

# Blockchain como fonte descentralizada de dados para LLMs

**Ana Regina Delazzari<sup>1</sup>, Lucas Machado da Palma<sup>2</sup>, Jean Everson Martina<sup>1</sup>**

<sup>1</sup>Universidade Federal de Santa Catarina (UFSC)  
Florianópolis – SC – Brasil

<sup>2</sup>Laboratório de Segurança em Computação (LabSEC)  
Universidade Federal de Santa Catarina (UFSC)  
Florianópolis – SC – Brasil

ana.regina.delazzari@grad.ufsc.br

lucaspalma.m@gmail.com

jean.martina@ufsc.br

**Abstract.** *Training large-scale language models (LLMs) on data from blockchain networks poses a fundamental architectural compatibility challenge. The practice of centralized training conflicts directly with the essential guarantees of blockchain technology (e.g., immutability, decentralization, and transparency) because its operation depends on the extraction and centralization of data for aggregation. In this article, we indicate that the solution to this challenge lies in a training architecture that preserves decentralization. We propose a model based on a permissioned blockchain, which combines decentralized federated learning with the application of differential privacy to offer guarantees against the inference of sensitive information by third parties.*

**Resumo.** *O treinamento de modelos de linguagem de grande escala (LLMs) sobre dados provenientes de redes blockchain impõe um desafio fundamental de compatibilidade arquitetônica. A prática de treinamento centralizado entra em conflito direto com as garantias essenciais da tecnologia blockchain (e.g., imutabilidade, descentralização e transparência) pois sua operação depende da extração e centralização dos dados para agregação. Neste artigo, indicamos que a solução para este desafio reside em uma arquitetura de treinamento que preserve a descentralização. Propomos um modelo baseado em uma blockchain permissionada, que combina aprendizado federado descentralizado com a aplicação de privacidade diferencial, para oferecer garantias contra a inferência de informações sensíveis por terceiros.*

## 1. Introdução

A tecnologia *blockchain* surgiu a partir de pesquisas em criptografia e sistemas distribuídos, consolidando-se como um registro descentralizado, transparente e imutável de informações. Seu marco inicial foi o artigo de Nakamoto [Nakamoto 2008], que apresentou o Bitcoin, e, desde então, a tecnologia evoluiu significativamente, especialmente com a introdução dos contratos inteligentes pela plataforma Ethereum [Buterin 2014]. Em termos gerais, uma rede *blockchain* funciona como um livro-razão distribuído, registrado em blocos encadeados e compartilhados entre diversos participantes [Xu et al. 2019].

Esses participantes seguem mecanismos de consenso que garantem a integridade das informações, mesmo quando não existe confiança prévia entre eles. Essa propriedade, aliada a outros avanços, permitiu que as aplicações da tecnologia *blockchain* se expandissem para além do setor financeiro, alcançando áreas como saúde, educação, segurança, transporte e administração pública [Sunny et al. 2022].

Paralelamente, a área de inteligência artificial também passou por uma revolução com o surgimento dos *large language models* (LLMs). Esses modelos, baseados na arquitetura Transformer [Vaswani et al. 2017], são capazes de compreender e gerar linguagem em um nível próximo ao humano, aprendendo a partir de enormes volumes de dados textuais. Exemplos conhecidos incluem o ChatGPT [OpenAI et al. 2024] e o LLaMa [Touvron et al. 2023], que já demonstraram sua versatilidade em tarefas como tradução automática, geração de código-fonte e apoio na resolução de problemas em diversos domínios [Minaee et al. 2024]. Um dos fatores que ampliam ainda mais essa versatilidade é o *fine-tuning*, processo em que um LLM pré-treinado é ajustado para um contexto ou conjunto de dados específico, reduzindo custos e permitindo aplicações personalizadas. Contudo, tanto o treinamento inicial quanto o *fine-tuning* dependem fortemente de grandes volumes de dados, cuja concentração em poucas organizações levanta preocupações sobre privacidade, integridade e transparência [Luo et al. 2024].

Nesse cenário, a tecnologia *blockchain* emerge como uma solução potencial ao oferecer uma infraestrutura descentralizada e confiável para armazenamento e compartilhamento de dados, endereçando as deficiências de confiança e centralização do treinamento de LLMs. Ainda assim, a utilização da *blockchain* como um repositório de dados não resolve a questão da privacidade durante o processo de treinamento. Técnicas como o aprendizado federado (*federated learning*, ou FL) surgiram para permitir o treinamento de modelos em fontes de dados distribuídas sem que os dados brutos precisem ser compartilhados [Luo et al. 2024, McMahan et al. 2017]. Entretanto, a arquitetura convencional de FL ainda depende de um elemento central para agregar as atualizações dos modelos locais, reintroduzindo um ponto único de falha. Quando essa etapa de agregação em um servidor é substituída por uma rede *blockchain*, o processo passa a ser chamado de aprendizado federado descentralizado (*decentralized federated learning*, ou DFL) [Wang et al. 2025].

A convergência entre essas tecnologias representa uma significativa oportunidade de pesquisa, dada a ausência de soluções consolidadas que considerem dados armazenados em redes *blockchains* para alimentar LLMs e, ao mesmo tempo, realizem o treinamento de maneira descentralizada. Desse modo, e visando explorar essa lacuna, este artigo propõe a integração da tecnologia *blockchain* com técnicas de aprendizado federado descentralizado para realizar o *fine-tuning* de LLMs. Trata-se de um trabalho em andamento, cuja implementação e avaliação serão abordadas em etapas futuras do trabalho.

A estrutura deste artigo é a seguinte: a seção 2 apresenta os trabalhos relacionados, identificados através de uma revisão da literatura, a seção 3 apresenta uma proposta de solução teórica e a seção 4 apresenta a conclusão.

## 2. Trabalhos Relacionados

A integração entre redes *blockchain* e LLMs é uma área de pesquisa emergente, com diferentes propostas que buscam utilizar a segurança e a transparência das *blockchains*

para mitigar os riscos de centralização no treinamento de LLMs. Esta seção analisa o estado da arte através de uma revisão estruturada da literatura, sintetizando as principais abordagens, desafios e as lacunas que motivam este trabalho.

Para garantir o rigor e a reproduzibilidade da busca, a revisão foi estruturada com base nas diretrizes propostas por [Kitchenham 2004]. O processo foi orientado pelas seguintes questões de pesquisa:

1. *Como as blockchains podem ser integradas com LLMs?*
2. *Como o armazenamento de dados baseado em blockchain impacta a eficiência e a escalabilidade dos LLMs em comparação com sistemas centralizados?*
3. *De que forma a integração da tecnologia blockchain aos LLMs impacta a privacidade e a integridade dos dados utilizados no processo de treinamento?*
4. *Quão eficazes são os LLMs na geração de respostas corretas para perguntas relacionadas a dados armazenados em blockchains?*

A busca foi realizada nas bases de dados ACM Digital Library, IEEE Xplore, ScienceDirect, SpringerLink e Google Scholar, cobrindo publicações entre o final de 2019 e o final de 2024. A *query* de busca utilizada foi:

```
("blockchain" OR "distributed ledger") AND  
("data" OR "data storage" OR "data warehouse") AND  
("large language models" OR "llms" OR "transformer model")
```

A seleção dos artigos ocorreu em etapas sequenciais, envolvendo a análise de títulos, resumos e, por fim, a leitura completa dos trabalhos para verificar sua relevância em relação aos objetivos da pesquisa.

A princípio, constata-se que as propostas variam significativamente em maturidade e domínio de aplicação. Enquanto trabalhos como [Gao et al. 2023] e [Kozgunov et al. 2024] permanecem no campo teórico, outros apresentam protótipos aplicados a casos de uso específicos, como veículos autônomos [Alam et al. 2024], detecção de anomalias em redes [Batool et al. 2022], processamento de dados médicos [Balija et al. 2024] e privacidade em *chatbots* [Su et al. 2024]. No entanto, mesmo os trabalhos com validação prática geralmente se limitam a ambientes simulados, modelos de menor escala (TinyLMs<sup>1</sup> ou variações de BERT<sup>2</sup>) e redes *blockchain* privadas (Hyperledger Fabric<sup>3</sup>/Sawtooth<sup>4</sup> ou Ethereum<sup>5</sup> local), deixando incertezas sobre a viabilidade em cenários de produção com grande volume de dados e participantes.

Além disso, as arquiteturas propostas se dividem principalmente em duas estratégias de armazenamento de modelos: *on-chain* e *off-chain*. Propostas teóricas como GradientCoin [Gao et al. 2023] e LinguaChain [Kozgunov et al. 2024] defendem o armazenamento dos parâmetros do modelo diretamente na *blockchain* (*on-chain*). Essa abordagem maximiza a integridade e a auditabilidade, mas enfrenta severos desafios de escalabilidade, latência e custo, tornando sua implementação prática para LLMs de grande

---

<sup>1</sup>Versões reduzidas de LLMs, otimizadas para tarefas específicas [Su et al. 2024].

<sup>2</sup>Modelo de linguagem baseado na arquitetura Transformer que comprehende o contexto bidirecional do texto [Jiao and Zhang 2021].

<sup>3</sup>*Blockchain* permissionada modular voltada a aplicações empresariais [Xu et al. 2019].

<sup>4</sup>*Blockchain* permissionada modular, desenvolvida parcialmente pela Intel [Ampel et al. 2019].

<sup>5</sup>*Blockchain* amplamente utilizada, permite criar contratos inteligentes em uma linguagem Turing completa [Xu et al. 2019].

porte questionável. Em contraste, soluções com validação experimental, como as de [Batoool et al. 2022] e [Wang et al. 2025], adotam uma abordagem híbrida, utilizando sistemas de armazenamento de arquivos descentralizados como o IPFS<sup>6</sup> para os dados e a *blockchain* apenas para registrar os *hashes* de referência. Embora mais escalável, essa estratégia introduz dependências externas e complexidades adicionais.

No que tange à privacidade, a maioria dos trabalhos utiliza o aprendizado federado como base. No entanto, o FL por si só não é imune a ataques de inferência que podem expor dados de treinamento [Gupta et al. 2022]. Poucos estudos avançam para além do FL básico. [Su et al. 2024] se destaca ao incorporar privacidade diferencial para adicionar ruído aos gradientes, [Wang et al. 2025] implementa um protocolo de agregação segura (Secure Aggregation) e [Kozgunov et al. 2024] sugere o uso de criptografia homomórfica. [Zuo et al. 2024], por outro lado, oferece aos usuários a possibilidade de solicitar a remoção de seus dados. Essas técnicas, embora robustas, frequentemente introduzem uma sobrecarga computacional e podem impactar a acurácia do modelo final, evidenciando um *trade-off* ainda em aberto entre privacidade e desempenho.

Portanto, o estado da arte demonstra um interesse crescente na relação entre a tecnologia *blockchain* e LLMs, mas as soluções existentes são imaturas, na maior parte dos casos. Uma comparação dos trabalhos selecionados pode ser encontrada na Tabela 1, onde o símbolo  $\checkmark^*$  no caso de uso denota que o trabalho oferece uma noção geral de sua aplicabilidade, mas não define um domínio de aplicação explicitamente.

A revisão conduzida permitiu responder parcialmente às questões de pesquisa propostas. Quanto à primeira, observou-se que a integração entre *blockchains* e LLMs ocorre predominantemente por meio de aprendizado federado descentralizado, embora também existam aplicações que utilizam LLMs para análise de transações e contratos inteligentes. Em relação à segunda, identificou-se que o armazenamento de dados diretamente na *blockchain* impõe sobrecarga significativa, o que leva alguns trabalhos a adotar estratégias híbridas baseadas em sistemas externos. A terceira questão revelou que a tecnologia *blockchain* contribui para assegurar a rastreabilidade e a auditabilidade das atualizações dos modelos, mas não elimina por completo os riscos de exposição de dados, exigindo o uso de técnicas complementares como privacidade diferencial e agregação segura. Por fim, não foram encontrados estudos que respondam à quarta questão, evidenciando a ausência de arquiteturas de LLMs voltadas à consulta de dados armazenados em redes *blockchain* e confirmando que se trata de um campo ainda em aberto para investigação.

**Tabela 1. Comparativo dos trabalhos relacionados.**

Trabalho	Caso de uso	Experimento	Armazenamento dos parâmetros	Tipo de rede	Tecnologia blockchain	LLM	Mecanismo adicional de segurança
[Alam et al. 2024]	$\checkmark$	$\checkmark$	On-chain	Desconhecido	Desconhecida	GPT-4V	-
[Balija et al. 2024]	$\checkmark$	$\checkmark$	On-chain	Privada	Hyperledger Sawtooth	BioBERT	-
[Batoool et al. 2022]	$\checkmark$	$\checkmark$	IPFS	Privada	Ethereum	Transformer	-
[Gao et al. 2023]	$\checkmark^*$	$\times$	On-chain	Desconhecido	Desconhecida	Desconhecido	-
[Kozgunov et al. 2024]	$\checkmark^*$	$\times$	On-chain	Desconhecido	Desconhecida	Desconhecido	Criptografia homomórfica
[Su et al. 2024]	$\checkmark$	$\checkmark$	IPFS	Privada	Ethereum	Falcon-7B	Privacidade diferencial
[Subrmanian et al. 2024]	$\checkmark^*$	$\checkmark$	-	Desconhecido	Desconhecida	Desconhecido	-
[Wang et al. 2025]	$\checkmark^*$	$\checkmark$	IPFS	Privada	Hyperledger Fabric	BERT	Agregação segura
[Zuo et al. 2024]	$\checkmark^*$	$\checkmark$	On-chain	Privada	Hyperledger Fabric	GPT-2	Remoção de dados

<sup>6</sup>Interplanetary File System ou IPFS é um sistema de arquivos distribuído para compartilhamento de grandes volumes de dados [Xu et al. 2019].

### 3. Proposta

A proposta consiste em uma arquitetura de aprendizado federado descentralizado aplicada ao *fine-tuning* de LLMs. Diferente das abordagens tradicionais, o processo é realizado a partir de dados armazenados em uma *blockchain* privada e permissionada, na qual há participação ativa dos nós da rede e coordenação das etapas por meio de contratos inteligentes, garantindo descentralização, auditabilidade e preservação da integridade dos dados.

Na prática, todos os participantes da rede podem contribuir com seus recursos computacionais para o ajuste e atualização dos modelos, dividindo a carga de trabalho e assegurando que o resultado final esteja ancorado nos dados da própria *blockchain*. Além disso, a proposta não se restringe a um único modelo: diferentes LLMs podem ser treinados em paralelo, desde que exista consenso entre os participantes.

A arquitetura contempla duas versões, de acordo com o nível de segurança exigido. A versão simples é adequada a contextos em que os participantes são confiáveis, enquanto a versão robusta foi projetada para ambientes potencialmente adversos, onde há possibilidade da existência de nós maliciosos (bizantinos). Esta versão eleva a resiliência do treinamento, incorporando mecanismos de defesa, como um conjunto de nós que faz a validação das atualizações submetidas e cálculo da reputação dos nós, para desestimular contribuições de baixa qualidade e combater ataques como o envenenamento do modelo, onde agentes mal-intencionados inserem dados falsos ou atualizações manipuladas para degradar o desempenho global [Moraes et al. 2024].

Outra característica central da proposta é a parametrização via consenso. Inicialmente, todos os participantes definem coletivamente os parâmetros fundamentais do processo de treinamento, que podem ser ajustados posteriormente. Isso inclui:

- **Local de armazenamento do modelo:** Devido às limitações de armazenamento da *blockchain* (como o tamanho dos blocos) [Xie et al. 2019], é inviável armazenar um LLM completo *on-chain* de forma eficiente. Então, os participantes devem escolher um sistema externo (como IPFS), e salvar apenas o *hash* do modelo na *blockchain* [Cai et al. 2024] para garantir a integridade.
- **Local de armazenamento das atualizações:** Se a *blockchain* permite blocos maiores ou existe uma frequência baixa de atualização, é possível armazenar as atualizações *on-chain*, caso contrário, é preciso escolher um armazenamento externo, com o *hash* correspondente registrado na *blockchain*.
- **Regras de governança:** Define-se a quantidade mínima de blocos exigida entre atualizações (para evitar atualizações muito frequentes e desperdício de recursos computacionais), o método de distribuição dos blocos entre os participantes para o treinamento (*round-robin*, por exemplo), e os critérios para a seleção de nós agregadores e validadores.
- **Limites de tempo:** Estabelecem-se limites de tempo para cada fase do processo síncrono, garantindo que o treinamento avance de maneira organizada.

O fluxo de atualização de um modelo é um processo cíclico e colaborativo:

- **Inicialização e distribuição dos blocos:** Um nó propõe uma atualização, e as faixas de blocos da *blockchain* contendo os dados de treinamento são distribuídas entre os participantes. Nesse momento, realiza-se um pré-processamento em

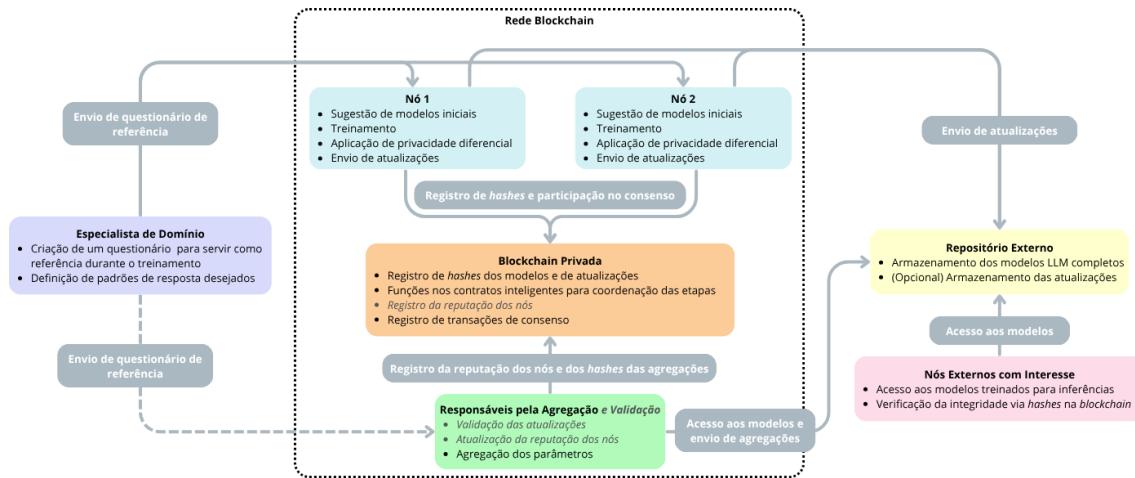
que cada bloco é associado ao seu respectivo *timestamp*, garantindo a correta ordenação temporal e permitindo ao modelo distinguir entre versões antigas e recentes dos dados, evitando inconsistências no treinamento.

- **Seleção dos responsáveis pela validação e agregação:** Na versão robusta, é necessário formar um comitê de nós para agregar e validar os modelos. Enquanto na versão simples, é necessário escolher um nó agregador.
- **Treinamento local e privacidade:** Cada nó realiza o *fine-tuning* em sua porção de dados. Para otimizar o uso de recursos, adota-se o algoritmo LoRA (*Low-Rank Adaptation*) [Hu et al. 2021]. Antes do envio, cada nó aplica técnicas de privacidade diferencial aos parâmetros do modelo, evitando que usuários externos que tenham acesso ao modelo publicado possam inferir os dados originais utilizados no treinamento.
- **Submissão e validação:** Cada participante submete sua atualização local (ou seu *hash*) à *blockchain*. Na versão robusta, um comitê de validação avalia a qualidade de cada contribuição, comparando a acurácia reportada com uma métrica de validação própria. As contribuições aceitas aumentam a reputação do nó, enquanto as rejeitadas a diminuem.
- **Agregação e consenso:** Os nós designados como agregadores combinam as atualizações locais validadas para criar um novo modelo global. Na versão robusta, múltiplos agregadores realizam o processo de forma independente e devem chegar a um consenso sobre o resultado final, adicionando uma camada extra de resiliência.
- **Finalização:** O *hash* do novo modelo global é registrado na *blockchain*, juntamente com metadados como a faixa de blocos utilizada e a acurácia final, concluindo o ciclo e criando um registro imutável e verificável da evolução do modelo.

Os nós da rede podem disponibilizar o modelo treinado a usuários externos que tenham interesse em utilizá-lo para tarefas específicas ou em avaliá-lo de forma independente. Esse acesso é concedido por meio do repositório externo definido pelo consenso da rede (IPFS, por exemplo), utilizando os *hashes* registrados na *blockchain* como referência de integridade e versão. Embora esses usuários não participem do processo de treinamento, eles podem acessar as versões publicadas do modelo de forma auditável. Por esse motivo, a aplicação de técnicas de privacidade diferencial durante o *fine-tuning* é essencial para impedir que informações sensíveis sobre os dados originais sejam inferidas a partir do modelo disponibilizado.

A arquitetura também prevê a participação de um especialista de domínio, responsável por criar um conjunto de perguntas e respostas de referência. Esse material é utilizado para o treinamento supervisionado e, crucialmente, pela comissão de validação para avaliar objetivamente o desempenho das atualizações de cada nó.

O diagrama da Figura 1 apresenta um exemplo das interações entre os diferentes elementos e suas funções em um cenário com dois participantes na rede *blockchain*. Esses participantes correspondem aos nós da rede, que executam localmente o aprendizado federado e participam da manutenção da *blockchain*. As funções destacadas em itálico e com fonte em tom de cinza representam funções exclusivas da versão mais robusta da arquitetura, enquanto as conexões tracejadas indicam interações que ocorrem apenas nessa configuração. Os balões em cinza posicionados sobre as setas indicam os tipos de dados



**Figura 1. Interação entre os componentes da arquitetura.** Fonte: Elaboração própria.

compartilhados entre os elementos durante as etapas de comunicação e atualização dos modelos.

Em síntese, a arquitetura proposta combina elementos já explorados na literatura com algumas inovações específicas para aprendizado federado apoiado em *blockchain*. Entre os aspectos aproveitados estão a coordenação do aprendizado federado via *blockchain* e o uso do LoRA para *fine-tuning*. Como contribuição original, a solução introduz mecanismos parametrizáveis definidos por consenso entre os nós participantes, incluindo a escolha do modelo base, o armazenamento dos parâmetros, a divisão do trabalho de treinamento e os critérios de seleção de nós agregadores e validadores. Além disso, permite a exposição controlada do modelo treinado a usuários externos autorizados, evitando que eles possam inferir os dados de treinamento por meio do mecanismo de privacidade diferencial. Porém, a principal diferença reside no fato de que o treinamento parte de dados armazenados em uma *blockchain* já existente, preservando desde a origem propriedades como descentralização, integridade e confiabilidade, que são mantidas ao longo de todo o ciclo de ajuste e atualização do modelo.

## 4. Conclusão

Este trabalho apresentou uma arquitetura que integra *blockchains*, aprendizado federado e LLMs, propondo o *fine-tuning* descentralizado como alternativa às abordagens centralizadas atuais. A proposta busca garantir descentralização, auditabilidade e resiliência, ao mesmo tempo em que preserva a privacidade dos dados. Apesar de ainda conceitual, acreditamos que essa convergência representa uma direção promissora de pesquisa, capaz de viabilizar modelos de linguagem treinados de forma colaborativa, transparente e segura. Vale ressaltar que o algoritmo de controle de reputação dos nós ainda não foi definido, constituindo uma lacuna a ser investigada. Como trabalho futuro, pretendemos desenvolver um protótipo para validar a viabilidade da solução em um cenário de uso real, no contexto de dados provenientes de instituições de ensino superior brasileiras.

## Referências

- Alam, T., Gupta, R., Ahamed, N. N., and Ullah, A. (2024). A decision-making model for self-driving vehicles based on GPT-4V, federated reinforcement learning, and blockchain. *Neural Computing and Applications*, 36(34):21545–21560.
- Ampel, B., Patton, M., and Chen, H. (2019). Performance modeling of hyperledger sawtooth blockchain. In *2019 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pages 59–61. IEEE.
- Balija, S. B., Nanda, A., and Sahoo, D. (2024). Building Communication Efficient Asynchronous Peer-to-Peer Federated LLMs with Blockchain. *Proceedings of the AAAI Symposium Series*, 3(1):288–292.
- Batool, Z., Zhang, K., Zhu, Z., Aravamuthan, S., and Aivodji, U. (2022). Block-FeST: A Blockchain-Based Federated Anomaly Detection framework with computation offloading using Transformers. In *2022 IEEE 1st Global Emerging Technology Blockchain Forum: Blockchain & Beyond (iGETblockchain)*, pages 1–6, Irvine, CA, USA. IEEE.
- Buterin, V. (2014). A Next Generation Smart Contract & Decentralized Application Platform.
- Cai, Z., Chen, J., Fan, Y., Zheng, Z., and Li, K. (2024). Blockchain-empowered Federated Learning: Benefits, Challenges, and Solutions. arXiv:2403.00873 [cs].
- Gao, Y., Song, Z., and Yin, J. (2023). GradientCoin: A Peer-to-Peer Decentralized Large Language Models. arXiv:2308.10502 [cs].
- Gupta, S., Huang, Y., Zhong, Z., Gao, T., Li, K., and Chen, D. (2022). Recovering Private Text in Federated Learning of Language Models.
- Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., and Chen, W. (2021). LoRA: Low-Rank Adaptation of Large Language Models. arXiv:2106.09685 [cs].
- Jiao, Q. and Zhang, S. (2021). A Brief Survey of Word Embedding and Its Recent Development. In *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, pages 1697–1701, Chongqing, China. IEEE.
- Kitchenham, B. (2004). Procedures for Performing Systematic Reviews.
- Kozgunov, N. V., Khalashi, M. H., Oliseenko, V. D., and Tulupyeva, T. V. (2024). Lingua-Chain: a Peer-to-peer Dynamic Decentralized Large Language Model with Coin-based Incentives. In *2024 XXVII International Conference on Soft Computing and Measurements (SCM)*, pages 178–181, Saint Petersburg, Russian Federation. IEEE.
- Luo, H., Luo, J., and Vasilakos, A. V. (2024). BC4LLM: A perspective of trusted artificial intelligence when blockchain meets large language models. *Neurocomputing*, 599:128089.
- McMahan, H. B., Moore, E., Ramage, D., and Hampson, S. (2017). Communication-Efficient Learning of Deep Networks from Decentralized Data.
- Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., and Gao, J. (2024). Large Language Models: A Survey. arXiv:2402.06196 [cs].

Morais, M., Costa, J., Gonzalez, L., Souza, A., and Villas, L. (2024). Mecanismo para mitigar ataques de envenenamento de modelo no aprendizado federado. In *Anais do VIII Workshop de Computação Urbana*, pages 224–237, Porto Alegre, RS, Brasil. SBC.

Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System.

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., Bello, I., Berdine, J., Bernadett-Shapiro, G., Berner, C., Bogdonoff, L., Boiko, O., Boyd, M., Brakman, A.-L., Brockman, G., Brooks, T., Brundage, M., Button, K., Cai, T., Campbell, R., Cann, A., Carey, B., Carlson, C., Carmichael, R., Chan, B., Chang, C., Chantzis, F., Chen, D., Chen, S., Chen, R., Chen, J., Chen, M., Chess, B., Cho, C., Chu, C., Chung, H. W., Cummings, D., Currier, J., Dai, Y., Decareaux, C., Degry, T., Deutsch, N., Deville, D., Dhar, A., Dohan, D., Dowling, S., Dunning, S., Ecoffet, A., Eleti, A., Eloundou, T., Farhi, D., Fedus, L., Felix, N., Fishman, S. P., Forte, J., Fulford, I., Gao, L., Georges, E., Gibson, C., Goel, V., Gogineni, T., Goh, G., Gontijo-Lopes, R., Gordon, J., Grafsstein, M., Gray, S., Greene, R., Gross, J., Gu, S. S., Guo, Y., Hallacy, C., Han, J., Harris, J., He, Y., Heaton, M., Heidecke, J., Hesse, C., Hickey, A., Hickey, W., Hoeschele, P., Houghton, B., Hsu, K., Hu, S., Hu, X., Huizinga, J., Jain, S., Jain, S., Jang, J., Jiang, A., Jiang, R., Jin, H., Jin, D., Jomoto, S., Jonn, B., Jun, H., Kaftan, T., Kaiser, L., Kamali, A., Kanitscheider, I., Keskar, N. S., Khan, T., Kilpatrick, L., Kim, J. W., Kim, C., Kim, Y., Kirchner, J. H., Kirov, J., Knight, M., Kokotajlo, D., Kondraciuk, L., Kondrich, A., Konstantinidis, A., Kosic, K., Krueger, G., Kuo, V., Lampe, M., Lan, I., Lee, T., Leike, J., Leung, J., Levy, D., Li, C. M., Lim, R., Lin, M., Lin, S., Litwin, M., Lopez, T., Lowe, R., Lue, P., Makanju, A., Malfacini, K., Manning, S., Markov, T., Markovski, Y., Martin, B., Mayer, K., Mayne, A., McGrew, B., McKinney, S. M., McLeavey, C., McMillan, P., McNeil, J., Medina, D., Mehta, A., Menick, J., Metz, L., Mishchenko, A., Mishkin, P., Monaco, V., Morikawa, E., Mossing, D., Mu, T., Murati, M., Murk, O., Mély, D., Nair, A., Nakano, R., Nayak, R., Neelakantan, A., Ngo, R., Noh, H., Ouyang, L., O’Keefe, C., Pachocki, J., Paino, A., Palermo, J., Pantuliano, A., Parascandolo, G., Parish, J., Parparita, E., Passos, A., Pavlov, M., Peng, A., Perelman, A., Peres, F. d. A. B., Petrov, M., Pinto, H. P. d. O., Michael, Pokorny, Pokrass, M., Pong, V. H., Powell, T., Power, A., Power, B., Proehl, E., Puri, R., Radford, A., Rae, J., Ramesh, A., Raymond, C., Real, F., Rimbach, K., Ross, C., Rotstet, B., Roussez, H., Ryder, N., Saltarelli, M., Sanders, T., Santurkar, S., Sastry, G., Schmidt, H., Schnurr, D., Schulman, J., Selsam, D., Sheppard, K., Sherbakov, T., Shieh, J., Shoker, S., Shyam, P., Sidor, S., Sigler, E., Simens, M., Sitkin, J., Slama, K., Sohl, I., Sokolowsky, B., Song, Y., Staudacher, N., Such, F. P., Summers, N., Sutskever, I., Tang, J., Tezak, N., Thompson, M. B., Tillet, P., Tootoonchian, A., Tseng, E., Tuggle, P., Turley, N., Tworek, J., Uribe, J. F. C., Vallone, A., Vijayvergiya, A., Voss, C., Wainwright, C., Wang, J. J., Wang, A., Wang, B., Ward, J., Wei, J., Weinmann, C. J., Welihinda, A., Welinder, P., Weng, J., Weng, L., Wiethoff, M., Willner, D., Winter, C., Wolrich, S., Wong, H., Workman, L., Wu, S., Wu, J., Wu, M., Xiao, K., Xu, T., Yoo, S., Yu, K., Yuan, Q., Zaremba, W., Zellers, R., Zhang, C., Zhang, M., Zhao, S., Zheng, T., Zhuang, J., Zhuk, W., and Zoph, B. (2024). GPT-4 Technical Report. arXiv:2303.08774 [cs].

- Su, H., Xiang, C., and Ramesh, B. (2024). Towards Confidential Chatbot Conversations: A Decentralised Federated Learning Framework. *The Journal of The British Blockchain Association*, 7(1):1–8.
- Subrmanian, K., Thangarasu, G., Yanyan, Z., and Kannan, K. N. (2024). Deep Learning Based Algorithm for Efficient Information Retrieval in Blockchain Transactions. In *2024 IEEE 6th Symposium on Computers & Informatics (ISCI)*, pages 264–269, Kuala Lumpur, Malaysia. IEEE.
- Sunny, F. A., Hajek, P., Munk, M., Abedin, M. Z., Satu, M. S., Efat, M. I. A., and Islam, M. J. (2022). A Systematic Review of Blockchain Applications. *IEEE Access*, 10:59155–59177.
- Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., and Lample, G. (2023). LLaMA: Open and Efficient Foundation Language Models. arXiv:2302.13971 [cs].
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is All you Need. *Advances in neural information processing systems*, 30.
- Wang, H., Cai, Y., Tao, Y., Wang, L., Li, Y., and Zhou, L. (2025). B2DFL: Bringing butterfly to decentralized federated learning assisted with blockchain. *Journal of Parallel and Distributed Computing*, 195:104978.
- Xie, J., Yu, F. R., Huang, T., Xie, R., Liu, J., and Liu, Y. (2019). A survey on the scalability of blockchain systems. *IEEE network*, 33(5):166–173.
- Xu, X., Weber, I., and Staples, M. (2019). *Architecture for Blockchain Applications*. Springer International Publishing, Cham.
- Zuo, X., Wang, M., Zhu, T., Zhang, L., Ye, D., Yu, S., and Zhou, W. (2024). Federated TrustChain: Blockchain-Enhanced LLM Training and Unlearning. arXiv:2406.04076 [cs].