. Although this is computationally expensive, such an approach is justified for protecting the most valuable information. The use of HE is selective and applies only to a small part of the computations, making it practically feasible.

Thus, the system dynamically creates a hybrid protection model that adapts to the data context.

4.3. Optimization Model

At the core of the AROHA framework lies a sophisticated optimization model that governs the selection and configuration of the privacy-enhancing technologies. Its purpose is to maximize the overall effectiveness of the system by finding the best possible balance among three competing, and often conflicting, objectives: the utility of the model, the level of privacy, and the computational costs. Here, model utility refers to its performance, such as predictive accuracy. The privacy level represents the strength of the safeguards protecting the data, while computational costs include the necessary processing power and time. The model operates on a principle of maximizing this utility while simultaneously minimizing both privacy risks and computational expenses. It achieves this by treating the task as a multi-objective optimization problem. The system assigns different levels of importance or weights to each of these three factors, which can be tuned according to the specific requirements of the application. For instance, in a medical diagnostics system, the weight for privacy would be set extremely high, whereas for a less sensitive application, the weights for utility and efficiency might take precedence. Crucially, this optimization does not happen in a vacuum. It is bound by a fundamental rule: for every category of data, the level of privacy must never fall below a pre-defined minimum threshold. This threshold is established by security policies in strict accordance with the legal mandates of GDPR, ensuring that the most sensitive data (e.g., High-Risk Level 3) always receives the robust, non-negotiable protection it requires. In practice, this optimization model functions as an intelligent controller. At each round of the federated learning process, or for each new batch of data, it assesses the context and decides on the most appropriate combination of PETs to deploy. This transforms the static, one-time choice of an architecture into a continuous, dynamic process of intelligent resource allocation, which is the key innovation of the proposed approach.

5. Results and Discussion. The main result of this study is the development and justification of a conceptual framework for an adaptive, risk-oriented hybrid architecture (AROHA). This framework is not just another configuration of existing technologies, but offers a new engineering philosophy for implementing Privacy-by-Design in complex AI systems. Unlike the vast majority of existing solutions that apply unified, static protection mechanisms to the entire data stream, the proposed architecture provides dynamic, granular, and context-aware privacy management. The novelty of the approach lies in the creation of an intelligent orchestration layer that makes real-time decisions about which protection tool, with which parameters, and to which specific data or computations to apply. This allows for a much more effective and flexible balance between three critical, often conflicting, goals: compliance with strict GDPR requirements, preservation of high model accuracy (utility), and minimization of computational costs. In essence, AROHA transforms privacy from a rigid constraint into a manageable, optimized system parameter.

To gain a deeper understanding of the contribution of the proposed approach, let us compare it with the results and conclusions obtained by other leading researchers in this field. The pioneering works that laid the foundation for private machine learning, particularly the work of R. Shokri and V. Shmatikov [4], brilliantly demonstrated the very possibility of effectively combining distributed deep learning with differential privacy. However, their approach, though revolutionary, applied a single protection mechanism and level to all participants and their updates.

AROHA develops this idea by adding a superstructure in the form of an intelligent controller that differentiates the level of protection depending on the properties of the data itself. This allows for a partial solution to the fundamental problem of the trade-off between privacy and accuracy, which was so sharply and convincingly highlighted by B. Jayaraman et al. [5] in their landmark study. They empirically showed that to achieve strong, formal privacy guarantees with a unified approach, the accuracy of complex models can drop to an unacceptable level. AROHA directly responds to this criticism: our architecture proceeds from the assumption that applying such strong guarantees everywhere is strategically wrong. Instead, it selectively applies them only where the high risk

justifies the inevitable loss of utility, allowing the system as a whole to function with much higher accuracy.

More modern hybrid models, similar to the one proposed by S. Truex et al. [8], which combine FL, DP, and cryptography (SMC), are a significant step forward compared to mono-technology approaches. However, even in their system, the choice and configuration of these tools remain static, determined at the design stage.

AROHA proposes the next logical step in evolution: to make the very combination of these tools dynamic and context-dependent. For example, instead of always using computationally expensive secure multiparty computation for aggregation, AROHA can activate this mechanism only for updates classified as belonging to medium and high-risk levels, using simpler and faster methods for low-risk updates. A similar argument applies to research in the field of homomorphic encryption. Works such as that of Y. Joo et al. [7] focus on optimizing the technology itself from within, trying to reduce its computational costs. AROHA offers a complementary approach: instead of trying to perform all model training using HE, which is currently impractical, our architecture proposes a "surgically precise", selective application of HE only to those computations where it is absolutely necessary. This significantly reduces the overall computational load on the system.

An approach similar to ours for granular privacy management can be found in the article by Z. He et al. [10], where they propose a system that adapts DP parameters depending on the level of trust in the participants of federated learning.

Our proposal is more general and fundamentally more consistent with the GDPR, as it is based not on the subjective category of trust, but on the objective properties of the data itself. The requirements of the GDPR, especially regarding special categories of data, are tied to the nature of the information, not the reputation of its processor.

Of course, the proposed architecture is not without its own challenges and limitations. The practical implementation of the Multi-level Sensitivity Assessment Module is a complex engineering task that requires the development of accurate, fast, and reliable algorithms for automatic data classification. In addition, the optimization controller itself creates additional computational overhead and could potentially become a bottleneck in the system. However, we hypothesize that with proper implementation, the gains in resource efficiency and preservation of model accuracy will far outweigh these costs, especially in large-scale, industrial systems.

Thus, the proposed architecture is not just a theoretical construct, but also a practical engineering guideline. It shifts the discussion in the professional community from the question of "whether to use PETs?" to the much more productive and relevant question of "how to use them intelligently, flexibly, and adaptively to build trusted AI systems?".

6. Conclusions. This study was dedicated to solving one of the most fundamental and complex problems at the intersection of artificial intelligence and law: the effective and pragmatic implementation of the Privacy-by-Design principle, enshrined in the GDPR, in the architectures of modern intelligent systems. The analysis of the state-of-the-art confirmed that existing technological tools federated learning, differential privacy, and homomorphic encryption are powerful, yet limited in their isolated application. We concluded that their static, unified use, where a single level of protection is applied to the entire system, inevitably leads to an inefficient and rigid trade-off within the "privacy-accuracy-efficiency" trilemma. This approach forces engineers to choose between insufficient protection for sensitive data or excessive protection for non-critical information, which in both cases is a suboptimal solution from both a technical and a regulatory perspective.

In response to this identified gap, the article proposed and thoroughly substantiated the conceptual framework of a new adaptive, risk-oriented hybrid architecture (AROHA). The key idea underlying our proposal is a radical departure from the static, monolithic protection paradigm in favor of flexible, intelligent, and dynamic management of privacy mechanisms. The proposed architecture, consisting of three synergistic components a module for multi-level data classification by sensitivity,

a module for the dynamic application of an optimal set of PETs for each risk level, and a governing optimization model transforms the problem of ensuring privacy from an immutable constraint into a manageable parameter. This allows the system to adapt to the context in real-time, applying a level of protection that is adequate to the risks associated with specific data and computations.

The main contribution of this work lies in the formulation of a new engineering approach that serves as a practical blueprint for building the next generation of trusted AI systems. We propose creating more flexible, efficient, and granularly managed systems that not only formally comply with GDPR requirements but also embody its spirit. This approach avoids unnecessary degradation of model accuracy by protecting non-sensitive data with lighter methods, while simultaneously concentrating powerful, albeit computationally expensive, tools like homomorphic encryption on protecting the most vulnerable categories of information. This is not only a technical but also an economic optimization, making the implementation of robust privacy guarantees more realistic for a wide range of applications.

Prospects for further research are broad and multifaceted.

Firstly, an urgent practical step is the development and implementation of a software prototype of the proposed architecture. Its creation will allow for comprehensive empirical studies to quantitatively assess the gains in accuracy and performance compared to existing static approaches on standard benchmark datasets.

Secondly, further development of algorithms for the Sensitivity Assessment Module requires significant attention. This is a complex scientific task that demands the creation of lightweight and effective methods for automatic semantic data classification and identification of vulnerable model components in real-time, possibly using meta-learning.

Thirdly, a deep theoretical investigation of the properties of the optimization model is an interesting direction, including formal proofs of convergence guarantees and analysis of its computational complexity when using various optimization methods, from classical to reinforcement learning-based. Finally, the most important direction is the extension of the AROHA philosophy itself beyond privacy. We hypothesize that the proposed risk-oriented, adaptive approach can be generalized to manage other ethical aspects of AI, such as fairness, transparency, and explainability. This opens the way to the creation of comprehensive frameworks for managing trust in AI (Trustworthy AI), where the system can dynamically balance between different, often conflicting, ethical requirements, which is one of the most difficult challenges on the path to creating responsible artificial intelligence.

REFERENCE

[1] “TechSonar report: Federated learning”, *European Data Protection Supervisor*, June 2024. [Online]. Available: https://www.edps.europa.eu/press-publications/publications/techsonar/federated-learning_en. Accessed on: May 19, 2025.

[2] K. Bonawitz et al., “Towards federated learning at scale: System design”, in *Proc. 2nd SysML Conf.*, Stanford, CA, USA, 2019, 15 p. doi: https://doi.org/10.48550/arXiv.1902.01046.

[3] C. Dwork, “Differential Privacy”, in *Encyclopedia of Cryptography and Security*, H.C.A. van Tilborg, S. Jajodia, Eds. Boston, MA, USA: Springer, 2011, pp.338-340. doi: https://doi.org/10.1007/978-1-4419-5906-5_752.

[4] R. Shokri, and V. Shmatikov, “Privacy-Preserving Deep Learning”, in *Proc. 22nd ACM SIGSAC Conf. on Comp. and Commun. Sec.*, Denver, CO, USA, 2015, pp. 1310-1321. doi: https://doi.org/10.1145/2810103.2813687.

[5] B. Jayaraman, and D. Evans, “Evaluating Differentially Private Machine Learning in Practice”, in *Proc. 28th USENIX Sec. Symp. (USENIX Security 19)*, Santa Clara, CA, USA, 2019, pp.1895-1912. doi: https://doi.org/10.48550/arXiv.1902.08874.

[6] R. Podschwadt, D. Takabi, P. Hu, M.H. Rafiei and Z. Cai, “A Survey of Deep Learning Architectures for Privacy-Preserving Machine Learning with Fully Homomorphic Encryption”, *IEEE Access*, vol. 10, pp. 117477-117500, 2022. https://doi.org/10.1109/ACCESS.2022.3219049.

[7] Y. Joo, S. Ha, H. Oh, and Y. Paek, “Efficient Keyset Design for Neural Networks Using Homomorphic Encryption”, *Sensors* 2025, vol. 25 (14), July 2025. doi: https://doi.org/10.3390/s25144320.

[8] S. Truex et al, “A Hybrid Approach to Privacy-Preserving Federated Learning”, in *Proc. 12th ACM Workshop on AI and Sec*., London, UK, 2019, pp. 1-11. doi: https://doi.org/10.1145/3338501.3357370.

[9] E. Shalabi, W. Khedr, E. Rushdy, and A. Salah, “A Comparative Study of Privacy-Preserving Techniques in Federated Learning: A Performance and Security Analysis”, *Information*, vol. 16, iss. 3, art. 244, 2025. doi: https://doi.org/10.3390/info16030244.

[10] Z. He, L. Wang, and Z. Cai, “Clustered Federated Learning with Adaptive Local Differential Privacy on Heterogeneous IoT Data”, *IEEE Internet of Things Journal*, vol. 11, no. 1, pp. 137-146, 2024. doi: https://doi.org/10.1109/JIOT.2023.3299947

[11] N. Papernot, M. Abadi, U. Erlingsson, I. Goodfellow, and K. Talwar, “Semi-supervised Knowledge Transfer for Deep Learning from Private Training Data”, in *Proc. Int. Conf. on Learning Representations (ICLR)*, Toulon, France, 2017, 16 p. doi: http://dx.doi.org/10.48550/arXiv.1610.05755.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] “TechSonar report: Federated learning”, *European Data Protection Supervisor*, June 2024. [Online]. Available: https://www.edps.europa.eu/press-publications/publications/techsonar/federated-learning_en. Accessed on: May 19, 2025.

[2] K. Bonawitz et al., “Towards federated learning at scale: System design”, in *Proc. 2nd SysML Conf*., Stanford, CA, USA, 2019, 15 p. doi: https://doi.org/10.48550/arXiv.1902.01046.

[3] C. Dwork, “Differential Privacy”, in *Encyclopedia of Cryptography and Security*, H.C.A. van Tilborg, S. Jajodia, Eds. Boston, MA, USA: Springer, 2011, pp.338-340. doi: https://doi.org/10.1007/978-1-4419-5906-5_752.

[4] R. Shokri, and V. Shmatikov, “Privacy-Preserving Deep Learning”, in *Proc. 22nd ACM SIGSAC Conf. on Comp. and Commun. Sec.*, Denver, CO, USA, 2015, pp. 1310-1321. doi: https://doi.org/10.1145/2810103.2813687.

[5] B. Jayaraman, and D. Evans, “Evaluating Differentially Private Machine Learning in Practice”, in *Proc. 28th USENIX Sec. Symp. (USENIX Security 19)*, Santa Clara, CA, USA, 2019, pp.1895-1912. doi: https://doi.org/10.48550/arXiv.1902.08874.

[6] R. Podschwadt, D. Takabi, P. Hu, M.H. Rafiei and Z. Cai, “A Survey of Deep Learning Architectures for Privacy-Preserving Machine Learning with Fully Homomorphic Encryption”, *IEEE Access*, vol. 10, pp. 117477-117500, 2022. https://doi.org/10.1109/ACCESS.2022.3219049.

[7] Y. Joo, S. Ha, H. Oh, and Y. Paek, “Efficient Keyset Design for Neural Networks Using Homomorphic Encryption”, *Sensors* 2025, vol. 25 (14), July 2025. doi: https://doi.org/10.3390/s25144320.

[8] S. Truex et al, “A Hybrid Approach to Privacy-Preserving Federated Learning”, in *Proc. 12th ACM Workshop on AI and Sec*., London, UK, 2019, pp. 1-11. doi: https://doi.org/10.1145/3338501.3357370.

[9] E. Shalabi, W. Khedr, E. Rushdy, and A. Salah, “A Comparative Study of Privacy-Preserving Techniques in Federated Learning: A Performance and Security Analysis”, *Information*, vol. 16, iss. 3, art. 244, 2025. doi: https://doi.org/10.3390/info16030244.

[10] Z. He, L. Wang, and Z. Cai, “Clustered Federated Learning with Adaptive Local Differential Privacy on Heterogeneous IoT Data”, *IEEE Internet of Things Journal*, vol. 11, no. 1, pp. 137-146, 2024. doi: https://doi.org/10.1109/JIOT.2023.3299947

[11] N. Papernot, M. Abadi, U. Erlingsson, I. Goodfellow, and K. Talwar, “Semi-supervised Knowledge Transfer for Deep Learning from Private Training Data”, in *Proc. Int. Conf. on Learning Representations (ICLR)*, Toulon, France, 2017, 16 p. doi: http://dx.doi.org/10.48550/arXiv.1610.05755.

The article was received 15.08.2025.

ОЛЕКСІЙ ШАМОВ

АДАПТИВНА АРХІТЕКТУРА ШІ ДЛЯ РЕАЛІЗАЦІЇ ПРИНЦИПУ ПРИВАТНОСТІ ЗА ЗАМОВЧУВАННЯМ ВІДПОВІДНО ДО GDPR

У статті розглядається одна з ключових проблем сучасної інженерії інтелектуальних систем: практична реалізація в архітектурах штучного інтелекту принципу приватності за замовчуванням (Privacy-by-Design), закріпленого у Загальному регламенті про захист даних (GDPR). Існуючі підходи, такі як федеративне навчання, диференційна приватність та гомоморфне шифрування, хоча й є ефективними інструментами, при їх статичному застосуванні створюють жорсткий компроміс між рівнем захисту персональних даних, корисністю (точністю) моделі та обчислювальною ефективністю. Такий уніфікований підхід «один розмір для всіх» є неефективним, оскільки призводить або до надмірного захисту нечутливих даних, що невиправдано знижує продуктивність, або до недостатнього захисту найбільш вразливих категорій інформації. Метою даного дослідження є розробка концептуальної рамки нової архітектури штучного інтелекту, яка вирішує цю проблему шляхом динамічного, ризик-орієнтованого управління механізмами приватності.

Результатом дослідження є запропонована адаптивна гібридна архітектура. Наукова новизна роботи полягає у відході від статичної моделі застосування технологій підвищення приватності (PETs) до гнучкої, багаторівневої системи. Ця система в режимі реального часу класифікує дані та компоненти моделі за рівнем чутливості та пов’язаних з ними ризиків. Залежно від рівня ризику, архітектура динамічно застосовує оптимальний набір інструментів захисту: від базового федеративного навчання з легкими гарантіями диференційної приватності для низькоризикових даних до застосування гомоморфного шифрування для найбільш критичних обчислень. В основі архітектури лежить модель оптимізації, що прагне максимізувати корисність моделі при мінімізації обчислювальних витрат, гарантуючи при цьому дотримання заздалегідь визначених порогових значень приватності для кожної категорії даних відповідно до вимог GDPR. Такий підхід дозволяє створити більш ефективні, безпечні та продуктивні інтелектуальні системи, що відповідають сучасним регуляторним вимогам.

Ключові слова: штучний інтелект, GDPR, приватність за замовчуванням, федеративне навчання, диференційна приватність, гомоморфне шифрування, адаптивна архітектура, технології підвищення приватності.

Shamov Oleksii, Intelligent systems researcher, head of Human Rights Educational Guild, Cherkasy, Ukraine. ORSID 0009-0009-5001-0526, shamov@hreg.org.ua

Шамов Олексій Анатолійович, дослідник інтелектуальних систем, голова ГО “Просвітня фундація з прав людини”, Черкаси, Україна.