Educational Administration: Theory and Practice
2023, 29(4), 5950-5958
ISSN: 2148-2403
https://kuey.net/

Research Article

Privacy-Preserving Machine Learning Models for Sensitive Customer Data in Insurance Systems

Keerthi Amistapuram[1*]

[1*]*Lead Software Developer* ORCID ID : 0009-0009-6408-1958

Citation: Keerthi Amistapuram (2023). Privacy-Preserving Machine Learning Models for Sensitive Customer Data in Insurance Systems., *Educational Administration: Theory and Practice*, 29(4) 5950-5958
Doi: 10.53555/kuey.v29i4.10965

ARTICLE INFO

ABSTRACT

The insurance industry is exploring the use of machine learning (ML) models to leverage the huge volume of customer data for of-the-moment business decisions. It is, however, extremely sensitive information. From a design perspective, data attribute utility should be carefully balanced with privacy guarantees, particularly when sensitive customer data is involved. Privacy risks can be mitigated by using techniques that reduce and control the amount of sensitive information exposed during the training and use of ML models. A wide spectrum of privacy-preserving machine learning solutions has been developed. They are based on a comprehensive view of data protection-impact assessments under privacy laws and reg- ulations, subsequently consolidating the specific requirements for both personal identifiable information (PII) and personal health identifiable (PHI) information. For sufficiently large datasets, fair ML solutions with differential privacy-DPIA compliance can be obtained without compromising model performance. Notably, certain ML tasks, such as risk scoring and underwriting, can be accomplished with very close-to-the-source data while preserving DP-compliance for protected attributes. Risk scoring and underwriting processes are performed under the control of one institution, while fraud detection and claims management procedures apply an anomaly-detection-based architecture. For sensitive attributes such as health data, disparity in training data volume can be solved by transferring knowledge through privacy-preserving federated learning. Sensitive attributes with low entropy are avoided at prediction time to mitigate the associated disclosure risk. For such features, privacy and risk evaluation techniques such as k-anonymity and ℓ-diversity are embedded into the data-governance step, ensuring that the data support radarized and risk-aware disclosures when exposed to third parties.

***Index Terms*—**Privacy-Preserving Machine Learning (PPML),Federated Learning,Differential Privacy,Secure Multi-Party Computation (SMPC),Insurance Data Analytics,Sensitive Customer Data Protection,Data Anonymization and Encryption,Regulatory Compliance (GDPR / HIPAA),Explainable Artificial Intelligence (XAI) in Insurance,Trustworthy and Ethical AI Systems.

I. INTRODUCTION

Balancing the conflicting requirements of privacy regu- lations and quality-driven data-hungry machine-learning ap- proaches presents a difficult challenge in insurance systems. Real-world experience shows that debt collection systems' decision procedures enjoy an insurable risk. While preventing the ecosystem's participation is less costly, the monitoring commitment to detect fraud remains a standard market as-sumption. Privacy-preserving machine-learning models ad- dress risk scoring, fraud detection, and claims management, enabling sensitive information sharing within privacy-sensitive channels. Privacy-preserving ML refers to the development and application of machine-learning techniques that allow the sharing and utilization of sensitive information and personal identifiable information (PII) without violating privacy regula- tions, such as the General Data Protection Regulation (GDPR) or local implementations of the Health Insurance Portability and Accountability Act. Privacy regulations generally identify two main categories of sensitive data: (1) personal health information (PHI), such as health status or health claims for individuals,

and (2) credit data information obtained through financial institutions, e.g., identification of credit card payment events, classified as a previous bad debt event. Agencies or institutions are usually rewarded for collecting samples describing such events. Nevertheless, real-world systems face participation risks despite those rewards.

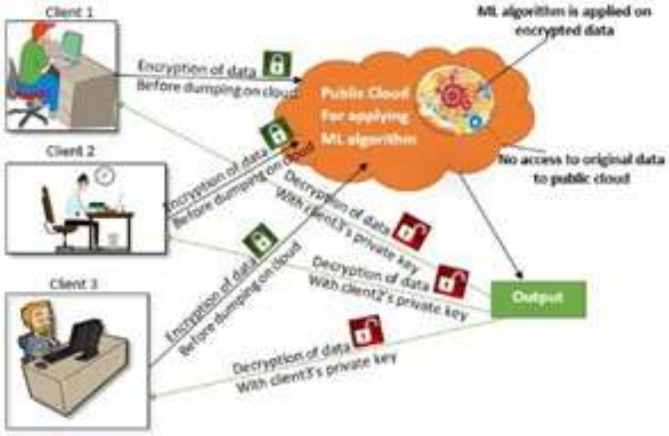

Fig. 1. Overview of Data Privacy in Machine Learning

A. Context and Motivation

The growing concern for privacy in the digital age has resulted in stricter regulations governing the processing of personal data, many of which are applicable to the insurance industry. These laws not only constrain the storage, retention, and sharing of sensitive customer data (for example, personally identifiable information (PII) and personally health informa- tion (PHI)), but also demand that entities deploying automated decision systems must mitigate algorithmic discrimination. However, these obligations introduce friction with common practices in the analysis of sensitive customer data in the insurance industry. Specifically, these constraints hinder the construction of large data sets to train powerful machine learning models, the synergistic sharing of data among collab- orating insurers to benefit all parties, and the use of modeling approaches that favour the attainment of high utility metrics over low risks of privacy loss. Nevertheless, the principles of data protection can also serve as a guide toward superior technical solutions. The risks associated with the lack of privacy-preserving measures are driving new approaches in machine learning that help to deliver more privacy-aware products, services, and decisions. Privacy-preserving machine learning refers to a collection of techniques aimed at protecting sensitive attributes during the training and execution phases of algorithms within the context of data-driven models potentially involving sensitive customer data.

B. Scope and Definitions

The Insurance market is highly data-driven, covering a considerable area of business across the globe. Predictive models based on Machine Learning (ML) help insurers in fact- based decision making and assist in managing risks. However, for most applications, sensitive customer data containing Per- sonally Identifiable Information (PII), e.g., names, addresses, phone numbers, email ids, etc., or Protected Health Informa- tion (PHI), e.g., health records, are used. Data-driven insurance systems must balance the utility of information against the seriousness of unintended disclosures and, therefore, Privacy- Preserving Machine Learning Models (PPMLM) are neces- sary to prevent misuse of sensitive customer data. A model or technique is Privacy-Preserving (PP) if it minimizes the disclosure risk against the data utility requirements of a given business application. Disclosure risk can be measured using several mathematical metrics, e.g., k-anonymity, l-diversity, t- closeness, differential privacy budgets, etc. It is possible to use these metrics in tandem with business-specific Response Variability Measures (RVM) to determine an acceptable level of risk. PPMLM aim to achieve adequately low disclosure risk so that the decision makers can choose to ignore the risk element when using the information for business decisions. Data Protection Impact Assessments (DPIA) must also be conducted to assess whether Machine Learning models yield sensitive features such as gender, race, age, etc.

II. REGULATORY AND ETHICAL CONSIDERATIONS

To protect sensitive label information, data protection laws have emerged or evolved in several regions. They require that appropriate trade-offs between model performance and transparency be established based on data types and subjects. Insurance datasets often contain the sensitive attributes that can be adjusted by using other variables. Using these features while concealing the sensitive or identity information can enhance the model's transparency. Hence, fairness, accountability, and transparency (FAT) are essential in AI/ML applications. Privacy, fairness, accountability, and transparency should be taken into serious consideration throughout the entire ML lifecycle (from data collection and preparation to model development and deployment) to comply

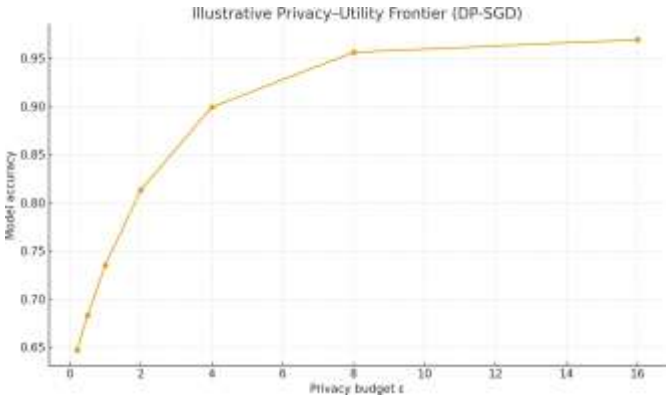

Fig. 2. Illustrative Privacy–Utility Frontier (DP-SGD)

epsilon	accuracy
0.2	0.647
0.5	0.683
1	0.735
2	0.813
4	0.899
8	0.956
16	0.969

TABLE I DP PRIVACY–UTILITY FRONTIER

with regulations such as the European Union's GDPR. Doing a data protection impact assessment (DPIA) is needed by law when the data-processing operation is likely to result in a high risk to the rights and freedoms of natural persons. The FAT framework can also be supplemented by audits that cover explainability and interpretability of model behavior. These aspects are critical for ensuring liberated machine data privacy. AI models that are perceived as biased, sexist, or racist are more likely to be disliked for these reasons and face calls for bans. Such factors can negatively affect the business. Hence, ensuring fairness and considering the influence of sensitive attributes on the decision-making process is becoming ever more crucial, if not lawfully mandated.

Equation 01: Differential Privacy (DP) — definitions & mechanisms

$$\Pr[M(D) \in S] \leq e^{\epsilon} \Pr[M(D') \in S] + \delta \quad (1)$$

ϵ ("privacy budget") controls multiplicative leakage; δ is a small failure probability

A randomized mechanism $M : D \to R$ is (ϵ, δ)-DP if for all adjacent datasets D, D' (differ by one individual) and all measurable S ⊆ R: $\Pr[M(D) \in S] \leq e^{\epsilon} \Pr[M(D') \in S] + \delta$

ϵ ("privacy budget") controls multiplicative leakage; δ is a small failure probability

A. Data Protection Laws and Compliance

The generic privacy-preserving ML concept maps to these requirements through different techniques and models for handling sensitive data within an insurance use-case. Risk scoring, underwriting, fraud detection, and claims management are key insurance scenarios that involve personal data. Such scenarios become useful through new techniques like federated learning, cross-silo-portable federated learning, homomorphic encryption, secure multi-party computation, and differential privacy. Inspecting data-protection requirements through the lens of fairness, accountability, and transparency shows that any essential predictive model must comply with legal and ethical principles. Above all, it should enable individuals to be magically included in the group of persons holding predictive models or classifiers capable of inferring sensitive attributes such as the risk of death. At present, this requirement is not yet achievable, but it remains a direction for future deployment. Organizations should indeed be able to provide individuals with evidence concerning the possibilities of data usage and follow the principle of justifiable use. Predictive models could even be explained in easy-to-understand language, showing how the model drives the flow of the data and why a specific decision was reached.

B. Fairness, Accountability, and Transparency

The GDPR's Article 9 prohibits processing of special cate- gories of personal data, including PII and PHI, unless specific conditions are fulfilled. Although many ML models could be legally deployed after passing a Data Protection Impact Assessment (DPIA), executing these models may nonetheless violate other fundamental rights unless they are supervised by humans. Equally applying these principles to capital markets, the European Commission recently proposed new artificial intelligence (AI) regulations. Advertised

as ensuring fairness, accountability, and transparency, the actors affected by these regulations expect explanations that allow for a transfer of trust. Trust enables insurers to access otherwise inaccessible personal data when pricing P3 offers. This ethical use of P3 data requires adequately modelling Protected Attributes such as age or sex. Such Fair ML must consider Disparate Impact as well as accuracy across Protected Attributes. Within the overall design of machine learning (ML) models, an appropriate Data Governance Framework is hence the first prerequisite to ensure Fair ML. The separation of insurance systems into independent silos does not prevent the joint training of models for Risk Scoring, Underwriting, and Claims Management. On the contrary, it allows collaboration across institutional boundaries. Information asymmetries are a prerequisite for low-cost corruption or low-cost fraud. Announcing the fraud detection during a carnival week could for example microtarget perpetrators if Claims Management covers the costs of the model retraining. However, during the risk-scoring and under- writing process, carriers receive highquality offers in staged processes that use Only what is Needed. However, proper pricing only occurs when risk informations can be shared across carriers. Thus, a proper design of the ML system for Risk Scoring and Underwriting contains two Data Governance layers. The first one removes identifiers and reduces the risk of re-identification as explained by Data Minimization. The Negative Policy assesses information loss. When performance degradation is tolerable, a second Data Governance layer explicitly governs shared P3 Data in a Federated Learning architecture.

III. PRIVACY-PRESERVING TECHNIQUES IN ML

Maintaining Intelligence in a Machine Learning Context While Data Minimization or Sensitive Data Increase Utility is often at the Core of Planning Phase 2 Data Processing, System Developers should Keep in Mind, or Look Up, the Abstract and Introduction. Privacy-preserving machine learn- ing is a technical design that enables modeling with privacy- preserving input while retaining high predictive performance. Two broad categories of privacy-preserving machine learning are federated learning, which allows for modelling while minimizing data-sharing data volume, and secure computa- tion, which creates equivalences between trained models and corresponding plaintext-trained models without sharing the private/training data. Within federated learning in areas where stronger utility is needed, template external collaboration can also occur, across non-collaborating organisation-specific data stores. Models deployed in a collaborative but external way cannot learn from or predict on the sensitive personal data. Feature flows governing privacy-preserving machine learning methods are also useful in the context of privacy- invading-model creation or use, enabling independent addition of privacy-preserving considerations to such parts. Federated Learning & Cross-Institution Collaboration When models are planned for model- or input-sharing purposes—particularly models that also use data from data subjects who are classed as children—data minimization on the model is mapped to k-anonymity. When data on sensitive attributes are collected but combination leaves the records unique, that attribute is a candidate for use in the model but not in sharing or learning. In areas where subjects are somewhat willing to share information, practical limits on these data but not on the model in a risk-, flow- or underwriter-scoring application life cycle, model predictions (target variable) and sensitive attributes can minimisation on data subjects who will supply these data when probed, create a double gamble for the data subjects named in that model's prediction.

A. Federated Learning and Cross-Institution Collaboration

Federation is beneficial when organizations with sensitive data can form an informal alliance, whether periodically or over a more extended period. To improve the granularity or range of a scoring model, for instance, an insurance company might agree to build a model with another company that has access to a different set of customers. No additional customer information should be sent to either organization, but only model updates are disclosed. Risk features such as controllable attributes, event features, or industry exposure can be expected to remain constant over the life of the policy.

Fig. 3. Techniques in Privacy-Preserving Machine Learning

Uncontrollable event features—an emerging fraud trend before a large claim—is a signal that could rupture the federation at the time the information was captured and hurt the disclosure risk more than usual. Other

parties might have the same objectives and, more importantly, redundant information that feeds the model but are less reliable. A third party attempts to detect colluding fraud networks but cannot*detect, identify, or understand*. Non-federated features related to those fraud networks could be useful for the third party and the other two organizations but represent an added risk for the information subjects. The models can be trained more robustly, but the computation must be executed with the weakest link in mind, such as through secure multi-party computation.

B. Secure Computation and Homomorphic Encryption

Secure computation, including secure multi-party computa- tion and homomorphic encryption, enables mutually distrustful parties to jointly compute desired functions over their inputs without revealing the inputs themselves or any additional in- formation. While a variety of functionalities can be represented in normal form for secure computation, the overhead of this approach is too high to apply to the insurance setting where a common sensitive attribute, such as the primary beneficiary status, is sufficient for a trusted third-party setup. A security- preserving service provider hosting an insurance database from multiple institutions can leverage this capability. Institute A can use the sensitive attributes of Institute A clients and/or Institute B clients available to Institute A to build a risk score predictive model against fraudsters who aim to falsely claim insurance benefits. Complex models and ensembles can be

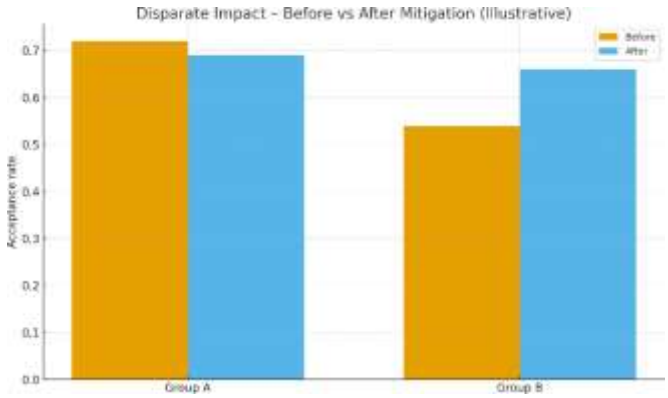

Fig. 4. Disparate Impact – Before vs After Mitigation (Illustrative)

IV. DATA GOVERNANCE AND FEATURE ENGINEERING

Privacy-preserving design needs to align with regulatory and ethical policies. The Data Protection Impact Assessment is a helpful tool. Reducing the use of PII limits disclosure risk and the degree of privacy protection needed. Pinkers, and Timothy, Santillán, and Starck provide guidance on how to mitigate the risk of sensitive attributes, and Cornelius and Nai also offer insights on non-sensitive attributes. These strategies should be considered when planning risk-scoring, underwriting, and fraud detection models. Privacy Data Minimization requires designs to restrict the use of PII data to the smallest quantity possible. Data Minimization should also aim to use identifiers with the least disclosure risk and the lowest frequency of presence in the dataset. Sensitive Data Anonymization aims to remove sensitive information from the training data. Sensitive attributes might have a substantial impact on the performance of smart contracts that process claims or evaluate fraud. For example, Health Indicators, Age Groups, and Claim History Category could strongly influence those models. When employing these sensitive attributes, careful consideration of the effects and a discussion about the omitted signal should be included.

Equation 02: Sensitivity and mechanisms

For a function $f : D \rightarrow R^k$ with ℓ_1-sensitivity employed, as the primary goal is to increase the score. Despite connection-preserving homomorphic encryption being able to calculate any arbitrary polynomial function at the cost of the$\Delta_1 f = \max_{D \sim D'}$ the Laplace mechanism

$$|f(D) - f(D')|_1 \qquad (2)$$

communication complexity, under-sampling can be applied appropriately to keep the communication overhead reasonable. Another approach would be to utilize a symmetric attribute as the index for a trusted third-party system in a cloud- based deployment. For example, an encryption key for the attribute pair (age, claim) that divides the whole data size into several sets can be generated so that these sets are functionally separate from each other. Even under the real-world cloud risk of bi-cryptographic secret discovery, this design can withstand the inside-outside attack model to a certain level.

For ℓ_2-sensitivity $\Delta_2 f$, the Gaussian mechanism with noise $\sigma \geq \epsilon 2 \ln(1.25/\delta)$ achieves (ϵ, δ)-DP.

A. Data Minimization and Anonymization

Although data protection laws do not prescribe an ex- haustive list of appropriate feature variables for modeling applications, compliance with data minimization requirements dictates that the likelihood of indirect identification should be actively reduced in any deployed model. Generalized models aimed at risk scoring or claims management warrant stronger data anonymization than models for scrutiny of individual claims decisions. Nevertheless, the predictive signal present in standard identification attributes should be retained wherever possible, with increasing levels of generalization applied as risk-unaware management requires a summary of the data distribution for a cohort. The modeling application therefore goes beyond mere utility and transforms into a form that is less dependent on specific data distributions of sensitive attributes, a transition that is particularly important for claims decisions and crucial for any form of fairness-aware modeling in a data-minimizing regime. In conventional risk scoring or underwriting models, insurance firms routinely use features associated with the sensitive attributes of persons represented in the data that may be deemed undesirable to any use- case in which the sensitive attributes are properly suppressed. Attributes such as age, health indications and previous claims record are considered strong indicators in pricing policies or detection of fraudulent activities. Such considerations also extend to the specialized use-cases of Fraud Detection and Management. However, it is also acknowledged in these spe- cialized use-cases that exposed sensitive attributes cannot be fully suppressed and used for fair risk scoring for persons that are explicitly represented in the modeling cohort.

TABLE II K-ANONYMITY VS APPROX. RE-ID RISK

k	approx reidentification risk		
1	1		
2	0.5		
3	0.333		
5	0.2		
10	0.1		
20	0.05		
50	0.02		
100	0.01		

B. Sensitive Attribute Handling

Minimizing the amount of data being processed during any workflow is often one of the primary aspects of data preservation in model development. However, some attributes bear direct personal information and should be removed or moved towards a more suitable implementation. Special categories in the GDPR such as “sensitive personal data” or “health data” require more attention than others since their presence increases the need for disclosure of trained or predicted models and expose them to higher scrutiny. In other contexts, the presence of records that unavoidably belong to a single individual should be dealt with using higher θ-differential privacy parameters, stricter k-anonymity definitions, or higher σ-relational anonymity. The insurance sector is no exception. For example, modelling a risk score for an insurance application using a confirmed age—an almost unique identifier in very small segments—or health indicators increases the chances of »model inversion« attacks. Therefore, proper risk mitigation is required. Nonetheless, both parties involved in the scoring process—the data supplier and data consumer—must adopt a balanced approach when it comes to the sensitive attribute category. For instance, the adoption of θ privacy towards health indicators would be an effective risk-compliance solution for insurers, while impacting the profitability of insurers with very few customers that match an insurance policy condition directly linked to the expiration of the insurance.

Fig. 5. Model Architectures for Insurance Use-Cases

V. MODEL ARCHITECTURES FOR INSURANCE USE-CASES

Data-driven solutions for risk scoring, underwriting, fraud detection, claims management, and customer service are now commonplace within the insurance sector. The models, how- ever, possess a single point of failure with respect to customer privacy, as external entities can possess sensitive identifiers or even very detailed information that can then be used to exploit weaknesses. Under potential future regulations, such a disclosure may also lead to useful information being leaked from a model's output or explainability methods. In a risk- scoring or underwriting task, sensitive customer attributes (e.g., health indicators, age, previous claims) are used to make predictions. Modelling must therefore be conducted in a way that minimizes the risk of leakage. If the model is trained on sensitive customer data, k-anonymity must be achieved. If a customer data point is protected through homo- morphic encryption or secure computation, a higher level of k-anonymity and/or a check for differential privacy must be conducted, although a single data point will be replaced by two datapoints sharing the same sensitive information. To further reduce the risk of any sensitive identifiers leaking through explainability methods, privacy-preserving integrated gradients can be computed on the encrypted model outputs. For any post-hoc model, differential privacy must also be checked.

A. Risk Scoring and Underwriting

Approaches to risk scoring and underwriting can retain their predictive power while addressing privacy concerns. As a general principle, identifiers should be reduced even when they are not explicitly protected. Such characteristics as name, zip code, nationality, and date of birth often have little bearing on risk but can dramatically ease re-identification. Other sensitive columns—those for which the insurance company's ability to know but not disclose would be desirable—should be modeled in such a manner that they do not necessarily form part of the data asymmetrically shared across institutions. When FL and horizontal collaboration are not viable, fraud detec- tion frequently can use anonymized or highly K-anonymized attributes; in this context, a genotypified or ancestral-age file may underline — through ML techniques — a label- distributed and thus coarsely privacy-preserving typology. Re- duced details on claims history, implemented as ML-distilled integrated-level distributions, also can add informative value without revealing crucial re-identification anchors.

B. Fraud Detection and Claims Management

Machine learning plays a key role in combating fraud in insurance systems. While the advantages of superior data and better models are clear, most institutions still focus on developing internal solutions, which creates an opportunity for fraudsters to exploit. Combining resources allows the development of better solutions, both to model the normal behaviour of clients and also to identify abnormal transactions that could indicate fraud. A good example of fraud detection that could be developed by multiple institutions that will benefit from each other's data is fraud detection in claims management, an area where private information, such as health-related information, must be carefully handled. Claims management is the process of evaluating whether a claim is legit and estimating the final amount to pay. Detecting fraud in a large number of claims is not easy, but clients usually follow the law in most claims, creating a better understanding of how claims should work. The model checks if the claim is following the pattern expected for open claims. When a claim does not respect the profile of all claims with the same attributes, it indicates that there could be something wrong with it, and it is then sent for a detailed review. A feasible solution for a scenario like this could be the development of a collaborative model that detects anomalies across all claims from different institutions. As health-related attributes are highly sensitive information, homomorphic encryption must be implemented in the model. The institutions send their claims through an additively homomorphic encryption scheme to the secrecy provider, which has no other information to decrypt the claims. The secrecy provider then performs the necessary computations to validate the services and decrypts the output with its secret key. Only the decrypted result is sent back to the institutions. If anything (support services, episode cost, diagnosis, or treatment) is classified as anomalous, the institution triggers human analysis for that claim.

VI. CONCLUSION

Natural language processing (NLP) has undergone rapid development in recent years, owing to the availability of vast amounts of textual information and advances in deep learning methods. First-order logic with quantifiers provides a convenient representation of a large range of applications in NLP and is expressively equivalent to second-order logic, albeit with a different modelling approach. Various problems of inference or decision-making can be mapped onto first-order

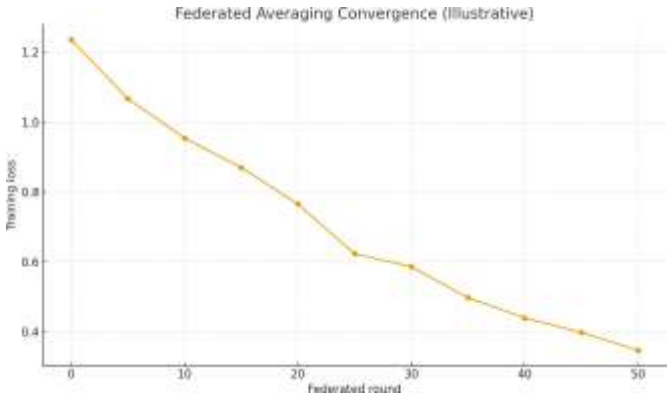

Fig. 6. Federated Averaging Convergence (Illustrative)

logical forms. Research on classes of first-order formulae appears important to NLP research. Such classes include Horn formulae, classes associated with the well-known description logics of knowledge representation, framework perspectives in defence of Web ontologies, and formulae generated by a modal logic associated with belief modelling. In dealing with an informative and open-ended range of questions, an oriented connectionism approach seems especially promising for natural language understanding, addressing the neglected aspect of understanding in the learning-unlearning dichotomy often invoked in artificial intelligence. Future advances in first-order, and its descendants, logical theory will improve their development of methods for NLP, and intelligent systems exploring such methods will hopefully continue to have the benefits of important properties often ignored in deep learning. A variety of other emerging advances in knowledge-based NLP processing are also of a prospective nature. The structural language grammar of Wilks has been developed and increasingly applied in recent years. As with the networks proposed by Rosenblatt and by McCulloch-Pitts, the Barker network has also inspired interest in its conceptual foundation and offers a theoretical link with natural language word order and with other forms of information modelling such as Beckett's structured languages.

Equation 03: HE-based scoring (claims/fraud))
With additively homomorphic scheme E for model vector **w** and feature vector **x**

$$E(w\top x) = j = 1 \oplus dE(wjxj) \qquad (3)$$

A. Future Trends
Amid heightened public awareness of exploitable personal data and complex regulatory requirements, recent years have seen increasing interest in technical solutions that mitigate the risk of sensitive data exposure while allowing its use for valuable analyses. Utilization of models built on privacy- preserving techniques, suitable for the insurance context, can help realize this goal. The designed systems strive to protect privacy while optimizing for multiple objectives.

round	training_loss	
0	1.2353	
5	1.067	
10	0.9541	
15	0.8696	
20	0.7652	
25	0.6228	
30	0.5858	
35	0.4972	
40	0.4394	

TABLE III FEDAVG CONVERGENCE (TRAINING LOSS BY ROUND)

The identified techniques are used together with data governance strategies that minimize the risk of disclosing sensitive data while maximizing the analysis signal. A DPIA informs the definition of a minimum dataset, the idea of working only on anonymized data and the generation of secondary attributes that provide better risk signals. The first privacy-preserving solutions for risk scoring, underwriting, fraud detection, claims manage- ment, and complaints analyses cover the basic insurance use cases that can be modeled with sensitive customer data, focusing on privacy-respecting model architectures. Future work is required to develop more complex use cases, refine the already identified use-case solutions further, and deploy them in a productive system. An industry-ready privacy-preserving insurance model acting primarily as a data processing engine could leverage both the inbound and outbound data flows of insurers to serve multiple

business units at the same time, generating secondary attributes while providing risk scores for different areas.

REFERENCES

[1] Gadi, A. L. The Role Of AI-Driven Predictive Analytics In Automotive R&D: Enhancing Vehicle Performance And Safety.

[2] Altman, E., Blanuša, J., von Niederhäusern, L., Egressy, B., Anghel, A., & Atasu, K. (2023). Realistic synthetic financial transactions for anti-money laundering models. NeurIPS 2023 Datasets and Benchmarks Track.

[3] Lahari Pandiri, "Leveraging AI and Machine Learning for Dynamic Risk Assessment in Auto and Property Insurance Markets," International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering (IJIREEICE), DOI 10.17148/IJIREE- ICE.2023.111212.

[4] Masrom, S., Tarmizi, M. A., Halid, S., Rahman, R. A., Abd Rahman, A. S., & Ibrahim, R. (2023). Machine learning in predicting anti- money laundering compliance with protection motivation theory among professional accountants. International Journal of Advanced and Applied Sciences, 10(7), 48–53.

[5] Nandan, B. P., & Chitta, S. S. (2023). Machine Learning Driven Metrology and Defect Detection in Extreme Ultraviolet (EUV) Lithog- raphy: A Paradigm Shift in Semiconductor Manufacturing. Educational Administration: Theory and Practice, 29 (4), 4555–4568.

[6] National Institute of Standards and Technology (NIST). (2023, May 9). US-UK PETs Prize Challenge.

[7] Koppolu, H. K. R., Sheelam, G. K., & Komaragiri, V. B. (2023). Au- tonomous Telecommunication Networks: The Convergence of Agentic AI and AI-Optimized Hardware. International Journal of Science and Research (IJSR), 12(12), 2253-2270.

[8] Cheng, D., Ye, Y., Xiang, S., Ma, Z., Zhang, Y., & Jiang, C. (2023). Anti-money laundering by group-aware deep graph learning. IEEE Transactions on Knowledge and Data Engineering.

[9] Kalisetty, S., & Singireddy, J. (2023). Agentic AI in Retail: A Paradigm Shift in Autonomous Customer Interaction and Supply Chain Automa- tion. American Advanced Journal for Emerging Disciplinaries (AAJED) ISSN: 3067-4190, 1(1).

[10] Tariq, H., & Hassani, M. (2023). Topology-agnostic detection of tem- poral money-laundering flows in billion-scale transactions (FaSTMAN).

[11] Lakkarasu, P. (2023). Generative AI in Financial Intelligence: Unrav- eling its Potential in Risk Assessment and Compliance. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 241-273.

[12] Egressy, B., Fischer, M., & Müller, N. (2023). Provably powerful graph neural networks for directed graphs with applications to financial crime analysis. arXiv preprint (arXiv:2306.11586).

[13] Kummari, D. N. (2023). Energy Consumption Optimization in Smart Factories Using AI-Based Analytics: Evidence from Automotive Plants. Journal for Reattach Therapy and Development Diversities. https://doi. org/10.53555/jrtdd. v6i10s (2), 3572.

[14] Guembe, B., Azeta, A., Osamor, V., & Ekpo, R. (2023). A federated machine learning approaches for anti-money laundering detection. SSRN/ ICISET 2023 proceedings

[15] Sheelam, G. K. (2023). Adaptive AI Workflows for Edge-to-Cloud Processing in Decentralized Mobile Infrastructure. Journal for Reattach Therapy and Development Diversities. https://doi. org/10.53555/jrtdd. v6i10s (2). 3570ugh Predictive Intelligence.

[16] Zhang, H., Pei, Z., Chang, P., & Zhang, D. (2023). A privacy-preserving hybrid federated learning framework for financial crime detection.

[17] Motamary, S. (2023). Integrating Intelligent BSS Solutions with Edge AI for Real-Time Retail Insights and Analytics. European Advanced Journal for Science & Engineering (EAJSE)-p-ISSN 3050-9696 en e-ISSN 3050-970X, 1(1).

[18] Investopedia (Attarwala, F.). (2023, June 21). Google Cloud launches AI-powered anti-money laundering tool for banks.

[19] Meda, R. (2023). Data Engineering Architectures for Scalable AI in Paint Manufacturing Operations. European Data Science Journal (EDSJ) p-ISSN 3050-9572 en e-ISSN 3050-9580, 1(1).

[20] Jensen, R. I. T., Rægaard, M., Pedersen, L., Engsig-Karup, A. P., & Jullum, M. (2023). A synthetic dataset to benchmark anti-money- laundering methods (SynthAML10). Scientific Data, 10, 620.

[21] Somu, B. (2023). Towards Self-Healing Bank IT Systems: The Emer- gence of Agentic AI in Infrastructure Monitoring and Management. American Advanced Journal for Emerging Disciplinaries (AAJED) ISSN: 3067-4190, 1(1).

[22] Johannessen, F., & Jullum, M. (2023). Finding money launderers using heterogeneous graph neural networks.

[23] Inala, R. Revolutionizing Customer Master Data in Insurance Technol- ogy Platforms: An AI and MDM Architecture Perspective.

[24] Karim, M. R., Hermsen, F., Chala, S. A., de Perthuis, P., & Mandal, A. (2023). Catch me if you can: Semi-supervised graph learning for spotting money laundering.