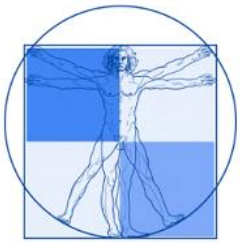

Web3 Journal: ML in Health Science

ISSN 2942-8726

2024. Vol. 1(1)

Original Article

TxnHash: 0x7db71354054baa52f4547cbad10e145a3e91fbc5e06c8a9f68f33d02e08dbdf1

DOI: 10.62487/yyx99243

Do You Consent to the Use of Your Biological Data for Training ML and AI Models? Online Survey Targeting Clinicians and Researchers.

Yury Rusinovich[1], Volha Rusinovich[1]

Abstract.

Aim: The majority of machine learning (ML) models in healthcare are built on retrospective data, much of which is collected without explicit patient consent for use in artificial intelligence (AI) and ML applications. The primary aim of this study was to evaluate whether clinicians and scientific researchers themselves consent to provide their own data for the training of ML models. **Materials and Methods:** This survey was conducted through an anonymous online survey, utilizing platforms such as Telegram, LinkedIn, and Viber. The target audience comprised specific international groups, primarily Russian, German, and English-speaking, of clinicians and scientific researchers. These participants ranged in their levels of expertise and experience, from beginners to veterans. The survey centered on a singular, pivotal question: "Do You Consent to the Use of Your Biological and Private Data for Training Machine Learning and AI Models?" Respondents had the option to choose from three responses: "Yes" and "No". **Results:** The survey was conducted in January 2024. A total of 119 unique and verified individuals participated in the survey. The results revealed that only 50% of respondents (63 persons) expressed consent to provide their own data for the training of ML and AI models. **Conclusion:** In the development of ML and AI models, particularly open-source ones, it is crucial to ascertain whether participants are willing to provide their private data. While ML algorithms can transform the nature of data, it is important to remember that the primary owner of this data is the individual. Our findings show that in 50% of the cases, even participants from scientific research and clinical backgrounds – individuals typically accountable for ensuring data quality in AI and ML model development – do not consent to the use of their data in AI and ML settings. This highlights the need for more stringent consent processes and ethical considerations in the utilization of personal data in AI and ML research.

Keywords: Human-Centered AI, AI Supervision, Trust in AI, Human-AI Collaboration, Healthcare Survey

Background:

The majority of machine learning (ML) models in healthcare are built on retrospective data, much of which is gathered without explicit patient consent for its use in artificial intelligence (AI) and ML applications. These models often utilize highly confidential information, such as disease patterns, age, place of birth, and biological markers. The reliance on retrospective data in health science research generally stems from a broad patient consent that permits the use of their data for research purposes. However, specific consent regarding the use of this data for ML and AI training was not typically obtained, largely due to the data being collected before the surge in AI and ML technologies. It is noteworthy that even today, explicit patient consent for the use of their data in AI and ML research is not always obtained.

[1]ML in Health Science, Leipzig, Germany

Corresponding author: Yury Rusinovich

Email: info@mlhs.ink

Hypothesis

In a 2023 study conducted in the UK, it was found that only 25% of patients consented to provide their data for commercial AI/ML research, whereas 78% were agreeable to their data being used for university-led research[1]. This disparity highlights a relatively low acceptance rate for commercial use, raising questions about the public's trust and willingness to share personal data for profit-driven endeavors. However, this leads to an intriguing question: Are clinicians and scientific researchers themselves willing to provide their own data for the purpose of AI and ML training?

Aim

To understand the root of this issue, we conducted the present research. The primary aim of this study was to assess whether clinicians and scientific researchers themselves would consent to provide their personal data for the training of AI and ML models.

Material and Methods:

This survey was conducted through an anonymous online survey, utilizing platforms such as Telegram, LinkedIn, and Viber. The target audience comprised specific international groups, primarily Russian, German, and English-speaking, of clinicians and scientific researchers. These participants ranged in their levels of expertise and experience, from beginners to veterans. The survey centered on a singular, pivotal question: “Do You Consent to the Use of Your Biological and Private Data for Training Open-Source Machine Learning and AI Models?” Respondents had the option to choose from three responses: “Yes” and “No”.

The methods of this survey are openly accessible on the official Telegram Channel of the Web3 Society: ML in Health Science, which can be visited at: https://t.me/MLinHS

Statistics

The data were analyzed with descriptive statistics.

Results:

The survey was conducted in January 2024. A total of 119 unique and verified individuals participated in the survey. The results revealed that only 50% of respondents (63 persons) expressed consent to provide their own data for the training of ML and AI models.

Table 1 and **Figure 1** summarize the survey results:

Variable	Respondents
Yes	63
No	56
Total	119

Table 1: Survey Results

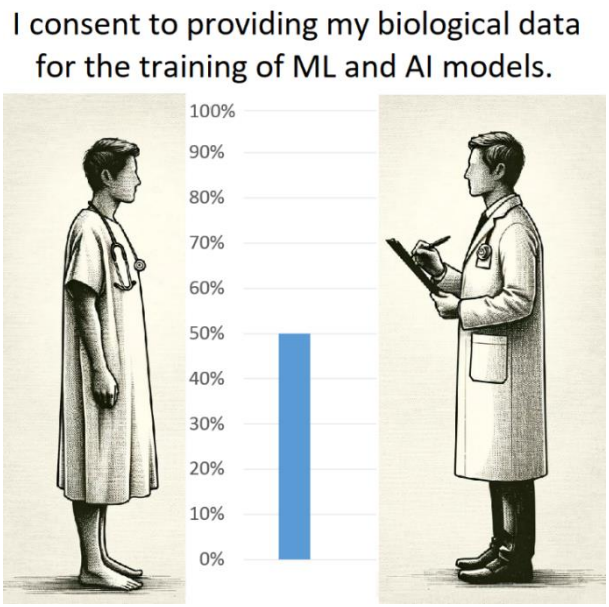

Figure 1: Survey Results

Discussion:

Practical standpoint

Our study reveals a notably low willingness among clinicians and scientific researchers to provide their personal data for AI and ML research. Interestingly, this acceptance rate is even lower than the generally reported willingness to participate in conventional clinical research, such as tissue banking and genetic studies[2]. A possible explanation for this finding is that data donors in AI and ML projects may experience apprehension about being merged into a mass dataset or fears regarding AI technology, such as loss of privacy and individuality, or feeling like a part of a global experiment.

Further research is needed to explore the sentiments and psychological behaviors of patients, clinicians, and scientific researchers in the context of ML and AI model building and training.

From a practical standpoint, it is crucial for clinicians and scientific researchers to acknowledge this low acceptance rate during model building. Particularly, models based on retrospective data should be mindful of this fact and strive to avoid using sensitive and private patient information, such as place of birth, living conditions, occupation, social status, or biological patterns, in their development process.

To the best of our knowledge, this is the first study to evaluate the willingness of clinicians and scientific researchers to provide their own data for the building and training of AI and ML models.

Limitations

The study employed an anonymous approach, which could potentially lead to a lack of purity in the cohort. Another limitation is the relatively limited size of the dataset.

Conclusion

In the development of ML and AI models, particularly open-source ones, it is crucial to ascertain whether participants are willing to provide their private data. While ML algorithms can transform the nature of data, it is important to remember that the primary owner of this data is the individual. Our findings show that in 50% of the cases, even participants from scientific research and clinical backgrounds – individuals typically accountable for ensuring data quality in AI and ML model development – do not consent to the use of their data in AI and ML settings. This highlights the need for more stringent consent processes and ethical considerations in the utilization of personal data in AI and ML research.

Conflict of Interest: YR and VR state that no conflict of interest exists.

Authorship: YR: Concept, data analysis, original draft, survey. YR, VR: Review and editing.

References

1 **Aggarwal R, Farag S, Martin G, Ashrafian H, Darzi A. Patient Perceptions on Data Sharing and Applying Artificial Intelligence to Health Care Data: Cross-sectional Survey. J Med Internet Res. 2021 Aug 26;23(8):e26162. doi: 10.2196/26162.**

2 **Moorcraft SY, Marriott C, Peckitt C, Cunningham D, Chau I, Starling N, Watkins D, Rao S. Patients' willingness to participate in clinical trials and their views on aspects of cancer research: results of a prospective patient survey. Trials. 2016 Jan 9;17:17. doi: 10.1186/s13063-015-1105-3.**