HALO: Uma Abordagem Quântica para Privacidade Diferencial com Robustez Geométrica e Alta Utilidade

Adriano Maia[1,2], Isys Sant'Anna [1,2], Marcus Freire[1,2], Thiago Mello[1,2], Fêlipe R. de Araújo[2], Anderson Tomkelski[1], João Marcelo[1], Bruno Santos[2], Cassio Prazeres[2], Gustavo Figueiredo[2], Maycon Peixoto[2]

[1]QuIIN Quantum Industrial Innovation,
Centro de Competência EMBRAPII CIMATEC em Tecnologias Quânticas
{marcus.elias,anderson.tomkelski,joao.marcelo}@fieb.org.br

[2]Universidade Federal da Bahia (UFBA)
{adriano.maia, isys.nogueira, thiagomello, feliperosario, maycon.leone}@ufba.br

{bruno.ps, prazeres, gustavobf}@ufba.br

Abstract. *The deployment of Differential Privacy in Federated Learning frameworks often imposes a severe utility trade-off, destabilizing the convergence of classical neural networks within the Cloud-Edge Continuum. This paper introduces HALO (Hybrid Algorithms Learning on Orbits), a novel approach based on Variational Quantum Circuits that leverages the periodic and bounded geometry of quantum parameters to mitigate stochastic noise. Comparative experiments demonstrate that, within the rigorous privacy regime (*$1.0 \leq \epsilon \leq 3.0$*), traditional architectures experience prediction collapse (*$\sim 67\%$ *accuracy), whereas HALO maintains superior stability with* 98.9% *accuracy. Furthermore, risk auditing confirms effective resilience against Membership Inference Attacks (MIA* < 0.60*). These results validate quantum geometric robustness as a superior alternative for reconciling high predictive performance with formal privacy guarantees in distributed environments.*

Resumo. *A integração de Privacidade Diferencial em sistemas de Aprendizado Federado impõe um dilema crítico de utilidade, frequentemente desestabilizando a convergência de arquiteturas clássicas no Cloud-Edge Continuum. Este trabalho introduz o HALO (Hybrid Algorithms Learning on Orbits), uma abordagem fundamentada em Circuitos Quânticos Variacionais que explora a geometria periódica e limitada dos parâmetros quânticos para mitigar o impacto do ruído estocástico. Experimentos comparativos revelam que, no intervalo rigoroso de privacidade (*$1.0 \leq \epsilon \leq 3.0$*), as redes neurais tradicionais sofrem colapso de predição (*$\sim 67\%$ *de acurácia), enquanto o HALO sustenta estabilidade superior com* 98.9% *de acurácia. Adicionalmente, a auditoria de risco confirma a eficácia da proteção contra ataques de inferência de membros (MIA* < 0.60*). Os resultados validam a robustez geométrica quântica como uma alternativa superior para conciliar alto desempenho preditivo e garantias formais de privacidade em ambientes distribuídos.*

1. Introdução

A consolidação do *Cloud-Edge Continuum* [Gkonis et al. 2023] transformou o Aprendizado Federado (*Federated Learning*– FL) em um componente central da *Edge Intelli-*

gence, ao viabilizar o treinamento colaborativo de modelos diretamente na borda, sem a centralização de dados sensíveis e com ganhos em latência e uso de largura de banda [McMahan et al. 2017]. Contudo, essa descentralização não implica garantias robustas de confidencialidade: atualizações de gradientes compartilhadas durante o treinamento federado carregam informações estatísticas suficientes para viabilizar ataques sofisticados, como o *Membership Inference Attack* (MIA), capazes de inferir a participação de registros individuais no conjunto de treinamento [Shokri et al. 2017, Nasr et al. 2019]. Em aplicações sensíveis, como saúde digital, redes veiculares [Maia et al. 2025, Peixoto 2024] e automação residencial, essa vulnerabilidade torna indispensável a adoção de mecanismos formais de privacidade [Freire et al. 2025].

A Privacidade Diferencial (*Differential Privacy* – DP) se estabeleceu como o padrão de referência para mitigar tais riscos, ao introduzir ruído estocástico calibrado nos parâmetros do modelo antes de sua comunicação [Dwork et al. 2006]. Contudo, a integração da DP com arquiteturas clássicas de *Deep Learning* na borda impõe um desafio estrutural. Redes neurais tradicionais operam em espaços euclidianos ilimitados, nos quais perturbações aditivas necessárias para assegurar baixos valores de ϵ tendem a deslocar os pesos para regiões de instabilidade, comprometendo a convergência e degradando a utilidade do modelo global [Abadi et al. 2016]. Esse fenômeno estabelece um impasse fundamental no *Cloud-Edge Continuum*: garantias rigorosas de privacidade frequentemente inviabilizam o desempenho preditivo requerido por sistemas reais.

Nesse cenário, o Aprendizado de Máquina Quântico (*Quantum Machine Learning* – QML) desponta como uma alternativa promissora, não apenas pela perspectiva de aceleração computacional, mas pela natureza geométrica de seus modelos. Circuitos Quânticos Variacionais (*Variational Quantum Circuits* – VQCs), adequados à era *Noisy Intermediate Scale Quantum* (NISQ), são parametrizados por rotações em espaços de Hilbert naturalmente periódicos e limitados [Mitarai et al. 2018]. Apesar do crescente interesse na expressividade e capacidade de generalização desses modelos, permanece amplamente inexplorado como essa geometria não euclidiana pode ser explorada como um mecanismo intrínseco de robustez à injeção deliberada de ruído, particularmente em cenários híbridos de Aprendizado Federado com agregação clássica [Schuld et al. 2021].

Motivado por essa lacuna, este trabalho apresenta o HALO (*Hybrid Algorithms Learning on Orbits*), uma arquitetura de Aprendizado Federado projetada para o *Cloud-Edge Continuum* que explora a periodicidade dos parâmetros quânticos como um mecanismo natural de regularização contra o ruído imposto pela Privacidade Diferencial. A hipótese central investigada é que a dinâmica orbital dos VQCs permite absorver perturbações estocásticas sem comprometer a estabilidade da convergência global.

As principais contribuições deste trabalho são: (i) a demonstração empírica de que clientes quânticos mantêm alta utilidade preditiva em regimes rigorosos de privacidade nos quais modelos clássicos sofrem colapso de convergência ($1.0 \leq \epsilon \leq 3.0$); (ii) a identificação de um *Sweet Spot Quântico*, no qual o equilíbrio entre acurácia global e risco de ataques de inferência é significativamente otimizado; e (iii) a validação de que a geometria não euclidiana dos circuitos variacionais oferece uma estabilidade estocástica superior, viabilizando a orquestração segura de FL em ambientes distribuídos.

2. Trabalhos Relacionados

A Privacidade Diferencial é amplamente adotada como mecanismo formal para mitigar vazamentos de informação no aprendizado de máquina. O trabalho de [Abadi et al. 2016] estabeleceu o DP-DL por meio de *gradient clipping* e ruído calibrado; contudo, estudos posteriores, como [Ponomareva et al. 2023], demonstram que tais mecanismos continuam impondo perdas expressivas de acurácia em modelos de larga escala. No Aprendizado Federado, essas limitações são amplificadas: o FedAvg [McMahan et al. 2017] pressupõe atualizações estáveis dos clientes, mas a combinação entre heterogeneidade estatística dos dados (non-IID) e injeção de ruído diferencial exacerba a instabilidade da convergência, como evidenciado por [Yu et al. 2026], comprometendo a fidelidade do modelo global em ambientes de borda.

A necessidade de mecanismos formais de proteção é reforçada pela viabilidade de ataques de inferência de membros. Estudos demonstraram que ataques MIA são eficazes tanto em modelos de caixa-preta [Shokri et al. 2017] quanto em cenários federados e de caixa-branca [Nasr et al. 2019], confirmando que atualizações de gradientes carregam informações suficientes para inferir a participação de dados individuais. Esses resultados evidenciam que a descentralização do Aprendizado Federado não constitui, por si só, uma garantia de sigilo.

No domínio da computação quântica, o *Quantum Circuit Learning* [Mitarai et al. 2018] demonstrou a capacidade de VQCS em aproximar funções complexas. A literatura subsequente se concentrou na expressividade dos *feature maps* e na busca por vantagem computacional [Schuld et al. 2021], enquanto análises de robustez frente a ruído e privacidade permanecem limitadas ou restritas a cenários centralizados, mesmo quando se considera a resiliência passiva induzida pelo ruído intrínseco do hardware quântico [Du et al. 2021]. Abordagens recentes de *Quantum Federated Learning* (QFL) [Li et al. 2021] e propostas voltadas ao contínuo nuvem-borda priorizam eficiência de comunicação ou latência de inferência, deixando em aberto a estabilidade da convergência e a incorporação de mecanismos formais de Privacidade Diferencial, conforme sintetizado na Tabela 1.

Tabela 1. Trabalhos Relacionados. O símbolo (•) indica suporte total, (○) suporte parcial e (−) ausência da característica.

Artigos	FL	ϵ-DP	VQC	Robustez Geométrica	Cloud-Edge
Ponomavera et al. [2023]	-	•	-	-	-
Yu et al. [2026]	•	○	-	-	•
Nasr et al. [2019]	•	-	-	-	•
Mitarai et al. [2018]	-	-	•	-	-
Du et al. [2021]	-	○	•	○	-
Li et al. [2021]	•	-	•	-	○
HALO (Proposto)	•	•	•	•	•

Abordagens clássicas sofrem degradação severa sob DP, enquanto propostas quânticas existentes carecem de mecanismos formais de proteção ou permanecem restritas a topologias centralizadas. O método proposto, HALO, diferencia-se ao explorar a geometria periódica dos circuitos quânticos variacionais como um mitigador intrínseco do ruído diferencial, conciliando Aprendizado Federado, Privacidade Diferencial rigorosa e estabilidade de convergência no *Cloud-Edge Continuum*.

3. HALO - *Hybrid Algorithms Learning on Orbits*

3.1. Visão Geral

O HALO é uma arquitetura de Aprendizado Federado projetada para operar de forma robusta no *Cloud-Edge Continuum*, conciliando garantias rigorosas de Privacidade Diferencial com estabilidade de convergência. A proposta parte da hipótese de que a degradação da utilidade observada em modelos clássicos sob ruído diferencial está fortemente associada à natureza ilimitada do espaço euclidiano em que seus parâmetros residem. Em contraste, circuitos quânticos variacionais operam em espaços paramétricos periódicos e limitados, oferecendo uma base geométrica naturalmente mais estável frente a perturbações estocásticas. O HALO explora essa propriedade ao integrar clientes quânticos na borda com agregação clássica na nuvem, utilizando a geometria orbital dos parâmetros quânticos como um mecanismo intrínseco de regularização contra o ruído imposto pela DP.

3.2. Definição do Problema

Considera-se um sistema de Aprendizado Federado composto por K clientes distribuídos, cada um possuindo um conjunto de dados locais D_k, e um servidor central responsável pela agregação do modelo global. O objetivo é minimizar a função de perda empírica global sem acesso direto aos dados locais, definida como $\min_\Theta F(\Theta) = \sum_{k=1}^{K} \frac{|D_k|}{|D|} L_k(\Theta)$, onde $|D| = \sum_{k=1}^{K} |D_k|$ e $L_k(\cdot)$ é a função de custo local do cliente k. O principal desafio é assegurar que as atualizações compartilhadas durante o processo federado satisfaçam o ruído da Privacidade Diferencial, sem induzir instabilidade de convergência ou colapso da utilidade preditiva.

3.3. Arquitetura do Sistema HALO

O HALO adota um modelo de orquestração centralizado composto por clientes quânticos na borda (*Quantum Edge*) e um agregador clássico na nuvem (*Cloud*). Cada cliente executa treinamento local por meio de VQC, enquanto o servidor central é responsável exclusivamente pela agregação dos parâmetros protegidos. Essa separação funcional permite que o processamento sensível a privacidade permaneça na borda, ao passo que a nuvem atua apenas como coordenadora do processo federado. A Figura 1 ilustra o fluxo operacional completo do sistema.

O ciclo de aprendizado é estruturado em três etapas: (i) treinamento local, no qual os dados sensíveis permanecem restritos à borda; (ii) aplicação de uma barreira de ofuscação por Privacidade Diferencial imediatamente antes da comunicação; e (iii) agregação global dos parâmetros perturbados, seguida da redistribuição do modelo atualizado. Nesse processo, os estados do modelo distinguem-se entre θ_{ideal}, correspondente aos parâmetros obtidos localmente após a otimização do VQC, e $\theta_{\text{noisy}} = \mathcal{M}(\theta_{\text{ideal}}, \epsilon)$, que representa a versão protegida transmitida ao servidor após a aplicação do mecanismo de privacidade.

A adoção dessa separação explícita entre parâmetros ideais e protegidos é fundamental para o modelo de segurança do HALO. Assume-se um adversário honesto-curioso com acesso às atualizações transmitidas ao servidor central, capaz de realizar ataques de inferência de membros. Para mitigar isso, o HALO emprega Privacidade Diferencial Local (LDP), garantindo que cada cliente aplique ruído estocástico antes da comunicação. Dessa forma, o mecanismo $\mathcal{M}$ assegura ϵ-Privacidade Diferencial ao limitar a contribuição

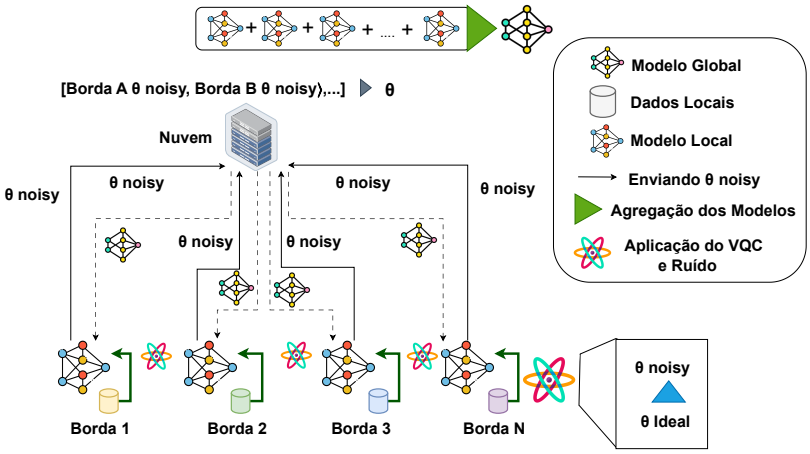


Figura 1. Esquema do protocolo HALO. O fluxo destaca a separação entre o modelo ideal (interno) e o ruidoso (transmitido), garantindo que a agregação na nuvem ocorra exclusivamente sobre dados perturbados.

informacional de qualquer amostra individual, impedindo que parâmetros não-ofuscados sejam compartilhados durante o processo federado.

Além da proteção estocástica, a robustez do HALO está diretamente associada à forma como os parâmetros do modelo são representados e manipulados. Em arquiteturas clássicas, os parâmetros residem em um espaço euclidiano ilimitado ($\mathbb{R}^d$), o que torna o modelo altamente sensível à injeção de ruído aditivo sob baixos valores de ϵ. No HALO, por outro lado, os parâmetros são interpretados como ângulos definidos no intervalo $[0, 2\pi]$, associados a rotações quânticas e representados geometricamente na Esfera de Bloch. A Figura 2 contrasta essas duas representações paramétricas.

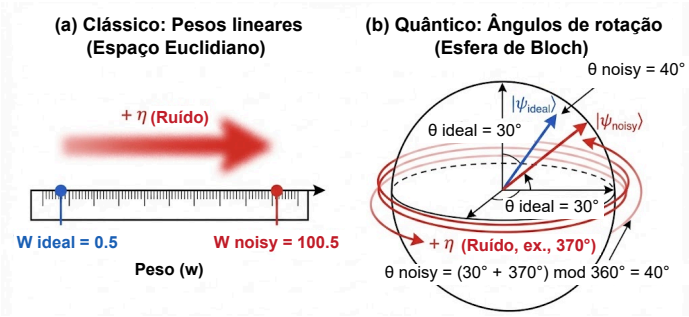


Figura 2. Estabilidade Geométrica. (a) modelo clássico diverge com ruído aditivo; (b) HALO utiliza a periodicidade angular para manter os parâmetros limitados e operacionais.

Assim, a aplicação do ruído de privacidade segue uma aritmética modular, caracterizando a propriedade de **Saturação Geométrica**, expressa por $\theta_{\text{final}} = (\theta_{\text{ideal}} + \text{Ruído}) \pmod{2\pi}$. Essa operação garante que, mesmo sob perturbações de grande magnitude, os parâmetros permaneçam em regiões fisicamente válidas do espaço de estados.

A dinâmica de Aprendizado Federado do HALO preserva a lógica do algoritmo FedAvg, porém adaptada a espaços paramétricos rotacionais. Em cada rodada t, o servidor distribui o modelo global $\theta^{(t)}$, e cada cliente realiza E épocas de SGD localmente, atualizando seus parâmetros segundo $\theta_k^{t+1} \leftarrow \theta_k^t - \eta \nabla \mathcal{L}(\theta_k^t; D_k)$. A agregação global

subsequente é realizada por média angular, respeitando a topologia periódica do espaço de parâmetros e mitigando a amplificação de ruído durante a sincronização.

Após o treinamento local, cada cliente aplica ruído Laplaciano, $\mathbf{n} \sim \text{Laplace}\left(0, \frac{\Delta}{\epsilon}\right)$, seguido de uma projeção modular no espaço angular. Essa combinação assegura ϵ-Privacidade Diferencial sem induzir divergência paramétrica, em contraste com o comportamento observado em arquiteturas clássicas. Para explicitar essa diferença operacional, o Algoritmo 1 descreve a abordagem de FL com DP, na qual a perturbação ocorre de forma linear em um espaço euclidiano ilimitado.

Algoritmo 1 FL Clássico com Privacidade Diferencial (Base)

Entrada: Clientes K, Rodadas T, Orçamento de Privacidade ϵ, Taxa de Aprendizado η
Inicializar: Pesos Globais $w_{global}^{(0)} \in \mathbb{R}^d$ (Aleatório)
Comportamento do Servidor:
for $t = 0$ **até** $T - 1$ **do**
  Transmitir $w_{global}^{(t)}$ para os clientes
  for cada cliente k **em paralelo do**
    $w_k \leftarrow w_{global}^{(t)}$
    for $lote \in D_k$ **do**
      Calcular gradientes ∇w_k
      $w_k \leftarrow w_k - \eta \nabla w_k$
    end for
    Amostrar ruído $\mathbf{n} \sim \text{Laplace}(0, \frac{\Delta}{\epsilon})$
    $\tilde{w}_k \leftarrow w_k + \mathbf{n}$
    Enviar $\tilde{w}_k$ para o Servidor
  end for
  $w_{global}^{(t+1)} \leftarrow \frac{1}{K} \sum_{k=1}^{K} \tilde{w}_k$
end for

Em contraste direto, o Algoritmo 2 apresenta a abordagem proposta pelo HALO, na qual a atualização dos parâmetros respeita explicitamente a periodicidade intrínseca do *ansatz* quântico.

Algoritmo 2 FL Quântico com Privacidade Diferencial (Proposto)

Entrada: Clientes K, Rodadas T, Orçamento de Privacidade ϵ, Taxa de Aprendizado η
Inicializar: Parâmetros Globais $\theta_{global}^{(0)} \in [0, 2\pi]^d$ (Uniforme)
Comportamento do Servidor:
for $t = 0$ **até** $T - 1$ **do**
  Transmitir $\theta_{global}^{(t)}$ para os clientes
  for cada cliente k **em paralelo do**
    $\theta_k \leftarrow \theta_{global}^{(t)}$
    for $lote \in D_k$ **do**
      Codificar $x \rightarrow |\psi_x\rangle$
      Aplicar $U(\theta_k)|\psi_x\rangle$
      Medir $\langle \hat{Z} \rangle$
      Calcular $\nabla \theta_k$
      $\theta_k \leftarrow \theta_k - \eta \nabla \theta_k$
    end for
    Amostrar ruído $\mathbf{n} \sim \text{Laplace}(0, \frac{\Delta}{\epsilon})$
    $\tilde{\theta}_k \leftarrow (\theta_k + \mathbf{n}) \pmod{2\pi}$
    Enviar $\tilde{\theta}_k$ para o Servidor
  end for
  $\theta_{global}^{(t+1)} \leftarrow \frac{1}{K} \sum_{k=1}^{K} \tilde{\theta}_k$
end for

Por fim, para viabilizar a execução prática em dispositivos NISQ, o HALO adota um pipeline híbrido de processamento quântico-clássico, ilustrado na Figura 3. Esse pipeline compreende a compressão dos dados por PCA ($\mathbb{R}^{64} \rightarrow \mathbb{R}^2$), a codificação angular por rotações $R_x(x)$ e a aplicação de um *ansatz* variacional com portões parametrizados e

entrelaçamento (CNOT), responsável por aprender a fronteira de decisão não linear.

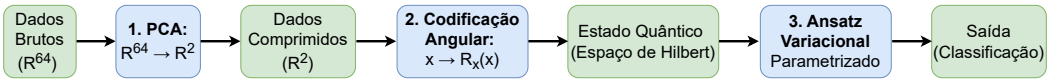


Figura 3. Pipeline Híbrido de Processamento Quântico-Clássico do HALO

Com base nessa arquitetura e nos mecanismos de otimização e privacidade descritos, a próxima seção apresenta a configuração experimental adotada para avaliar o desempenho, a estabilidade de convergência e as garantias de privacidade do HALO.

4. Configuração Experimental

O protocolo experimental foi concebido para avaliar a robustez geométrica de modelos quânticos sob regimes de DP, analisando o impacto da injeção deliberada de ruído.

4.1. Conjunto de Dados e Redução de Dimensionalidade

Os experimentos utilizaram o *Optical Recognition of Handwritten Digits Dataset* (Digits) [Alpaydin and Kaynak]), disponibilizado na biblioteca Scikit-learn, composto por imagens em escala de cinza de 8×8 pixels, as quais são vetorizadas em um espaço de entrada de dimensão ($d = 64$). Para isolar a análise da fronteira de decisão e evitar efeitos associados à classificação multiclasse, o escopo experimental foi restrito à tarefa binária entre as classes "Dígito 0" e "Dígito 1".

Devido às limitações dos dispositivos quânticos da era NISQ, em particular o número reduzido de qubits e a profundidade limitada dos circuitos, a codificação direta das 64 *features* torna-se impraticável. Assim, aplicou-se a Análise de Componentes Principais (PCA) [Bishop and Nasrabadi 2006] como etapa de pré-processamento, projetando os dados em um subespaço latente $\mathbb{R}^2$. Conforme ilustrado na Figura 4, apesar da redução de 96,8% na dimensionalidade, a estrutura topológica dos dados é preservada, mantendo a separabilidade linear entre as classes. As componentes resultantes foram normalizadas por *MinMax Scaling* para o intervalo $[0, \pi]$, garantindo a aplicação correta do *Angle Embedding* e rotações válidas ($R_x(\theta)$) na esfera de Bloch.

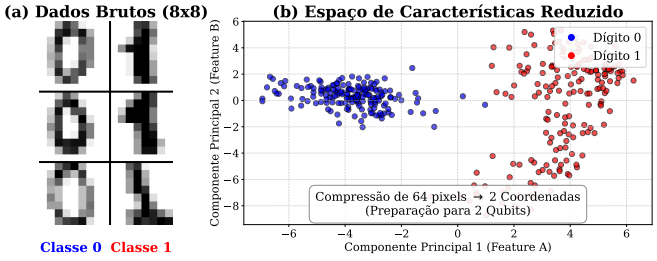


Figura 4. Pré-proc. e codificação. (a) Amostras originais. (b) Projeção PCA 2D.

4.2. Parâmetros dos Modelos

Com o espaço de entrada uniformizado, a análise comparativa concentrou-se em dois agentes de aprendizado com complexidades de tarefa equivalentes, porém baseados em representações geométricas distintas. Como *baseline* clássico, adotou-se uma Rede Neural *Feedforward* do tipo *Multilayer Perceptron* (MLP), amplamente reconhecida por sua eficácia em tarefas de aprendizado supervisionado sobre dados tabulares [LeCun et al. 2015], permitindo contrastar diretamente a dinâmica de aprendizado em espaços euclidianos

com a abordagem quântica baseada no Espaço de Hilbert. A arquitetura clássica consiste em uma camada de entrada com 2 neurônios correspondentes às componentes principais do PCA, uma camada oculta com 4 neurônios e ativação ReLU ($f(x) = \max(0, x)$), introduzindo não linearidade eficiente, e uma camada de saída com 1 neurônio e ativação tangente hiperbólica ($\tanh$), mapeando a decisão para o intervalo $[-1, 1]$.

O modelo totaliza 17 parâmetros treináveis, inicializados segundo a distribuição uniforme $\mathcal{U}(-0.5, 0.5)$. Em contraposição, o modelo quântico consiste em um Classificador Quântico Variacional (*Variational Quantum Classifier* – VQC), seguindo o framework introduzido por Mitarai et al. [Mitarai et al. 2018]. O circuito opera com dois qubits e é estruturado em três estágios complementares: (i) codificação de estado por *Angle Embedding*, mapeando os dados normalizados para rotações $R_x(x_i')$ [Schuld et al. 2021]; (ii) um *Hardware Efficient Ansatz* [Kandala et al. 2017], composto por camadas de rotações parametrizadas intercaladas por portões CNOT; e (iii) a etapa de medição, na qual a classificação é derivada do valor esperado do operador Pauli-Z ($\langle Z \rangle$) no primeiro qubit. O circuito resultante possui apenas quatro parâmetros treináveis, permitindo avaliar a eficiência da representação quântica com baixa complexidade paramétrica.

4.3. Parametrização do Aprendizado Federado e Privacidade Diferencial

A simulação do ecossistema federado considerou quatro clientes colaborativos sob a coordenação de um servidor central de agregação. O processo de otimização seguiu o algoritmo *Federated Averaging* (FedAvg) [McMahan et al. 2017], estruturado em $R = 5$ rodadas de comunicação. Em cada rodada, os clientes realizam cinco épocas de treinamento local via Descida do Gradiente Estocástica (SGD) antes de submeterem seus parâmetros atualizados ao servidor.

A proteção das atualizações compartilhadas foi assegurada por meio do Mecanismo de Laplace [Dwork et al. 2006], garantindo ϵ-Privacidade Diferencial nos parâmetros transmitidos. Imediatamente após o treinamento local e antes da agregação, foi injetado ruído aleatório extraído de uma distribuição $Lap(0, \Delta f/\epsilon)$, onde Δf representa a sensibilidade do modelo e ϵ o orçamento de privacidade. Os experimentos exploraram o espectro completo de proteção por meio do conjunto $\epsilon \in \{0.5, 1.0, 2.0, 3.0, 10.0, \infty\}$, cobrindo desde regimes de alta privacidade, caracterizados por ruído severo, até o cenário de referência sem proteção estocástica.

4.4. Definição das Métricas de Segurança e Convergência

Para garantir significância estatística, cada configuração de orçamento de privacidade (ϵ) foi avaliada em 10 execuções independentes com sementes aleatórias distintas. Essa estratégia possibilitou a construção de intervalos de confiança robustos e fundamentou a análise por meio da Função de Distribuição Acumulada (CDF), utilizada para capturar a variabilidade do processo de aprendizado sob ruído estocástico.

A avaliação de privacidade foi realizada por meio do *Membership Inference Attack* (MIA), baseado na técnica de *Shadow Models* [Shokri et al. 2017], na qual um adversário treina modelos substitutos para inferir a participação de registros no conjunto de treinamento a partir dos *confidence scores*. O desempenho do ataque, quantificado pelo *MIA Score*, foi adotado como métrica central de vazamento de informação. Com os protocolos de treinamento e auditoria definidos, a execução experimental seguiu os Algoritmos 1 e 2, e

a próxima seção analisa quantitativamente o impacto dessas escolhas sobre a convergência, a utilidade preditiva e a resiliência a ataques de inferência.

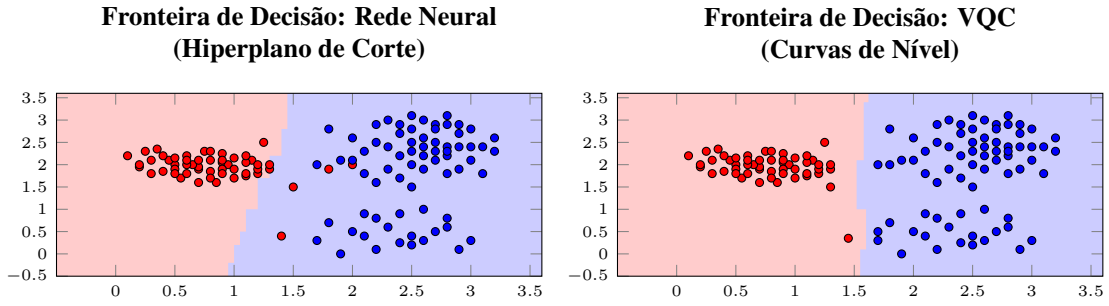
5. Resultados

Esta seção apresenta a avaliação comparativa entre um modelo clássico baseado em Rede Neural Feedforward (FNN) [Ilonen et al. 2003] e o modelo quântico proposto, fundamentado em VQC [Biamonte 2021], no contexto de Aprendizado Federado com DP.

5.1. Paridade de Capacidade Sem Ruído

A primeira etapa do estudo consistiu em medir o potencial de aprendizado de ambas as arquiteturas sem restrições de privacidade $\epsilon = \infty$. Essa base de comparação é essencial para confirmar que tanto o modelo clássico quanto o VQC possuem expressividade suficiente para classificar o dataset processado via a Análise de Componentes Principais (PCA) para redução de dimensionalidade [Bishop and Nasrabadi 2006], garantindo que os testes posteriores partam de um mesmo patamar de eficiência. O PCA projetou o espaço original $\mathbb{R}^{64}$ em um subespaço latente $\mathbb{R}^2$, preservando as componentes de maior variância.

As fronteiras de decisão após a convergência na Figura 5 revelam a natureza distinta de cada abordagem. Enquanto a Rede Neural Clássica representada na Figura 5(a), soluciona o problema traçando um hiperplano linear no espaço euclidiano, o VQC representado na Figura 5(b), projeta fronteiras não-lineares. Através de rotações parametrizadas na Esfera de Bloch, o modelo quântico gera curvas suaves e periódicas que se moldam à distribuição das classes.



(a) Abordagem Clássica **(b) Abordagem Quântica**

Figura 5. Fronteiras de decisão no cenário base ($\epsilon = \infty$). Enquanto a Rede Neural (a) gera um hiperplano linear, o VQC (b) produz fronteiras não-lineares. Ambos atingem paridade de desempenho na ausência de ruído.

Apesar de se basearem em representações geométricas distintas, ambos os modelos atingiram acurácia superior a 99% no cenário sem privacidade, estabelecendo uma paridade inicial de desempenho. Esse resultado valida a premissa do estudo, assegurando que eventuais degradações observadas nos experimentos com Privacidade Diferencial decorrem exclusivamente da sensibilidade de cada arquitetura à injeção de ruído, e não de limitações na capacidade de aprendizado.

A aplicação do mecanismo de DP aos pesos locais mostra o impacto do ruído Laplaciano [Gholipour et al. 2025] sobre utilidade e segurança, conforme sintetizado na Figura 6 para orçamentos de privacidade entre $\epsilon = 0.5$ e $\epsilon = \infty$. Em regimes de alta privacidade ($\epsilon < 1.0$), observa-se uma degradação generalizada da acurácia, enquanto no

intervalo intermediário $\epsilon \in [1.0, 3.0]$ emerge uma distinção clara entre as arquiteturas: o modelo clássico apresenta recuperação lenta e elevada variabilidade, alcançando apenas $67{,}3\%$ de acurácia em $\epsilon = 2.0$ com alto desvio padrão ($\sigma \approx 0.22$), ao passo que o VQC mantém desempenho estável, atingindo $98{,}9\%$ de acurácia com variação mínima, evidenciando maior robustez a perturbações aditivas nos parâmetros.

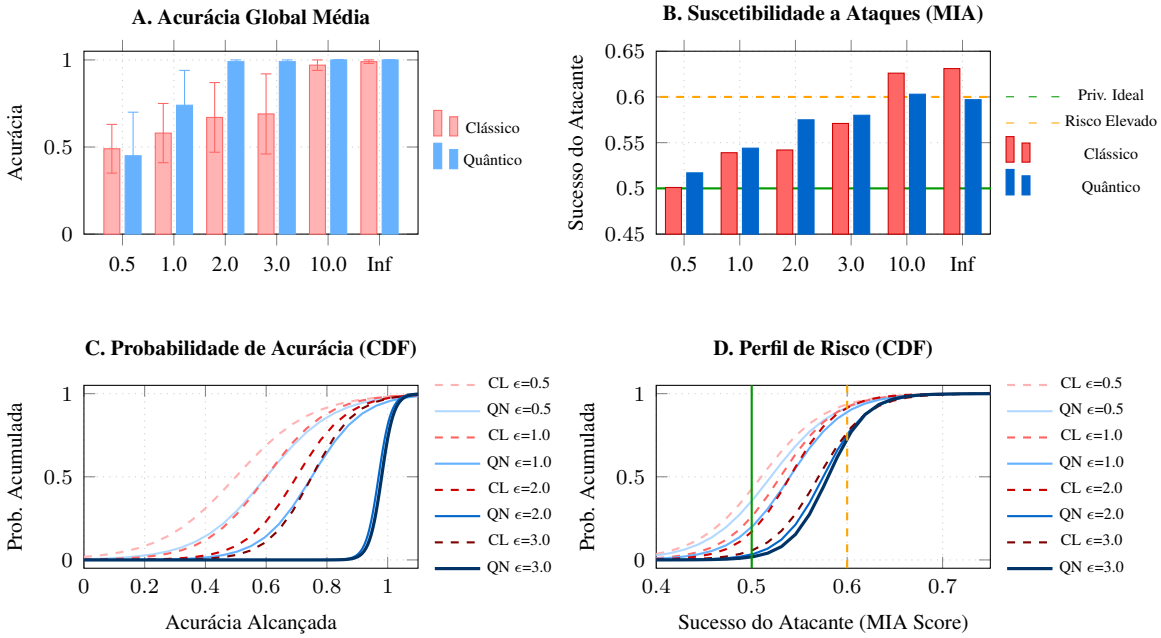


Figura 6. Impacto da DP. (A) O modelo quântico (azul) elimina a instabilidade de acurácia observada no clássico (vermelho). (B) Ambos respeitam o limiar de risco MIA 0.60 para $\epsilon \leq 3.0$. (C) A inclinação vertical das curvas quânticas na CDF confirma maior estabilidade de convergência. (D) O ganho de performance quântico preserva as garantias de segurança do *baseline*.

Essa disparidade nos desvios padrão aponta para uma diferença fundamental na natureza da estabilidade de cada modelo. Enquanto as métricas médias oferecem um panorama geral, elas não revelam o comportamento probabilístico das execuções individuais sob ruído. Para compreender se a variância clássica decorre de falhas esporádicas ou de uma instabilidade sistêmica, e como a geometria quântica mitiga esse risco, é necessário analisar a distribuição acumulada de probabilidade.

A Função de Distribuição Acumulada (CDF) das acurácias demonstrada na Figura 6(c), ajuda a explicar a variância supracitada. As curvas tracejadas do modelo clássico possuem inclinações suaves, o que caracteriza uma distribuição mais achatada na visualização e indica instabilidade estocástica: para um mesmo ϵ, o sucesso do treinamento depende da sorte na semente aleatória do ruído. Em contraste, as curvas do VQC, especialmente para $\epsilon \geq 2.0$, apresentam uma transição de fase abrupta, quase vertical. Esse perfil indica robustez determinística, onde a chance de o modelo performar abaixo de 90% é muito baixa. Esse fenômeno decorre da saturação natural dos portões de rotação (R_x, R_y, R_z) [Schuld et al. 2021]. Como esses portões são periódicos, eles limitam o impacto de ruídos de grande magnitude, impedindo que os pesos desviem para regiões de divergência.

Para quantificar o risco de privacidade, foi implementado o Ataque de Inferência de

Membros *Membership Inference Attack* (MIA) seguindo a metodologia de modelos sombra proposta por Shokri et al. [Shokri et al. 2017]. Esse tipo de ataque explora a discrepância nas distribuições de confiança do modelo entre amostras de treinamento e de teste para inferir a pertinência dos dados. A estabilidade do VQC não comprometeu a segurança dos dados. A Figura 6(b) detalha o sucesso MIA, onde ambos os modelos permaneceram em zona segura, com score < 0.60 para $\epsilon \leq 3.0$. O perfil de risco acumulado demonstrado na Figura 6(d) mostra que a curva do VQC para $\epsilon = 2.0$ está ligeiramente à direita da clássica, sugerindo um vazamento de informação marginalmente superior. Contudo, esse incremento é muito baixo, de $+0.03$ no score MIA, frente ao salto de utilidade de $+30\%$ de acurácia, o que valida a eficiência entre privacidade e desempenho na abordagem quântica.

5.2. Dinâmica de Convergência e Estabilidade do Gradiente

A instabilidade observada nas métricas globais é explicada pela evolução do erro quadrático médio (MSE) ao longo das rodadas federadas, apresentada na Figura 7 para o cenário crítico de $\epsilon = 2.0$. O modelo clássico (linha tracejada vermelha) exibe comportamento altamente oscilatório e sem tendência clara de convergência, indicando que o ruído diferencial desestabiliza o processo de otimização. Em contraste, o VQC (linha sólida azul) converge de forma rápida e estável, atingindo MSE inferior a 0.5 já na quinta rodada e mantendo esse patamar até o final, evidenciando maior capacidade de absorção do ruído estocástico pela parametrização quântica.

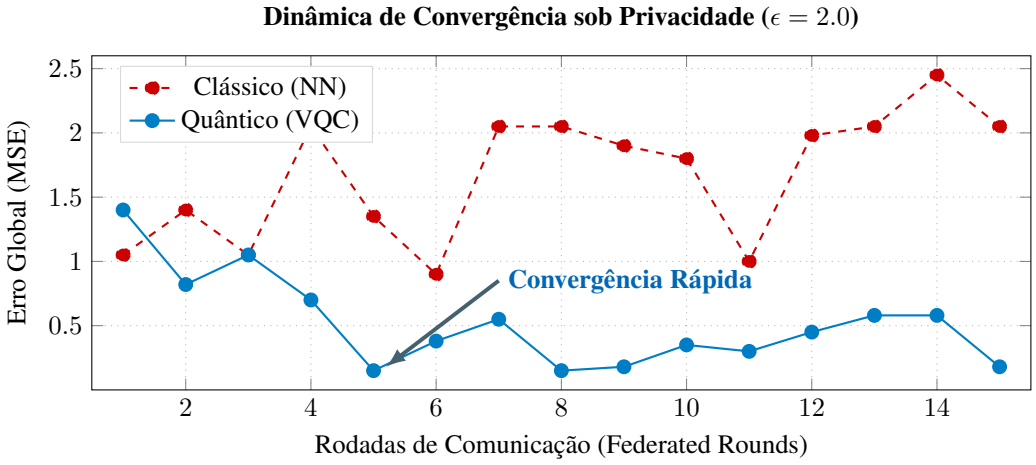


Figura 7. Dinâmica de convergência ($\epsilon = 2.0$). O modelo quântico estabiliza em menos de 10 rodadas, superando as oscilações severas do clássico causadas pela perturbação dos gradientes.

5.3. Análise Qualitativa via Matrizes de Confusão

As matrizes de confusão da Figura 8, obtidas para $\epsilon = 2.0$, mostram os comportamentos qualitativamente distintos entre as arquiteturas sob ruído diferencial. A rede clássica (Figura 8(a)) apresenta colapso preditivo, convergindo para uma solução trivial que classifica quase todas as amostras como "Dígito 0", com 46 acertos e 44 falsos positivos, indicando perda da fronteira de decisão e adoção de uma estratégia de classe majoritária. Em contraste, o VQC (Figura 8(b)) preserva a separabilidade entre classes, exibindo uma diagonal dominante com apenas um erro total, o que confirma que a geometria do *feature map* quântico sustenta a capacidade discriminativa do modelo mesmo sob regimes rigorosos de DP.

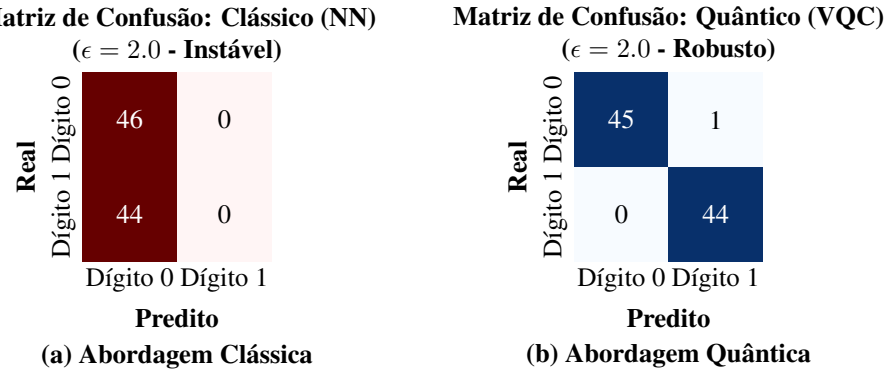


Figura 8. Comparativo de Robustez ($\epsilon = 2.0$). (a) abordagem clássica; (b) VQC.

5.4. Fronteira de Eficiência: O Ponto de Operação Ótimo Quântico

A Figura 9 sintetiza o compromisso entre utilidade e privacidade por meio de um Gráfico de Pareto, no qual se busca maximizar a acurácia (eixo Y) e minimizar o risco de privacidade (eixo X). A trajetória do modelo clássico (linha tracejada vermelha) evidencia uma degradação praticamente linear: o aumento do rigor de privacidade implica perdas proporcionais de desempenho, inviabilizando um ponto de operação que concilie segurança elevada e alta utilidade.

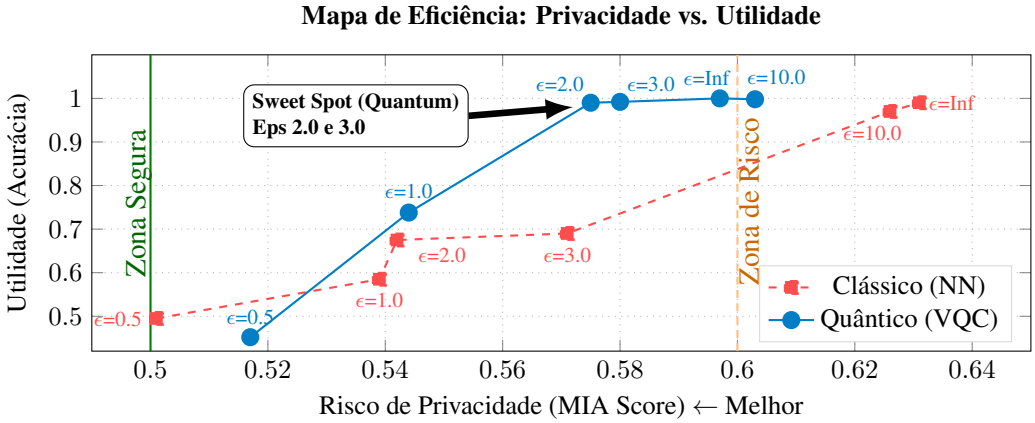


Figura 9. Análise de Eficiência: ↑ Melhor. O *Sweet Spot* evidencia a robustez quântica em regimes restritivos ($\epsilon = 2.0$) frente à queda acentuada do método clássico.

Em contraste, o VQC apresenta um *sweet spot* quântico no intervalo $\epsilon \in [2.0, 3.0]$, no qual sustenta acurácia superior a 98% com risco de MIA próximo ao limiar ideal (≈ 0.50). Como indicado na figura, em $\epsilon = 2.0$ a transição do modelo clássico para o quântico resulta em um ganho de utilidade de cerca de 30% com aumento marginal de risco, demonstrando a viabilidade prática de circuitos quânticos variacionais em cenários que exigem forte proteção de dados sem comprometer o desempenho preditivo.

6. Conclusão

Este trabalho demonstrou que a geometria paramétrica de Circuitos Quânticos Variacionais pode ser explorada como um mecanismo intrínseco de robustez ao ruído imposto pela Privacidade Diferencial em cenários de Aprendizado Federado. Os resultados evidenciam que, enquanto arquiteturas clássicas sofrem degradação severa ou colapso de convergência

sob orçamentos rigorosos de privacidade, o HALO mantém alta utilidade preditiva e estabilidade estatística, preservando simultaneamente as garantias formais contra ataques de inferência de membros. A identificação de um ponto de operação quântico ótimo, no qual desempenho e risco permanecem equilibrados, valida empiricamente a hipótese central do trabalho e posiciona o HALO como uma alternativa viável para aprendizado colaborativo seguro no Cloud-Edge Continuum, especialmente em aplicações sensíveis onde privacidade e desempenho não podem ser tratados como objetivos conflitantes

Disponibilidade de Artefatos

Em aderência aos princípios da Ciência Aberta, o código-fonte e o dataset utilizados neste trabalho podem ser acessados em: https://github.com/adianohum/HALO.

Agradecimentos

Este trabalho foi parcialmente financiado pelo projeto QuIIN "Integração CV-QKD com Redes Clássicas", apoiado pelo QuIIN *Quantum Industrial Innovation*, Centro de Competência EMBRAPII CIMATEC em Tecnologias Quânticas. O apoio contou com recursos financeiros do programa PPI IoT/Manufatura 4.0, no âmbito da chamada MCTI nº 053/2023, estabelecida com a EMBRAPII. Este estudo também foi financiado, em parte, pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, e pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brasil, sob a concessão nº 403231/2023-0.

Referências

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., and Zhang, L. (2016). Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 308–318. ACM.

Alpaydin, E. and Kaynak, C. Optical recognition of handwritten digits data set. uci machine learning repository (1998). *URL https://archive. ics. uci. edu/ml/datasets/Optical+ Recognition+ of+ Handwritten+ Digits*.

Biamonte, J. (2021). Universal variational quantum computation. *Physical Review A*, 103(3):L030401.

Bishop, C. M. and Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*, volume 4. Springer.

Du, Y., Hsieh, M.-H., Liu, T., and Tao, D. (2021). Quantum noise protects quantum classifiers against adversarial attacks. *Physical Review Research*, 3(2):023153.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer.

Freire, M., Mello, T. L., Sant'Anna, I., Maia, A., Moreira, R., Rivelino, R., and Peixoto, M. (2025). Rana: Uma abordagem híbrida para qkd bb84 com expansão e encapsulamento de chave. In *Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC)*, pages 938–951. SBC.

Gholipour, H., Bozorgnia, F., Hambarde, K., Mohammadigheymasi, H., Mancilla, J., Sequeira, A., Neves, J., Proença, H., and Challenger, M. (2025). A laplacian-based

quantum graph neural networks for quantum semi-supervised learning. *Quantum Information Processing*, 24(4):106.

Gkonis, P. K., Trakadas, P., Karkazis, P., and Leligou, H. C. (2023). A survey on iot-edge-cloud continuum systems: Status, challenges, use cases, and open issues. *Future Internet*, 15(12):383.

Ilonen, J., Kamarainen, J.-K., and Lampinen, J. (2003). Differential evolution training algorithm for feed-forward neural networks. *Neural Processing Letters*, 17(1):93–105.

Kandala, A., Mezzacapo, A., Temme, K., Takita, M., Brink, M., Chow, J. M., and Gambetta, J. M. (2017). Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, 549(7671):242–246.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436–444.

Li, W., Lu, S., and Deng, D.-L. (2021). Quantum federated learning through blind quantum computing. *Science China Physics, Mechanics & Astronomy*, 64(10):100312.

Maia, A., Freire, M., Mello, T., Rodrigues-Filho, R., Almeida, E., Prazeres, C., Figueiredo, G., and Peixoto, M. (2025). Q-edge: Leveraging quantum computing for enhanced software engineering in vehicular networks. In *Proceedings of the 40th ACM/SIGAPP Symposium on Applied Computing*, pages 1457–1467.

McMahan, B., Moore, E., Ramage, D., Hampson, S., and y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR.

Mitarai, K., Negoro, M., Kitagawa, M., and Fujii, K. (2018). Quantum circuit learning. *Physical Review A*, 98(3):032309.

Nasr, M., Shokri, R., and Houmansadr, A. (2019). Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning. In *2019 IEEE Symposium on Security and Privacy (SP)*, pages 739–753. IEEE.

Peixoto, M. L. M. (2024). Quantum edge computing for data analysis in connected autonomous vehicles. In *2024 IEEE Symposium on Computers and Communications (ISCC)*, pages 1–6. IEEE.

Ponomareva, N., Hazimeh, H., Kurakin, A., Xu, Z., Denison, C., McMahan, H. B., Vassilvitskii, S., Chien, S., and Thakurta, A. G. (2023). How to dp-fy ml: A practical guide to machine learning with differential privacy. *Journal of Artificial Intelligence Research*, 77:1113–1201.

Schuld, M., Sweke, R., and Meyer, J. J. (2021). Effect of data encoding on the expressive power of variational quantum-machine-learning models. *Physical Review A*, 103(3):032430.

Shokri, R., Stronati, M., Song, C., and Shmatikov, V. (2017). Membership inference attacks against machine learning models. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 3–18. IEEE.

Yu, S., Zhu, K., Liang, F., Wang, J., Kant, K., and Yin, L. (2026). Robust multimodal federated learning for non-iid multimodal data with incompleteness. *Future Generation Computer Systems*, 174:107948.