TrustBCI: A Trustworthy Platform for Collaborative Brain-Computer Interface Data Exchange

Jianyu Yang
Tsinghua University
jyyang@mail.tsinghua.edu.cn

Jiannan Wang
Tsinghua University
jnwang@tsinghua.edu.cn

Guoliang Li
Tsinghua University
liguoliang@tsinghua.edu.cn

ABSTRACT

High-quality brain-computer interface (BCI) datasets are essential for developing accurate and generalizable BCI models, but in practice they are often isolated across institutions, heterogeneous in format, and difficult to share under privacy constraints. Existing BCI data platforms address only part of this problem, and do not provide unified support for data contribution, privacy-preserving computation, and downstream use. To this end, we present TrustBCI, a trustworthy platform for collaborative BCI data exchange. TrustBCI combines three key components: assetization tools for preparing raw BCI datasets, incentive mechanisms for sustained data contribution, and privacy-aware services for using protected data without exposing raw records. Our demonstration focuses on two representative workflows: controllable data synthesis and privacy-preserving data query. These workflows show how TrustBCI supports both dataset-level and query-level access to protected BCI data through a functional web prototype. More broadly, TrustBCI illustrates how protected domain data can be turned into usable data services through a unified data platform.

PVLDB Reference Format:
Jianyu Yang, Jiannan Wang, and Guoliang Li. TrustBCI: A Trustworthy Platform for Collaborative Brain-Computer Interface Data Exchange. PVLDB, 19(12): 4854 - 4857, 2026.
doi:10.14778/3827998.3828139

PVLDB Artifact Availability:
The source code, data, and other artifacts have been made available at https://github.com/yangjianyu26/trustbci.

1 INTRODUCTION

Brain-computer interface (BCI) technology is advancing rapidly in areas such as neural rehabilitation, cognitive analysis, and human-computer interaction. A key driver of this progress is access to high-quality BCI data, including neural signals, medical images, and multimodal annotations. Such data are important for developing accurate and generalizable BCI models. In practice, however, BCI datasets are often isolated across institutions, stored in different formats, and subject to strict privacy constraints, which makes cross-institutional sharing and reuse difficult. A trustworthy platform for BCI data exchange is therefore increasingly needed.

Proceedings of the VLDB Endowment, Vol. 19, No. 12 ISSN 2150-8097.
doi:10.14778/3827998.3828139

Table 1: Comparison of Existing BCI Data Platforms

Name	Modalities	Assetization	Incentive	Private Computation
OpenNeuro [1]	Multi-modal	✓	✓	✗
PhysioNet [2]	Multi-modal	✗	✓	✗
MetaBCI [6]	Single-modal	✓	✗	✗
TrustBCI	Multi-modal	✓	✓	✓

Building such a platform is challenging for three reasons. First, raw BCI datasets are often not ready for use in their original form. They usually need to be transformed into standardized data assets with clear metadata and usage policies. Second, sustained data sharing is difficult in practice. High-quality datasets are costly to produce, so contributors need clear returns for continued participation. Third, data consumers need useful services over protected data, while direct access to raw BCI data may create substantial privacy risks. Existing BCI data platforms [1, 2, 6] and related data management systems [4, 5, 7, 9–11] address only part of this problem, such as domain-specific data management, and do not provide unified support for data contribution, privacy-preserving computation, and downstream use (see Table 1 for a detailed comparison).

To address these challenges, we present TrustBCI, a trustworthy platform for collaborative BCI data exchange. TrustBCI is organized around three roles: the platform, contributors, and consumers. The platform provides the infrastructure for managing data assets, controlling access, and executing services. For contributors, TrustBCI helps turn raw BCI datasets into standardized assets and provides incentive mechanisms that encourage continued participation. For consumers, it offers privacy-aware services that support useful access to protected data without exposing raw sensitive records. In this way, TrustBCI supports data preparation, sustained contribution, and protected data access within a unified platform.

This demonstration focuses on two representative workflows in TrustBCI because they capture two common ways of using protected BCI data. The first is *controllable data synthesis*, where the platform generates a synthetic dataset under a user-specified privacy setting. This workflow shows how TrustBCI supports dataset-level access without exposing the original data. The second is *privacy-preserving data query*, where the platform executes an SQL-style query and returns privacy-protected results. This workflow shows how TrustBCI supports interactive access to protected data through controlled computation. Together, these two workflows illustrate how TrustBCI turns protected BCI datasets into usable data services under privacy constraints.

The demo is based on a functional web prototype. Attendees can walk through dataset upload, privacy-level selection, synthetic data generation, query submission, and protected result return. The demonstration shows not only the overall design of TrustBCI, but

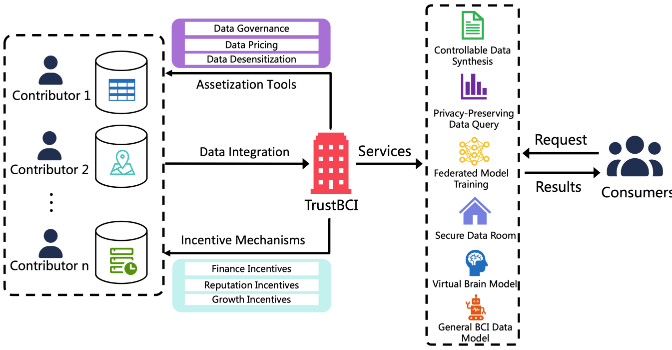

Figure 1: Overview of `TrustBCI`

also how its main components support concrete usage scenarios. The online prototype is publicly available at the project website[1].

In summary, this demo paper makes three contributions. It presents `TrustBCI`, a platform for collaborative BCI data exchange; it shows how the platform supports protected data use through a unified service model; and it demonstrates these ideas through two interactive workflows based on a functional web prototype.

2 SYSTEM OVERVIEW OF TRUSTBCI

This section presents an overview of `TrustBCI`. We first describe the platform architecture and the interactions among the platform, contributors, and consumers. We then introduce three key parts of the system: assetization tools for preparing raw BCI datasets, incentive mechanisms for sustained data contribution, and privacy-aware services for using protected data.

2.1 Architecture Overview

Figure 1 illustrates the overall architecture of `TrustBCI`. The platform involves three roles:

- **`TrustBCI`** acts as the trusted platform that manages data assets, enforces access control, and executes service requests.
- **Contributors** provide raw BCI datasets, such as EEG, ECG, EOG recordings, and brain imaging data.
- **Consumers** invoke platform services to use these data assets under controlled access policies.

Specifically, `TrustBCI` provides contributors with assetization tools, including data governance, pricing, and desensitization. Contributors may download these tools offline or use online services to process their data. After that, contributors upload the processed data to the platform. Moreover, `TrustBCI` designs finance, reputation, and growth incentive mechanisms to encourage contributors to share more data. We will describe the data assetization tools and incentive mechanisms in Section 2.2 and Section 2.3, respectively.

For consumers, `TrustBCI` offers six typical services, including controllable data synthesis, privacy-preserving data query, federated model training, secure data room, virtual brain model, and general BCI data model. These services involve charging consumers based on their actual data utilization, rather than a flat fee. Upon receiving service requests from consumers, the platform executes corresponding computation jobs and returns the results. We will discuss these usage-based services in detail in Section 2.4.

[1]https://yangjianyu26.github.io/trustbci

2.2 Assetization Tools

Before contributors can share BCI datasets through `TrustBCI`, raw data often need further processing to become usable platform assets. A typical dataset may contain multichannel signal records, task annotations, and subject-related identifiers. For example, an EEG dataset may include timestamped channel values together with labels such as motor imagery and fields such as subject identifiers. In practice, such datasets may differ in file format, schema, metadata completeness, and privacy sensitivity. To address this issue, `TrustBCI` provides three assetization tools that support dataset preparation before upload: data governance, data pricing, and data desensitization.

Data Governance. This tool standardizes raw BCI datasets before they are uploaded to the platform. For signal data, it normalizes schema elements such as channel names, sampling timestamps, label fields, and subject-level metadata. For multimodal datasets, it aligns metadata across signal, image, and annotation files. This step improves dataset consistency and makes downstream services, such as querying, synthesis, and model training, easier to support.

Data Pricing. This tool estimates a suggested usage price for each uploaded dataset. The estimate considers dataset characteristics that are particularly relevant to BCI applications, such as modality type, recording scale, annotation quality, signal quality, and task specificity. This provides a basis for usage-based charging and for allocating rewards to contributors when their datasets are used by downstream services.

Data Desensitization. This tool reduces privacy risks in raw BCI datasets before upload. It detects and removes sensitive fields such as subject identifiers, device identifiers, and contact-related information. It also supports anonymization methods such as k-anonymity [8] for structured metadata. For example, in the dataset preview of Figure 2, the synthetic output removes the subject_id field while retaining the signal content and task label needed for downstream use.

2.3 Incentive Mechanisms

High-quality BCI data contribution is costly and difficult to sustain in practice. Compared with general tabular data, BCI datasets often require subject recruitment, repeated recording sessions, expert annotation, and careful metadata curation before they become reusable on a shared platform. This makes incentive alignment a central challenge for `TrustBCI`. To address it, `TrustBCI` designs three complementary incentive mechanisms, namely finance incentives, reputation incentives, and growth incentives.

Finance Incentives. Finance incentives provide direct rewards for datasets that are actually used on the platform. In `TrustBCI`, contributor return is tied to downstream usage rather than dataset upload alone. For example, when a dataset is used by controllable data synthesis, privacy-preserving data query, or federated model training, the corresponding usage is recorded and mapped to contributor rewards. This mechanism addresses the practical challenge that BCI data collection is expensive, while one-time data upload alone does not provide sufficient return for sustained sharing.

Reputation Incentives. Reputation incentives distinguish contributors based on the quality and usefulness of the BCI assets they provide, as reflected in contribution records that cover modality coverage, annotation completeness, signal quality, task diversity, and usage statistics. These records serve as a platform-level signal of dataset reliability and reusability, which enables the platform to award honorary certifications that enhance contributors' industry influence, thereby helping consumers identify well-curated datasets in a highly heterogeneous BCI landscape.

Growth Incentives. Growth incentives encourage long-term contribution instead of isolated uploads. `TrustBCI` accumulates contributor credits based on repeated contributions and continued dataset usage, and uses these credits to determine contributor status and platform benefits, such as reduced service fees or prioritized access to selected services. This mechanism addresses the challenge that a trustworthy BCI platform depends not only on initial data contribution but also on continued dataset maintenance, refinement, and expansion over time.

2.4 Services

`TrustBCI` provides six services over protected BCI data assets. These services are designed for common downstream uses of BCI data while avoiding direct release of raw sensitive records. They support different access patterns, including dataset-level access through controlled transformation, query-level access through protected computation, and model-level access through collaborative training or deployment.

Controllable Data Synthesis. This service generates a synthetic dataset from a protected BCI dataset under a user-specified privacy setting. The generated dataset preserves the schema and key statistical properties of the source data while avoiding direct exposure of the original records. This service supports dataset-level use when consumers need data-like inputs for downstream development but raw data cannot be released.

Privacy-Preserving Data Query. It allows consumers to submit data query statements (e.g., SQL) to `TrustBCI`. The platform executes the query to compute results and then applies privacy-enhancing techniques [3, 8] to sanitize them. Only the sanitized outputs are returned to the consumers, which ensures that sensitive information is protected throughout the process. The service is designed for fast response, making it well-suited for real-time or ad-hoc query scenarios.

Federated Model Training. In this service, consumers initiate model training by uploading its configurations to `TrustBCI`. Then, the platform coordinates multiple contributors to jointly train the model, with each contributor sharing only necessary updates rather than raw data. Thus, this service preserves data privacy while leveraging diverse datasets, making it especially suitable for building generalizable models without compromising data ownership.

Secure Data Room. It provides a privacy-preserving computing environment built on Trusted Execution Environment (TEE) technology that enables consumers to upload executable code for custom analysis without direct access to raw data. `TrustBCI` runs the submitted code on corresponding BCI data within TEE, and returns only privacy-protected results.

Virtual Brain Model. It allows consumers to upload a research subject' brain state along with relevant parameters to `TrustBCI`. The platform uses this information to train a virtual brain model, and produces a digital twin that helps consumers better understand the subject's brain. For instance, a neurologist can submit the clinical information of a patient with Parkinson's disease to simulate the patient's brain, which assists in diagnosis and treatment.

General BCI Data Model. This service provides a specialized large language model (LLM) tailored to the BCI domain, which is fine-tuned from a foundational LLM using the BCI data aggregated by `TrustBCI`. The model is offered as a service, allowing consumers to utilize it online or download it for local deployment.

3 DEMONSTRATION

We demonstrate `TrustBCI` through two representative workflows for protected BCI data access. The first, controllable data synthesis, supports dataset-level access by generating a privacy-controlled synthetic dataset instead of releasing the original records. The second, privacy-preserving data query, supports query-level access by executing SQL-style analysis over protected data and returning sanitized results. These workflows show how `TrustBCI` supports different access patterns over protected BCI data.

3.1 Controllable Data Synthesis

Figure 2 presents the controllable data synthesis workflow. This scenario highlights two key aspects of `TrustBCI`: first, the platform associates dataset contribution with explicit incentives; second, it enables dataset-level use of protected BCI data through synthetic data generation under adjustable differential privacy guarantees. Attendees can observe how a contributed dataset is uploaded, how a consumer specifies the desired privacy level, and how the platform returns a synthetic dataset that preserves utility-relevant structure while reducing privacy risk. The workflow consists of three steps:

Step 1: Data Contribution. A contributor uploads a dataset named "bci_dataset_1" to the platform. After the upload is completed, `TrustBCI` issues a finance incentive in the form of compute vouchers, simulated as 200 credits. This step demonstrates that the platform does not treat contribution as passive storage only; instead, contribution is explicitly connected to incentive mechanisms.

Step 2: Generation Request. A consumer selects the uploaded dataset and requests a synthetic version by choosing one of three privacy levels: *High*, *Medium*, or *Low*. These levels correspond to different differential privacy settings:

- High: $\epsilon = 0.5$, strong privacy with more noise injection.
- Medium: $\epsilon = 2.0$, trade-off between privacy and utility.
- Low: $\epsilon = 5.0$, high fidelity with less noise.

This step makes the privacy–utility trade-off explicit in the user interface and shows how `TrustBCI` exposes privacy control.

Step 3: Synthesizing Dataset. `TrustBCI` generates a synthetic dataset named "bci_dataset_1-syn" and returns it to the consumer. The prototype presents sample records from both the original and synthetic datasets, allowing attendees to directly observe that sensitive identifiers are removed while the schema and signal-like value patterns are preserved.

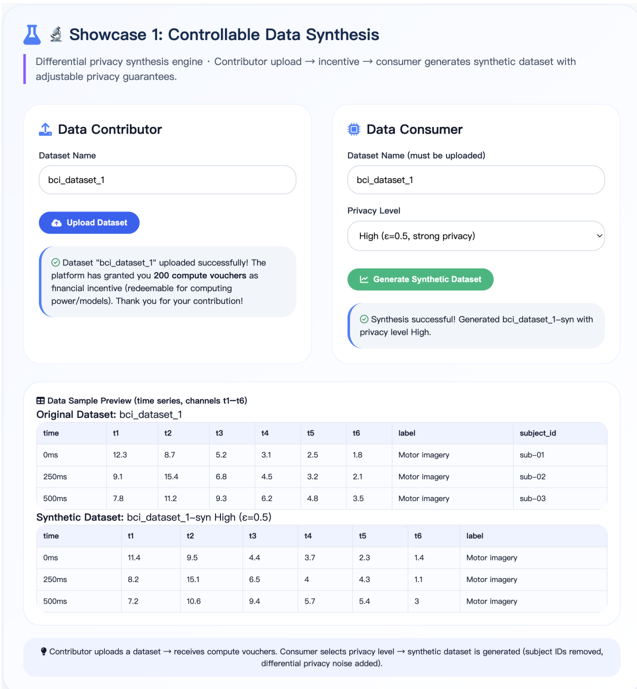

Figure 2: Controllable Data Synthesis

For VLDB attendees, this workflow provides a concrete experience of dataset-level access to protected BCI data without raw-data release: attendees can directly observe how adjustable privacy guarantees affect the returned synthetic dataset.

3.2 Privacy-Preserving Data Query

Figure 3 presents the privacy-preserving data query workflow. It shows how `TrustBCI` supports interactive analysis over protected BCI data through SQL queries. Instead of releasing record-level data, the platform executes the query on the protected dataset and returns a sanitized aggregate result. For demonstration purposes, the prototype also shows the true result together with the noisy result, allowing attendees to directly observe the privacy effect and the associated utility loss. The workflow consists of three steps:

Step 1: Dataset Selection. A consumer specifies a previously uploaded dataset (e.g., "bci_dataset_1"). The platform first checks whether the dataset exists in the repository. If not, it returns an error message. This step shows that queries are issued over governed platform assets rather than over arbitrary external files.

Step 2: Query Formulation. The consumer submits an aggregate query for statistical analysis. In the example shown in Figure 3, the query counts records in "bci_dataset_1" satisfying a condition on the label field. This step demonstrates that `TrustBCI` supports a familiar database-style interface for exploring protected BCI data.

Step 3: Private Result Generation. `TrustBCI` computes the true count that satisfies the SQL conditions and then adds Laplace noise to produce the final released result. In the prototype, both the true count and the noisy count are displayed for verification. This step demonstrates how the platform supports query-level access to protected BCI data while preventing direct exposure of raw records.

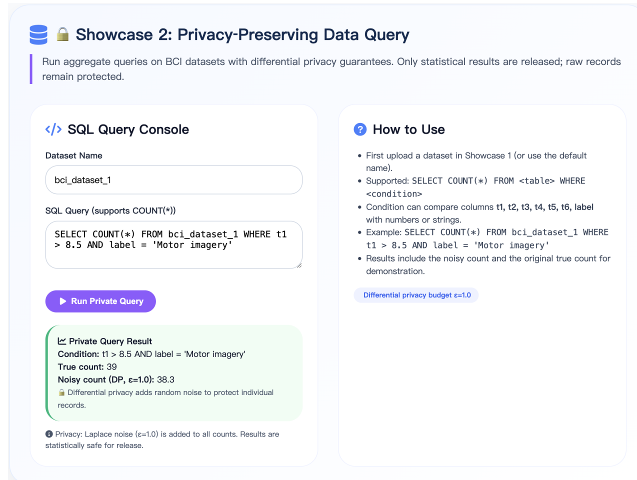

Figure 3: Privacy-Preserving Data Query

For VLDB attendees, this workflow provides a concrete experience of interactive querying over protected BCI data: attendees can issue SQL-style queries and see that the platform executes them over protected data while returning only privacy-protected results.

ACKNOWLEDGMENTS

Guoliang Li is the corresponding author. This paper was supported by National Key R&D Program of China (2023YFB4503600), NSF of China (62525202, 62232009), Postdoctoral Fellowship Program of CPSF (GZC20231297), Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (JYB2025XDXM509), Shenzhen (CJGJZD20230724093403007), Science and Technology Research and Development Plan of China Railway & National Engineering Research Center of System Technology for High-Speed Railway and Urban Rail Transit (L2024W001), Zhongguancun Lab and Huawei.

REFERENCES

[1] 2026-03. Openneuro Platform. https://openneuro.org/.
[2] 2026-03. Physionet Platform. https://physionet.org/.
[3] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam D. Smith. 2006. Calibrating Noise to Sensitivity in Private Data Analysis. In *TCC (Lecture Notes in Computer Science)*, Shai Halevi and Tal Rabin (Eds.), Vol. 3876. Springer, 265–284.
[4] Yuwei Huang and Guoliang Li. 2024. Laser: Buffer-Aware Learned Query Scheduling in Master-Standby Databases. *Proc. VLDB Endow.* 18, 3 (2024), 743–755.
[5] Guoliang Li, Ji Sun, James Pan, Jiang Wang, Yongqing Xie, Ruicheng Liu, and Wen Nie. 2025. GaussDB-Vector: A Large-Scale Persistent Real-Time Vector Database for LLM Applications. *Proc. VLDB Endow.* 18, 12 (2025), 4951–4963.
[6] Jie Mei, Ruixin Luo, Lichao Xu, Wei Zhao, Shengfu Wen, Kun Wang, Xiaolin Xiao, Jiayuan Meng, Yongzhi Huang, Jiabei Tang, Longlong Cheng, Minpeng Xu, and Dong Ming. 2024. MetaBCI: An open-source platform for brain-computer interfaces. *Comput. Biol. Medicine* 168 (2024), 107806.
[7] Zhaoyan Sun, Xuanhe Zhou, Guoliang Li, Xiang Yu, Jianhua Feng, and Yong Zhang. 2025. R-Bot: An LLM-based Query Rewrite System. *Proc. VLDB Endow.* 18, 12 (2025), 5031–5044.
[8] Latanya Sweeney. 2002. k-Anonymity: A Model for Protecting Privacy. *Int. J. Uncertain. Fuzziness Knowl. Based Syst.* 10, 5 (2002), 557–570.
[9] Jiayi Wang, Yuan Li, Jianming Wu, Shihui Xu, and Guoliang Li. 2025. Unify: A System For Unstructured Data Analytics. *Proc. VLDB Endow.* 18, 12 (2025), 5287–5290.
[10] Xuanhe Zhou, Wei Zhou, Liguo Qi, Hao Zhang, Dihao Chen, Bingsheng He, Mian Lu, Guoliang Li, Fan Wu, and Yuqiang Chen. 2025. OpenMLDB: A Real-Time Relational Data Feature Computation System for Online ML. In *SIGMOD*. ACM, 729–742.
[11] Yaoyu Zhu, Guoliang Li, Jianhua Feng, and Nan Tang. 2026. Synthesizing tables for supervised learning. *Frontiers Comput. Sci.* 20, 3 (2026), 2003603.