Extras

For cpu stress:

sudo apt install stress-ng

stress-ng --cpu 4 --vm 2 --vm-bytes 1G --timeout 60s

**Citation Paragraph:**

Membership inference attacks (MIAs) were first systematically studied by Shokri et al. (2017), who demonstrated that machine learning models can unintentionally leak whether a specific data point was included in the training set, especially when models are overconfident on such inputs. Their work introduced the use of prediction confidence thresholds and true/false membership labeling to assess privacy leakage. In line with their methodology, this dissertation implements a confidence-based thresholding approach to simulate MIAs, comparing the model’s responses to member and non-member samples. We calculate a metric called *Membership Inference Advantage* (MIA), defined as the difference in prediction confidence between training and test sets, to quantify privacy leakage. Furthermore, this implementation incorporates differential privacy using the Opacus library, and demonstrates—both visually and quantitatively—that enabling DP significantly reduces MIA. This approach provides an interpretable, non-black-box validation of the system’s privacy guarantees, closely aligning with the original findings of Shokri et al. while remaining computationally efficient for edge use cases.

**Suggested Reference (APA Style):**

Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). *Membership Inference Attacks Against Machine Learning Models*. In IEEE Symposium on Security and Privacy. arXiv:1610.05820.