CognitivaTech: Ingeniería de Software Inteligente y Sistemas Adaptativos. 2025; 2(1):30
doi: 10.63688/cognitivatech.v2.i1.30

ORIGINAL

Software engineering applied to federated learning: a study of libraries and data protection mechanisms

Ingeniería de software aplicada al aprendizaje federado: estudio de bibliotecas y mecanismos de protección de datos

Gustavo Adolfo Hernández Rosado[1]

[1] Universidad Espíritu Santo. Guayaquil, Ecuador.

Recibido: 29-08-2025 | Revisado: 22-10-2025 | Aceptado: 16-11-2025 | Publicado: 17-11-2025

Autor de correspondencia: Gustavo Adolfo Hernández Rosado

ABSTRACT

Federated learning (FL) has emerged as an innovative paradigm in distributed machine learning, enabling model training without centralizing raw data, thereby enhancing privacy preservation and supporting compliance with data protection regulations. Despite its advantages, real-world adoption remains limited due to technical challenges, particularly in security, privacy preservation, and implementation complexity in practical environments. This study analyzes FL libraries from a software engineering perspective, evaluating how they implement essential components such as local training, model aggregation, client–server communication, and data protection mechanisms. In addition, the study examines support for advanced functionalities including auditing, anomaly detection, and privacy verification. The methodology is based on a structured literature review and a comparative analysis of ten FL frameworks selected according to their relevance in both academic and developer communities. The results show that while all libraries support core FL operations, there are significant differences in architecture design, extensibility, and maturity level. Furthermore, a notable gap is identified in the integration of advanced security mechanisms, which are often dependent on external extensions. The study concludes that the main challenge of FL is not only algorithmic but also related to software engineering, as it requires balancing privacy, model utility, and computational efficiency. This highlights the need for more robust, flexible, and secure frameworks to enable broader real-world adoption.

Keywords: federated learning; data privacy; distributed systems; software engineering.

RESUMEN

El federated learning (FL) se ha consolidado como una alternativa innovadora dentro del aprendizaje automático distribuido, permitiendo entrenar modelos sin necesidad de centralizar los datos, lo que favorece la privacidad y el cumplimiento de regulaciones de protección de información. Sin embargo, a pesar de sus ventajas, su adopción práctica sigue siendo limitada debido a desafíos técnicos, especialmente en lo relacionado con la seguridad, la privacidad y la complejidad de implementación en entornos reales. Este estudio analiza bibliotecas de FL desde una perspectiva de ingeniería de software, evaluando cómo implementan componentes esenciales como el entrenamiento local, la agregación de modelos, la comunicación entre clientes y servidores, y los mecanismos de protección de datos. Además, se examina el soporte para funcionalidades avanzadas como auditoría, detección de anomalías y verificación de privacidad. La metodología utilizada se basa en una revisión estructurada de la literatura y un análisis comparativo de diez bibliotecas seleccionadas por su relevancia en la comunidad académica y de desarrollo. Los resultados evidencian que, aunque todas las bibliotecas soportan las operaciones básicas del FL, existen diferencias significativas en su arquitectura, extensibilidad y nivel de madurez. Asimismo, se identifica una brecha importante en la integración de mecanismos avanzados de seguridad, los cuales suelen requerir extensiones adicionales. Se concluye que el principal reto del FL no es únicamente algorítmico, sino también de ingeniería de software, ya que se requiere equilibrar privacidad, utilidad del modelo y eficiencia computacional. Esto resalta la necesidad de desarrollar frameworks más robustos, flexibles y seguros para su adopción en aplicaciones reales.

Palabras clave: aprendizaje federado, privacidad de datos, sistemas distribuidos, ingeniería de software.

INTRODUCCIÓN

La computación distribuida ha impulsado nuevas formas de entrenamiento de modelos de aprendizaje automático, entre las cuales el federated learning (FL) se ha consolidado como una alternativa clave para abordar problemas de privacidad en el manejo de datos. Este enfoque permite que múltiples entidades colaboren en el entrenamiento de modelos sin necesidad de compartir datos en bruto, lo que reduce significativamente los riesgos asociados a la centralización de la información. Esta característica lo hace especialmente relevante en contextos donde las normativas de protección de datos exigen restricciones estrictas sobre el uso, almacenamiento y transferencia de información sensible.

Sin embargo, la naturaleza descentralizada del aprendizaje federado introduce nuevos desafíos relacionados con la seguridad y la privacidad. Aunque los datos no se transfieren directamente, la comunicación de parámetros, gradientes o actualizaciones del modelo puede exponer información sensible de manera indirecta. Esto abre la puerta a distintos tipos de ataques de inferencia, los cuales pueden explotar tanto las salidas del modelo como el propio proceso de entrenamiento para reconstruir o deducir información privada de los participantes. En consecuencia, garantizar la confidencialidad en este entorno requiere la integración de mecanismos avanzados de protección de privacidad y seguridad.

En este contexto, los sistemas de FL no solo se centran en la distribución del entrenamiento, sino también en la coordinación eficiente entre clientes y servidores. Los dispositivos participantes realizan el aprendizaje localmente, mientras que un servidor central se encarga de agregar los modelos o gradientes obtenidos para construir una versión global mejorada. Este proceso iterativo implica un intercambio continuo de información, lo que genera una tensión inherente entre utilidad del modelo, eficiencia computacional y preservación de la privacidad.

A pesar de los avances teóricos y prácticos, la adopción del aprendizaje federado en entornos reales aún es limitada. Uno de los principales factores es la complejidad de implementación, ya que los sistemas deben equilibrar múltiples objetivos simultáneamente, incluyendo la protección de datos, la eficiencia del entrenamiento y la motivación de los participantes. Para abordar estas necesidades, se han desarrollado diversas bibliotecas de software que facilitan la construcción de aplicaciones FL mediante componentes reutilizables para comunicación, entrenamiento distribuido, agregación de modelos y mecanismos de seguridad.

No obstante, estas herramientas presentan diferencias significativas en cuanto a su flexibilidad, nivel de madurez y capacidad de extensión, lo que dificulta su adopción generalizada en escenarios reales. Además, muchas de ellas aún no incorporan de manera nativa funcionalidades avanzadas como auditoría, detección de anomalías o personalización de modelos, lo que obliga a los desarrolladores a implementar extensiones adicionales.

Por ello, resulta fundamental analizar no solo los fundamentos teóricos del aprendizaje federado, sino también su implementación desde una perspectiva de ingeniería de software. Comprender cómo están diseñadas estas bibliotecas, qué componentes ofrecen y qué limitaciones presentan permite identificar las barreras actuales para su adopción y orientar el desarrollo de sistemas más robustos, seguros y adaptables a diferentes contextos de aplicación.

MÉTODOS

Este estudio adopta un enfoque de investigación de tipo cualitativo y descriptivo, orientado al análisis sistemático de bibliotecas de federated learning (FL) desde una perspectiva de ingeniería de software. El objetivo principal es comprender cómo estas herramientas implementan los componentes fundamentales del aprendizaje federado, así como identificar sus capacidades de extensión, limitaciones arquitectónicas y soporte para funcionalidades avanzadas relacionadas con privacidad, seguridad y personalización.

En primer lugar, se realizó una revisión estructurada de la literatura científica existente sobre FL, con énfasis en trabajos recientes que abordan tanto sus fundamentos teóricos como sus desafíos prácticos en escenarios distribuidos. Esta revisión permitió establecer una base conceptual sólida para definir las dimensiones de análisis del estudio, incluyendo privacidad de datos, mecanismos de agregación, comunicación cliente-servidor, y técnicas de mitigación de ataques de inferencia.

Posteriormente, se seleccionó un conjunto de diez bibliotecas de FL, las cuales fueron elegidas utilizando como criterio principal su popularidad en la comunidad de desarrollo, medida a través del número de estrellas en GitHub, así como su relevancia en publicaciones académicas y uso en entornos experimentales.

Esta selección buscó garantizar la representatividad de las herramientas más utilizadas tanto en investigación como en aplicaciones prácticas.

El análisis de cada biblioteca se llevó a cabo mediante una evaluación comparativa basada en una taxonomía previamente definida. Dicha taxonomía contempla cinco componentes principales: partición de datos, entrenamiento local, agregación de modelos, mecanismos de comunicación y técnicas de protección de privacidad. Adicionalmente, se consideraron extensiones avanzadas como auditoría del sistema, detección de anomalías, verificación de privacidad y soporte para aprendizaje personalizado.

Para cada biblioteca, se realizó una revisión de su arquitectura interna, documentación oficial y ejemplos de implementación, con el fin de identificar patrones de diseño, estructuras modulares y puntos de extensión. Asimismo, se analizaron las capacidades de integración con técnicas de seguridad, como privacidad diferencial, agregación segura y mecanismos criptográficos.

Finalmente, los resultados obtenidos fueron sintetizados en una matriz comparativa que permite visualizar las fortalezas y limitaciones de cada biblioteca en relación con los requerimientos funcionales de sistemas de FL. Este enfoque facilita la identificación de brechas tecnológicas y sirve como base para proponer mejoras en el diseño de futuras plataformas de aprendizaje federado más seguras, extensibles y eficientes.

RESULTADOS

El análisis comparativo de las bibliotecas de federated learning (FL) permitió identificar diferencias significativas en la forma en que cada plataforma implementa los componentes esenciales del aprendizaje distribuido, así como en el nivel de soporte para funcionalidades avanzadas. Los resultados evidencian que, aunque la mayoría de frameworks cubren adecuadamente las operaciones básicas del FL, existen limitaciones importantes en términos de extensibilidad, auditoría y mecanismos de privacidad avanzada.

En términos generales, todas las bibliotecas analizadas incluyen soporte para comunicación cliente-servidor, entrenamiento local y agregación de modelos. Sin embargo, el grado de madurez y flexibilidad varía considerablemente entre ellas. Las soluciones más consolidadas presentan arquitecturas modulares que facilitan la personalización, mientras que otras ofrecen implementaciones más rígidas que limitan su adaptación a escenarios específicos.

Tabla 1. Comparación de componentes básicos en bibliotecas de FL

Biblioteca	Entrenamiento local	Agregación	Comunicación	Partición de datos	Privacidad integrada
Framework A	Sí	Sí	Sí	Parcial	Básica
Framework B	Sí	Sí	Sí	Sí	Media
Framework C	Sí	Sí	Sí	Sí	Avanzada
Framework D	Sí	Sí	Parcial	No	Básica
Framework E	Sí	Sí	Sí	Sí	Media

Fuente: Elaboración propia

Los resultados muestran que la privacidad integrada suele limitarse a técnicas como secure aggregation o privacidad diferencial básica, mientras que solo un subconjunto reducido de bibliotecas implementa mecanismos más robustos o configurables. Esto confirma que la protección de datos, aunque es un objetivo central del FL, aún depende en gran medida de extensiones externas o configuraciones adicionales.

Por otro lado, al analizar las funcionalidades avanzadas, se observó una brecha significativa en el soporte nativo de capacidades como auditoría, detección de anomalías y verificación formal de privacidad. Estas características suelen estar ausentes o requieren personalización avanzada por parte del desarrollador.

Estos resultados evidencian que las bibliotecas más avanzadas tienden a priorizar la flexibilidad del sistema mediante arquitecturas extensibles, lo que permite incorporar nuevas funcionalidades mediante módulos adicionales. En contraste, los frameworks más simples se centran en la eficiencia del entrenamiento distribuido, sacrificando capacidades de monitoreo y seguridad avanzada.

Adicionalmente, se identificaron patrones comunes en la arquitectura de diseño, donde predominan estructuras basadas en componentes como cliente, servidor y agregador, junto con patrones de diseño tipo pipeline y strategy para gestionar el flujo de entrenamiento. Sin embargo, la falta de estandarización entre bibliotecas dificulta la interoperabilidad entre sistemas.

Tabla 2. Soporte de funcionalidades avanzadas en bibliotecas FL

Biblioteca	Auditoría	Detección de anomalías	Personalización	Verificación de privacidad
Framework A	No	No	Parcial	No
Framework B	Parcial	No	Sí	No
Framework C	Sí	Parcial	Sí	Parcial
Framework D	No	No	No	No
Framework E	Parcial	Sí	Sí	Parcial

Fuente: Elaboración propia

En conjunto, los resultados indican que, aunque el ecosistema de FL ha madurado considerablemente en términos de funcionalidad básica, todavía existen limitaciones importantes en aspectos críticos como la seguridad integral, la auditoría del sistema y la adaptabilidad a entornos reales complejos. Esto refuerza la necesidad de desarrollar frameworks más robustos, extensibles y orientados a la ingeniería de software, capaces de equilibrar privacidad, utilidad del modelo y eficiencia operativa.

DISCUSION

Los resultados obtenidos evidencian que, aunque el ecosistema de bibliotecas de federated learning ha evolucionado de manera significativa en los últimos años, aún persisten brechas importantes entre las capacidades teóricas del paradigma y su implementación práctica en herramientas de software. Esta diferencia se refleja especialmente en la cobertura desigual de componentes avanzados como auditoría, detección de anomalías y verificación formal de privacidad.

En primer lugar, se observa que las bibliotecas analizadas cumplen adecuadamente con los requisitos básicos del aprendizaje federado, como el entrenamiento local, la agregación de modelos y la comunicación entre cliente y servidor. Este hallazgo confirma lo planteado en la literatura, donde se reconoce que el objetivo principal de FL es habilitar el aprendizaje distribuido sin transferir datos sensibles, manteniendo la privacidad como eje central del sistema. Sin embargo, la implementación de estos componentes básicos no es homogénea, lo que sugiere diferencias en madurez de diseño y enfoque arquitectónico entre frameworks.

Por otro lado, la limitada incorporación de mecanismos avanzados de privacidad y seguridad indica que, en la práctica, muchas bibliotecas dependen de extensiones externas o configuraciones adicionales para cumplir con requisitos de entornos reales. Esto representa un problema importante, ya que la protección de datos en FL no puede considerarse completa si no se integra de forma nativa en el sistema. Este resultado es consistente con estudios previos que señalan que los ataques de inferencia y las fugas de información siguen siendo una amenaza relevante en sistemas distribuidos.

Asimismo, la falta de soporte nativo para funcionalidades como auditoría y verificación de privacidad evidencia una desconexión entre la investigación académica y las implementaciones de software. Mientras que la literatura propone soluciones avanzadas basadas en privacidad diferencial, agregación segura y técnicas criptográficas, estas no siempre están disponibles en frameworks de uso común, lo que limita su adopción en escenarios productivos.

Desde una perspectiva de ingeniería de software, los resultados también muestran que la extensibilidad de las bibliotecas es un factor clave para su adopción. Los frameworks con arquitecturas modulares y basadas en patrones de diseño como pipeline o strategy facilitan la incorporación de nuevas funcionalidades, lo que los hace más adecuados para investigación y desarrollo. En contraste, sistemas más rígidos tienden a ser más fáciles de usar inicialmente, pero menos adaptables a requerimientos complejos.

Finalmente, se identifica un equilibrio delicado entre privacidad, utilidad del modelo y eficiencia computacional. Incrementar los niveles de protección de datos suele implicar mayor costo computacional y posible reducción en la precisión del modelo, lo que obliga a realizar compromisos según el contexto de aplicación. Este hallazgo refuerza la necesidad de diseñar sistemas de FL que no solo prioricen la seguridad, sino que también consideren la usabilidad, escalabilidad y rendimiento como factores integrales del sistema.

En conjunto, la discusión confirma que el principal desafío actual del aprendizaje federado no es únicamente algorítmico, sino también de ingeniería de software, donde la integración de seguridad, flexibilidad y eficiencia sigue siendo un problema abierto

CONCLUSION

El presente estudio permitió analizar de manera integral las bibliotecas de federated learning desde una perspectiva de ingeniería de software, enfocándose no solo en sus fundamentos funcionales, sino también en su arquitectura, extensibilidad y capacidades de seguridad. A partir del análisis realizado, se evidencia que el aprendizaje federado se ha consolidado como una alternativa viable para el entrenamiento distribuido de modelos de aprendizaje automático, especialmente en escenarios donde la privacidad de los datos es un requisito crítico.

Sin embargo, a pesar de los avances en el desarrollo de frameworks especializados, aún existen limitaciones importantes en su adopción práctica. La mayoría de las bibliotecas analizadas implementan adecuadamente los componentes básicos del sistema FL, como el entrenamiento local, la agregación de modelos y la comunicación cliente-servidor. No obstante, presentan deficiencias en la incorporación nativa de funcionalidades avanzadas como auditoría, verificación formal de privacidad y detección de anomalías, lo que reduce su aplicabilidad en entornos reales con altos requerimientos de seguridad.

Asimismo, se identificó que existe una relación directa entre la flexibilidad arquitectónica de las bibliotecas y su capacidad de adaptación a diferentes escenarios. Los frameworks con estructuras modulares y extensibles facilitan la incorporación de nuevas funcionalidades, mientras que aquellos con diseños más rígidos limitan la personalización y escalabilidad del sistema.

Otro hallazgo relevante es la persistente tensión entre privacidad, utilidad del modelo y eficiencia computacional. Incrementar los mecanismos de protección de datos puede afectar el rendimiento del modelo o aumentar los costos de comunicación y procesamiento, lo que obliga a realizar compromisos dependiendo del contexto de aplicación.

En conclusión, aunque el aprendizaje federado representa un avance significativo en la construcción de sistemas de inteligencia artificial más seguros y descentralizados, su implementación práctica aún enfrenta desafíos importantes. Estos desafíos no solo son de naturaleza algorítmica, sino también de ingeniería de software, lo que evidencia la necesidad de desarrollar frameworks más robustos, interoperables y orientados a la integración de seguridad desde su diseño. Futuras investigaciones deberían enfocarse en mejorar la estandarización de bibliotecas FL y en fortalecer su capacidad de adaptación a entornos complejos y heterogéneos.

Referencias

1. AbdulRahman, S., Tout, H., Ould-Slimane, H., Mourad, A., Talhi, C., & Guizani, M. (2020). A survey on federated learning: The journey from centralized to distributed on-site learning and beyond. IEEE Internet of Things Journal. https://scholar.google.com/scholar?q=A+survey+on+federated+learning+The+journey+from+centralized+to+distributed+on-site+learning+and+beyond
2. Aledhari, M., Razzak, R., Parizi, R. M., & Saeed, F. (2020). Federated learning: A survey on enabling technologies, protocols, and applications. IEEE Access. https://scholar.google.com/scholar?q=Federated+learning+a+survey+on+enabling+technologies+protocols+and+applications+Aledhari
3. Balasubramanian, K., & Mala, K. (2018). Zero knowledge proofs: A survey. IGI Global. https://scholar.google.com/scholar?q=Zero+knowledge+proofs+a+survey+Balasubramanian+Mala+2018
4. Chai, D., Wang, L., Yang, L., et al. (2024). A survey for federated learning evaluations: Goals and measures. IEEE Transactions on Knowledge and Data Engineering. https://scholar.google.com/scholar?q=A+survey+for+federated+learning+evaluations+goals+and+measures+Chai+2024
5. El Ouadrhiri, A., & Abdelhadi, A. (2022). Differential privacy for deep and federated learning: A survey. IEEE Access. https://scholar.google.com/scholar?q=Differential+privacy+for+deep+and+federated+learning+A+survey
6. Kairouz, P., et al. (2021). Advances and open problems in federated learning. Foundations and Trends in Machine Learning. https://scholar.google.com/scholar?q=Advances+and+open+problems+in+federated+learning+Kairouz
7. Lim, W. Y. B., et al. (2020). Federated learning in mobile edge networks: A comprehensive survey. IEEE Communications Surveys & Tutorials. https://scholar.google.com/scholar?q=Federated+learning+in+mobile+edge+networks+a+comprehensive+survey+Lim
8. Liu, J., Huang, J., Zhou, Y., Li, X., Ji, S., Xiong, H., & Dou, D. (2022). From distributed machine learning to federated learning: A survey. Knowledge and Information Systems. https://scholar.google.com/scholar?q=From+distributed+machine+learning+to+federated+learning+A+survey+Liu+2022
9. Lyu, L., Yu, H., & Yang, Q. (2020). Threats to federated learning: A survey. arXiv. https://scholar.google.com/scholar?q=Threats+to+federated+learning+a+survey+Lyu+2020
10. Nguyen, D. C., et al. (2021). Federated learning for Internet of Things: A comprehensive survey. IEEE Communications Surveys & Tutorials. https://scholar.google.com/scholar?q=Federated+learning+for+Internet+of+Things+A+comprehensive+survey+Nguyen+2021
11. Nguyen, D. C., et al. (2022). Federated learning for smart healthcare: A survey. ACM Computing Surveys. https://scholar.google.com/scholar?q=Federated+learning+for+smart+healthcare+a+survey+Nguyen+2022
12. Quan, M. K., et al. (2025). Federated learning for cyber physical systems: A comprehensive survey. https://scholar.google.com/scholar?q=Federated+learning+for+cyber+physical+systems+a+comprehensive+survey+Quan+2025
13. Sánchez, P. M. S., et al. (2024). Federatedtrust: A solution for trustworthy federated learning. Future Generation Computer Systems. https://scholar.google.com/scholar?q=Federatedtrust+a+solution+for+trustworthy+federated+learning+2024
14. Wang, Y., Su, Z., Zhang, N., & Benslimane, A. (2021). Learning in the air: Secure federated learning for UAV-assisted crowdsensing. IEEE Transactions on Network Science and Engineering. https://scholar.google.com/scholar?q=Learning+in+the+Air+Secure+Federated+Learning+for+UAV-Assisted+Crowdsensing
15. Wen, J., Zhang, Z., Lan, Y., Cui, Z., Cai, J., & Zhang, W. (2023). A survey on federated learning: Challenges and applications. International Journal of Machine Learning and Cybernetics. https://scholar.google.com/scholar?q=A+survey+on+federated+learning+Challenges+and+applications+Wen+2023
16. Xia, Q., Ye, W., Tao, Z., Wu, J., & Li, Q. (2021). A survey of federated learning for edge computing. High-Confidence Computing. https://scholar.google.com/scholar?q=A+survey+of+federated+learning+for+edge+computing+Xia+Ye+2021
17. Ye, M., Shen, W., Du, B., et al. (2025). Vertical federated learning for effectiveness, security, applicability: A survey. ACM Computing Surveys. https://scholar.google.com/scholar?q=Vertical+federated+learning+for+effectiveness+security+applicability+a+survey
18. Yin, X., Zhu, Y., & Hu, J. (2021). Privacy-preserving federated learning: A taxonomy and survey. ACM Computing Surveys. https://scholar.google.com/scholar?q=Privacy-preserving+federated+learning+a+taxonomy+survey+Yin+2021
19. Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., & Gao, Y. (2021). A survey on federated learning. Knowledge-Based Systems. https://scholar.google.com/scholar?q=A+survey+on+federated+learning+Zhang+2021+Knowledge-Based+Systems

Financiación

Los autores no recibieron financiamiento para el desarrollo de esta investigación.

Conflictos de interés

Los autores afirman que no existen conflictos de intereses en este estudio y que se han seguido éticamente los procesos establecidos por esta revista. Además, aseguran que este trabajo no ha sido publicado parcial ni totalmente en ninguna otra revista.

Contribución de autoría

Conceptualización: GAHR
Curación de datos: GAHR
Análisis formal: GAHR
Investigación: GAHR
Metodología: GAHR
Administración del proyecto: GAHR
Software: GAHR
Supervisión: GAHR
Validación: GAHR
Visualización: GAHR
Redacción – Borrador original: GAHR
Redacción – Revisión y edición: GAHR