From NSIGHT, we can see that GPU matrix multiplication has a runtime of 24,953.884 μs whereas OPT matrix multiplication has a runtime of 10,750.562 μs.

So, Speed Up Between GPU and Optimized GPU: 2.2

A screenshot of a cell phone

Description automatically generated

A screenshot of a social media post

Description automatically generated