Global Journal of Engineering and Technology Advances
eISSN: 2582-5003
Cross Ref DOI: 10.30574/gjeta
Journal homepage: https://gjeta.com/

(REVIEW ARTICLE)

Check for updates

Collaborative Intelligence Databases (CID): Harnessing AI for Privacy-Preserving Multi-Source Data Management

Oluwatoyin Olawale Akadiri [1, *], Ololade Babatunde [2], Adam Adebayo Jamiu [3], Olamotse Roland Igbape [4], and Bibilari Oladipupo Samson [5]

[1] Department of Information Sciences, School of Information Sciences and Engineering, Bay Atlantic University, United States.
[2] Department of Computer Engineering, Izmir Institute of Technology, Izmir, Turkey.
[3] Department of Educational Technology, Faculty of Education, University of Ilorin, Nigeria.
[4] Department of Computer Science, College of Computing, Georgia Institute of Technology, Atlanta, GA, USA.
[5] Department of Business Studies, Greater Manchester Business School, University of Greater Manchester, Bolton, United Kingdom,

Global Journal of Engineering and Technology Advances, 2025, 24(03), 406-416

Publication history: Received on 17 August 2025; revised on 23 September 2025; accepted on 26 September 2025

Article DOI: https://doi.org/10.30574/gjeta.2025.24.3.0289

Abstract

The proliferation of data in distributed platforms has brought enormous opportunities for collaborative intelligence and, at the same time, generated a major concern in the area of privacy. The concept that will be presented in this article is the notion of Collaborative Intelligence Databases (CID), which is a new model that uses artificial intelligence to facilitate privacy-conscious and secure data management across jurisdictions. CID structures enable institutions to combine the use of heterogeneous data sources to derive insights by applying a suite of sophisticated cryptographic techniques, including homomorphic encryption, secure multi-party computation, differential privacy, as well as a federated learning architecture.

The survey provides an overview of the existing state of the art in privacy-conscious collaborative data management, defines the basic elements of technology, and classifies possible implementation plans in various sectors such as health care, banking, transportation, and industrial Internet of Things. As our analysis has shown, deep technical challenges still exist, but recent developments have proven that it is possible to establish large-scale collaborative intelligence and, at the same time, provide high privacy guarantees.

Keywords: Federated Learning; Privacy-Preserving Computing; Multi-Party Computation; Homomorphic Encryption; Collaborative Intelligence; Data Federation

1. Introduction

Organizational digitization created massive volumes of data that were created and stored within the distributed systems. There are strengths and threats to this development in the collaborative intelligence efforts. Traditional forms of multi-organization data sharing tend to be centralized in their aggregation, where there are extreme privacy, compliance, and competition measures [1]. The Collaborative Intelligence Databases (CID) is the concept that signifies the new paradigm - the delivery of safe and privacy-conscious collaboration across organizational boundaries with the capability of utilizing the power of distributed data.

* Corresponding author: Oluwatoyin, Olawale Akadiri

Copyright © 2025 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution Liscense 4.0.

CID systems are an assortment of technologies and methodologies that help make safe calculations on delicate datasets. Based on the most recent cryptographic systems, federated learning frameworks, and AI-driven optimization approaches, these types of systems enable companies to attract useful insights to be created collectively, without the need to disclose proprietary or sensitive individual details [2]. In essence, CID solutions offer the possibility of enabling all participants to cooperate and mutually gain ground on the joint intelligence without the need to release their data.

The increase in the importance of CID applications is observable in different high-stakes fields. Healthcare multi-institutional research must have good privacy-sensitive measures to ensure compliance and facilitate medical discoveries [3], [4]. Similarly, financial institutions are in a predicament of cooperating to identify fraud and controlling it, and at the same time satisfying the high privacy demands [2]. Transportation and Smart city activities have also included intra-vehicle communication systems and urban data-sharing ecosystems, also [5], [6].

This paper will proceed to make an exhaustive discussion of the systems CID, the state of the art in the market, challenges to its implementation, and the development patterns in the future. Our contribution will be as follows: the systematic taxonomy of privacy-relevant procedures, assessment of needs in the considered problem, and critical analysis of the available frameworks and constraints.

2. Theoretical Framework and Conceptual Foundations.

2.1. Theoretical Framework

The Collaborative Intelligence Databases (CID) are, in theory, based on the distributed systems theory, cryptographic security, and information-theoretic privacy models. The framework is based on three large paradigms: the secure multi-party computation model presented by Yao in 1982, the federated learning paradigm, and the differential privacy model proposed by Dwork in 2006.

In this context, some key elements can be singled out. The involved parties are the organizations or parties that are involved in the collaboration. All these bodies have their own private databases, and therefore, the sensitive information will be locally owned and managed. The collaborative role establishes the activity or computation that is carried out by the system between parties. To allow that, a collection of privacy-sensitive protocols is used, which offers the means of safe and secret data exchange. Lastly, the model includes the assurances of security and robustness, which guarantee that the collaborative process becomes reliable and resistant to possible adversarial behaviors.

2.2. Conceptual Review

According to the development of collaborative intelligence, it has undergone various phases. First, information exchange had to be physically carried over and had to be processed in a central place, posing significant regulatory and privacy concerns. Privacy-preserving computation increased the effectiveness of distributed collaboration, but initial implementations were limited in scalability.

2.2.1. The elements of the modern CID system combine several technological streams:

- Privacy-Centric Design Principles: The modern CID designs incorporate privacy as an integral part, rather than a supplement to, based on homomorphic encryption, secure multi-party computation, and differential privacy [7].
- Federated Learning Integration: Federated learning has been demonstrated to be scalable to distributed machine learning and does not require central aggregation, but minimizes privacy risks and limits bandwidth [8].
- Hybrid Cryptographic Solutions: There is no single cryptographic solution that can support all privacy requirements, and, therefore, hybrid solutions strategically combine techniques according to the use-case requirements [9].
- Domain-Specific Customizations: Industry-specific CID frameworks are required by different industries- such as healthcare [3], finance [2], and transportation [5].

2.3. Architecture Conceptual Models.

2.3.1. CID systems have developed three key architectural models:

- Hierarchical Federation Model: Group members are arranged in a tree structure that has specific access controls and levels of trust. The model boosts scalability and also supports fine-grained access control [10].
- Peer-to-peer Collaborative Model: This is a decentralized system in which all the participants are equal. It has consensus mechanisms and distributed protocols that regulate calculations, which make the model resilient and remove single points of failure [7].
- Hybrid Orchestration Model: This is a coordination of computation that is both central and decentralized. Collaborative procedures are managed by trusted coordinators, and privacy is ensured by cryptographic protocols [11].

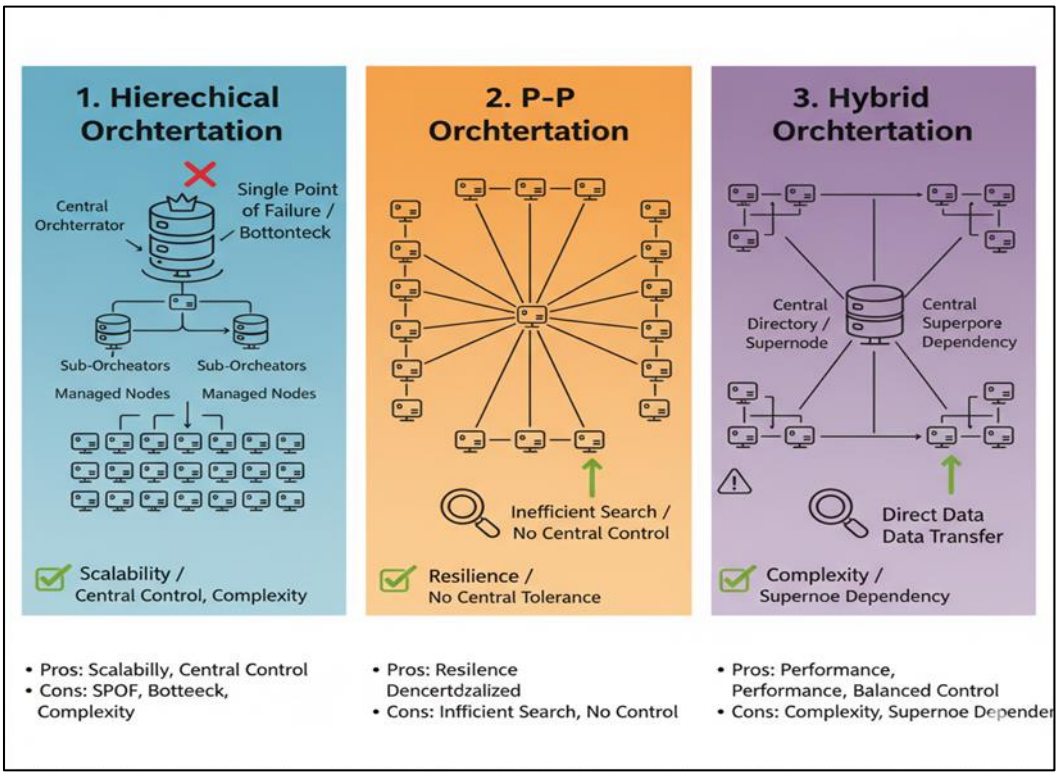


Figure 1 CID Conceptual Architecture Models – Comparison of hierarchical, peer-to-peer, and hybrid orchestration approaches

3. Technological Foundations

3.1. Cryptographic Privacy-Preserving Techniques

CID systems utilize a lot of cryptography to guarantee distributed computation. The first of these is Homomorphic Encryption (HE), which permits the processing of encrypted data without the need to decrypt it [12], [9]. New hybrid HE methods have enhanced scalability without compromising on the high privacy guarantees, and as such, HE can be used in large-scale applications [9].

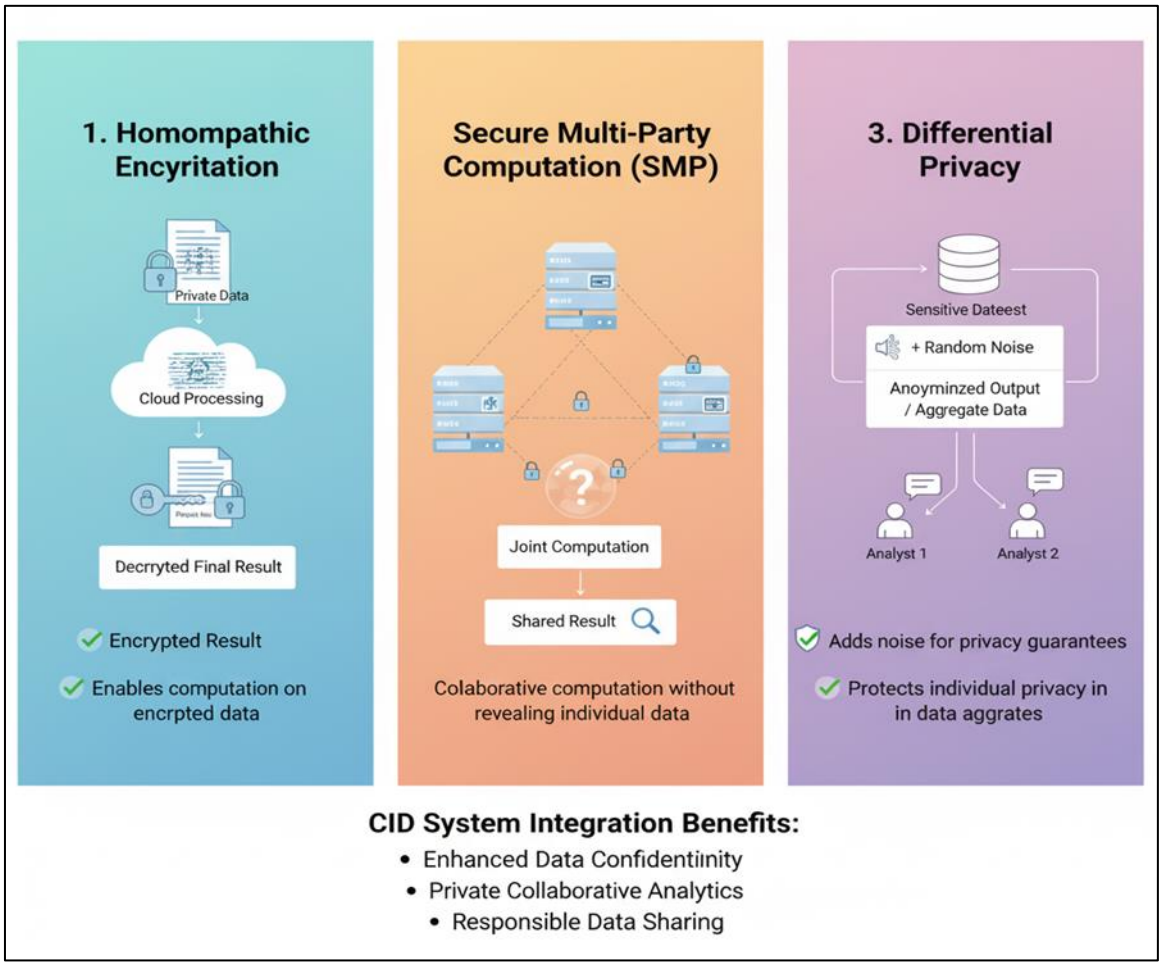


Figure 2 Cryptographic Techniques in CID Systems – Illustration of homomorphic encryption, SMPC, and differential privacy integration

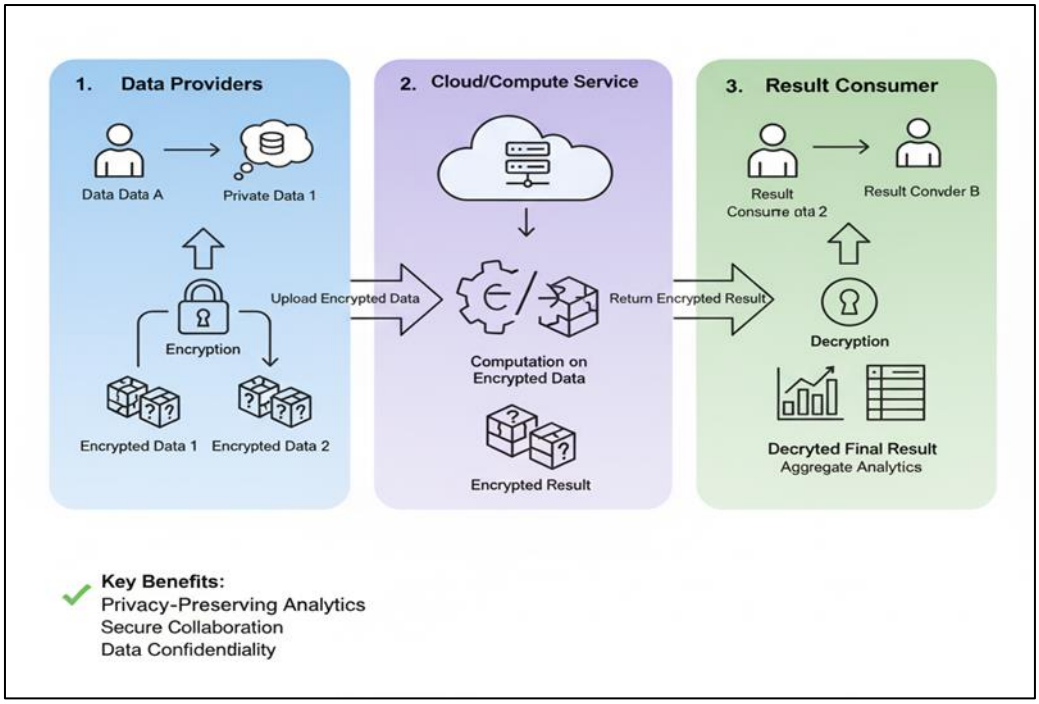


Figure 3 Homomorphic Encryption Workflow – Encryption, computation, and result aggregation in collaborative environments

Secure Multi-Party Computation (SMPC) is another crucial technology that enables several parties to collaboratively execute functions without disclosing the input [7], [13]. SMPC is improved with federated learning, especially when particular attention is given to privacy-sensitive tasks.

Differential Privacy (DP) guarantees individual protection by adding noise to outputs controlled to avoid any inference of a particular piece of data [14], [11]. Other privacy approaches are typically combined with DP in highly controlled areas.

3.2. Federated Learning Structures.

A key component of collaboration in machine learning on distributed data is Federated Learning (FL). Global models in FL are trained by means of aggregation of local updates; hence, no centralized data storage is required [8]. Multi-party FL that is scalable has been used to deal with the problem of non-homogeneous data and unequal computational capabilities amongst participants [8].

Byzantine attacks are also a problem in FL systems, especially when applied to healthcare, where the contributions of malicious examples can compromise the models. New developments also comprise strong aggregation techniques with local differential privacy [15].

Also, a federated unlearning phenomenon has been introduced, which grants subjects the ability to take contributions out of the system without destabilizing it [16]. Federated environments assist in selective data removal with the help of ontology-based methods.

3.3. Federation and Integration of Data.

Data federation supports queries and integration of different sources without the need to move them physically [17]. In contemporary designs, focus is made on semantic integration, access control, and mature query optimization across organizational borders.

Optimization schemes that are based on AI (distributed query planning, adaptive resource allocation, and machine learning-based optimization) greatly optimize performance and scale [1].

4. Domain-Specific Applications

Healthcare and medical research are a particular type of research involving medical care and drug development (Stern, 2018).

The development of CID has attracted attention to the healthcare sector because it has stringent privacy needs, and its research must be multi-institutional. Its applications can be in clinical research, drug development, and epidemiology, each of which demands sensitive collaboration of data [3].

Recent applications show privacy-sensitive collaborative medical imaging, in which hospitals do not share patient records, but they jointly train diagnostic models [18]. These systems combine the protection of differential privacy, safe aggregation, and exclusive access controls.

One of them is DeCaPH (Decentralised, Collaborative, and Privacy-Preserving Machine Learning to Multi-Hospital Data), a framework that allows learning securely across multiple hospitals and overcomes the problem of scalability, heterogeneity, and compliance issues [3].

- *Fintech and Financial Services.*

Competition, high regulation, and sensitivity of data present specific challenges to financial institutions in the collaboration of data. In finance, CID helps to detect fraud, manage risks, monitor compliance, and protect customer confidentiality [2].

Combining blockchain and federated learning enhances credibility, transparency, and accountability among institutions [19]. These hybrid structures reduce the difficulties of verification in cross-organization partnerships.

- *Transportation and Smart Cities.*

CID has a crucial role in the transport sector in autonomous cars and smart cities. Vehicle-to-vehicle and vehicle-to-infrastructure communication needs privacy-preserving data exchange, as sensitive location and behavioral data need to be safeguarded [5].

CID solutions with digital twin technologies help streamline the transportation system by incorporating real-time information of vehicles, sensors, and traffic networks [6].

New directions are UAV-based Internet of Vehicles, where CID frameworks are modified to address aerial and ground coordination issues. Multi-party contracts are based on incentives to maximize cooperation [20].

- *Industrial IoT (IIoT)*

CID helps organisations in industrial environments to make predictions in maintenance, supply chain, and process improvements. The industrial IoT generates large amounts of sensor data, which needs privacy-sensitive analysis [21].

IIoT Federated learning has been implemented successfully in the context of data space ecosystems, which have allowed sharing knowledge securely between partners of the manufacturing industry [21]. Such systems are normally equipped with a high level of semantic integration and rigorous access control.

- *Energy and Smart Grids*

CID is used in the energy sector in the optimization of generation, distribution, and consumption. Smart grids are based on privacy-sensitive infrastructures to withhold consumer usage information and ensure efficiency of the system as a whole [22].

Federated learning is useful in demand forecasting, fault detection, and energy trading- balancing privacy and operational stability.

5. Privacy-Preserving Techniques Taxonomy

5.1. Cryptographic Approaches

Table 1 Comparison of Cryptographic Privacy-Preserving Techniques in CID Systems

Technique	Description	Advantages	Limitations	Application Domains
Homomorphic Encryption	Computation on encrypted data	Strong mathematical privacy guarantees	High overhead, limited operation types	Healthcare, Finance
Secure Multi-Party Computation	Joint computation without revealing inputs	Flexible, supports complex operations	Communication-intensive, scalability issues	Multi-institutional research
Differential Privacy	Adds calibrated noise to preserve privacy	Formal guarantees, regulatory approval	Utility-privacy tradeoff, parameter sensitivity	Public health, Census data
Hybrid Approaches	Combines multiple cryptographic techniques	Balanced privacy-performance trade-offs	Integration complexity, higher implementation demands	Large-scale, multi-domain applications

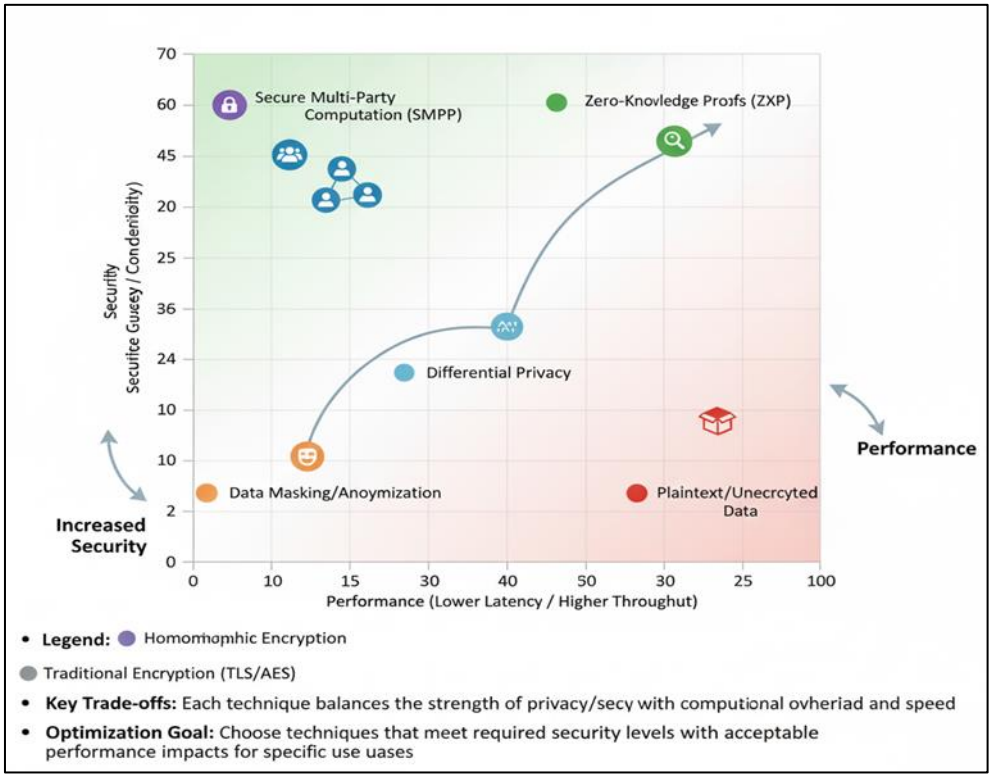


Figure 4 Cryptographic Techniques Comparison Matrix – Performance vs. security trade-offs

5.2. Trade-offs between Performance and Security.

There are trade-offs between privacy guarantees and system scalability, which are found in various privacy-preserving techniques [23].

Homomorphic Encryption is a costly and highly secure encryption method and is well applicable in applications where low latency and high security are required.

Differential Privacy scales and parameter tunings are good, but privacy budgets should be maintained.

SMPC can perform advanced calculations; it suffers in the area of communication overhead. Recent advances in threshold cryptography and verifiable secret sharing have allowed it to be more practical.

5.3. Byzantine Fault Tolerance and Robustness.

This has enhanced the power of CID systems against malicious people. Byzantine-resistant federated learning algorithms are layers unified with successful aggregation protocols that filter and reject malicious updates and maintain privacy [15].

Layered defense mechanisms (a combination of outlier detection, secure aggregation, and cryptographic validation of contributions) are also proposed in other current works as the means to ensure privacy and robustness [24].

6. System Implementation and Architecture.

6.1. Layered Federation Models

The layered federation models are used to arrange collaborative intelligence into various levels of abstraction - starting with raw data, followed by feature extraction and, finally, high-level knowledge representation [10]. It supports flexible participation with both privacy protection and computational contributions as a hierarchical approach.

An interesting innovation is the Libertas architecture that integrates decentralized personal data stores with secure multi-party computation to facilitate privacy-preserving collaboration between user groups [7]. It specifically deals with the issues of data sovereignty and personal privacy on a massive scale.

6.2. Practice Implementation Models.

Several real-world frameworks combine a variety of privacy-conserving tools to solve practical deployment challenges. Indicatively, FedLabX offers a framework that is modular with privacy models, secure aggregation, and Byzantine fault tolerance features [11]. It is modular by nature and enables organizations to choose the technique to suit their needs.

The FLoBC model presents the combination of blockchain with federated learning to improve decentralization and trust [19]. The immutability of blockchain contributes to transparency and accountability, whereas cryptography algorithms ensure privacy.

6.3. Network Integration and Edge Computing.

A combination of edge computing and high-end networking (e.g., 6G) with CID increases the potential of scalable joint intelligence. Privacy-preserving methods have now been expanded to distributed ecosystems of IoT scale and are used to tackle heterogeneity with a combination of homomorphic encryption and differential privacy [14].

New federated intelligence systems combine giant AI models with federated fine-tuning and edge-based reasoning [25]. Such systems overcome resource constraints, network congestion, and device heterogeneity with strong privacy protection

7. Performance Evaluation and Benchmarking

7.1. Scalability Analysis

Table 2 Scalability Characteristics of Privacy-Preserving Techniques

System Component	Small Scale (10–100 participants)	Medium Scale (100–1,000 participants)	Large Scale (1,000+ participants)
Homomorphic Encryption	Acceptable performance	Significant overhead	Prohibitive for real-time use
Differential Privacy	Minimal impact	Low overhead	Scales well with tuning

Secure Aggregation	Good performance	Moderate overhead	Communication bottlenecks
Hybrid Approaches	Balanced trade-offs	Optimized performance	Needs careful design

Scalability is significantly different in evaluations. Differential privacy has the best scaling property with a small incremental cost as the number of participants increases [23].

Homomorphic encryption is also associated with high computational costs, and secure multi-party computation is also associated with communication overhead costs. Hybrid schemes are useful in striking a balance between privacy and performance, but take conscious planning.

7.1.1. Communication and Computational Overhead.

CID systems' architecture decouples the computational overhead and the communication overhead.

Cryptographic schemes have different computational overheads, and homomorphic encryption is the most costly one.

A large-scale deployment is still limited by communication overhead, particularly in SMPC protocols. Cryptography threshold and verifiable secret sharing have made these costs less, but they are still limited.

Hybrid solutions apply various techniques selectively, which reduce the overall system overhead but maintain privacy.

8. Regulatory and Compliance Problems.

8.1. Data Protection and GDPR

CID should be adopted according to the evolving laws regarding data protection, such as the General Data Protection Regulation (GDPR). Among these problems is the so-called right to be forgotten that requires an advanced federated unlearning architecture capable of erasing specific contributions without negatively influencing the integrity of the entire model [16].

CID applications are also privacy by design, with privacy-sensitive decisions on the architecture and lifecycle throughout the lifecycle of the system, encompassing development, deployment, and operation.

8.2. Sector-Specific Regulatory Requirements.

Healthcare: Is required to comply with HIPAA, HITECH, and other medical privacy requirements.

Finance The business is subject to such laws as the anti-money laundering (AML), know-your-customer (KYC), and privacy laws concerning financial data [2].

Cross-Border Data Sharing: CID systems must manage different regulatory settings, and, therefore, international data transfer is a complex task.

The criteria that influence the feasibility of cross-border CID applications include privacy adequacy agreements and jurisdictional harmonization.

9. Future Projections and Research Problems.

9.1. New Technology and Integration.

The development of quantum computing is accompanied by threats and challenges. The CID systems will also need quantum-resistant cryptography to ensure protection against future attackers. Meanwhile, quantum algorithms can support new privacy-preserving computation schemes to overcome existing scalability challenges.

The other frontier is connecting CID with large AI models, including foundation models. Among the obstacles are model size, resource needs, and privacy-preserving fine-tuning of different participants [25].

The standardization and interoperability rely on the compatibility of the methods and tools used to conduct the study.<|human|>Interoperability and Standardization

9.2. It is based on the compatibility of methods and tools used to conduct the study.

The absence of standardized protocols and interfaces is one of the significant obstacles. Despite the current advances in the direction of shared frameworks and reference architectures, additional work is required to ensure vendor and implementation interoperability.

The process of cross-domain collaboration requires semantic integration, as well as the alignment of ontologies, which are required to ensure the presence of seamless interoperability between industries and organizational environments.

10. Incentive Models and Economic Models.

Good economic incentive systems are essential to make CID systems sustainable. The current partnerships tend to presuppose the voluntary nature, whereas the real-life situations require other advanced cost-sharing and benefit-distribution designs.

Blockchain-based incentive programs and smart contracts can be proposed as possible solutions and can guarantee equitable and transparent reward methods. Further studies, however, are necessary to develop scaled and viable economic systems.

10. Conclusion

The model of Collaborative Intelligence Databases (CID) is a revolutionary model of inter-organizational data collaboration that allows joint intelligence and high levels of privacy. CID systems create the stable foundation of privacy-conscious data sharing by means of combining advanced cryptography, federated learning, and optimizing the situation using AI.

Through our analysis, we have noted that, although issues persist, CID applications in healthcare, finance, transportation, and industrial IoT have already shown scalability. Further adoption will be motivated by ongoing research into privacy-safe methods, as well as a consideration of regulatory compliance, economic, and scalability of the system.

10.1. Key takeaways include:

- The value of domain-specific hybrid cryptography strategies.
- The need to have elaborate frameworks that consider technical, regulatory, and economic factors.
- The possibility of having quantum-resistant and AI-enhanced approaches to inform the future of CID.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] A. Bharadwaj and G. Cormode, "Federated computation: A survey of concepts and challenges," *Distributed and Parallel Databases*, 2024.

[2] A. Kumar et al., "Privacy-preserving federated learning in healthcare, e-commerce, and finance: A taxonomy of security threats and mitigation strategies," *EPJ Web of Conferences*, 2025.

[3] H. Fang, A. Dziedzic, R. Zhang, J. Oliva, A. Verma, Z. Razak, N. Papernot, and B. Wang, "Decentralised, collaborative, and privacy-preserving machine learning for multi-hospital data (DeCaPH)," *eBioMedicine*, 2024.

[4] S. Ivanov, "Collaborative federated learning techniques for enhancing privacy-preserving medical data exchange," *Hong Kong Journal of AI & Medicine*, 2024.

[5] A. Castro et al., "Privacy-preserving data sharing mechanisms for autonomous vehicle collaboration," *Journal of AI in Healthcare and Medicine*, 2023.

[6] H. Zhang et al., "Design and application of intelligent transportation multi-source data collaboration framework based on digital twins," *Applied Sciences*, 2023.

[7] J. Zhao, A. Goel, M. Agrawal, S. Stein, D. Albayaydh, R. Verborgh, R. Binns, T. Berners-Lee, and N. Shadbolt, "Libertas: Privacy-preserving collaborative computation for decentralised personal data stores," 2025.

[8] L. Wang et al., "Scalable multi-party collaborative data mining based on federated learning," *Preprints.org*, 2025.

[9] T. D. Nguyen, S. Khan, E. Abdinasibfar, and A. Michalas, "A privacy-centric approach: Scalable and secure federated learning enabled by hybrid homomorphic encryption," *arXiv*, 2025.

[10] H. Li, D. Meng, et al., "Knowledge Federation: A unified and hierarchical privacy-preserving AI framework," 2020.

[11] A. Alshawki, M. Alshawki, et al., "Fedlabx: A practical and privacy-preserving framework for federated learning," *Complex & Intelligent Systems*, 2024.

[12] D. Rathod, U. Baid, V. Cherukuri, et al., "Encrypted federated search using homomorphic encryption," *arXiv*, 2025.

[13] I. Jarin and B. Eshete, "PRICURE: Privacy-preserving collaborative inference in a multi-party setting," *arXiv*, 2021.

[14] U. Ruby et al., "Advanced privacy-preserving federated learning in 6G networks using differential privacy and homomorphic encryption," 2024.

[15] FedCCW: A privacy-preserving Byzantine-robust federated learning with local differential privacy for healthcare," *Cluster Computing*, 2025.

[16] R. Ghannam and Y. Mahareek, "Privacy-preserving federated unlearning with ontology-guided relevance modeling," *Future Internet*, 2025.

[17] A. Hogan et al., "A systematic overview of data federation systems," *Semantic Web*, 2024.

[18] Collaborative privacy-preserving approaches for distributed deep learning using multi-institutional data," *Radiographics*, 2023.

[19] A privacy-preserving federated learning framework for blockchain networks (FLoBC)," *Cluster Computing*, 2024.

[20] W. Y. B. Lim, J. Huang, et al., "Towards federated learning in UAV-enabled internet of vehicles: A multi-dimensional contract-matching approach," *arXiv*, 2020.

[21] Smart and collaborative industrial IoT: A federated learning and data space approach," *ScienceDirect*, 2023.

[22] A privacy-preserving framework for federated learning in smart grids," *INN-DS&A*, 2024.

[23] M. Shalabi, A. Khedr, A. Rushdy, and A. Salah, "A comparative study of privacy-preserving techniques in federated learning: A performance and security analysis," *Information*, 2025.

[24] Robust and privacy-preserving collaborative training: A comprehensive survey," *Artificial Intelligence Review*, 2024.

[25] Z. Ni, Y. Sun, X. Ao, H. Tian, et al., "Federated intelligence: When large AI models meet federated fine-tuning and collaborative reasoning at the network edge," 2025.