

# CryptoLLM

## Large Language Models for Cryptography Assistance

---

Manpa Barman

April 25, 2024

MSc. in Information Technology (INFOTECH), University of Stuttgart

[www.perceptualui.org](http://www.perceptualui.org) 

Introduction

Methodology

Progress and Results

Future Work



- Suboptimal performance of LLMs in mathematical reasoning task [1] [2]
- Arithmetic problems have a single solution with multistep reasoning.
- Recent LLMs perform well for low-difficulty mathematics but suffer in medium and higher-difficulty challenges [3; 4; 5].



## Motivation - Why cryptographic LLMs?

- Cryptography is essential for secure information and transactions.
- Cryptography is a complex field that requires a strong foundation in mathematics and computer science.
- Cryptography is one of the most difficult fields to learn.
- There exist no AI assistants, only limited mathematical tools.



# Motivation - Why cryptographic LLMs?



You

what is dishonest majority in cryptography?



ChatGPT

I'm not aware of a specific term or concept called "dishonest majority" in cryptography. It's possible that there might be a misunderstanding or a miscommunication of the term. Cryptography involves the study of secure communication techniques, and concepts like trust, honesty, and security play crucial roles in this field.

If you have a specific context or if there's a different term you're referring to, please provide more details, and I'll do my best to assist you. Alternatively, if it's a newer concept or terminology that has emerged after my last training data in January 2022, I might not be aware of it.



You

Is  $2+2+22+222+2=300$ ?

< 3 / 3 >



ChatGPT

Let's add the numbers:

$$2 + 2 + 22 + 222 + 2 = 250$$

So,  $2 + 2 + 22 + 222 + 2$  equals 250, not 300.



You

Is  $02+2+22+0222+2=300$ ?

< 4 / 4 >



ChatGPT

Let's add the numbers:

$$02 + 2 + 22 + 0222 + 2 = 220$$

So,  $02 + 2 + 22 + 0222 + 2$  equals 220, not 300.



We aim to develop:

- CryptoLLM - an LLM assistant for cryptographic problems using language chaining and fine-tuned model.
- Crypto dataset collection to train the LLM.

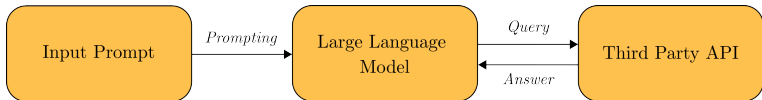


Figure: Language Chaining Illustration



Introduction

**Methodology**

Progress and Results

Future Work



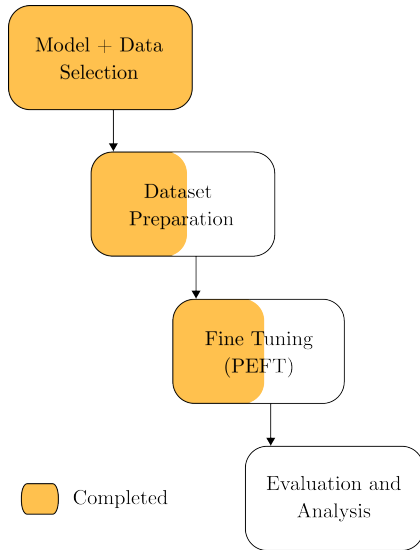


Figure: Methodology of the design





- Llama-2-Chat 7B model [6]
  - Open Source
  - At par in some criteria with popular closed-source models like ChatGPT [7] and PaLM [8].
  - Proven results for basic math datasets [9]
  - Reasonable compute requirements for fine-tuning - 28GB GPU for full precision (7B model)
- Mistral 7B [10] and CodeLlama[11] will be tested parallelly for results.



- Main cryptographic primitives:
  - **Number Theory**
  - **Symmetric Cipher Model**
  - Block Ciphers
  - Finite Fields
  - Random Bit Generation and Stream Ciphers
  - Asymmetric Ciphers... etc.



- Well-established academic textbooks
  - Cryptography and Network Security, William Stallings [12]
  - Number Theory and Cryptography, Neal Koblitz [13]
  - A Gentle introduction to number theory and cryptography, Luis Finotti [14]
  - Mathematics Aptitude Test of Heuristics (MATH) [15]



- Tokenization
- Standardization
  - Numerical form in latex common standard.
  - Chain of thought approach [16]
- Monolingual Data - English only
- Manual Inspection



- Labels: for a good train-test split and markers
  - question
  - answer
  - type: *word* or *math*
  - category: *orig* or *back*
  - topic: *numbertheory* or *cipher*
- Manual Annotation (using prompt engineering in gpt-3.5)
- Initial target: 1k samples



- Backward vs. Forward reasoning answers [17]
  - Deduce missing values in question
  - Only numerical masks - no phrase mask
- Input Space Augmentation



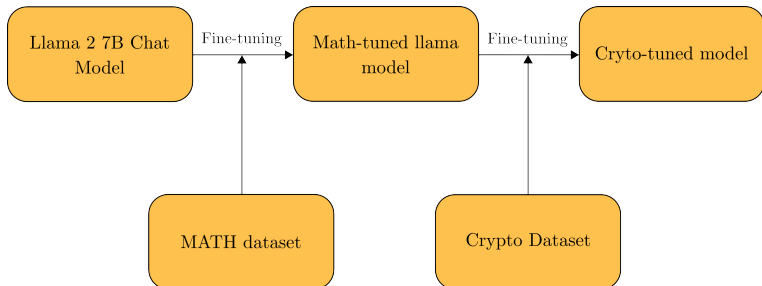
```
<s>[INST] <<SYS>>  
{{ system_prompt }}  
<</SYS>>
```

```
{{ user_message }} [/INST]
```

- Ensures high-quality dataset
- Other styles include ChatML and Alpaca style prompt formatting.



# Fine Tuning Flowchart





Introduction

Methodology

**Progress and Results**

Future Work



- 300/1000 samples prepared till now
  - Sample data(csv file):  
question, answer, type, category, topic, source
- 

"Use the Euclidean Algorithm to compute the GCD of the following numbers: \$300\$ and \$222\$", "To compute the GCD of \$300\$ and \$222\$ using the Euclidean Algorithm, we perform a series of divisions:  $300 = 222 \cdot 1 + 78$ ,  $222 = 78 \cdot 2 + 66$ ,  $78 = 66 \cdot 1 + 12$ ,  $66 = 12 \cdot 5 + 6$ ,  $12 = 6 \cdot 2 + 0$ . The last non-zero remainder is \$6\$. Therefore, the GCD of \$300\$ and \$222\$ is \$6\$.", orig, math, numbertheory, grad.pdf



- Parameter efficient fine tuning (PEFT) [18; 19].
- 12GB GPU RAM for 7B parameter model
- *paged-adamw-32bit* optimizer
- Supervised Fine Tuning (SFT)
- 7.5k samples of basic (high school level) math problems



- Fine Tuning
  - Massive Compute Requirements:
    - QLoRA [20]: 1x A100 (80GB GPU) or 2x RTX 3090
    - Full fine-tuning: 16x A100s
- Dataset Preparation
  - Manual inspection for verifying correctness of computations
  - Scarcity of documented solutions (to relevant numerical problems)



Introduction

Methodology

Progress and Results

**Future Work**



- Complete crypto-dataset
- Crypto fine-tuned model
- Analysis on language chaining vs. fine-tuned model
- Analysis on prompt engineering vs. fine-tuning model output



# Timeline

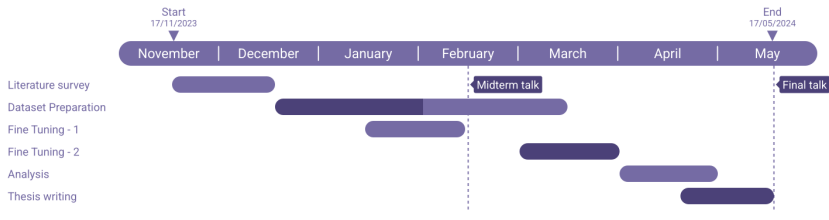


Figure: Timeline of the project



- [1] Zhen Yang, Ming Ding, Qingsong Lv, Zhihuan Jiang, Zehai He, Yuyi Guo, Jinfeng Bai, and Jie Tang. Gpt can solve mathematical problems without a calculator, 2023.
- [2] Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, Christopher Hesse, and John Schulman. Training verifiers to solve math word problems, 2021.
- [3] D. Noever. Large language models for ciphers. *International Journal of Artificial Intelligence Applications*, 14:1–20, 05 2023.
- [4] Wesley Tann, Yuancheng Liu, Jun Heng Sim, Choon Meng Seah, and Ee-Chien Chang. Using large language models for cybersecurity capture-the-flag challenges and certification questions, 2023.
- [5] Vasily Mikhalev, Nils Kopal, and Bernhard Esslinger. Evaluating gpt-4’s proficiency in addressing cryptography examinations. Cryptology ePrint Archive, Paper 2023/1551, 2023. <https://eprint.iacr.org/2023/1551>.





- [6] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. Llama 2: Open foundation and fine-tuned chat models, 2023.



- [7] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners, 2020.
- [8] Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, Parker Schuh, Kensen Shi, Sasha Tsvyashchenko, Joshua Maynez, Abhishek Rao, Parker Barnes, Yi Tay, Noam Shazeer, Vinodkumar Prabhakaran, Emily Reif, Nan Du, Ben Hutchinson, Reiner Pope, James Bradbury, Jacob Austin, Michael Isard, Guy Gur-Ari, Pengcheng Yin, Toju Duke, Anselm Levskaya, Sanjay Ghemawat, Sunipa Dev, Henryk Michalewski, Xavier Garcia, Vedant Misra, Kevin Robinson, Liam Fedus, Denny Zhou, Daphne Ippolito, David Luan, Hyeontaek Lim, Barret Zoph, Alexander Spiridonov, Ryan Sepassi, David Dohan, Shivani Agrawal, Mark Omernick, Andrew M. Dai, Thanumalayan Sankaranarayanan Pillai, Marie



- Pellat, Aitor Lewkowycz, Erica Moreira, Rewon Child, Oleksandr Polozov, Katherine Lee, Zongwei Zhou, Xuezhi Wang, Brennan Saeta, Mark Diaz, Orhan Firat, Michele Catasta, Jason Wei, Kathy Meier-Hellstern, Douglas Eck, Jeff Dean, Slav Petrov, and Noah Fiedel. Palm: Scaling language modeling with pathways, 2022.
- [9] Zhangir Azerbayev, Hailey Schoelkopf, Keiran Paster, Marco Dos Santos, Stephen Marcus McAleer, Albert Q. Jiang, Jia Deng, Stella Biderman, and Sean Welleck. Llemma: An open language model for mathematics. In *The Twelfth International Conference on Learning Representations*, 2024.
- [10] Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, L  lio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timoth  e Lacroix, and William El Sayed. Mistral 7b, 2023.



- [11] Baptiste Rozière, Jonas Gehring, Fabian Gloeckle, Sten Sootla, Itai Gat, Xiaoqing Ellen Tan, Yossi Adi, Jingyu Liu, Romain Sauvestre, Tal Remez, Jérémy Rapin, Artyom Kozhevnikov, Ivan Evtimov, Joanna Bitton, Manish Bhatt, Cristian Canton Ferrer, Aaron Grattafiori, Wenhan Xiong, Alexandre Défossez, Jade Copet, Faisal Azhar, Hugo Touvron, Louis Martin, Nicolas Usunier, Thomas Scialom, and Gabriel Synnaeve. Code llama: Open foundation models for code, 2024.
- [12] William Stallings. *Cryptography and Network Security: Principles and Practice*. Prentice Hall Press, USA, 6th edition, 2013.
- [13] Neal Koblitz. *A course in number theory and cryptography*. Springer-Verlag, Berlin, Heidelberg, 1987.
- [14] Luis Finotti. A gentle introduction to number theory and cryptography. <https://web.math.utk.edu/~finotti/papers/grad.pdf>, 2009. [Accessed 14-02-2024].
- [15] Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. Measuring mathematical problem solving with the math dataset, 2021.



- [16] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models, 2023.
- [17] Aniruddha Deb, Neeva Hareshbhai Oza, Sarthak Singla, Dinesh Khandelwal, Dinesh Garg, and Parag Singla. Fill in the blank: Exploring and enhancing LLM capabilities for backward reasoning in math word problems, 2024.
- [18] Lingling Xu, Haoran Xie, Si-Zhao Joe Qin, Xiaohui Tao, and Fu Lee Wang. Parameter-efficient fine-tuning methods for pretrained language models: A critical review and assessment, 2023.
- [19] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models, 2021.
- [20] Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. Qlora: Efficient finetuning of quantized llms, 2023.



Thank you!

