Transfer Deep Learning using Homomorphic Encryption

Péter **Mészáros**[1]

1. Budapest University of Technology and Economics, Budapest, Hungary

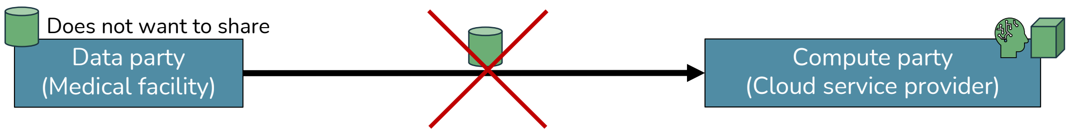
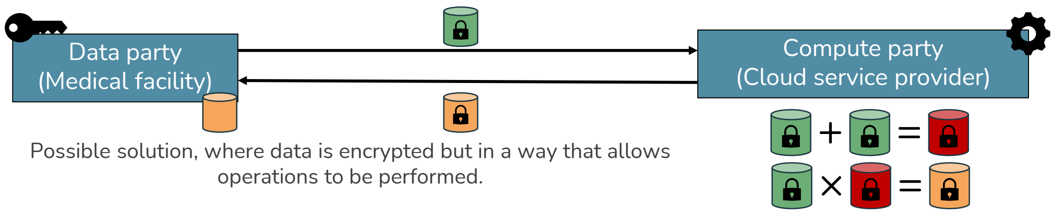


Illustration of the problem where the data party (in this case a medical facility) being reluctant to share the sensitive patient data (medical records) to a compute party (a cloud service provider) directly, but wants to be able to profit from it by making predictions (possible patient conditions).



Possible solution, where data is encrypted but in a way that allows operations to be performed.

Abstract. This work explores the use of homomorphic encryption (HE) for privacy-preserving deep learning, particularly in transfer learning scenarios, where a model is only trained on a part of a model. However, applying HE to deep learning is challenging due to computational overhead and limited support for complex operations. This study investigates possible limitations and effects on the models.

Introduction

Deep learning applications require large amounts of data and extensive computational power. The data and computational power along with expertise in AI are usually at different parties, requiring the data to be exposed. However, directly sharing sensitive data poses significant privacy risks, especially in domains like healthcare and finance. A possible solution is homomorphic encryption (HE), which allows computations on encrypted data without revealing them [1,2], **ensuring that privacy is preserved at the compute party.** Current HE schemes support only a limited set of operations and often a limited number of times they can be performed. A promising fully homomorphic encryption scheme currently is CKKS [3], which supports arbitrary number of addition and multiplication. **It is not straightforward to apply it to deep learning, due to model complexity and lack of openly available solutions to encrypted learning**. The proposed solution looks at performing inference on a larger model and training on a smaller part of the model, where it is feasable to do without losing accuracy and without using excessive computational resources.

Related works

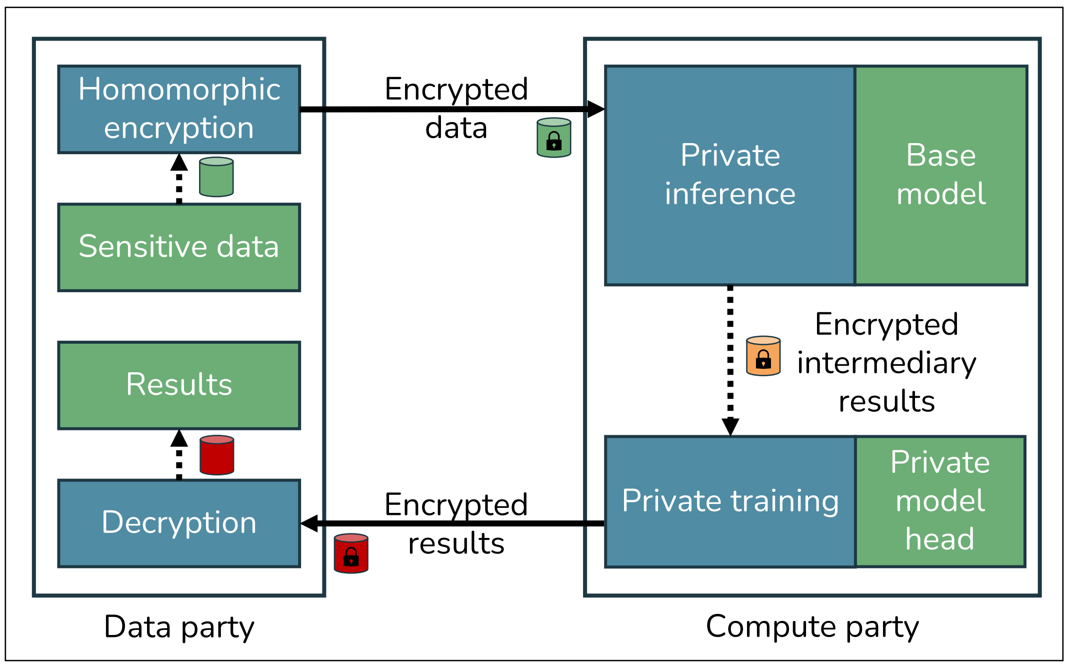
Works implementing HE with deep learning mainly focus on private inference (and also image processing neural networks) [4], not private training. So the problem is that the sensitive data has to be sent to the other party for training or the data party has to train the model, but may lack computational power.

Other approaches for this problem include secure multi-party computation [6], federated learning [5] and differencial privacy [7], with each having its trade-offs in terms of privacy, efficiency and applicability, they can also differ in the number of parties involved.

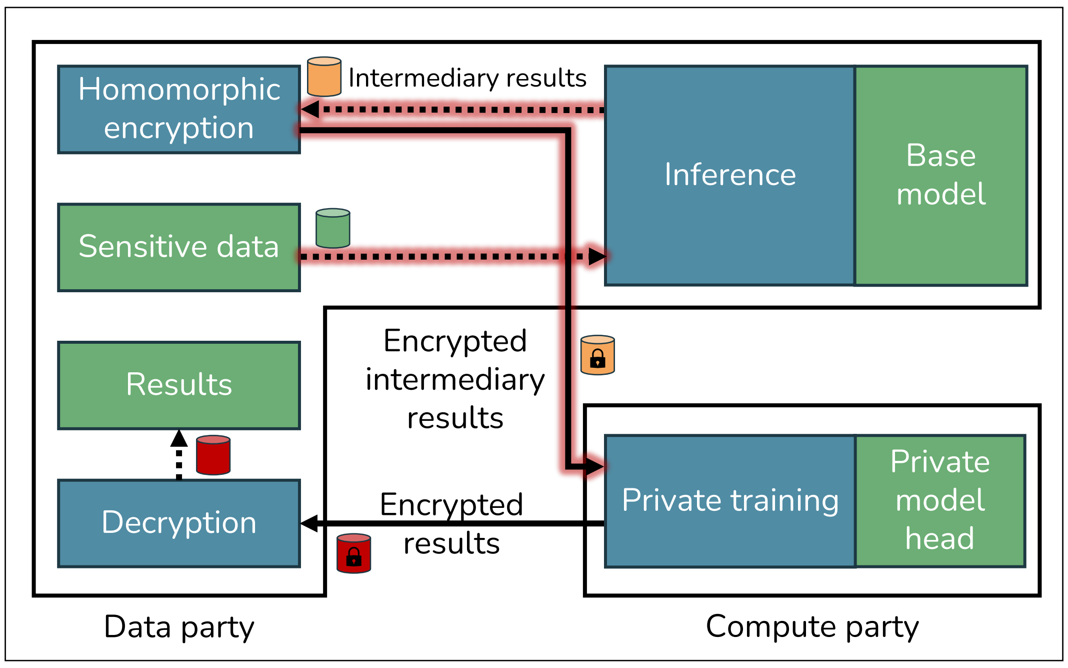
Research questions

- Assess what limitations are imposed by HE on deep learning models. (Number of layers in the model, used activation functions, approximated activation functions.)
- How does HE impact accuracy in training?
- Does private inference reach the same level of accuracy?
- What are the computational overheads of using HE in transfer learning?
- Is HE a suitable approach for privacy-preserving deep learning.

Setup 1. Private inference and training



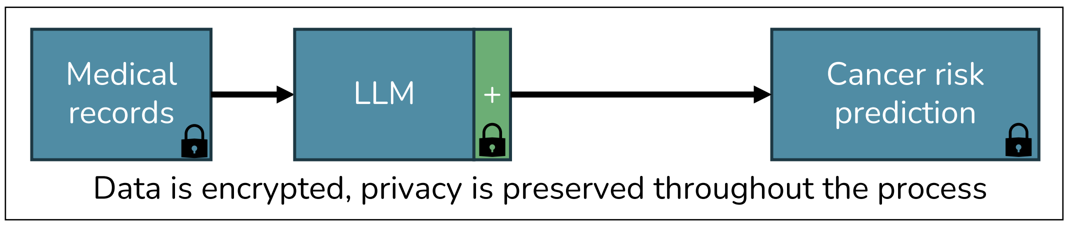
Setup 2. Only private training



Proposed solutions preverving privacy. On the left, the data party encrypts its data and the compute party performs inference *privately* on a base model and *trains* a model head *privately* for the data party's needs. On the right the inference is performed openly by the data party and only the intermediary representations are encrypted and then trained on by the compute party. (Dotted lines represent data flow within a party, solid lines represent data flow between the parties.)

Healthcare case study

Suppose an application, where a hospital wants to analyze medical records with an LLM-based solution to classify patients according to their risk of having cancer. The compute party has no access to the data directly, only does computations.



Data is encrypted, privacy is preserved throughout the process

Conclusions

Homomorphic encryption presents a promising approach for privacy-preserving deep learning by enabling computations on encrypted data without exposing sensitive information. However, its practical application seems challenging due to computational overhead and constraints on supported operations. This study explores two setups, leveraging CKKS for inference on a larger model while training a smaller submodel or just train on the smaller model. The findings will highlight the complexity and performance limitations.

References.

[1] C. Marcolla, V. Sucasas, M. Manzano, R. Bassoli, F. Fitzek, and N. Aaraj, "Survey on fully homomorphic encryption, theory, and applications," *Proceedings of the IEEE*, pp. 1–38, 2022.

[2] A. Acar, H. Aksu, A. S. Uluagac, and M. Conti, "A survey on homomorphic encryption schemes: Theory and implementation," *ACM Computing Surveys*, vol. 51, no. 4, Article 79, Jul. 2019.

[3] J. H. Cheon, A. Kim, M. Kim, and Y. Song, "Homomorphic encryption for arithmetic of approximate numbers," *Advances in Cryptology – ASIACRYPT 2017*, T. Takagi and T. Peyrin, Eds. Cham: Springer, 2017, vol. 10624.

[4] A. Ebel, K. Garimella, and B. Reagen, Orion: A Fully Homomorphic Encryption Framework for Deep Learning. 2024. [Online]. Available https://arxiv.org/abs/2311.03470 Accessed on: 30-01-2025

[5] C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, and Y. Gao, "A survey on federated learning," *Knowledge-Based Systems*, vol. 216, p. 106775, 2021.

[6] I. Zhou, F. Tofigh, M. Piccardi, M. Abolhasan, D. Franklin, and J. Lipman, "Secure Multi-Party Computation for Machine Learning: A Survey," *IEEE Access*, vol. 12, pp. 53881–53899, 2024.

[7] M. Yang, T. Guo, T. Zhu, I. Tjuawinata, J. Zhao, and K.-Y. Lam, "Local differential privacy and its applications: A comprehensive survey," *Computer Standards & Interfaces*, vol. 89, p. 103827, 2024.

Acknowledgements. The HCAIM (the Human-Centred AI Master's Programme) Project is Co-Financed by the Connecting Europe Facility of the European Union Under Grant №CEF-TC-2020-1 Digital Skills 2020-EU-IA-0068. This poster was created as part of the Blended Intensive Programme organized under the Erasmus + Programme of the European Union