Revista Hambatu Science

Ambato – Ecuador / ISSN 3151-815X (en línea) / octubre - diciembre 2026

Volumen 1, Número 4

https://doi.org/10.63862/rhs-v1n4-356-381-2026

# Privacidad y seguridad en sistemas de aprendizaje federado: revisión de métodos y desafíos computacionales

## *Privacy and security in federated learning systems: a review of methods and computational challenges*

**Francisco Ricardo Baños Solis**
**Tecnológico Nacional de México - Instituto Tecnológico de Oaxaca**

**Adelina Martínez Nieto**
**Tecnológico Nacional de México - Instituto Tecnológico de Oaxaca**

**Maricarmen Montserrat Velázquez Hernández**
**Tecnológico Nacional de México - Instituto Tecnológico de Oaxaca**

CC BY

Revista Hambatu Science
Periodicidad: octubre - diciembre
Volumen: 1, Número: 4, Año: 2026

**DOI : https://doi.org/10.63862/rhs-v1n4-356-381-2026**

# Privacidad y seguridad en sistemas de aprendizaje federado: revisión de métodos y desafíos computacionales

**Francisco Ricardo Baños Solis**
Tecnológico Nacional de México
Instituto Tecnológico de Oaxaca
francisco.solis@itoaxaca.edu.mx
https://orcid.org/0000-0002-0935-9779
México

**Adelina Martínez Nieto**
Tecnológico Nacional de México
Instituto Tecnológico de Oaxaca
adelina.martinez@itoaxaca.edu.mx
https://orcid.org/0009-0002-9101-7870
México

**Maricarmen Montserrat Velázquez Hernández**
Tecnológico Nacional de México
Instituto Tecnológico de Oaxaca
maricarmen.mv@itoaxaca.edu.mx
https://orcid.org/0009-0009-5519-3165
México

**Recibido:** 2026-15-09

**Aceptado:** 2026-10-11

**Publicado:** 2026-20-11

© The Author(s). This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Resumen**

El aprendizaje federado permite entrenar modelos mediante información distribuida sin concentrar necesariamente los registros originales. Sin embargo, el intercambio de parámetros, la coordinación entre participantes y la exposición a actores maliciosos introducen riesgos que no desaparecen por mantener los datos en origen. El objetivo de este trabajo es examinar los principales métodos de protección de la privacidad y la seguridad en sistemas de aprendizaje federado, así como sus limitaciones y desafíos computacionales. Se realizó una revisión narrativa estructurada con búsqueda bibliográfica orientada en fuentes académicas de acceso verificable y análisis temático de investigaciones fundacionales, artículos metodológicos y revisiones. Los resultados distinguen ataques de inferencia y reconstrucción, envenenamiento de datos o modelos, puertas traseras y amenazas al servidor; asimismo, comparan privacidad diferencial, agregación segura, optimización robusta y mecanismos de verificación. La literatura indica que ninguna técnica aislada protege de manera completa la confidencialidad, integridad y disponibilidad del entrenamiento. La protección depende del modelo de amenaza, la arquitectura y los costos de comunicación, precisión y cómputo. Se concluye que el diseño de sistemas confiables exige defensas complementarias, métricas explícitas y experimentos reproducibles ajustados a las características de cada despliegue.

**Palabras clave:** Aprendizaje federado; privacidad diferencial; agregación segura; seguridad informática; ataques de inferencia; aprendizaje distribuido.

**Revista Hambatu Science**
Periodicidad: octubre - diciembre
Volumen: 1, Número: 4, Año: 2026

# *Privacy and security in federated learning systems: a review of methods and computational challenges*

**Abstract**

Federated learning enables model training on distributed information without necessarily centralizing raw records. Nevertheless, parameter exchanges, participant coordination, and malicious behavior create threats that do not disappear when data remain at their source. This paper examines privacy and security safeguards in federated learning systems and their computational trade-offs. A structured narrative literature review was conducted through a targeted search of verifiable academic sources and thematic analysis of foundational studies, methodological contributions, and surveys. Findings distinguish inference and reconstruction attacks, model poisoning, backdoors, and server-related threats, and compare differential privacy, secure aggregation, robust optimization, and verification mechanisms. The reviewed literature suggests that no single method simultaneously guarantees confidentiality, integrity, and availability. Appropriate protection depends on the threat model, architecture, accuracy requirements, communication overhead, and computational constraints. Trustworthy deployment, therefore, requires complementary safeguards, explicit evaluation metrics, and reproducible tests tailored to each use case.

**Keywords:** Federated learning; differential privacy; secure aggregation; cybersecurity; inference attacks; distributed machine learning.

## Introducción

El uso de aprendizaje automático en dispositivos personales, instituciones sanitarias y redes empresariales ha incrementado el interés por arquitecturas capaces de explotar información distribuida. McMahan et al. (2017) mostraron que un modelo común podía entrenarse mediante actualizaciones calculadas localmente y agregadas iterativamente, reduciendo la necesidad de transferir datos de entrenamiento a un repositorio central. Este procedimiento, conocido como aprendizaje federado, modificó el lugar donde se computan los gradientes, pero no eliminó los intercambios de información ni las dependencias de coordinación. Desde una perspectiva arquitectónica, es necesario separar el beneficio de la minimización del movimiento de datos de una garantía formal de protección de la información.

Kairouz et al. (2021) identifican problemas abiertos asociados con heterogeneidad estadística, recursos desiguales, privacidad y robustez. La diversidad de entornos implica que una arquitectura apropiada para teléfonos móviles puede ser insuficiente para consorcios hospitalarios, donde intervienen otras reglas de acceso, responsabilidades y escalas de participación. De igual manera, la imposibilidad de inspeccionar íntegramente las bases locales limita los procesos de auditoría del entrenamiento. Los riesgos no dependen únicamente del lugar donde residen los datos, sino de qué mensajes se comparten, qué actor puede observarlos y qué inferencias se derivan de ellos.

La pregunta que orienta la revisión es: ¿qué métodos de protección de la privacidad y la seguridad se utilizan en sistemas de aprendizaje federado, frente a qué amenazas ofrecen protección y cuáles son sus principales desafíos computacionales? El objetivo general consiste en analizar las estrategias documentadas, sus supuestos de operación y sus restricciones, contrastando la evidencia sobre confidencialidad, integridad y eficiencia. Se busca responder mediante una síntesis conceptual de estudios publicados y no mediante la estimación de frecuencias de publicación o indicadores de impacto científico

## Estado del arte

### Arquitecturas y operación del aprendizaje federado

El esquema horizontal considera participantes con estructuras de variables comparables pero conjuntos distintos de observaciones. Cada cliente recibe un modelo, calcula actualizaciones con registros propios y transmite información al coordinador para producir un nuevo modelo global. El algoritmo FedAvg, descrito por McMahan et al. (2017), promedia parámetros o actualizaciones provenientes de múltiples participantes y constituye un punto de referencia para contrastar alternativas. Su desempeño, sin embargo, depende de la cantidad de trabajo local, la proporción de dispositivos participantes y la similitud entre distribuciones de datos. No deben generalizarse resultados obtenidos con condiciones homogéneas a instituciones o dispositivos con perfiles divergentes.

La distribución no independiente e idéntica de los datos, denominada con frecuencia non-IID, no es una anomalía periférica, sino una propiedad esperable de muchos sistemas distribuidos. Karimireddy et al. (2020) explican que las diferencias entre clientes generan desplazamientos en las actualizaciones locales, lo que puede dificultar la convergencia; su método SCAFFOLD utiliza variables de control para corregir esa deriva. Esta observación relaciona directamente la confiabilidad del sistema con la optimización: un mecanismo de defensa que altere actualizaciones también puede modificar el comportamiento del modelo, en especial cuando unos participantes aportan distribuciones minoritarias.

El entorno de amenazas puede incluir clientes honestos, dispositivos comprometidos, coordinadores curiosos, atacantes externos sobre canales de comunicación y actores que presentan actualizaciones engañosas. La literatura distingue escenarios entre organizaciones con pocos participantes y escenarios masivos de dispositivos, pero una misma etiqueta de aprendizaje federado no define por sí sola quién es confiable. Kairouz et al. (2021) señalan precisamente que la arquitectura, las hipótesis de confianza y la capacidad computacional constituyen condiciones de diseño que deben describirse antes de formular garantías.

**Privacidad, seguridad y modelos de amenaza**

La privacidad se refiere a limitar la divulgación de información sobre individuos, registros o entidades participantes. La seguridad incluye además la integridad del proceso de entrenamiento y la disponibilidad de los servicios. Esta distinción es importante porque impedir la lectura directa de una actualización no garantiza que la actualización sea válida o benigna. Una agregación cifrada puede reducir la visibilidad de contribuciones individuales, mientras que un cliente malicioso puede intentar introducir un patrón adversarial en la suma. Por tanto, las propiedades de confidencialidad e integridad deben evaluarse por separado, con hipótesis verificables.

En un ataque de inferencia, el adversario intenta deducir atributos, pertenencia o propiedades de los datos mediante la observación del entrenamiento. En un ataque de reconstrucción, busca obtener aproximaciones de los ejemplos originales a partir de gradientes o parámetros disponibles. Estas posibilidades cuestionan la interpretación simplificada de que no transmitir registros originales equivale a impedir filtraciones. La revisión de Saha et al. (2024) identifica reconstrucción e inferencia como amenazas destacadas para el aprendizaje federado confiable. La severidad de una amenaza depende de la información observada, la estructura del modelo, el tamaño de lote y el conocimiento previo del atacante.

Los ataques de envenenamiento son diferentes: su propósito principal es alterar el comportamiento del modelo. Pueden modificar etiquetas, ejemplos de entrenamiento o actualizaciones transmitidas, y en ciertos casos perseguir una puerta trasera activada por un patrón específico. Es importante evitar equiparar cualquier degradación de precisión con un ataque, dado que datos heterogéneos o fallos de sincronización también pueden perjudicar el desempeño. La revisión de seguridad y privacidad recogida por Mothukuri et al. (2021) aborda ambas familias de riesgos y hace evidente la necesidad de especificar tanto objetivos del adversario como recursos bajo su control.

**Privacidad diferencial**

La privacidad diferencial introduce una condición formal sobre cuánto puede variar la distribución de las salidas de un procedimiento cuando se modifica un registro de entrada. En aprendizaje federado puede aplicarse sobre el cálculo local, las actualizaciones transmitidas o

determinadas estadísticas agregadas; esas variantes no son intercambiables. La elección del nivel de protección, sea por ejemplo registro o participante, determina qué interpretación puede atribuirse a los parámetros de privacidad. Zheng et al. (2021) desarrollan una formulación de privacidad diferencial federada centrada en los registros y muestran que el escenario requiere atender el mecanismo de composición de garantías entre iteraciones.

El mecanismo de ruido permite limitar cierta información observable, aunque también puede deteriorar la calidad del aprendizaje cuando el presupuesto de privacidad es muy restrictivo. Kerkouche et al. (2021) exploran una configuración para dispositivos con ancho de banda limitado, relacionando privacidad diferencial, selección de pesos y eficiencia de comunicación. Esta evidencia indica que el impacto de la protección depende del modelo, el mecanismo y la tarea, por lo que no debe presentarse una cifra de pérdida de precisión como una constante aplicable a todas las implementaciones.

Un error metodológico frecuente consiste en hablar de privacidad diferencial sin informar la unidad protegida, la configuración de ruido, el presupuesto acumulado o las rondas de entrenamiento. En sistemas federados también importa si el coordinador puede inspeccionar contribuciones individuales, si existen abandonos de clientes y cómo se administra el muestreo. Lowy et al. (2023) examinan aprendizaje federado privado sin asumir un servidor confiable, situación que ilustra el valor de declarar los supuestos de amenaza antes de prometer confidencialidad.

**Agregación segura y defensas de integridad**

La agregación segura busca hacer posible el cálculo de una suma o promedio colectivo sin revelar, en condiciones definidas, las contribuciones individuales al coordinador. Su finalidad principal es reducir la exposición de actualizaciones, no detectar automáticamente todos los ejemplos maliciosos ni certificar la legitimidad de cada cliente. Chen et al. (2022) analizan los límites de comunicación asociados con combinar agregación segura y privacidad diferencial y muestran que las garantías de confidencialidad tienen costos técnicos que deben cuantificarse. El efecto sobre la escalabilidad depende de dimensiones del modelo, cantidad de participantes y esquema criptográfico concreto.

La combinación de mecanismos plantea un problema operativo: cuanto menos puede observar un coordinador de las actualizaciones individuales, más difícil puede resultar aplicar verificaciones que requieran inspeccionarlas. El aprendizaje federado confiable demanda, por ello, métodos de validación y reglas de participación compatibles con la arquitectura de privacidad. Saha et al. (2024) organizan estos desafíos dentro de una agenda más amplia de fiabilidad, robustez y seguridad. No hay contradicción en que una tecnología resuelva una amenaza y simultáneamente complique la mitigación de otra; la tensión debe documentarse como resultado del análisis.

Los controles sobre integridad pueden incluir filtrado de comportamientos anómalos, agregación robusta, límites a magnitudes de actualización y evaluaciones de desempeño bajo escenarios adversariales. Su conveniencia depende de supuestos como la proporción de clientes comprometidos y la heterogeneidad esperable. La heterogeneidad legítima puede parecer una anomalía frente a un detector configurado para datos uniformes. En consecuencia, una defensa debe demostrar que distingue, dentro de parámetros razonables, entre divergencias estadísticas válidas y contribuciones intencionadamente dañinas.

## Metodología

### Diseño y delimitación

Se adoptó un diseño documental de revisión narrativa estructurada, con apoyo en búsqueda bibliográfica y análisis temático comparativo. Este trabajo no se declara revisión sistemática ni estudio bibliométrico exhaustivo porque no se ejecutó un protocolo integral de búsqueda, deduplicación, cribado, evaluación del riesgo de sesgo y análisis de redes de publicaciones. Tampoco se reportan recuentos de documentos recuperados o excluidos, debido a que no se dispone de una exportación y registro auditables de todas las consultas. Esta precisión busca que el alcance de los resultados sea congruente con la evidencia efectivamente examinada.

La búsqueda orientada priorizó catálogos y páginas editoriales verificables, incluyendo Proceedings of Machine Learning Research, ACM Digital Library y la plataforma editorial de Foundations and Trends in Machine Learning. Se emplearon combinaciones conceptuales de federated learning, privacy, security, secure aggregation, differential privacy, poisoning, inference attacks, robustness y non-IID. Se privilegiaron contribuciones seminales, trabajos

metodológicos y revisiones capaces de describir mecanismos, amenazas y restricciones. La selección no tiene pretensión de representar estadísticamente toda la producción científica sobre el tema.

Se consideraron pertinentes documentos que abordaran, de forma directa, operaciones de aprendizaje federado, ataques sobre datos o actualizaciones, métodos de privacidad, integridad de modelos o costos computacionales. Se excluyeron del argumento central las fuentes sin relación sustantiva con los mecanismos examinados y se evitaron afirmaciones sustentadas únicamente en descripciones comerciales. La revisión se delimitó conceptualmente al entrenamiento colaborativo de modelos y no a todos los sistemas de cómputo distribuido o a todas las modalidades criptográficas existentes.

**Extracción y análisis de información**

El procedimiento analítico consistió en identificar para cada estudio su problema técnico, el mecanismo investigado, el escenario de confianza, la amenaza tratada, las variables de desempeño y los límites de generalización. Posteriormente se organizaron hallazgos en cuatro categorías: amenazas a la privacidad; ataques contra integridad; mecanismos de protección; y restricciones de cómputo, comunicación y gobernanza. Se compararon los estudios por su contribución conceptual y por las condiciones de aplicación, no por un conteo de citaciones, factor de impacto o cuartil de revista.

La pregunta de investigación se operacionalizó mediante tres interrogantes: qué método se propone, frente a qué amenaza brinda protección y cuál es el costo o limitación para desplegarlo. En la categoría de privacidad se revisaron inferencias, reconstrucciones y garantías formales; en seguridad se distinguieron envenenamiento, puertas traseras y exposición del servidor. En las condiciones operativas se consideraron heterogeneidad estadística, comunicación, participantes intermitentes y precisión predictiva. Esta estructura permitió elaborar una síntesis comparable sin afirmar una metaevaluación cuantitativa inexistente.

La validez del análisis depende de la verificabilidad documental y de la correspondencia entre citas y argumentos. Por ello se incluyeron enlaces directos a registros editoriales o actas académicas en el listado final. No se recopilaron datos personales, se desplegaron experimentos ni se simuló un sistema federado; cualquier recomendación sobre combinaciones de defensas

debe leerse como interpretación basada en bibliografía, pendiente de verificación en un entorno específico. Este reconocimiento delimita la inferencia científica y previene atribuir resultados empíricos originales a una investigación únicamente documental.

## Resultados

### Amenazas identificadas y propiedades comprometidas

La primera conclusión documental es que la localización de los datos no basta para asegurar su confidencialidad. El entrenamiento requiere mensajes que pueden contener señales correlacionadas con observaciones privadas, especialmente cuando un adversario controla el punto de observación o dispone de conocimiento auxiliar. Saha et al. (2024) distinguen mecanismos de inferencia y reconstrucción que no necesitan acceder directamente al archivo de datos original. La protección debe definirse respecto de un atacante concreto y de la información que realmente puede obtener en cada ronda; declarar que los datos nunca abandonan el dispositivo resulta insuficiente como descripción de seguridad.

Una segunda familia agrupa las amenazas contra la integridad de los modelos. Las modificaciones maliciosas de ejemplos o parámetros pueden afectar el resultado global aun cuando la información personal permanezca oculta. Las revisiones recientes de aprendizaje federado confiable consideran estas amenazas junto con los riesgos de privacidad porque convergen en el mismo procedimiento de coordinación, pero demandan defensas diferentes (Saha et al., 2024). Una evaluación adecuada debe registrar tanto la precisión de la tarea legítima como el éxito de ataques dirigidos, la tasa de falsos positivos de las defensas y su sensibilidad a clientes con distribuciones atípicas.

La tercera familia de problemas corresponde a la operación. Las actualizaciones pueden llegar en distintos momentos y los clientes pueden abandonar una ronda. Nguyen et al. (2022) estudian agregación asíncrona amortiguada y su compatibilidad con mecanismos de privacidad, evidenciando que la forma de coordinar mensajes tiene consecuencias sobre escalabilidad y confidencialidad. Una arquitectura resistente no se evalúa únicamente en una situación ideal de dispositivos permanentes, sino ante fallos parciales, variabilidad de redes y recursos heterogéneos.

**Métodos de defensa: alcances y condiciones de aplicación**

La privacidad diferencial constituye una alternativa pertinente cuando se busca definir una garantía formal de limitación de información individual. Zheng et al. (2021) vinculan el análisis con privacidad a nivel de registro, mientras que Lowy et al. (2023) examinan escenarios que prescinden de confianza plena en el servidor. Sus contribuciones no significan que cualquier implementación federada sea diferencialmente privada; implican mecanismos, parámetros y supuestos determinados. En particular, la garantía formal solo corresponde al procedimiento concreto cuya privacidad se demostró o verificó bajo sus condiciones.

La agregación segura protege la confidencialidad de contribuciones en el momento de combinarlas y puede complementarse con privacidad diferencial para reducir distintos tipos de exposición. Chen et al. (2022) indican que las restricciones de comunicación forman parte del costo de alcanzar niveles definidos de privacidad. El aporte de esta familia de técnicas consiste en limitar la información disponible para el agregador, pero no debe confundirse con una capacidad general de identificar clientes deshonestos o reconocer actualizaciones orientadas a manipular el modelo.

Los mecanismos de robustez necesitan operar bajo hipótesis sobre la cantidad de adversarios, el comportamiento de las actualizaciones y la variabilidad legítima. La desviación estadística entre clientes es un problema estructural: Karimireddy et al. (2020) muestran que puede alterar el curso de la optimización incluso sin adversarios. Por tanto, el análisis de una defensa robusta requiere baselines no atacados con diferentes grados de heterogeneidad para separar cambios causados por la defensa, por el entorno y por la acción maliciosa. Ese contraste es necesario para no adjudicar a la seguridad fallas que pertenecen a la optimización.

## Matriz comparativa de métodos y desafíos

**Tabla 1.** *Síntesis analítica de mecanismos y desafíos.*

| Método | Amenaza principal | Fortaleza documentada | Restricción a evaluar |
|---|---|---|---|
| Privacidad diferencial | Inferencia sobre registros | Garantía formal bajo condiciones explícitas | Presupuesto de privacidad y precisión |
| Agregación segura | Visibilidad de aportaciones individuales | Oculta actualizaciones frente al agregador | Comunicación y compatibilidad con auditoría |
| Agregación robusta | Envenenamiento de actualizaciones | Restringe influencia de valores adversariales | No-IID y porcentaje de clientes adversos |
| Optimización para datos heterogéneos | Deriva y convergencia inestable | Atenúa diferencias entre participantes | Costo de estado y rondas de comunicación |
| Agregación asíncrona protegida | Intermitencia y escalabilidad | Gestiona llegadas desiguales de clientes | Protección al agregar grupos reducidos |

**Fuente:** Elaboración propia, 2026, a partir de McMahan et al. (2017), Karimireddy et al. (2020), Zheng et al. (2021), Chen et al. (2022), Nguyen et al. (2022) y Lowy et al. (2023).

La comparación evidencia que no existe una correspondencia uno a uno entre la etiqueta "privacidad" y una protección integral del sistema. Cada método actúa sobre una superficie de ataque particular y puede dejar sin resolver otras propiedades. Por ejemplo, impedir la inspección de una actualización individual no impide necesariamente que una contribución maliciosa altere el promedio colectivo. De manera complementaria, rechazar contribuciones anómalas puede afectar clientes honestos cuyas observaciones sean estadísticamente minoritarias. La decisión de arquitectura debe identificar la amenaza que se considera prioritaria y justificar por qué la protección elegida cubre esa amenaza.

## Desafíos computacionales: precisión, comunicación y heterogeneidad

El costo de comunicación no es un componente accesorio del aprendizaje federado: determina la viabilidad de intercambiar modelos en redes con capacidades y disponibilidad desiguales. McMahan et al. (2017) sitúan la eficiencia comunicacional entre las motivaciones de este paradigma. Aumentar mensajes de verificación, incorporar protocolos criptográficos o añadir rondas de protección puede modificar ese balance. Chen et al. (2022) estudian formalmente el precio de la agregación segura bajo privacidad diferencial, lo que confirma que la confidencialidad debe discutirse junto con recursos, dimensiones del modelo y cantidad de participantes.

La heterogeneidad estadística condiciona la convergencia y la evaluación de mecanismos defensivos. SCAFFOLD propone corregir la deriva de los clientes mediante variables de control (Karimireddy et al., 2020). Pero la heterogeneidad de sistemas incluye también procesadores distintos, conectividad irregular y tiempos variables de actualización. Nguyen et al. (2022) muestran que la coordinación asíncrona puede aportar alternativas para manejar esa variabilidad, aunque deben mantenerse las propiedades de privacidad en la forma de agrupar mensajes. La mejor opción, en consecuencia, no depende de una sola métrica de precisión.

La privacidad diferencial incorpora una relación entre utilidad y grado de protección que demanda pruebas en conjuntos de datos específicos. Kerkouche et al. (2021) vinculan protección y reducción del volumen comunicado mediante una propuesta para dispositivos de bajo ancho de banda. Su resultado constituye evidencia contextual de una arquitectura, no un porcentaje universal transferible a otras aplicaciones. Del mismo modo, los análisis de Lowy et al. (2023) destacan que abandonar el supuesto de un servidor confiable cambia las condiciones de diseño y evaluación. La comparación entre soluciones exige documentar esos supuestos para evitar conclusiones engañosas.

**Respuesta a la pregunta de investigación**

La literatura examinada permite responder que los principales métodos de protección se organizan en tres grupos complementarios. La privacidad diferencial reduce, bajo parámetros especificados, la información inferible acerca de registros; la agregación segura limita la observación de actualizaciones individuales por parte del agregador; y los métodos robustos, junto con controles operativos, buscan preservar integridad y continuidad ante participantes maliciosos o inestables. Estos grupos responden a amenazas distintas y sus propiedades no pueden intercambiarse por simple analogía. La evidencia que sustenta la clasificación procede de trabajos metodológicos y revisiones sobre aprendizaje federado (Kairouz et al., 2021; Saha et al., 2024; Zheng et al., 2021).

Respecto de los desafíos computacionales, las investigaciones muestran dependencias entre confidencialidad, precisión, recursos de comunicación, sincronización y heterogeneidad. Chen et al. (2022) analizan costos de proteger mensajes agregados; Karimireddy et al. (2020) explican problemas de convergencia inducidos por datos heterogéneos; y Nguyen et al. (2022) estudian

la coordinación asíncrona compatible con herramientas de privacidad. A partir de este contraste se infiere que la configuración recomendable debe partir de un modelo de amenaza y después seleccionar y validar las combinaciones de defensas que conserven un desempeño aceptable en el entorno propuesto.

Una respuesta metodológicamente prudente no consiste en declarar una técnica como la mejor de manera universal. Una institución con pocos participantes y comunicación estable puede tener prioridades diferentes a una red de dispositivos móviles intermitentes. Además, la eficacia frente a un atacante que conoce la arquitectura no equivale a la eficacia frente a un servidor curioso o un atacante externo. Los resultados obtenidos en los estudios consultados deben interpretarse dentro de sus diseños originales y no como magnitudes agregadas de una metaevaluación que este trabajo no realizó.

## Discusión

Los El hallazgo conceptual más relevante es la separación entre aprendizaje distribuido y garantía de confidencialidad. El paradigma descrito por McMahan et al. (2017) reduce la centralización de datos, pero su arquitectura no establece por sí misma privacidad diferencial ni protección absoluta frente a inferencias. Las revisiones de Kairouz et al. (2021) y Saha et al. (2024) ponen en relación el potencial operativo del aprendizaje federado con ataques y restricciones técnicas que persisten después de distribuir el entrenamiento. Esta tensión obliga a revisar los argumentos habituales de adopción: la preservación del dato local es una condición arquitectónica favorable, no una demostración suficiente de seguridad.

Otra tensión emerge entre privacidad e inspeccionabilidad. La agregación segura disminuye la información que ve el coordinador, mientras que ciertas defensas de integridad se apoyan en observar anomalías en contribuciones individuales. Por ello, la protección de la confidencialidad puede obligar a rediseñar el control de calidad de actualizaciones, antes que simplemente añadir dos componentes independientes. El análisis de costos de Chen et al. (2022) y la discusión de problemas abiertos de Kairouz et al. (2021) respaldan esta interpretación. La compatibilidad entre defensas debería verificarse mediante pruebas en lugar de suponerse a partir de sus propiedades aisladas.

También se identifica una limitación de comparabilidad. Los artículos pueden diferir en unidad protegida, proporción de actores maliciosos, complejidad del modelo y condiciones de distribución de datos. Los resultados de Karimireddy et al. (2020) muestran que la heterogeneidad afecta la optimización incluso sin ataques; Nguyen et al. (2022) añaden que la modalidad de agregación modifica la escala operativa. Una afirmación causal sobre superioridad de una defensa requiere, por tanto, diseños experimentales que mantengan constantes los factores pertinentes o modelen explícitamente sus diferencias. Esta revisión no permite establecer tal superioridad global.

## Discusión

En contextos interinstitucionales, como colaboración entre hospitales o empresas, el diseño requiere acuerdos acerca de quién controla la infraestructura, quién puede actualizar modelos y qué información puede observar cada participante. Las aproximaciones de privacidad sin servidor plenamente confiable analizadas por Lowy et al. (2023) muestran que no todas las organizaciones pueden asumir una entidad coordinadora ajena a conflictos de interés. La seguridad se relaciona así con decisiones organizativas, aunque las propiedades técnicas continúan siendo indispensables. El control de acceso y la responsabilidad operativa no sustituyen pruebas formales de privacidad.

La literatura sobre aprendizaje federado confiable permite formular una agenda de evaluación más precisa: una defensa debe probarse contra amenazas definidas, con medidas de utilidad, costos y robustez bajo heterogeneidad. Los trabajos de Zheng et al. (2021), Chen et al. (2022) y Nguyen et al. (2022) ilustran dimensiones diferentes de esta evaluación, que no se reducen a reportar exactitud predictiva. La recomendación derivada es documentar, además, estabilidad por cliente, presupuesto de privacidad acumulado, fallos tolerados, latencia y degradación frente a ataques, siempre que tales variables correspondan al escenario de aplicación.

Las limitaciones de esta investigación también deben reconocerse. La búsqueda fue orientada y la selección tuvo un propósito analítico, de modo que no se puede inferir exhaustividad temática ni distribución temporal de la evidencia. No se calcularon indicadores bibliométricos ni tamaños de efecto agrupados; tampoco se aplicó una evaluación formal del riesgo de sesgo de todos los estudios. Algunas técnicas están mejor representadas por trabajos metodológicos

que por implementaciones en producción, lo cual restringe conclusiones sobre desempeño operativo. Una revisión sistemática futura podría definir bases, ecuaciones, fechas, criterios de exclusión y un protocolo de evaluación reproducible.

**Modelos de amenaza y supuestos de confianza**

Una lectura comparada de las investigaciones revisadas indica que la selección de una defensa comienza antes de escoger el algoritmo de protección. Es necesario identificar qué participante podría actuar de manera adversarial, qué información observa, qué capacidad de coordinación posee y en qué fase del entrenamiento puede intervenir. La amenaza puede provenir de un servidor que intenta inferir información, de clientes que envían actualizaciones manipuladas o de participantes externos que observan comunicaciones. Kairouz et al. (2021) y Mothukuri et al. (2021) muestran que esos escenarios no admiten una misma respuesta técnica. Por ello, una arquitectura descrita únicamente como federada informa dónde se entrenan los datos, pero no acredita por sí misma confidencialidad, integridad ni disponibilidad.

El servidor constituye un punto analítico particularmente delicado. En el esquema clásico de agregación, coordina rondas, recibe contribuciones y distribuye parámetros actualizados; sin controles adicionales, esa función puede conferirle capacidad de observación sobre señales derivadas de los datos de origen. La agregación segura modifica esa posibilidad al permitir operaciones sobre contribuciones protegidas, aunque no necesariamente impide que el modelo agregado revele tendencias de subgrupos pequeños. Chen et al. (2022) explican la relación entre agregación segura y aprendizaje federado diferencialmente privado, mientras que Lowy et al. (2023) estudian aprendizaje privado sin presuponer un servidor confiable. La diferencia entre confiar en la infraestructura y ofrecer garantías que reduzcan la necesidad de confianza debe hacerse explícita en el diseño.

La seguridad frente a clientes maliciosos presenta otra dificultad: los mecanismos que ocultan actualizaciones individuales también pueden limitar la inspección destinada a detectar contribuciones anómalas. Una protección orientada a confidencialidad podría entrar en tensión con una auditoría detallada de integridad. Manzoor et al. (2024) organizan distintas estrategias de defensa del modelo y de los datos, y Ma y Yan (2024) describen amenazas que atraviesan varias capas del aprendizaje federado. De ahí que la integración entre detección de anomalías

y agregación protegida no deba darse por resuelta: requiere definir qué señales de diagnóstico permanecen disponibles y qué consecuencias tiene su exposición para los participantes.

**Privacidad diferencial y unidad de protección**

La formulación de una garantía de privacidad diferencial requiere especificar qué se entiende por registro vecino y, por tanto, qué entidad resulta protegida. En aplicaciones con múltiples observaciones por participante, no equivale proteger un registro aislado a limitar la influencia completa de un dispositivo o institución. Esta distinción cambia el tamaño potencial de la contribución, las condiciones de muestreo y el presupuesto de privacidad. Kairouz et al. (2021) presentan la privacidad diferencial como una de las líneas relevantes de investigación, y Zheng et al. (2021) estudian su formulación mediante privacidad f-diferencial. En consecuencia, no basta con mencionar que un sistema incorpora ruido: debe especificarse el alcance de la garantía y las condiciones bajo las cuales es válida.

La acumulación de rondas de entrenamiento introduce un segundo problema. Una intervención que parece acotada en una sola comunicación puede implicar un consumo acumulado del presupuesto de privacidad cuando el entrenamiento repite consultas sobre contribuciones relacionadas. Los resultados de los estudios sobre mecanismos distribuidos recuerdan que los parámetros de protección tienen sentido dentro de una composición de operaciones, no como etiquetas aisladas. Girgis et al. (2021) examinan el modelo mezclado de privacidad diferencial, mientras que Chen et al. (2024) estudian el diseño de experimentos federados bajo privacidad diferencial distribuida. Estas contribuciones permiten reconocer que la garantía depende tanto del mecanismo como de su implementación y del procedimiento de evaluación.

El costo estadístico de la privacidad tampoco puede describirse mediante una regla simple aplicable a cualquier conjunto de datos. La perturbación y el recorte de actualizaciones pueden reducir la información útil transmitida, pero el efecto observado depende de la arquitectura, de la distribución de datos, del número de participantes y del objetivo de optimización. Kerkouche et al. (2021) estudian restricciones de aprendizaje diferencialmente privado en dispositivos de bajo ancho de banda, mientras que Chen et al. (2022) identifican relaciones entre los costos de la agregación segura y la protección diferencial. Por consiguiente, cualquier afirmación sobre

sacrificio de precisión necesita una comparación bajo condiciones de entrenamiento equivalentes; esta revisión no produce estimaciones propias de esa pérdida.

**Heterogeneidad, participación y sesgos de evaluación**

La heterogeneidad entre participantes influye simultáneamente en convergencia, calidad predictiva y viabilidad de las protecciones. Una red con clientes de distintas capacidades técnicas y con datos distribuidos de forma desigual enfrenta restricciones diferentes a las de una simulación homogénea. McMahan et al. (2017) presentan una estrategia de aprendizaje con datos descentralizados orientada a eficiencia de comunicación; Karimireddy et al. (2020) desarrollan SCAFFOLD para corregir desplazamientos ocasionados por heterogeneidad. Estas aportaciones no pueden interpretarse como soluciones universales al problema de seguridad, pero sí revelan que la evaluación de mecanismos defensivos debe controlar la dinámica de optimización que les sirve de base.

La participación parcial genera otro mecanismo de selección. Cuando solamente un subconjunto de dispositivos participa en cada ronda, el sistema aprende de contribuciones condicionadas por conectividad, disponibilidad de energía y reglas de convocatoria. Esa selección puede producir disparidades entre usuarios o grupos de dispositivos, incluso cuando la arquitectura no centraliza observaciones originales. Kairouz et al. (2021) identifican la disponibilidad irregular de clientes entre los problemas abiertos del campo. Desde una perspectiva metodológica, conviene reportar distribución del número de participaciones, composición de clientes por ronda y diferencias en desempeño entre subpoblaciones; sin estos datos, una métrica global elevada podría ocultar inestabilidad o cobertura desigual.

Los esquemas asíncronos intentan reducir la dependencia de rondas rígidamente sincronizadas, aunque introducen desafíos adicionales de obsolescencia de actualizaciones y coordinación. Nguyen et al. (2022) analizan agregación asíncrona con búfer; su propuesta permite discutir cómo la velocidad de llegada de los clientes puede interactuar con los mecanismos de protección. La pertinencia de la estrategia depende de la aplicación: una modificación destinada a mejorar disponibilidad no constituye automáticamente una defensa contra un adversario. Además, el uso de búferes y de diferentes periodos de acumulación plantea decisiones sobre qué contribuciones se combinan y bajo qué condiciones se preservan las garantías declaradas.

El problema de la heterogeneidad no se limita a la convergencia. También afecta la evaluación de ataques, porque participantes con distribuciones genuinamente distintas pueden producir gradientes que parecen anómalos sin ser maliciosos. Un detector de actualizaciones atípicas podría penalizar clientes minoritarios o territorios con condiciones de observación particulares. Las revisiones de Mothukuri et al. (2021) y Manzoor et al. (2024) permiten sostener que la robustez depende del escenario adversarial y de los supuestos operativos. La consecuencia analítica es que la tasa de detección debe leerse junto con falsos positivos, diferencias entre clientes y costos de exclusión. Sin esa evaluación conjunta, mayor severidad defensiva no equivale necesariamente a mayor confiabilidad del sistema.

**Criterios mínimos para comparar mecanismos de protección**

Para que dos investigaciones sean comparables deben describir, como mínimo, la unidad protegida, el atacante supuesto, el tipo de datos, la topología de comunicación, la regla de participación y el mecanismo de agregación. También conviene especificar las condiciones experimentales relativas a particiones no independientes, número de rondas y capacidad de los dispositivos. Ma y Yan (2024) y Saha et al. (2024) muestran la amplitud de configuraciones que se agrupan bajo el término aprendizaje federado. Esta diversidad vuelve problemático ordenar defensas a partir de una sola métrica. Una revisión documental puede identificar criterios compartidos, pero no transformar estudios heterogéneos en una clasificación cuantitativa sin un protocolo de armonización y evaluación.

La precisión predictiva debe reportarse de forma complementaria con evidencia sobre filtración de información o resistencia al ataque que la intervención busca disminuir. Una defensa puede obtener resultados adecuados de clasificación y, sin embargo, mantener oportunidades de inferencia; de modo inverso, puede imponer una restricción rigurosa de privacidad sin ofrecer calidad suficiente para la finalidad operativa. Chen et al. (2022) y Girgis et al. (2021) ilustran, desde perspectivas distintas, la necesidad de distinguir entre utilidad y garantías de privacidad. En estudios posteriores sería pertinente analizar sensibilidad a parámetros, estabilidad entre ejecuciones y pérdida de desempeño relativa a una línea de base común, evitando equiparar una mejora de exactitud con una demostración de seguridad.

El costo computacional tampoco debería reducirse al tiempo total de entrenamiento. Es necesario separar procesamiento local, tamaño de mensajes, cantidad de rondas, memoria, consumo energético cuando se mide y tiempo de recuperación ante fallos. Las limitaciones de ancho de banda descritas por Kerkouche et al. (2021) y la motivación de eficiencia comunicacional en McMahan et al. (2017) muestran que la factibilidad está condicionada por recursos disponibles. En contextos con dispositivos limitados, una defensa teóricamente apropiada podría ser inviable si exige demasiada comunicación o sincronización. La comparación debe incluir condiciones reales o claramente definidas de infraestructura, sin inferir desempeños que la literatura examinada no reporta.

Las decisiones sobre parámetros deben presentarse de manera transparente. En privacidad diferencial, corresponde declarar el presupuesto y el procedimiento de contabilización; en agregación segura, el umbral de participantes, el modelo de confianza y las consecuencias de desconexión; en defensas de integridad, los tipos de ataques usados y la proporción de clientes adversariales simulada. Kairouz et al. (2021) y Manzoor et al. (2024) sostienen la necesidad de examinar estos problemas de forma conjunta. La validez externa de una evaluación no depende únicamente de reproducir un resultado promedio: exige saber si el escenario representa la red y las amenazas para las cuales se pretende recomendar la solución.

**Implicaciones metodológicas y agenda de investigación**

El conjunto de fuentes revisadas permite formular una agenda de investigación orientada a combinaciones de defensas y no exclusivamente a mecanismos aislados. Un protocolo experimental plausible compararía una línea de base federada, un tratamiento con agregación segura, otro con privacidad diferencial y una combinación de ambos, manteniendo constantes los datos, el presupuesto de cómputo y los esquemas de participación. La comparación incluiría precisión, riesgo de filtración bajo ataques definidos, tráfico de red y fallos por desconexión. Este planteamiento constituye una propuesta metodológica derivada de la literatura, no un experimento realizado en el presente artículo. Su utilidad consiste en hacer explícitas las decisiones que deberían observarse para valorar los compromisos entre protección y desempeño.

Una segunda línea requiere estudiar modelos de confianza con y sin servidor honesto. Lowy et al. (2023) ofrecen una contribución sobre aprendizaje federado privado sin servidor confiable; Chen et al. (2022) muestran restricciones relevantes para la articulación entre agregación segura y privacidad diferencial. Comparar estos enfoques exigiría precisar qué actor conserva claves, qué mensajes pueden ser observados y cómo se gestionan los abandonos de clientes. El hallazgo analítico es que la distribución física del entrenamiento y la distribución de la confianza son dimensiones distintas. La primera describe dónde se procesan datos; la segunda define quién puede observarlos, manipularlos o inferir información a partir de los intercambios.

Una tercera agenda consiste en estudiar la robustez junto con los efectos diferenciales sobre usuarios heterogéneos. Sería necesario construir escenarios donde algunos clientes presenten distribuciones legítimamente minoritarias y otros envíen mensajes adversariales. La evaluación debería distinguir entre ambos grupos y reportar no solo el desempeño agregado sino también los errores de exclusión y sus consecuencias. Las discusiones de Karimireddy et al. (2020), Mothukuri et al. (2021) y Saha et al. (2024) ofrecen bases para esta distinción entre heterogeneidad estadística y amenaza deliberada. Aun así, los trabajos revisados no permiten concluir que una defensa específica sea la mejor para todos los sistemas y conjuntos de datos.

Finalmente, el análisis documental señala una necesidad de estándares descriptivos que favorezcan la reproducibilidad. Cada publicación debería distinguir con claridad qué se demuestra formalmente, qué se observa experimentalmente y qué se propone como posibilidad de implementación. Una garantía matemática definida bajo determinados supuestos no verifica que un despliegue real satisfaga esos supuestos; igualmente, un ataque exitoso en una simulación no establece automáticamente su frecuencia en aplicaciones productivas. La revisión de Kairouz et al. (2021), Ma y Yan (2024) y Manzoor et al. (2024) respalda la pertinencia de examinar amenazas, defensas y contexto de aplicación conjuntamente. Esta delimitación constituye una condición para trasladar resultados académicos a decisiones técnicas justificables.

**Alcances de la evidencia y transferencia a sistemas reales**

La extrapolación desde entornos experimentales a aplicaciones institucionales requiere comprobar que las condiciones de conectividad, gestión de identidades y supervisión coincidan

con las asumidas por el mecanismo. En un despliegue con varios centros, por ejemplo, la capacidad de gestionar pérdidas de conexión y la identificación de contribuciones autorizadas puede resultar tan importante como la elección del algoritmo de agregación. La bibliografía revisada trata estas dimensiones desde distintos problemas: Kairouz et al. (2021) sistematizan desafíos abiertos del aprendizaje federado y Nguyen et al. (2022) analizan agregación con participación asíncrona. Sin embargo, no sería correcto trasladar automáticamente sus resultados a una institución determinada. Se requerirían mediciones locales, pruebas de resistencia y procedimientos explícitos de mantenimiento para establecer si una configuración satisface necesidades operativas concretas.

La validez de las afirmaciones de seguridad depende, además, de mantener una distinción entre protección durante la comunicación, protección de las contribuciones y protección del modelo final. El cifrado del canal puede restringir la observación de un tercero en tránsito, pero no impide por sí mismo que quien recibe una actualización legítima intente extraer información. De manera semejante, limitar la observación individual mediante agregación segura no certifica que un modelo terminado no conserve señales susceptibles de inferencia. Mothukuri et al. (2021) y Saha et al. (2024) ofrecen antecedentes para reconocer esa diversidad de superficies de ataque. El diseño de un sistema debería declarar con precisión la frontera cubierta por cada defensa y evitar expresiones generales como sistema completamente privado cuando no se dispone de garantías que sostengan esa afirmación.

La gobernanza del sistema agrega decisiones técnicas que frecuentemente quedan fuera de las métricas predictivas. Entre ellas se encuentran la selección y revocación de participantes, la gestión de claves, el registro de cambios de configuración, el tratamiento de fallos y la auditoría de modificaciones del modelo. Ninguna de estas operaciones garantiza privacidad por sí misma, pero su ausencia puede invalidar supuestos relevantes de la arquitectura. Los análisis de Ma y Yan (2024) y Manzoor et al. (2024) permiten situar estos controles dentro de una perspectiva amplia de riesgos. En este artículo se consideran implicaciones inferidas de la síntesis bibliográfica, no resultados de observación directa sobre organizaciones. Para contrastarlas empíricamente sería preciso analizar implementaciones concretas y documentar sus configuraciones.

En suma, la comparación de defensas debe responder una secuencia de preguntas: ¿qué información se pretende preservar?, ¿frente a quién?, ¿con qué garantías?, ¿qué condiciones de operación exige el mecanismo? y ¿qué efectos produce sobre utilidad y disponibilidad? Esta secuencia articula los hallazgos de McMahan et al. (2017), Chen et al. (2022), Kairouz et al. (2021) y Lowy et al. (2023) sin presentarlos como resultados equivalentes. También delimita el alcance de la presente revisión: la literatura permite identificar familias de mecanismos, sus supuestos y tensiones observadas o teóricas, pero no determinar con validez universal el nivel de protección de cualquier sistema federado. Esa determinación requiere una evaluación situada del riesgo y de los recursos

## Conclusiones

La revisión permite concluir que la privacidad y la seguridad en aprendizaje federado constituyen problemas relacionados, pero analíticamente diferentes. La privacidad diferencial establece garantías formales respecto de información individual bajo parámetros determinados; la agregación segura reduce la observabilidad de contribuciones; y los métodos de robustez buscan preservar la integridad del entrenamiento. La coexistencia de estos mecanismos no implica cobertura completa y automática. La investigación documental muestra que deben seleccionarse de acuerdo con la amenaza concreta, el nivel de confianza entre actores y las condiciones de participación.

Los desafíos computacionales más visibles se vinculan con el costo de comunicación, la pérdida potencial de precisión, la heterogeneidad de datos y equipos y la compatibilidad entre mecanismos de protección. La evidencia de trabajos fundacionales y contribuciones recientes permite sostener que no existe una configuración universalmente superior. En consecuencia, se propone evaluar arquitecturas mediante matrices que relacionen adversario, activo protegido, técnica, garantía, métrica y costo, complementadas con ensayos que distingan efectos de datos heterogéneos y ataques maliciosos.

Como aportación, el artículo organiza una síntesis de amenazas y defensas que responde explícitamente la pregunta de investigación y evita confundir descentralización con confidencialidad garantizada. Su límite principal es documental: no cuantifica la producción científica global ni prueba de manera experimental nuevas defensas. Las investigaciones

posteriores podrían comparar combinaciones de privacidad diferencial, agregación segura y robustez en escenarios no-IID, reportando tanto eficacia defensiva como costo computacional. La generalización de esas pruebas deberá quedar circunscrita a sus datos, hipótesis de amenaza y condiciones técnicas.

## Referencias bibliográficas

Chen, W.-N., Choo, C. A. C., Kairouz, P., & Suresh, A. T. (2022). The fundamental price of secure aggregation in differentially private federated learning. Proceedings of Machine Learning Research, 162, 3056–3089. https://proceedings.mlr.press/v162/chen22c.html

Chen, W.-N., Cormode, G., Bharadwaj, A., Romov, P., & Ozgur, A. (2024). Federated experiment design under distributed differential privacy. Proceedings of Machine Learning Research, 238, 2458–2466. https://proceedings.mlr.press/v238/chen24c.html

Girgis, A., Data, D., Diggavi, S., Kairouz, P., & Suresh, A. T. (2021). Shuffled model of differential privacy in federated learning. Proceedings of Machine Learning Research, 130, 2521–2529. https://proceedings.mlr.press/v130/girgis21a.html

Kairouz, P., McMahan, H. B., Avent, B., et al. (2021). Advances and open problems in federated learning. Foundations and Trends in Machine Learning, 14(1–2), 1–210. https://doi.org/10.1561/2200000083

Karimireddy, S. P., Kale, S., Mohri, M., Reddi, S., Stich, S., & Suresh, A. T. (2020). SCAFFOLD: Stochastic controlled averaging for federated learning. Proceedings of Machine Learning Research, 119, 5132–5143. https://proceedings.mlr.press/v119/karimireddy20a.html

Kerkouche, R., Ács, G., Castelluccia, C., & Genevès, P. (2021). Constrained differentially private federated learning for low-bandwidth devices. Proceedings of Machine Learning Research, 161, 1756–1765. https://proceedings.mlr.press/v161/kerkouche21a.html

Lowy, A., Ghafelebashi, A., & Razaviyayn, M. (2023). Private non-convex federated learning without a trusted server. Proceedings of Machine Learning Research, 206, 5749–5786. https://proceedings.mlr.press/v206/lowy23a.html

Ma, X., & Yan, M. (2024). Research progress on security and privacy of federated learning: A survey. Wireless Personal Communications, 136(4), 2201–2242. https://doi.org/10.1007/s11277-024-11372-0

Mangold, P., Durmus, A., Dieuleveut, A., & Moulines, E. (2025). Scaffold with stochastic gradients: New analysis with linear speed-up. Proceedings of Machine Learning Research, 267, 42902–42946. https://proceedings.mlr.press/v267/mangold25a.html

Manzoor, H. U., Shabbir, A., Chen, A., Flynn, D., & Zoha, A. (2024). A survey of security strategies in federated learning: Defending models, data, and privacy. Future Internet, 16(10), Article 374. https://doi.org/10.3390/fi16100374

McMahan, B., Moore, E., Ramage, D., Hampson, S., & Agüera y Arcas, B. (2017). Communication-efficient learning of deep networks from decentralized data. Proceedings of Machine Learning Research, 54, 1273–1282. https://proceedings.mlr.press/v54/mcmahan17a.html

Mothukuri, V., Parizi, R. M., Pouriyeh, S., Huang, Y., Dehghantanha, A., & Srivastava, G. (2021). A survey on security and privacy of federated learning. Future Generation Computer Systems, 115, 619–640. https://doi.org/10.1016/j.future.2020.10.007

Nguyen, J., Malik, K., Zhan, H., Yousefpour, A., Rabbat, M., Malek, M., & Huba, D. (2022). Federated learning with buffered asynchronous aggregation. Proceedings of Machine Learning Research, 151, 3581–3607. https://proceedings.mlr.press/v151/nguyen22b.html

Saha, S., Hota, A., Chattopadhyay, A. K., Nag, A., & Nandi, S. (2024). A multifaceted survey on privacy preservation of federated learning: Progress, challenges, and opportunities. Artificial Intelligence Review, 57, Article 184. https://doi.org/10.1007/s10462-024-10766-7

Zheng, Q., Chen, S., Long, Q., & Su, W. (2021). Federated f-differential privacy. Proceedings of Machine Learning Research, 130, 2251–2259. https://proceedings.mlr.press/v130/zheng21a.html

## Declaraciones finales

**Conflicto de intereses:** El autor declara no tener ningún conflicto de interés relacionado con el presente trabajo.

**Financiamiento:** Esta investigación no recibió financiamiento externo de agencias públicas, comerciales o sin fines de lucro.

**Agradecimiento:** Los autores agradecen al Tecnológico Nacional de México Campus Oaxaca por el respaldo académico e institucional brindado para el desarrollo de esta investigación. Al primer autor por la dirección del trabajo de investigación y al resto de autores por sus aportes y la investigación de campo.

**Uso de inteligencia artificial generativa:** NA.

**Nota editorial:** El presente artículo es original e inédito y no es producto de una publicación anterior. Dirección de la investigación - Francisco Ricardo Baños Solis. Elaboración de la investigación: Emilio Raúl Peña Pérez, Mario Sergio Lerín Cruz, Maria Fernanda Cortes Martinez, Manuel de Jesús Melo Monterrey.