Proceedings of the 2nd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/97/20241388

Privacy by Design in Machine Learning Data Collection: An Experiment on Enhancing User Experience

Yang Xu[1,5,*,†], Yingchia Liu[2,†], Jiang Wu[3], Xiaoan Zhan[4]

[1]Interactive Telecommunications Program, New York University, NY, USA
[2]Parsons School of Design, MFA Design and Technology, NY, USA
[3]Computer Science, University of Southern California, Los Angeles, CA, USA
[4]Electrical Engineering, New York University, NY, USA

[5]rexcarry036@gmail.com
*corresponding author
[†]Both authors contributed equally to this work and are co-first authors.

Abstract. With the development of artificial intelligence (AI) technology, predictive analytics and data-driven decision making are becoming increasingly common across industries. While the widespread use of AI has brought many conveniences, it has also led to profound discussions about user privacy and data protection. This study aims to explore the application of privacy protection design principles to machine learning data collection and its impact on user experience. The study found that embedding privacy protection measures early in system design can significantly reduce privacy management vulnerabilities and enhance users' trust and satisfaction with the system. This study provides theoretical support and practical guidance for privacy protection in machine learning data analysis and provides a feasible design framework for future user experience optimization.

Keywords: Artificial intelligence, Privacy data design, User experience, UI/UX.

1. Introduction

The terms data protection and data privacy are often used interchangeably, but there are important differences between the two. Data privacy defines who has access to data, while data protection provides tools and policies to actually restrict access to data. Compliance regulations help ensure that users' privacy requests are carried out by companies, which are responsible for taking steps to protect private user data. Data protection and privacy generally apply to Personal Health information (PHI) and personally Identifiable Information (PII)[1].

While predictive analytics of AI as a powerful tool has many potential social and economic benefits, its current state of abuse and lack of regulation also poses significant risks. In order to maximize the positive effects of predictive analytics while reducing its negative impacts, the legal regulation of these technologies needs to be strengthened to ensure that they meet ethical standards in terms of data protection and anti-discrimination [2]. In addition, attention should be paid to the social inequalities that predictive analytics can generate, and measures should be taken to promote fairness and transparency in data-driven decision-making.

DOI: 10.54254/2755-2721/97/20241388

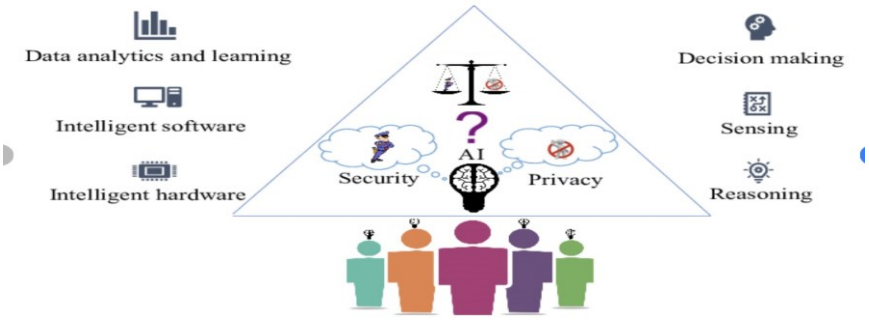

Figure 1. Relational architecture of privacy protection and machine learning data analysis

Predictive analytics is conducted according to the principle of "pattern matching," which, through learning algorithms, compares known auxiliary data (such as social media usage data, browsing history, geolocation data) of a target individual with data from thousands of other users. Therefore, in today's data-driven digital environment, the relationship between privacy protection and machine learning data analysis and user experience design is becoming increasingly important, and as the design and user experience of Internet interactive pages becomes a large part of the key privacy. With the widespread application of machine learning technology, the platform can optimize UI/UX [3] design by analyzing user behavior data to improve the user experience.

2. Related Work

2.1. Privacy interactions in human-machine design

When designing the system, it is crucial to have privacy protection as a core consideration. Failure to embed privacy protections early in system development often leads to poor privacy management, creating a gap between user privacy and actual privacy controls (Anthonyamy, Greenwood, and Rashid, 2013). The government regulates the use and management of personal information through various guidelines and laws, such as the Australian Privacy Act (APPs)[4]. This requires systems to be designed not only to meet privacy protection requirements, but also to ensure the best possible user experience.

In organizations such as banks, academic institutions, businesses, and government departments, information collection can have serious privacy concerns. For example, it is unethical to fraudulently collect personal information, even if the information is used for analytical and reporting purposes, and institutions must evaluate its ethics (Turilli and Floridi, 2009). Therefore, these organizations need to implement fully transparent mechanisms in the information collection process to ensure that users can make informed choices and reduce the risk of privacy disclosure.

2.2. User Experience Design and Machine Learning

User Experience Design (UXD) aims to create interactive experiences that meet user needs, and a human-centered design approach focuses on user expectations rather than relying solely on technical possibilities. David Benyon's User Experience Design argues that systems should be designed with human emotions and diversity in mind, not just technical possibilities. Rule-based systems, while efficient at handling specific tasks, rely on static sets of rules, which are often defined by experienced experts[5]. Therefore, when using machine learning models, designers still need to pay attention to the explanatory power of the model and the understanding experience of the user.

2.3. Implement privacy protections in machine learning data collection

Machine learning offers a unique opportunity to monitor user sentiment and satisfaction across a wide range of domains, including industrial, civil, military, and social media [6]. With this technology, user feedback can be obtained from a variety of platforms to make more informed decisions based on

DOI: 10.54254/2755-2721/97/20241388

experience and opinions. Figure 1 shows a sample of QOUEs on a typical platform that helps to understand user feedback and reviews.

Privacy protection is a key issue in machine learning data collection. With the advancement of data collection and analysis technology, the protection of users' personal information has become particularly important. By implementing data de-identification and anonymization techniques, researchers can use the data for model training without exposing the user's identity. In addition, the adoption of encryption technologies and privacy-protecting algorithms, such as differential privacy and secure multi-party computing, can further ensure the security of user data and prevent unauthorized data access and disclosure.

3. Methodology

In this study, we propose an innovative approach to evaluating user Privacy by Design (PbD) that aims to address the limitations that exist in traditional data collection methods. we first propose an innovative approach to User Experience (UX) evaluation by addressing the limitations inherent in traditional UX evaluation (UXE) methodologies. This methodology aims to gain insight into the overall impact of user behavior sequences on privacy protection and provide valuable insights for improving privacy design and user experience.

3.1. Dataset

In our online shopping evaluation experiment, users were asked to visit Web pages A, B, C, and D in a fixed order. However, this fixed order design fails to truly reflect the real user experience, because actual users are free to choose the order of accessing the pages according to their personal needs[7]."

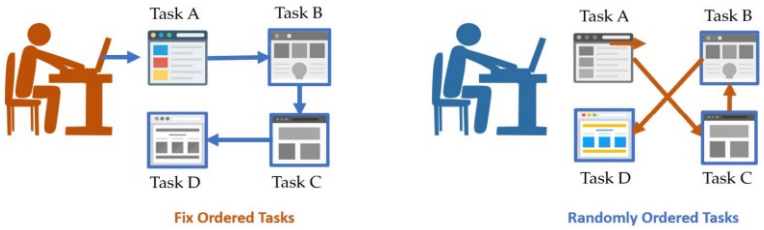

Figure 2. comparison of fixed and random order dual tasks for online shopping sites Use.

To better simulate real-world user behavior, our experimental dataset includes UX data collected when users freely choose the order in which they visit web pages. This random and ordered data set can more accurately reflect the behavior patterns of users when they actually use the product, thus improving the accuracy of user satisfaction predictions. During the construction of the dataset, we paid special attention to how to handle user interactions in this random order to ensure that the research results truly and effectively reflect the user experience.

3.2. Experiments

At present, many experiments are targeted, and the data privacy and protection design of products and services when users perform fixed tasks are carried out. Therefore, in this study, we conducted privacy, protection design, and user experience by disturbing the order of task data in fixed-order experiments[8-9]. The purpose of these experiments was mainly to analyze the real results of fixed order and random order. Three experiments were mainly implemented in the study. First, the main experiment users were carried out to perform all the tasks in a random order, and the operation order and main facilities were recorded.

We conducted preliminary experiments to determine the impact of task order on the user experience of privacy-protected design and how it affects the ultimate user satisfaction. To do this, we simulated the user experience by scrambling the order of tasks in our previous experiments. (preliminary experiment I) or a Google Nest Mini smart speaker (preliminary Experiment II). Responses are used to

DOI: 10.54254/2755-2721/97/20241388

draw customized user experience curves. Participants were also asked to complete a final satisfaction questionnaire to assess the privacy design of the product or service.

3.3. Preliminary experiment I: Travel agency website

In preliminary experiment 1, we used user data from a review study that asked participants to be evaluated by tracking a travel agency website[10]. In the experiment, users were asked to answer seven satisfaction survey questions about travel agency websites. And before the mission, they need to be reminded in advance that our goal is to find the place they want to go in their life through this website. All the participants completed the task conscientiously.

Dataset W1: Shuffled order
Shuffled order of tasks
(Satisfaction scores were shuffled)

	Shuffled order of tasks			
User01	−1	2	0	...
User02	−1	0	1	...
User03	−2	0	−1	...

Dataset W2: Fixed Order
Actual order of tasks
(Satisfaction scores were sorted by actual order of fixed tasks)

	Actual order of tasks			
User01	2 →	−1 →	0 →	...
User02	−1 →	1 →	0 →	...
User03	0 →	−2 →	−1 →	...

(→ = Next Task)

Figure 3. Example of the structure in each dataset

There were 50 participants in this experiment, and the steps they needed to complete the task and achieve the goal were six respectively. First, a curve conforming to user experience was developed for each participant. The experimental results show that the random sequence data sets can more truly reflect the user in the actual use of the privacy protection in the process of behavior, compared with the traditional fixed order method, provides a more accurate prediction of the user experience.

4. Conclusion

This study explores how privacy protection design principles can be effectively incorporated into the machine learning data collection process to optimize the user experience. Through systematic analysis, the study shows how to find a balance between user privacy protection and data analysis to improve user experience. The experimental results show that combining privacy protection technologies such as data de-identification, encryption, and differential privacy can significantly improve user experience. These technologies ensure that user data is effectively protected during collection and analysis, thereby reducing the risk of privacy breaches while enhancing user trust and satisfaction with the system. Through the comprehensive application of privacy protection technology and user experience optimization strategies in data analysis, this study not only promotes the theoretical development of privacy protection, but also provides a feasible framework for the design and implementation of practical applications, thus contributing to the protection of user privacy and the improvement of user experience in a data-driven digital environment.

References

[1] Abbas, Abdallah MH, Khairil Imran Ghauth, and Choo-Yee Ting. "User experience design using machine learning: a systematic review." IEEE Access 10 (2022): 51501-51514.

[2] Yu, K., Bao, Q., Xu, H., Cao, G., & Xia, S. (2024). An Extreme Learning Machine Stock Price Prediction Algorithm Based on the Optimisation of the Crown Porcupine Optimisation Algorithm with an Adaptive Bandwidth Kernel Function Density Estimation Algorithm.

[3] Carmona, K., Finley, E., & Li, M. (2018). The relationship between user experience and machine learning. Available at SSRN 3173932.

DOI: 10.54254/2755-2721/97/20241388

[4] Li A, Zhuang S, Yang T, Lu W, Xu J. Optimization of logistics cargo tracking and transportation efficiency based on data science deep learning models. Applied and Computational Engineering. 2024 Jul 8;69:71-7.

[5] Nwakanma, C. I., Hossain, M. S., Lee, J. M., & Kim, D. S. (2020). Towards machine learning based analysis of quality of user experience (QoUE). International Journal of Machine Learning and Computing, 10(6), 752-758.

[6] Xu, J., Yang, T., Zhuang, S., Li, H. and Lu, W., 2024. AI-based financial transaction monitoring and fraud prevention with behaviour prediction. Applied and Computational Engineering, 77, pp.218-224.

[7] Majd, M., & Safabakhsh, R. (2017, October). Impact of machine learning on improvement of user experience in museums. In 2017 Artificial Intelligence and Signal Processing Conference (AISP) (pp. 195-200). IEEE.

[8] Carmona, K., Finley, E., & Li, M. (2018). The relationship between user experience and machine learning. Available at SSRN 3173932.

[9] Ling, Z., Xin, Q., Lin, Y., Su, G. and Shui, Z., 2024. Optimization of autonomous driving image detection based on RFAConv and triplet attention. Applied and Computational Engineering, 77, pp.210-217.

[10] Shah, A., & Nasnodkar, S. (2021). The Impacts of User Experience Metrics on Click-Through Rate (CTR) in Digital Advertising: A Machine Learning Approach. Sage Science Review of Applied Machine Learning, 4(1), 27-44.