We acquired data from the Cybersecurity and Infrastructure Security Agency’s repository of cyber threat reports. We used PostgreSQL format with Docker for secure data storage and easier processing.

By training LLMs with CISA Vulnerabilities, we can simulate and summarize real adversarial scenarios.

Leveraging LLMs to create reports using CISA Vulnerabilities allows reports to be released with an incredibly efficient, end-in-depth analysis.

Evaluation Criteria
When evaluating our model, we decided upon the following evaluation metrics:
• Accuracy of data provided
• Relevance prompted information
• Coherence and clarity of report
• Completeness of details
These metrics are important as they ensure the model generates credible, relevant information.

Conclusion
We were successfully able to train our LLM through fine-tuning and RAG with our data, however, limited computer power stunted our results. Our LLM would give us an output, but we were never satisfied with the result. With more resources and time we would continue our process of training, testing, and evaluating until we had a worthy product.

Future Goals
• Implement real-time threat reporting and fully automate the threat analysis process
• Expand product to include in-depth user interface
• Chose BLOOM for its smaller LLM model with 560 million parameters
• Despite its size, localizing BLOOM for training and testing on our personal devices was challenging and time-consuming
• Tested other models to compare and contrast how different models react to our input
• In a professional setting with better computational resources, opting for a larger model would expedite product development

References