

# Benchmarking GPU Performance in Vulnerability Severity Model Training

CIRCL Team

2025-11-26

## Contents

Preface	1
GPU Architectures Used for Benchmarking	2
Dataset	2
Model	2
Training Hyperparameters . . . . .	2
Framework Versions . . . . .	3
Benchmark Comparisons	3
Duration . . . . .	3
Energy . . . . .	4
Emissions . . . . .	6
GPU Power . . . . .	7
Energy vs. Duration . . . . .	8
GPU Power vs. Duration . . . . .	9
GPU Power vs. Energy . . . . .	10
Feedback	10
Funding	10

---

## Preface

This document summarizes the benchmarking, training configuration, and performance results obtained while generating the **Vulnerability Severity Classification** model across different GPU architectures.

The **VLAI Vulnerability Severity Classification** model developed at CIRCL is regularly updated and shared on Hugging Face. It has been presented in:

Bonhomme, C., & Dulaunoy, A. (2025). *VLAI: A RoBERTa-Based Model for Automated Vulnerability Severity Classification* (Version 1.4.0) [Computer software].  
<https://doi.org/10.48550/arXiv.2507.03607><sup>1</sup>

---

<sup>1</sup><https://arxiv.org/abs/2507.03607>

# GPU Architectures Used for Benchmarking

The benchmarks in the following sections were performed on the GPU architectures listed below.

System	CPU Cores	GPU(s)	RAM
A	64 (AMD EPYC 9124 16-Core Processor)	2 × NVIDIA L40S	251.5 GB
B	224 (Intel Xeon Platinum 8480+)	2 × NVIDIA H100 NVL	2,014 GB
C	224 (Intel Xeon Platinum 8480+)	4 × NVIDIA L40S	2,014 GB

Before presenting the benchmark results, we first describe the dataset and the training parameters used.

## Dataset

The dataset used for training and evaluation is available on Hugging Face:

<https://huggingface.co/datasets/CIRCL/vulnerability-scores>

at the commit `cbb05f48e20e2186a80284de138cafe56b6544c`<sup>2</sup>.

This is the updated version of the dataset referenced in [arXiv.2507.03607](https://arxiv.org/abs/2507.03607).

Dataset statistics:

- Number of rows: 657,024
- Downloaded size: 162 MB
- Auto-converted Parquet size: 162 MB

This dataset is periodically updated with data collected with Vulnerability-Lookup.

## Model

The resulting model is available at on Hugging Face<sup>3</sup>. This model is a fine-tuned version of roberta-base on the dataset CIRCL/vulnerability-scores.

## Training Hyperparameters

The following hyperparameters were used during training:

- **Learning rate:** 3e-05
- **Train batch size:** 16
- **Eval batch size:** 16
- **Seed:** 42
- **Optimizer:** ADAMW\_TORCH\_FUSED
  - betas: (0.9, 0.999)
  - epsilon: 1e-08
  - optimizer\_args: *none*
- **Scheduler:** linear
- **Epochs:** 5

<sup>2</sup><https://huggingface.co/datasets/CIRCL/vulnerability-scores/tree/cbb05f48e20e2186a80284de138cafe56b6544c>

<sup>3</sup><https://huggingface.co/CIRCL/vulnerability-severity-classification-roberta-base>

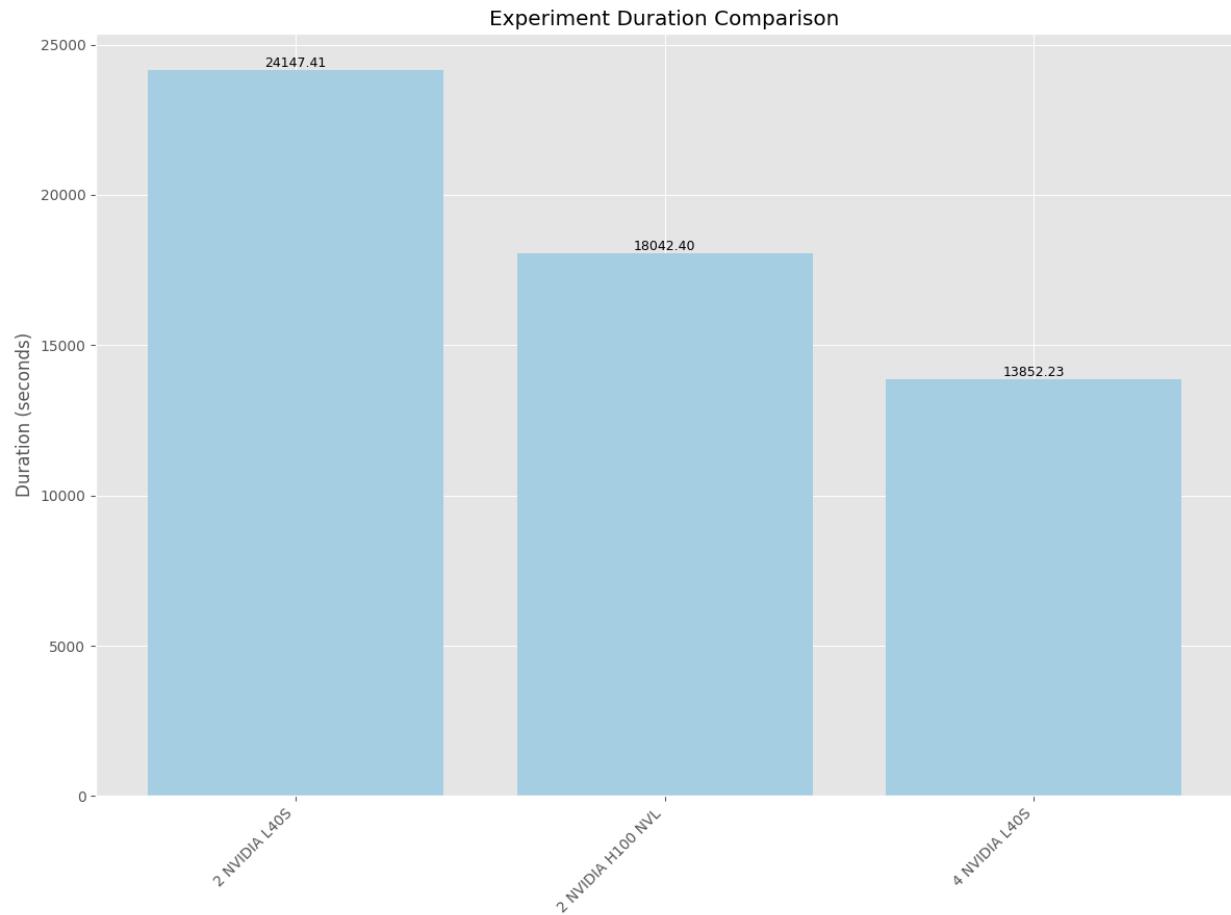
## Framework Versions

The environment used for training:

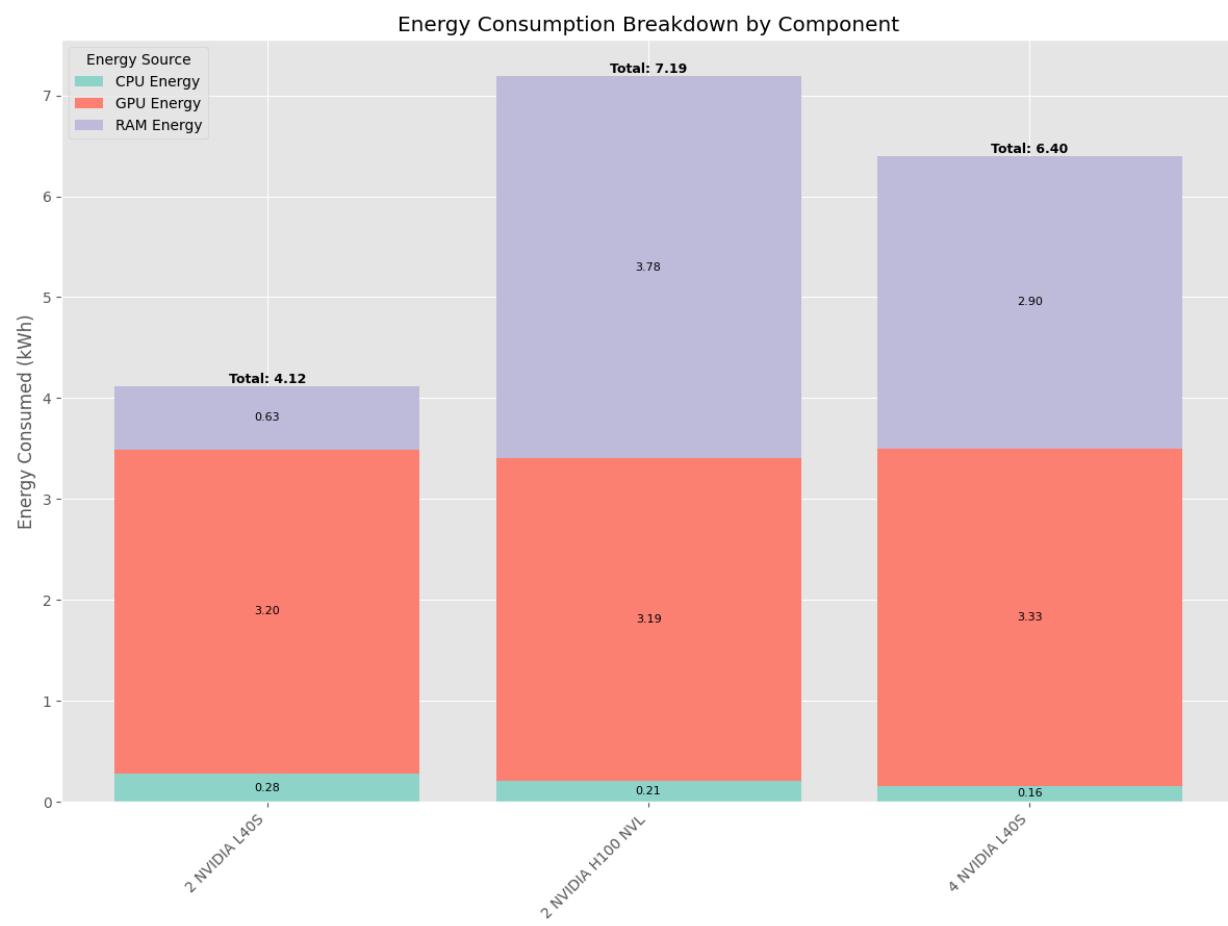
- **Python:** 3.12.3
- **Transformers:** 4.57.1
- **PyTorch:** 2.9.1+cu128
- **Datasets:** 4.4.1
- **Tokenizers:** 0.22.1

## Benchmark Comparisons

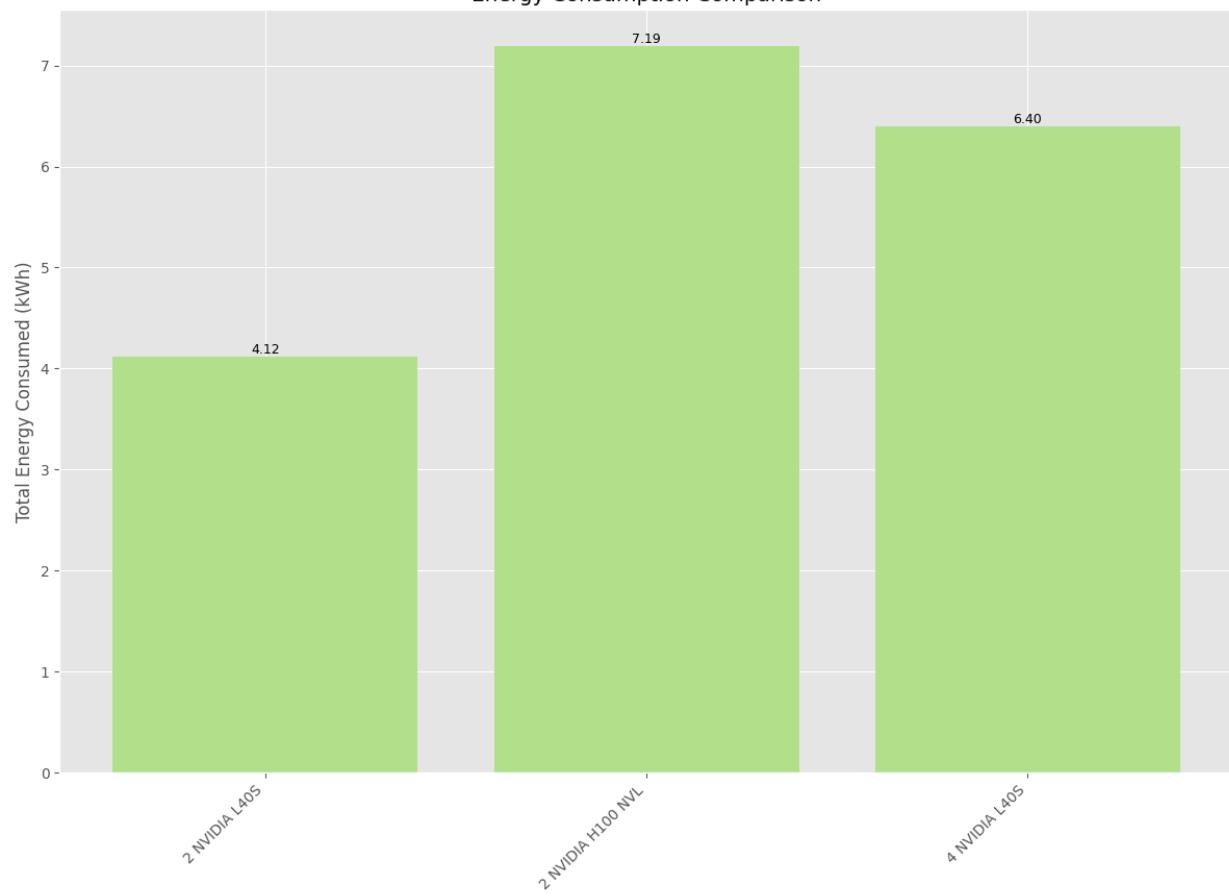
### Duration



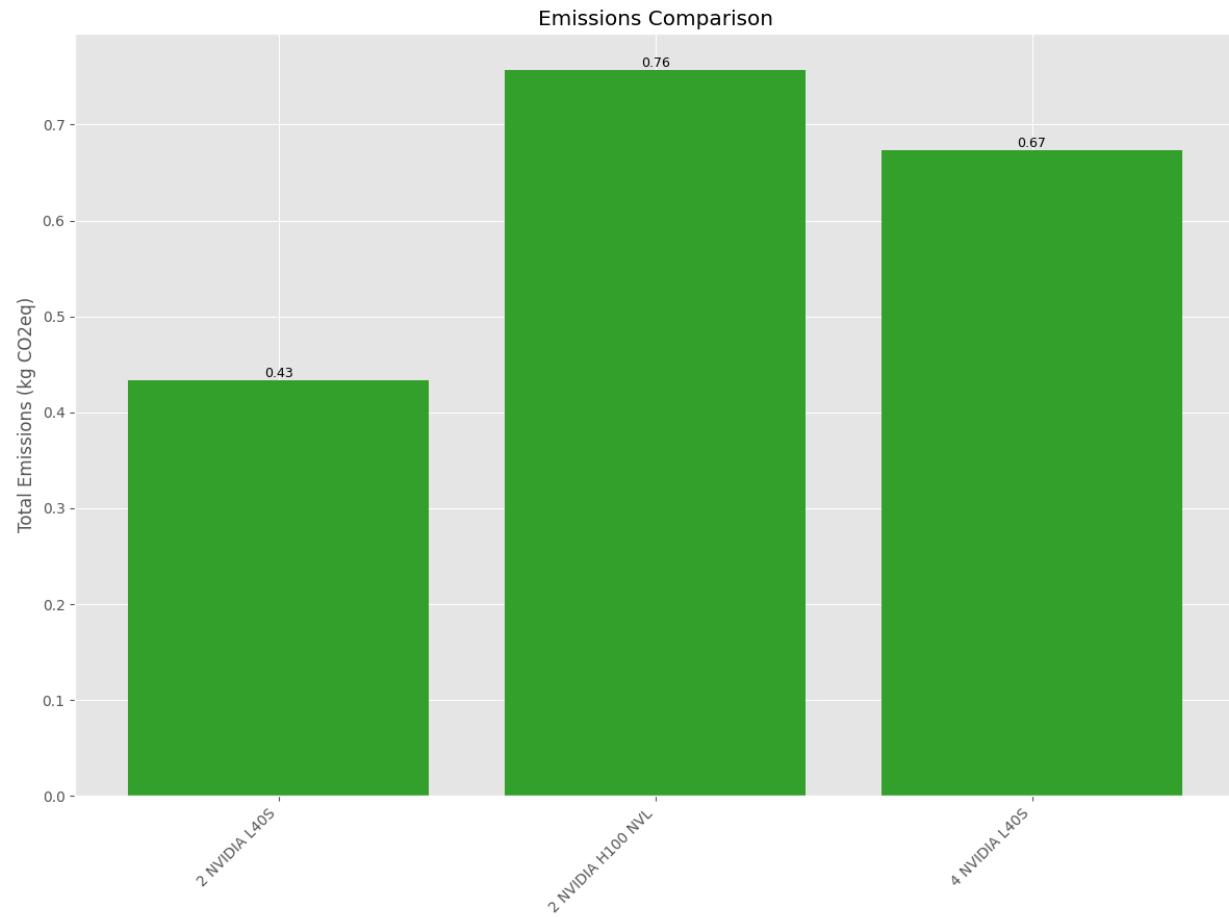
## Energy



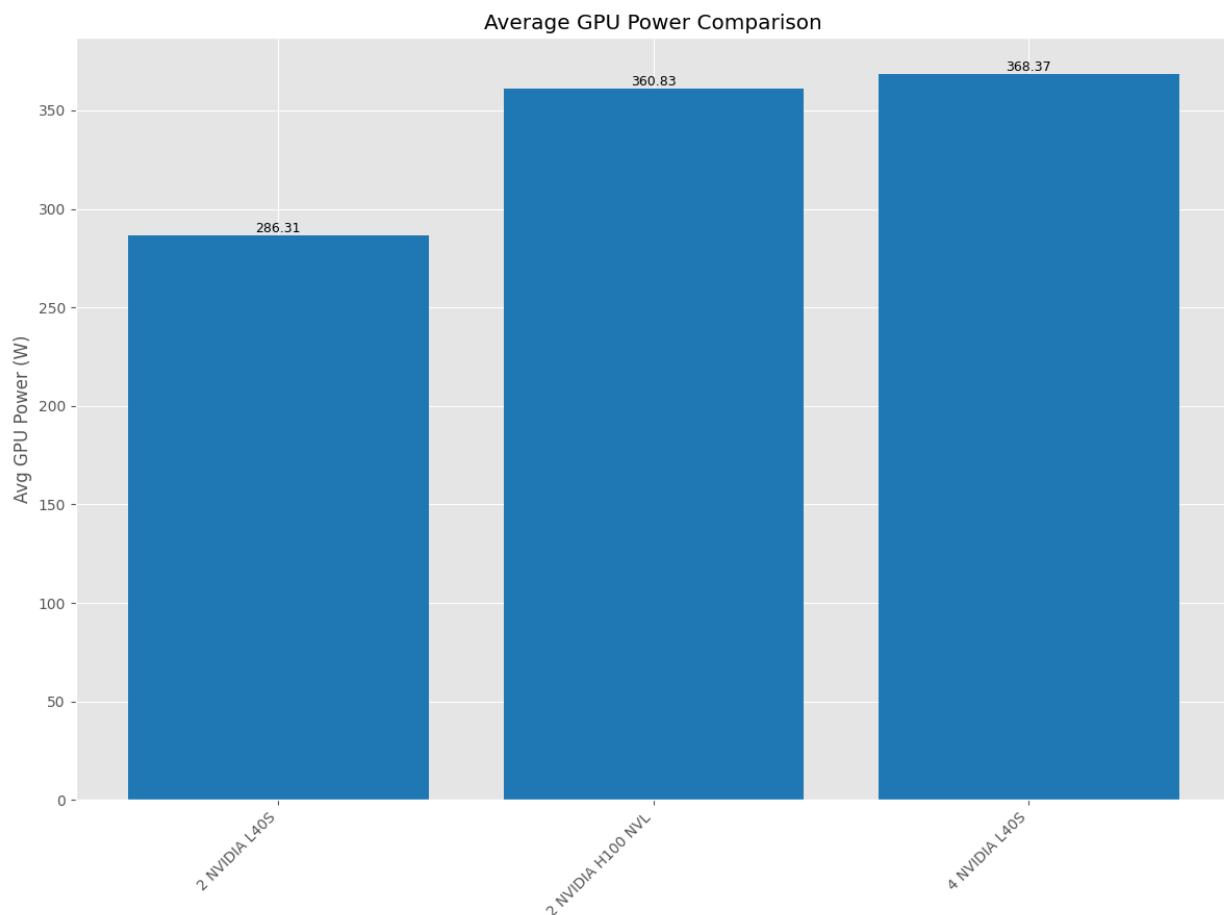
Energy Consumption Comparison



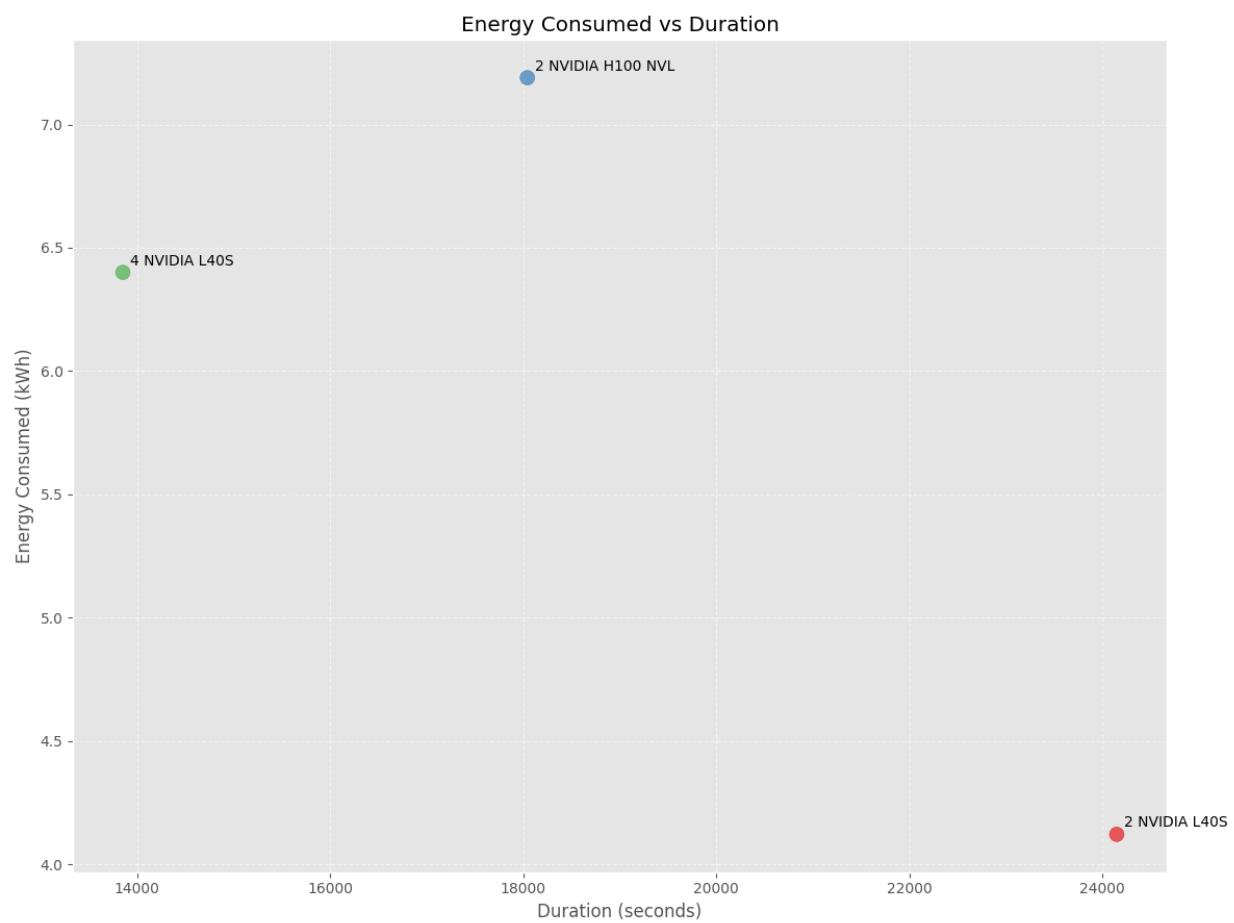
## Emissions



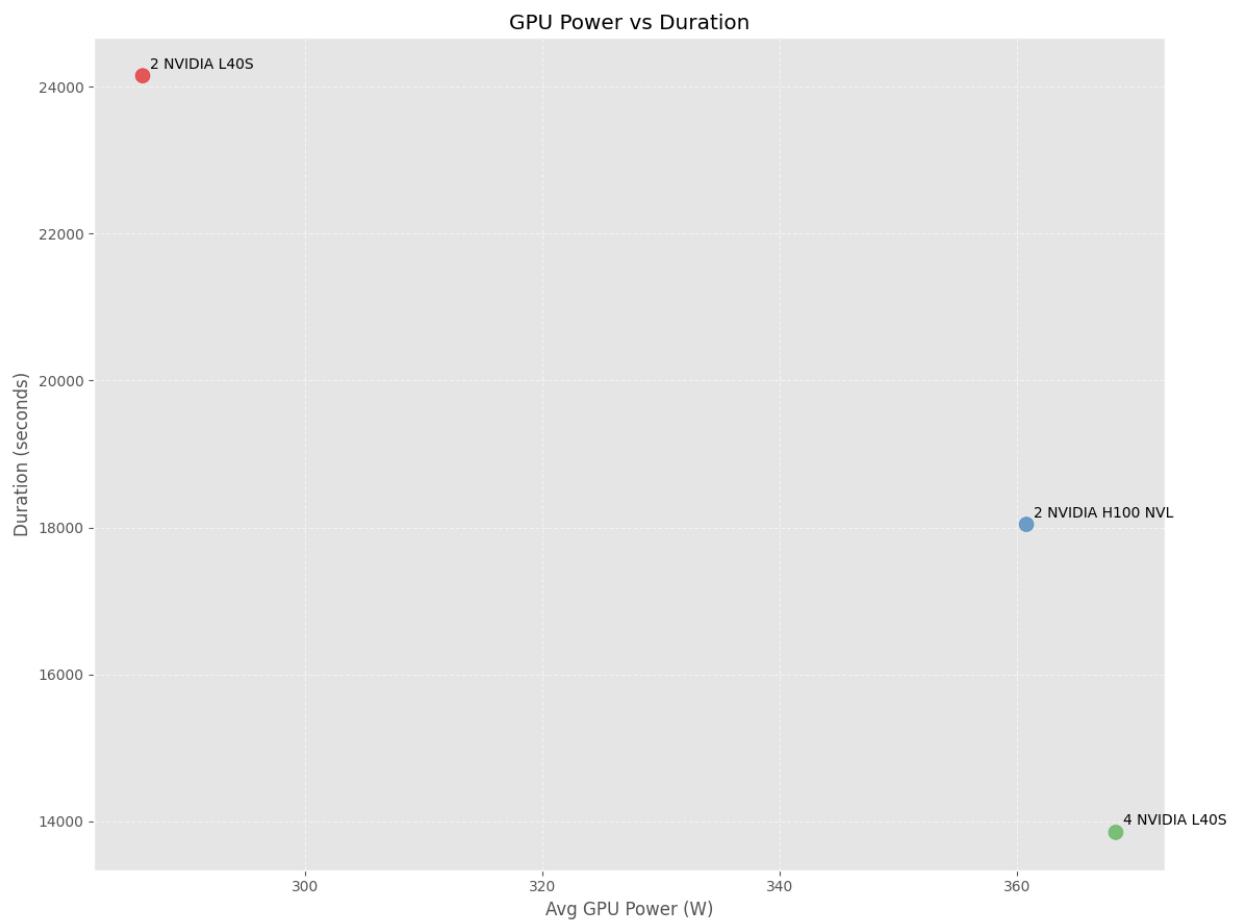
## GPU Power



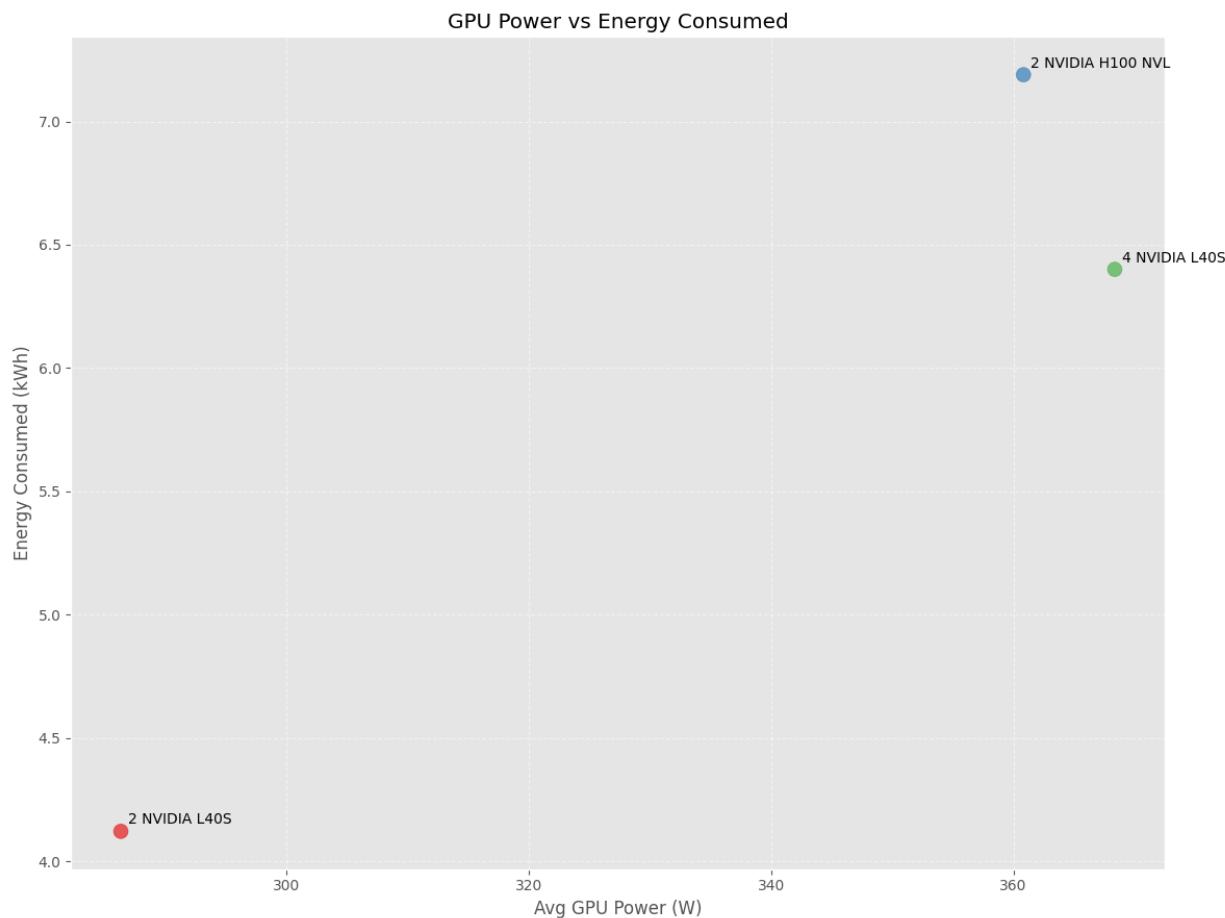
## Energy vs. Duration



## GPU Power vs. Duration



## GPU Power vs. Energy



## Feedback

Feel free to share your feedback at [info@circl.lu](mailto:info@circl.lu).

## Funding



**Co-funded by  
the European Union**

The project aims to create advanced artificial intelligence-based tools that will improve the operations of cybersecurity teams (e.g., sector-level or national-level CSIRT teams, or SOC teams in companies and institutions). These tools will enable faster detection, analysis, and neutralization of threats. We plan to develop intelligent early warning systems, an AI chatbot for analyzing incident reports, and datasets representing current cyberattacks. Work on these solutions has just begun, involving teams from Poland, Luxembourg, the Netherlands, and Italy.

AIPITCH aims to create a comprehensive set of tools supporting key operational services in cyber defense. These include technologies for early threat detection, automatic malware classification, and improvement of analytical processes through the integration of Large Language Models (LLM). The project has the potential to set new standards in the cybersecurity industry.

The project leader is NASK National Research Institute. The international consortium includes: - CIRCL (Computer Incident Response Center Luxembourg), Luxembourg - The Shadowserver Foundation, Netherlands - NCBJ (National Centre for Nuclear Research), Poland - ABI LAB (Centre of Research and Innovation for Banks), Italy

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Cybersecurity Competence Centre. Neither the European Union nor the European Cybersecurity Competence Centre can be held responsible for them.