Arquitetura de Aprendizado Federado para Adaptação Colaborativa de LLMs de Saúde: Estruturação de Dados Clínicos em HL7 FHIR

Fernando B. Rosito, Silvio C. Cazella, Muriel F. Franco

[1]Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA)
Porto Alegre – RS – Brasil

{fernando.rosito, silvioc, muriel.franco}@ufcspa.edu.br

Abstract. *Healthcare collaboration and the sharing of clinical intelligence are hindered by data silos, privacy regulations, and computational constraints. This paper presents a research design for a Federated Learning (FL) architecture to enable secure collaboration for training Large Language Models (LLMs) for clinical data structuring. By integrating semantic interoperability, such as HL7 FHIR, and efficient quantized adaptation using QLoRA, the proposal supports trust among participating institutions and enables cooperation among heterogeneous institutions, thereby addressing a key challenge in collaborative systems. The architecture focuses on preserving data sovereignty while ensuring semantic consistency and mitigating hallucinations.*

Resumo. *A colaboração em saúde e o compartilhamento de inteligência clínica são limitados por silos de dados, regulações de privacidade e restrições computacionais. Este trabalho apresenta um desenho de pesquisa que propõe uma arquitetura de Aprendizado Federado (FL) para viabilizar a colaboração segura no treinamento de Grandes Modelos de Linguagem (LLMs) para a estruturação de dados. Integrando interoperabilidade semântica, como o HL7 FHIR, e adaptação eficiente por meio do QLoRA, a arquitetura proposta permite estabelecer confiança em redes distribuídas e promover a cooperação entre instituições de saúde heterogêneas, contribuindo para sistemas colaborativos. A arquitetura tem como foco preservar a soberania dos dados, validar a consistência semântica e mitigar alucinações.*

1. Introdução

A colaboração efetiva entre instituições de saúde é fundamental para a medicina baseada em evidências e para a construção de sistemas de saúde resilientes [Franco et al. 2025]. Contudo, esse processo enfrenta um dilema sociotécnico: a necessidade de compartilhar conhecimento clínico contrasta com exigências legais de proteção de dados e soberania institucional [Rajpurkar et al. 2022]. Como resultado, hospitais operam como silos de dados isolados, nos quais o conhecimento contido em narrativas clínicas não estruturadas permanece pouco acessível ao uso colaborativo [Sedlakova et al. 2023, Li et al. 2022].

Embora a centralização e o compartilhamento de dados possam potencializar o treinamento de Grandes Modelos de Linguagem (LLMs), essa estratégia é inviável devido a riscos de segurança, restrições regulatórias e dependência de infraestruturas externas [Franco et al. 2025]. No contexto brasileiro, essa dependência compromete os

princípios de soberania informacional e entra em conflito com as diretrizes da Rede Nacional de Dados em Saúde (RNDS). Ademais, o refinamento local de LLMs demanda recursos computacionais frequentemente incompatíveis com a infraestrutura das instituições [Dettmers et al. 2023], enquanto a heterogeneidade semântica dos registros clínicos dificulta a cooperação interinstitucional sem o uso de padrões como o HL7 FHIR [Amar et al. 2024].

Nesse cenário, o Aprendizado Federado (FL) surge como um potencial habilitador de sistemas colaborativos em saúde, ao permitir o treinamento de modelos globais sem a transferência de dados brutos [Rieke et al. 2020]. Entretanto, revisões recentes indicam que sua adoção prática permanece limitada não apenas por desafios técnicos, mas, sobretudo, pela ausência de mecanismos robustos de governança, auditoria e confiança entre os participantes da federação [Eden et al. 2025, Wei and Guan 2024].

Portanto, neste trabalho, apresentamos um desenho de pesquisa de uma arquitetura baseada em FL para a adaptação colaborativa de LLMs em saúde, com foco na geração de dados interoperáveis a partir de narrativas clínicas em texto livre. Diferentemente de abordagens existentes que assumem a disponibilidade de informações já estruturadas, a proposta atua diretamente na transformação de registros clínicos não estruturados em recursos padronizados do HL7 FHIR, sem a necessidade de compartilhamento de dados sensíveis entre instituições. Para viabilizar essa colaboração, a arquitetura integra a adaptação eficiente via QLoRA [Dettmers et al. 2023], mecanismos de validação semântica para mitigação de alucinações [Ali et al. 2024] e um modelo de governança federada voltado à preservação da soberania dos dados. Com isso, busca-se oferecer uma base sociotécnica para o desenvolvimento de sistemas colaborativos seguros e escaláveis no contexto da saúde digital.

2. Trabalhos Relacionados

A literatura aponta para a convergência entre IA generativa e sistemas distribuídos, embora a tradução clínica enfrente barreiras sociotécnicas [Teo et al. 2024, Ali et al. 2024]. Revisões sistemáticas indicam que, apesar do potencial de privacidade, cerca de 95% dos estudos de FL falham na implementação prática devido à complexidade regulatória [Wei and Guan 2024]. Enquanto a maioria das soluções foca em otimização algorítmica, há poucos trabalhos voltados à governança de redes de saúde [Eden et al. 2025].

No domínio da estruturação, o *FHIR-Former* [Engelke et al. 2025] valida o uso de LLMs em dados padronizados, mas pressupõe interoperabilidade prévia. Recentemente, pipelines como o *Infherno* e o *SMART Text2FHIR* exploram a geração de FHIR a partir de texto, porém operam de forma centralizada, ignorando a diversidade semântica e as restrições de soberania de dados impostas pela Lei Geral de Proteção de Dados Pessoais (LGPD) [Li et al. 2023].

Paralelamente, estratégias de adaptação eficiente (*PEFT*) vêm sendo integradas ao cenário federado. O *FedLoRA* [Yi et al. 2023] reduz custos de comunicação ajustando apenas matrizes de baixa dimensão, enquanto o *FedLoDrop* [Xie et al. 2025] aplica *dropout* estruturado para mitigar heterogeneidade estatística. Contudo, esses métodos são avaliados predominantemente em benchmarks genéricos de PLN. Nossa proposta busca preencher essa lacuna ao aplicar uma adaptação eficiente para a geração estruturada de objetos clínicos complexos, integrando colaboração federada, eficiência computacional e

rigor semântico.

3. Arquitetura Federada Semântica Proposta

Para mitigar os dilemas de privacidade e a fragmentação dos dados, propõe-se uma arquitetura sociotécnica onde o conhecimento (parâmetros do modelo) é compartilhado pelos participantes , enquanto os dados sensíveis permanecem sob soberania local.

A Figura 1 ilustra o cenário atual, caracterizado por silos de dados isolados. A colaboração em saúde é limitada, por exemplo, não apenas pela LGPD [Corrêa and Franco 2025], mas também pela ausência de mecanismos de confiança que garantam às instituições o controle sobre seus ativos informacionais [Ali et al. 2024].

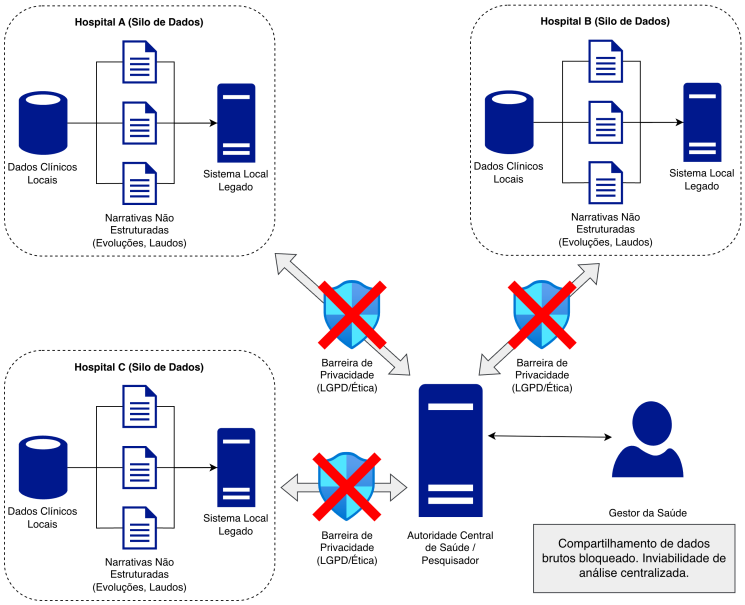

Figura 1. Representação do Cenário Atual com Silos de Dados Isolados e Limitações para Criação de Modelos Colaborativos.

A arquitetura proposta é apresentada na Figura 2 e divide-se em dois níveis para habilitar a colaboração segura. No nível Local (Agente), cada instituição opera um nó autônomo baseado em três pilares: (*i*) Interoperabilidade Semântica para conversão de narrativas para HL7 FHIR R4, promovendo o aprendizado de estruturas válidas e reduzindo ambiguidades; (*ii*) Adaptação Quantizada (QLoRA) para ajuste fino em GPUs de consumo via quantização de 4-bits [Dettmers et al. 2023], assim facilitando a colaboração e participação; e (*iii*) Aplicação de Privacidade Diferencial e atuação do nó como auditor (”Data Steward”) do conhecimento compartilhado. No nível Global (Orquestrador), um servidor central agrega os pesos dos adaptadores (via *FedAvg*), gerenciando o modelo global sem acessar dados de pacientes, assim estabelecendo a governança algorítmica.

4. Metodologia de Avaliação

A validação seguirá uma abordagem experimental quantitativa simulada. Inicialmente, será realizado um mapeamento de bases de dados públicas com narrativas clínicas, particionando-as em N silos com distribuição *Non-IID* para reproduzir a heterogeneidade real das redes de saúde [Ali et al. 2024]. Como tarefas iniciais, serão consideradas

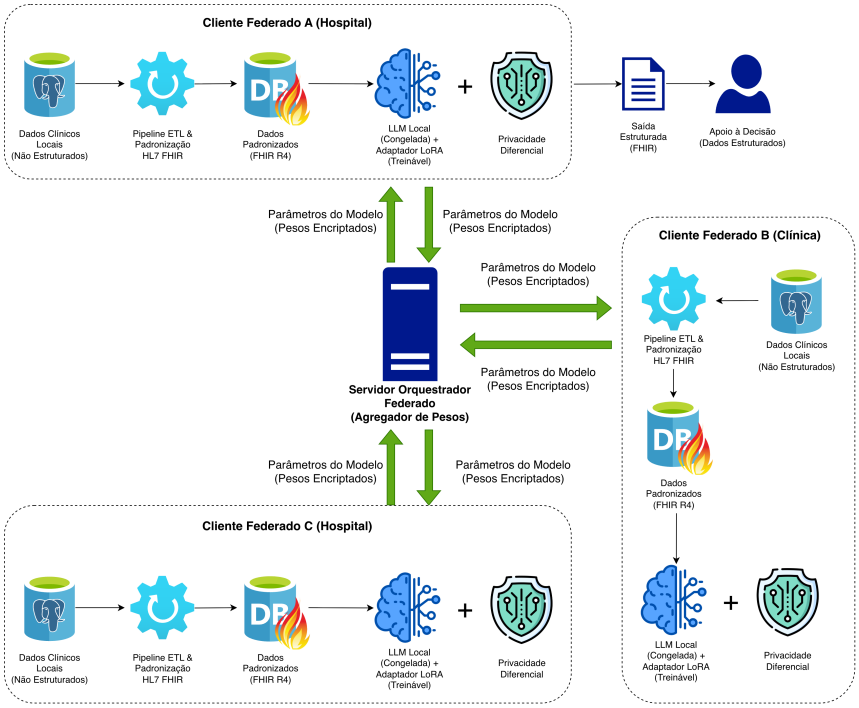

Figura 2. Visão Geral da Arquitetura Proposta para Colaboração Federada com Padronização FHIR e Privacidade Diferencial.

a extração e a normalização de entidades clínicas para recursos FHIR, como Condition, Observation e MedicationStatement, permitindo medir, de forma objetiva, a qualidade estrutural dos objetos gerados.

Para a avaliação, serão comparados três cenários de estruturação de diagnósticos, incluindo benchmarks para avaliação da arquitetura proposta: (*i*) Baseline Local: Treinamento isolado (silo); (*ii*) Federado Textual: Colaboração com texto bruto sem padronização; e (*iii*) Federado Semântico (Proposto): Colaboração com normalização HL7 FHIR e QLoRA, viabilizando o aprendizado sobre estruturas validadas.

A avaliação focará na robustez semântica por meio da taxa de alucinação (por exemplo, códigos inexistentes ou domínio incorreto) [Liu et al. 2025] e do F1-Score. Também, será avaliado o trade-off entre o ganho de eficiência e o custo de comunicação em termos de tempo e processamento [Xie et al. 2025]. Por fim, o consumo de recursos (por exemplo, VRAM e uso de CPU/GPU) será verificado para validar a viabilidade técnica e também a democratização do acesso via QLoRA [Dettmers et al. 2023].

5. Conclusão

Este trabalho propõe uma arquitetura de aprendizado federado para habilitar a colaboração em sistemas de saúde, explorando mecanismos que conciliam a cooperação clínica interinstitucional com a soberania nacional sobre os dados. A integração de FL, HL7 FHIR e QLoRA visa oferecer bases sociotécnicas para a governança necessária [Eden et al. 2025], reduzindo a dependência de tecnologias externas. Como trabalhos futuros, pretende-se implementar um protótipo funcional da arquitetura, conduzir experimentos em bases clínicas públicas e avaliar a viabilidade de implantação em um ambiente de colaboração entre instituições brasileiras de saúde.

Referências

Ali, M. S. et al. (2024). Federated learning in healthcare: Model misconducts, security, challenges, applications, and future research directions - a systematic review. *arXiv preprint arXiv:2405.13832*.

Amar, F., April, A., and Abran, A. (2024). Electronic health record and semantic issues using fast healthcare interoperability resources: Systematic mapping review. *Journal of Medical Internet Research*, 26:e45209.

Corrêa, H. and Franco, M. (2025). Lei geral de proteção de dados (lgpd) na saúde: Tendências, desafios e oportunidades. In *Anais da XXII Escola Regional de Redes de Computadores*, pages 123–129, Porto Alegre, RS, Brasil. SBC.

Dettmers, T. et al. (2023). Qlora: Efficient finetuning of quantized large language models. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Eden, R. et al. (2025). A scoping review of the governance of federated learning in healthcare. *npj Digital Medicine*, 8:427.

Engelke, M. et al. (2025). Fhir-former: enhancing clinical predictions through fast healthcare interoperability resources and large language models. *Journal of the American Medical Informatics Association*, 32(12):1793–1801.

Franco, M. F., Soares, L. R., and Nobre, J. C. (2025). Saúde sob ataque: Da avaliação de riscos ao desenvolvimento de estratégias de investimentos em cibersegurança na área da saúde. In *Anais do XXV SBCAS*, pages 1–44. SBC.

Li, I. et al. (2022). Neural natural language processing for unstructured data in electronic health records: A review. *Computer Science Review*, 46:100511.

Li, Y. et al. (2023). Enhancing health data interoperability with large language models: A fhir study. *arXiv preprint arXiv:2310.12989*.

Liu, Z. et al. (2025). Large language models for automating clinical criteria conversion: validation and evaluation study. *JMIR Medical Informatics*, 13:e71252.

Rajpurkar, P. et al. (2022). Ai in health and medicine. *Nature Medicine*, 28:31–38.

Rieke, N. et al. (2020). The future of digital health with federated learning. *NPJ Digital Medicine*, 3(1):119.

Sedlakova, J. et al. (2023). Challenges and best practices for digital unstructured data enrichment in health research: A systematic narrative review. *PLOS Digital Health*, 2(10):e0000347.

Teo, Z. L. et al. (2024). Federated machine learning in healthcare: A systematic review of clinical applications and technical architecture. *Cell Reports Medicine*.

Wei, C. and Guan, H. (2024). Privacy-preserving federated learning in medical ai: A systematic review of techniques, challenges, and the clinical deployment gap. *Artificial Intelligence and Machine Learning Review*.

Xie, S. et al. (2025). Fedlodrop: Federated lora with dropout for generalized llm fine-tuning. *arXiv preprint arXiv:2510.12078*.

Yi, L. et al. (2023). Pfedlora: Model-heterogeneous personalized federated learning with lora tuning. *arXiv preprint arXiv:2310.13283*.