2025 Google Gemma3

AI and Machine Learning

Tony Shen
Data Communications Labs
7/12/2025

Contents

Introduction

On March 10, 2025, Google officially announced Gemma3. It marked a major upgrade in their open model lineup, introducing multimodal capabilities, a 128K-token context window, and support for over 140 languages. Two features stand out in Gemma3. First, Gemma 3 is a multi-modal LLM model capable of processing not only texts but also images. Second, Gemma3 comes in four small sizes, 1B, 4B, 12B, and 27B, suitable for running on edge devices, robots, and modest computer systems with limited resources locally.

Gemma3, however, is not completely "open source". Gemma3 GitHub repo includes .bin or .safetensors files for inference and fine-tuning. It also includes setting up instructions and deployment scripts. It does not include the architecture source code, training data and training scripts. Gemma3, therefore, is not truly Open Source. Rather, its source model is "Open Weights", meaning Gemma3 is free and open for anyone to download and set it up anywhere to use but not telling you how it is made.

In this short article, we show you how to conduct a couple of simple performance tests with Gemma3 on a Windows PC locally. In the tests, we compare Gemma3 with DeepSeek, which remains completely Open Source. DeepSeek R1 was released in January 2025. Since then, DeepSeek has not made any significant changes, nor has it released a new version of the model.

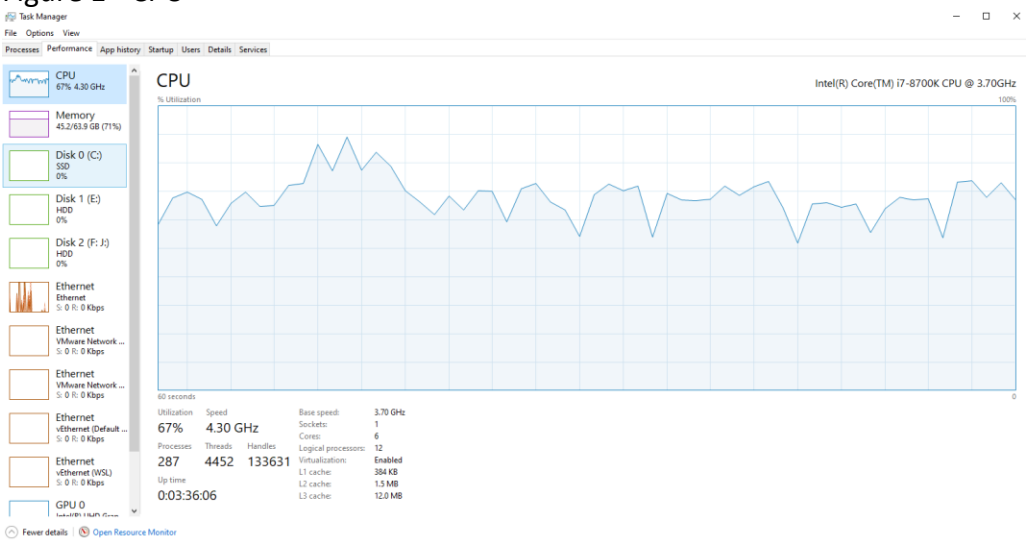
Test Environment

A Windows 10 PC provides the test environment. The PC is equipped with 64GB of memory, an RTX-4070 (a lower-end Nvidia GPU card), and sufficient disk space to accommodate a few selected Opensource LLM models. The test run was conducted locally on the PC. The models were accessed via Ollama, and Open Web UI provided the user interface.

Hardware Components

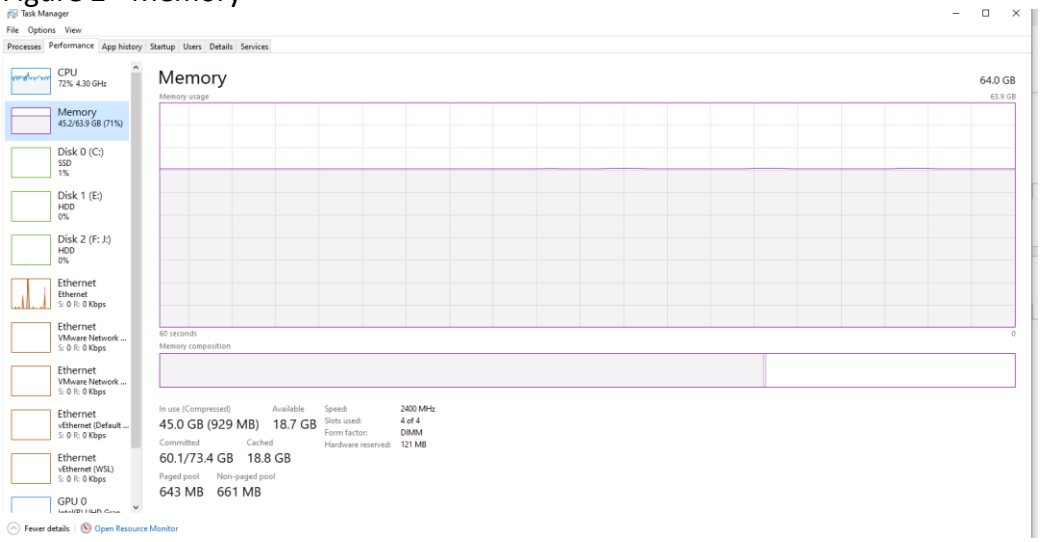
CPU – Intel Core i7-8700K CPU @ 3.70Ghz

Figure 1 - CPU



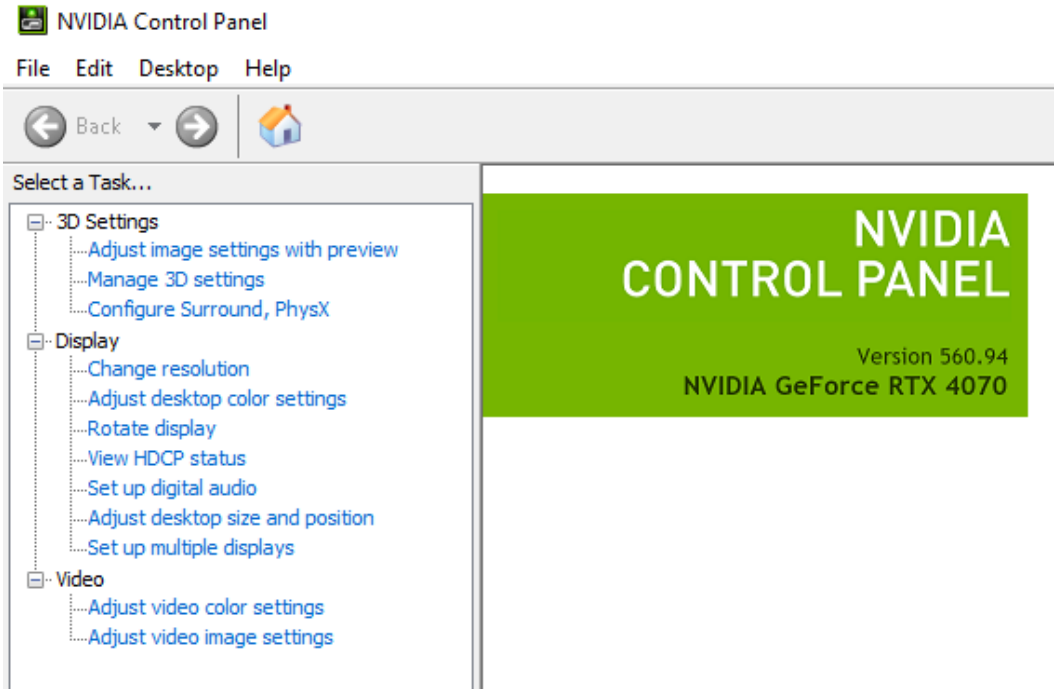
Memory – 64GB

Figure 2 - Memory



GPU – Nvidia GeForce RTX 4070

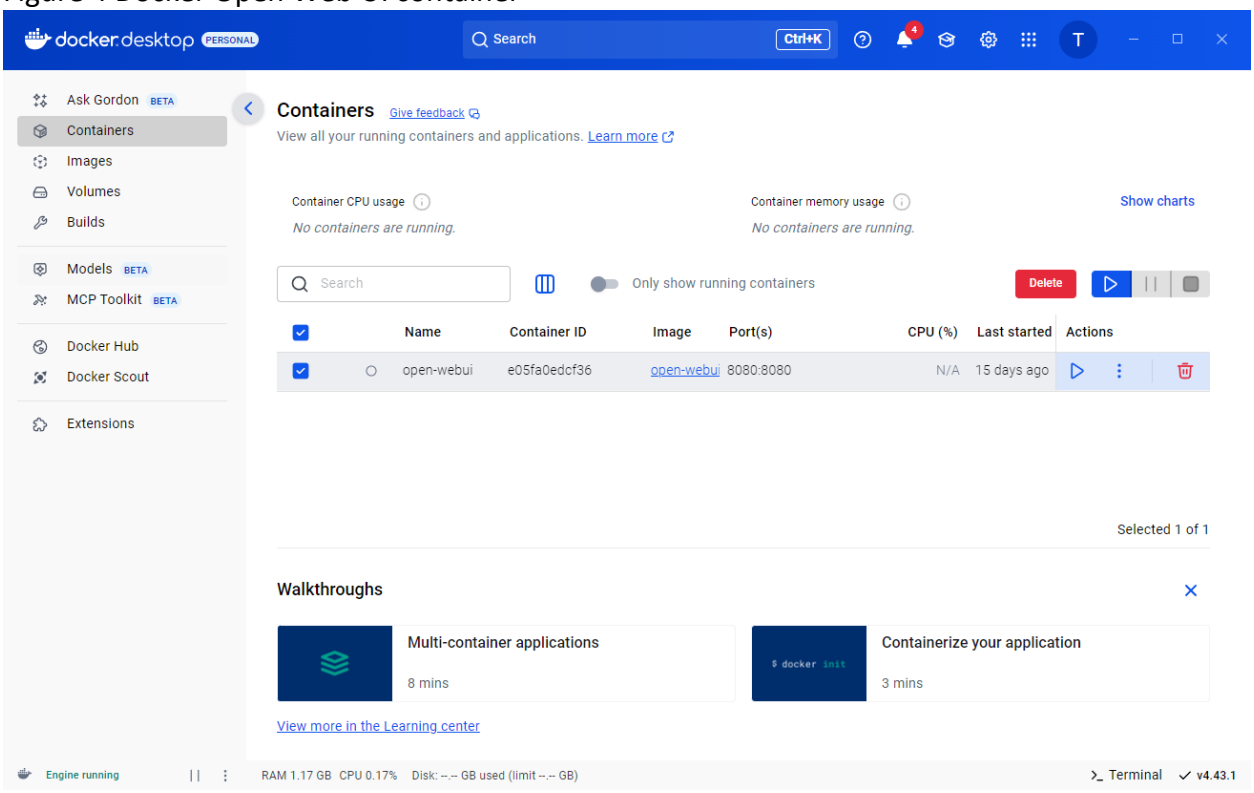
Figure 3 - GPU



Software Components

1. Ollama for Windows
2. Docker Desktop for Windows
3. Open Web UI that runs as a Docker container web server (See Figure 4 below)

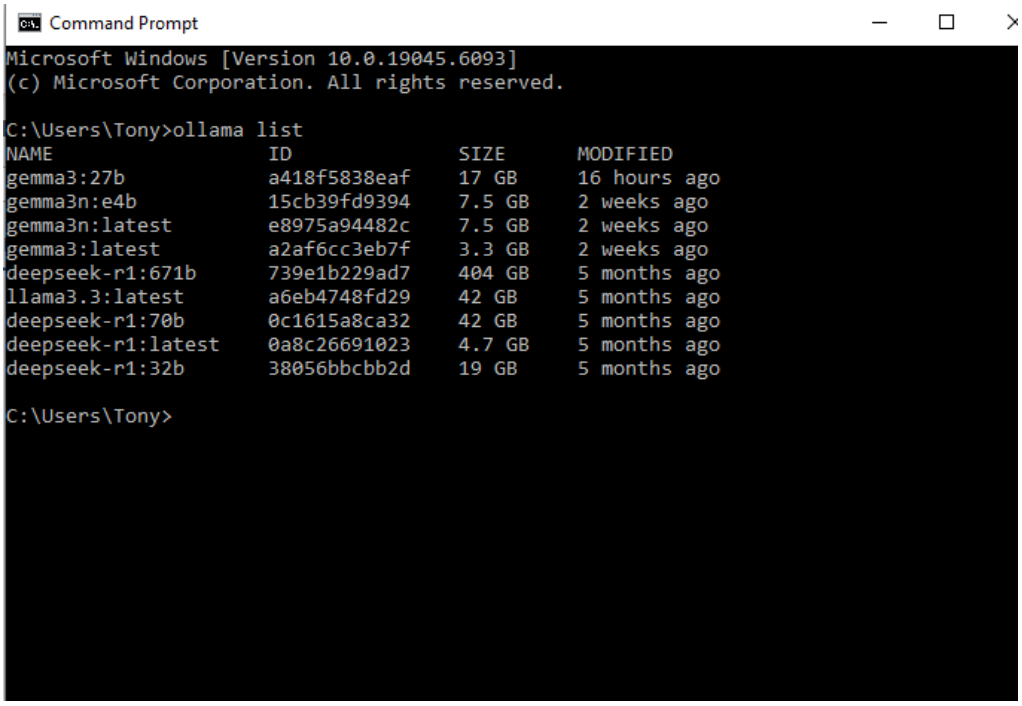
Figure 4 Docker Open Web UI container



Selected Models

Gemma3:27b
Deepseek-r1:32b

Figure 5 – Available models



Gemma3

Test 1 – Text

Figure 6 – Started at **7:09**

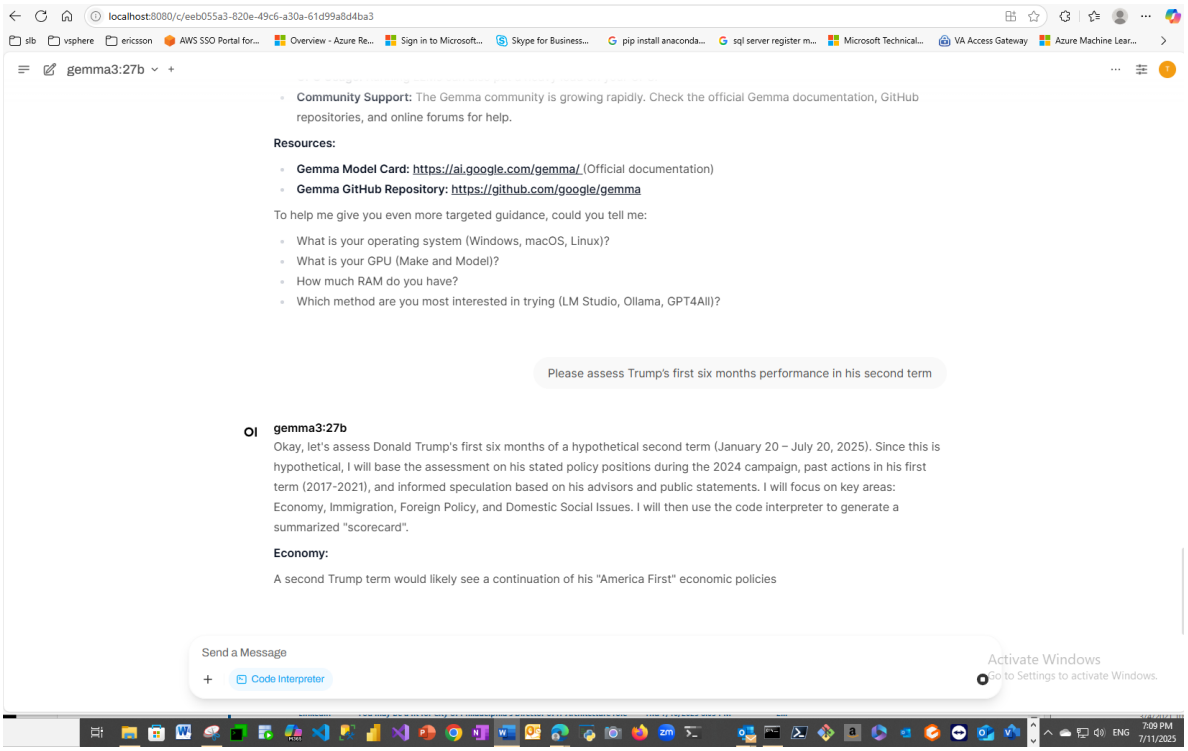


Figure 7 - Continued

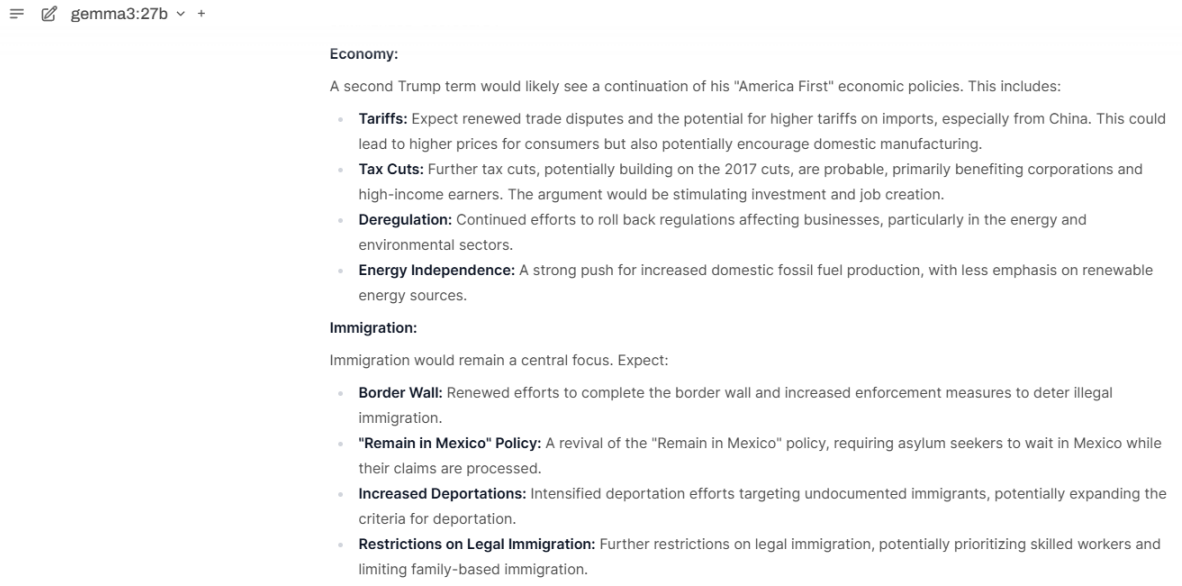


Figure 8 – Continued

gemma3:27b +

Foreign Policy:

A second Trump term would likely be characterized by a more isolationist and transactional foreign policy:

- **NATO:** Continued pressure on NATO allies to increase defense spending and potentially questioning the alliance's future.
- **China:** A continuation of the trade war and increased competition with China, potentially escalating tensions in the South China Sea.
- **Iran:** A more hawkish stance towards Iran, potentially withdrawing from the nuclear deal and increasing sanctions.
- **Ukraine:** A more ambivalent approach to the war in Ukraine, potentially seeking a negotiated settlement that favors Russia.
- **International Agreements:** A continued skepticism of international agreements and organizations, potentially withdrawing from more treaties.

Domestic Social Issues:

Expect a continuation of the culture war battles and efforts to appeal to his base:

- **Supreme Court Appointments:** Efforts to appoint conservative judges to federal courts, potentially solidifying a conservative majority on the Supreme Court for decades.
- **"Woke" Policies:** Attacks on "woke" policies and critical race theory, potentially restricting discussions of race and gender in schools.
- **Gun Control:** Opposition to gun control measures and support for Second Amendment rights.
- **Abortion:** Continued efforts to restrict abortion access and appoint judges who oppose abortion rights.

Figure 9 – Continued

gemma3:27b +

Code Interpreter – Scorecard:

Now, let's use the code interpreter to generate a scorecard summarizing the likely impact of these policies across various sectors. I'll assign scores from -3 (highly negative) to +3 (highly positive), with 0 representing neutral impact.

python Collapse Run Save Copy

```python
import pandas as pd

data = {
    'Sector': ['Economy', 'Immigration', 'Foreign Policy', 'Social Issues', 'Environmental Regulations', 'Internatio
    'Score': [1, -2, -1, -1, -3, -2, 0, -1],
    'Justification': [
        'Tax cuts & deregulation could stimulate growth, but tariffs could harm trade.',
        'Stricter policies could reduce illegal immigration, but harm economy & human rights.',
        'Transactional approach could isolate US & weaken alliances.',
        'Policies likely to exacerbate social divisions & limit rights.',
        'Deregulation would harm environmental protection efforts.',
        'Likely to damage relationships with key allies.',
        'Potential for changes to the Affordable Care Act.',
        'Focus on limiting certain topics & promoting conservative values.'
    ]
}

df = pd.DataFrame(data)
print(df)
```

Figure 10 – Continued

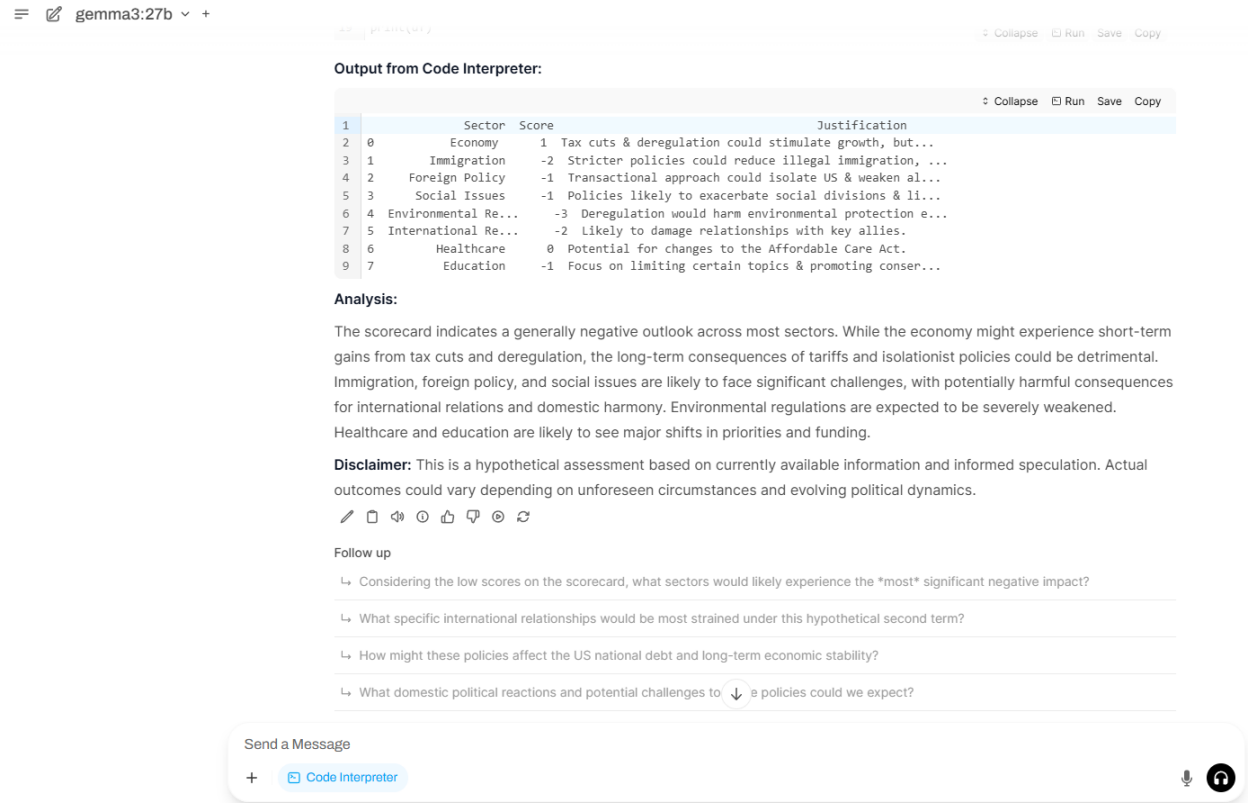


Figure 11 – Completed at **7:17**

Figure 12 – Computing Power Consumption – CPU

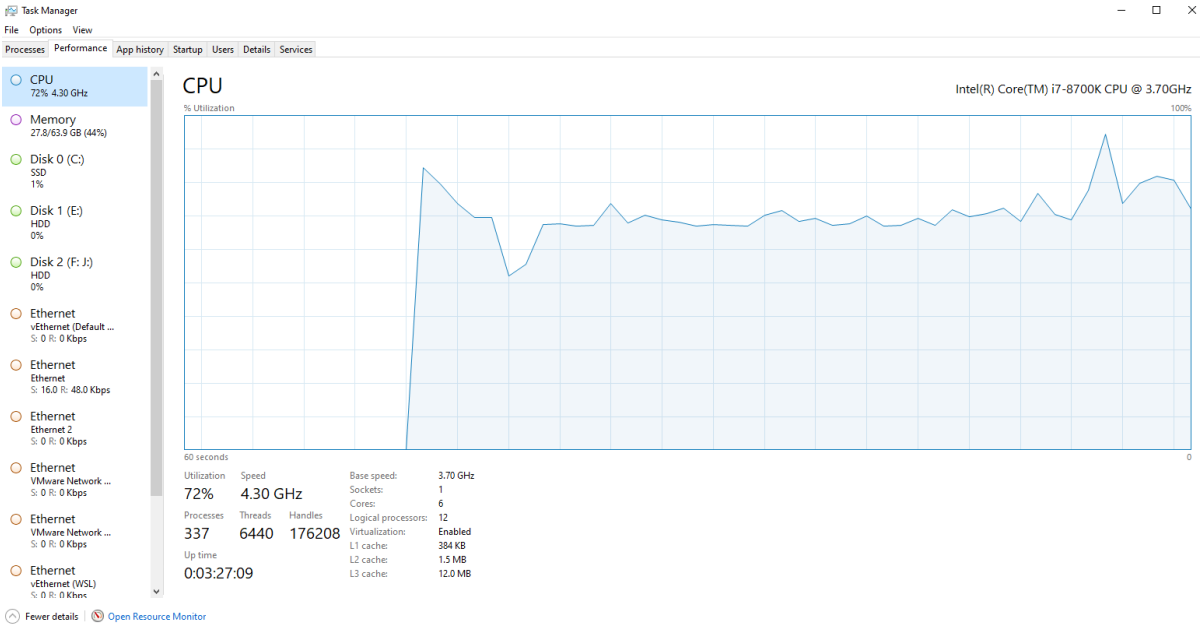


Figure 13 – Computing Power Consumption – GPU0 – Display Card

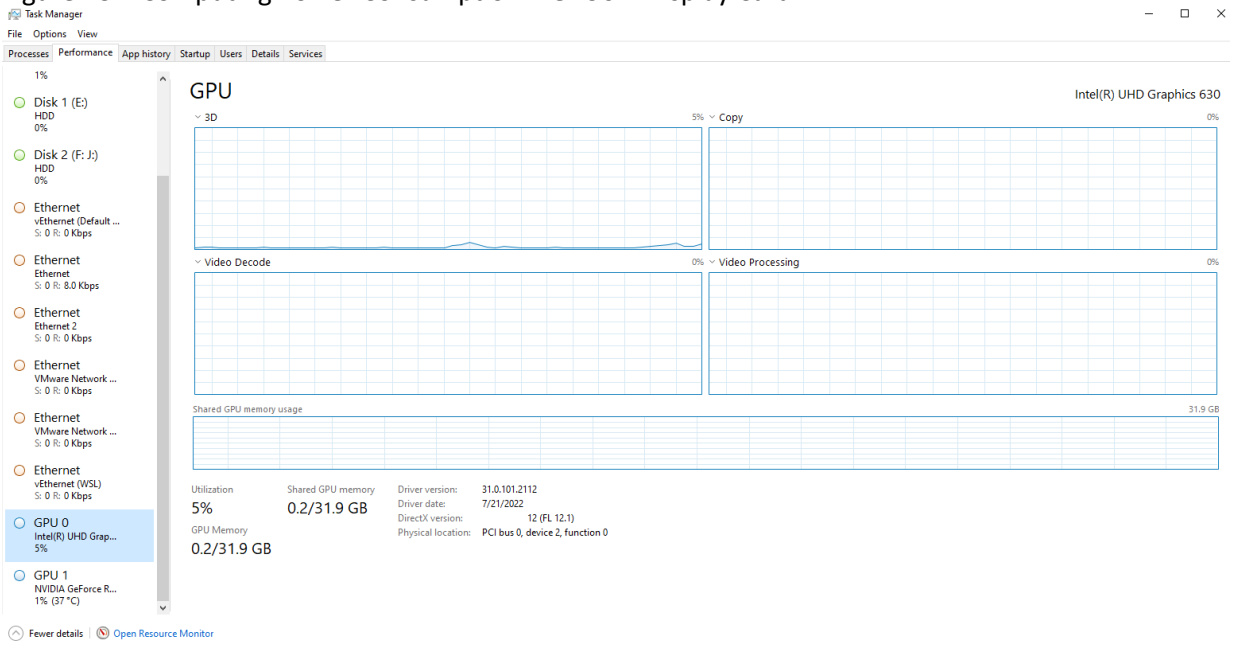
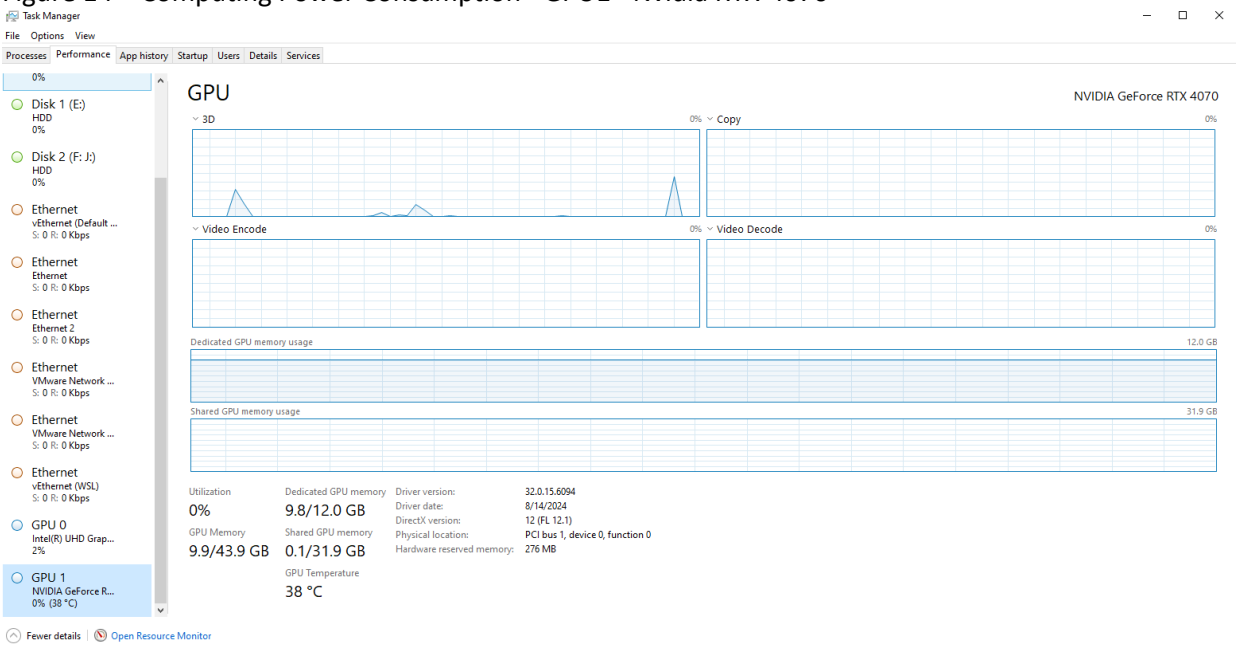


Figure 14 – Computing Power Consumption - GPU1 - Nvidia RTX-4070



Test 2 – Image

Figure 15 – The image question started at **11:02**

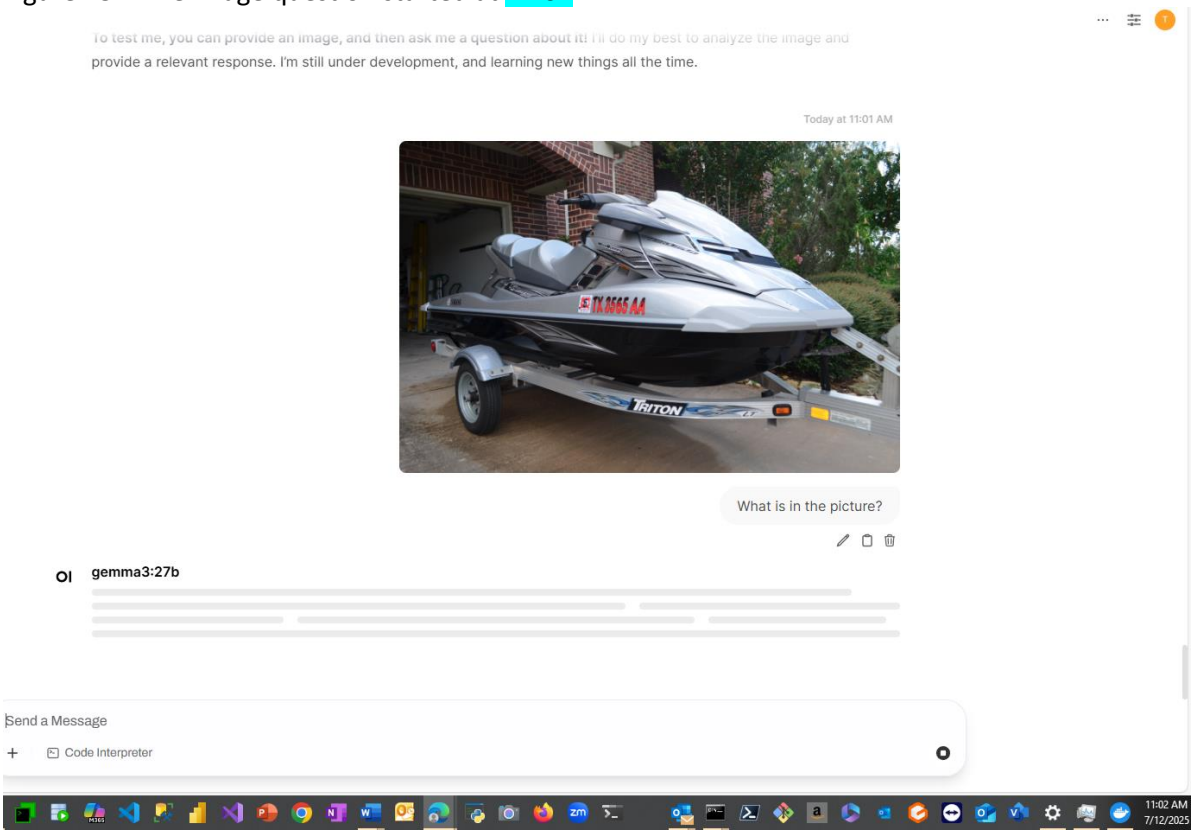
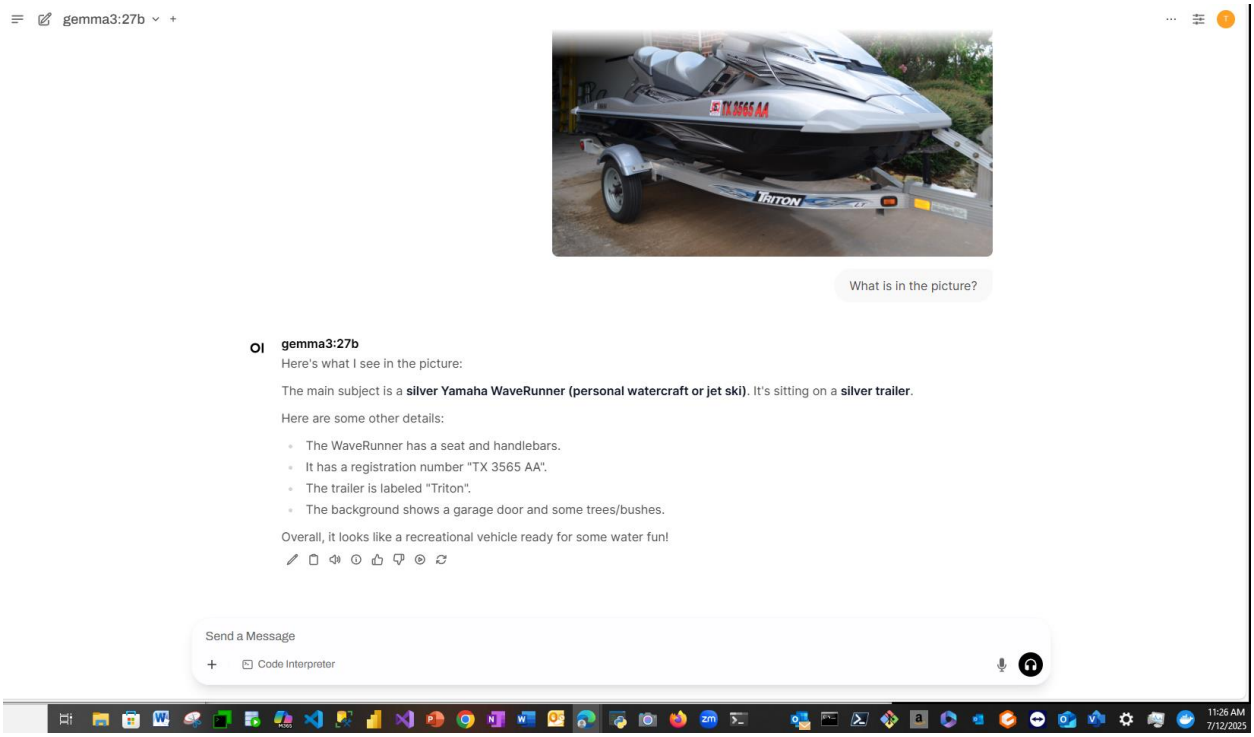


Figure 16 – The image question response completed at **11:26**



For comparison, let's switch to DeepSeek R1:32B and run the same test. DeepSeek R1:32B is the one most comparable to Gemma3:27B in size.

DeepSeek R1

Test 1 – Text

Figure 17 – Started at **7:24**

Figure 18 – Continued

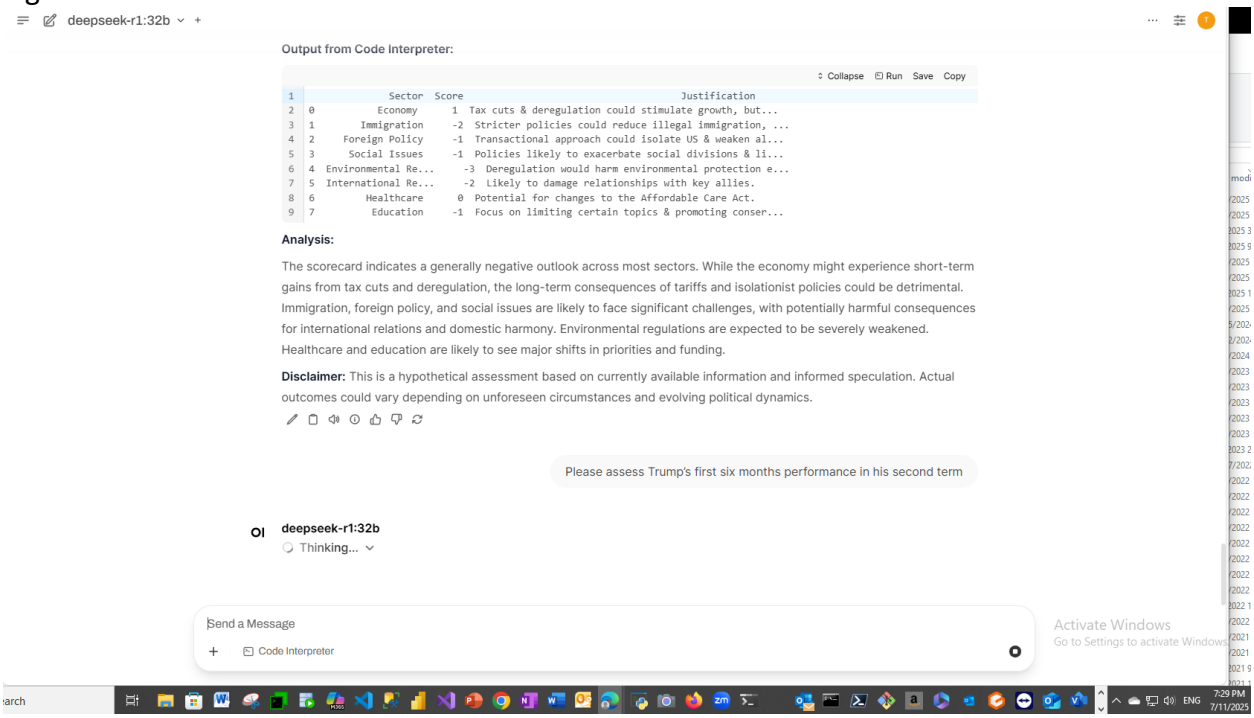


Figure 19 – Continued

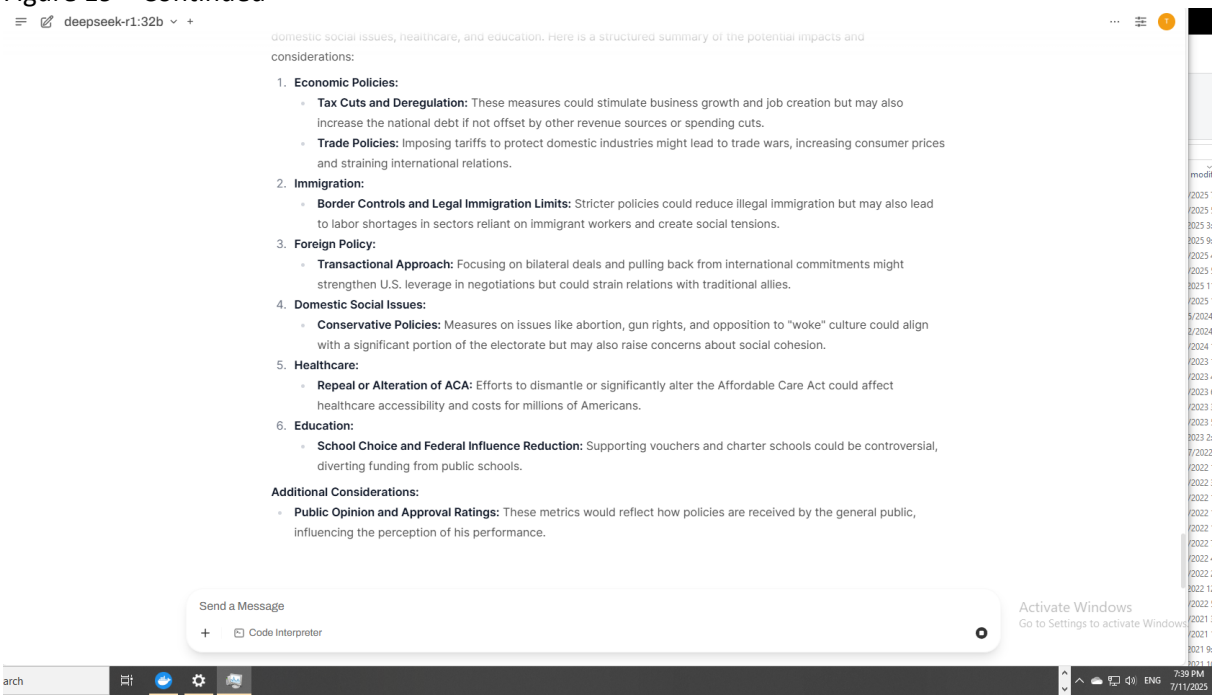


Figure 20 – Continued

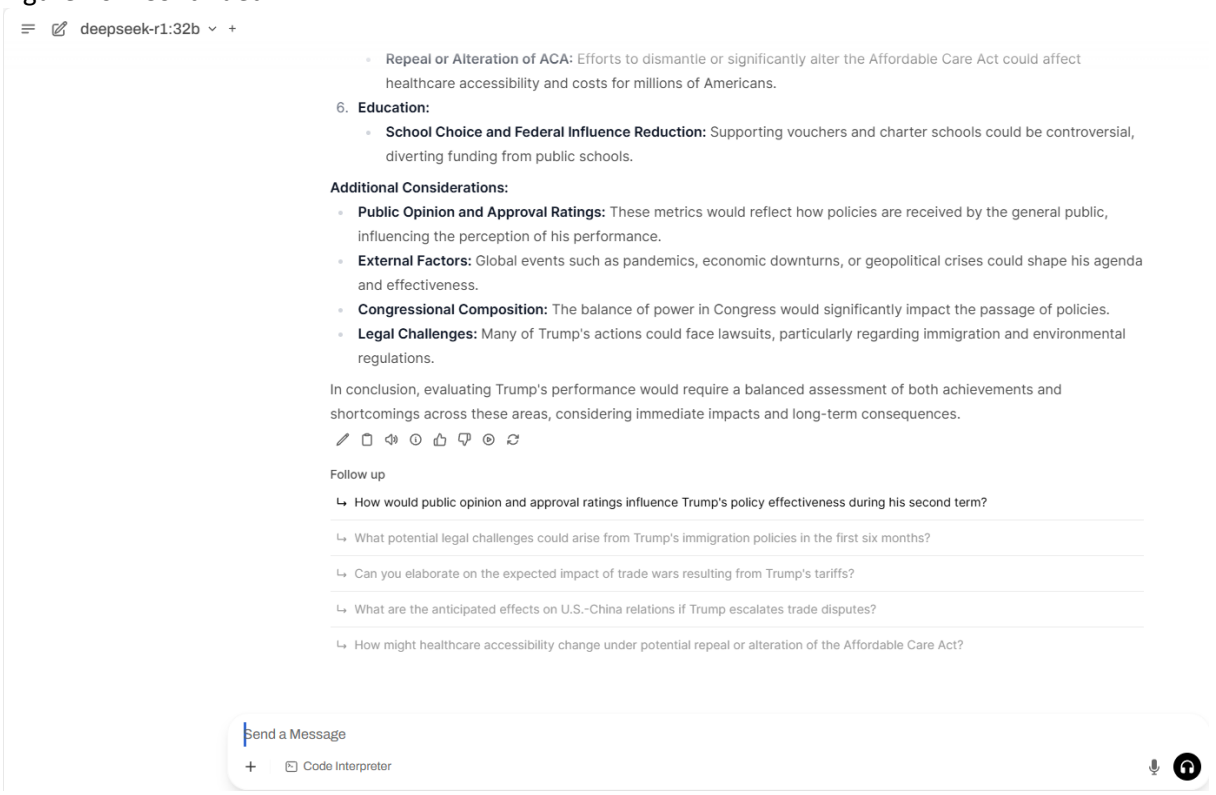


Figure 21 – **Thought for 9 minutes**

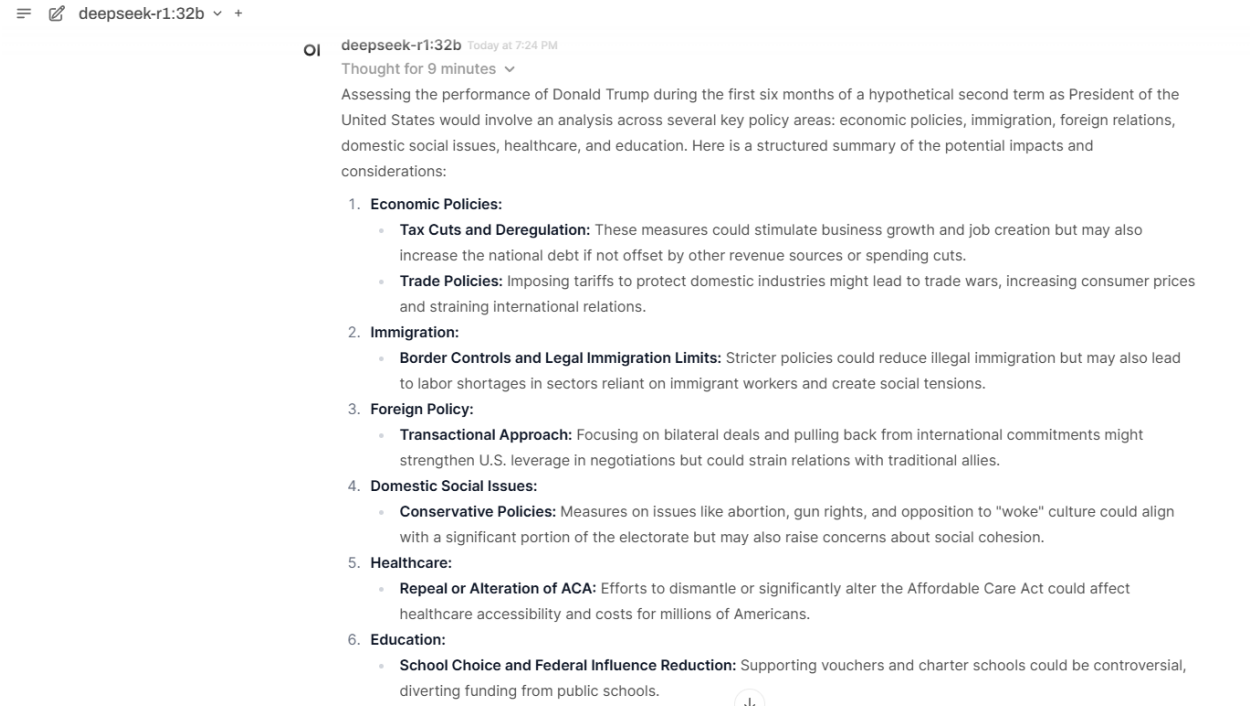


Figure 22 – Computing Power Consumption – CPU

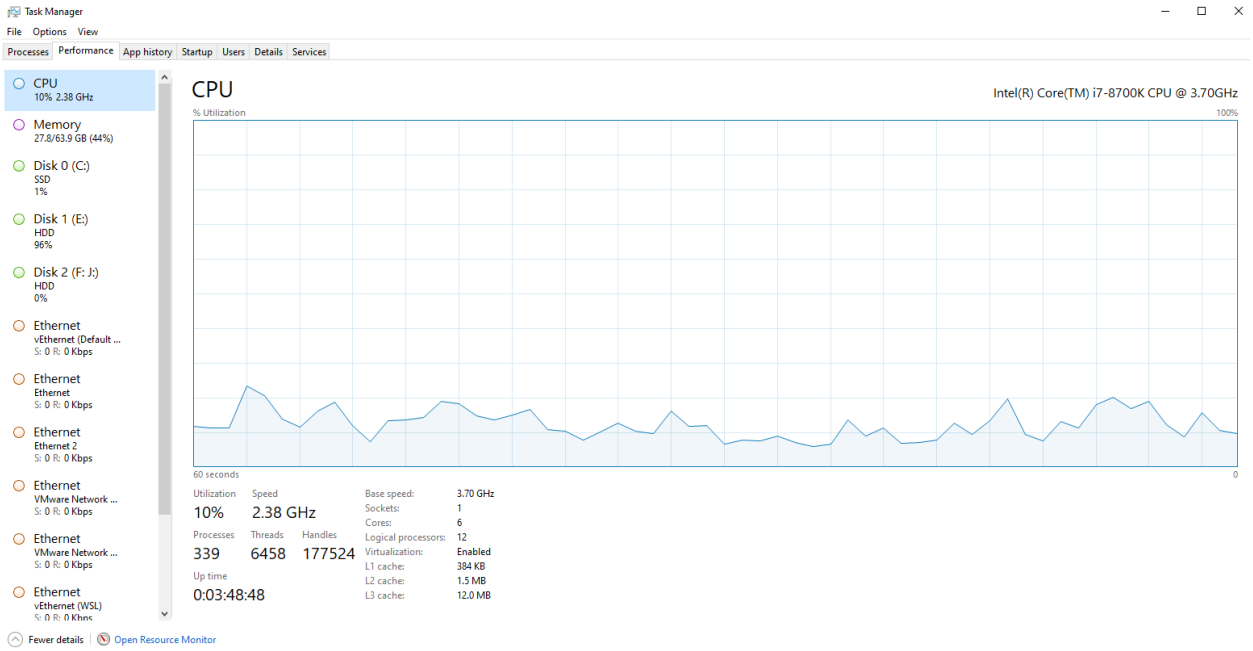


Figure 23 – Computing Power Consumption – Memory

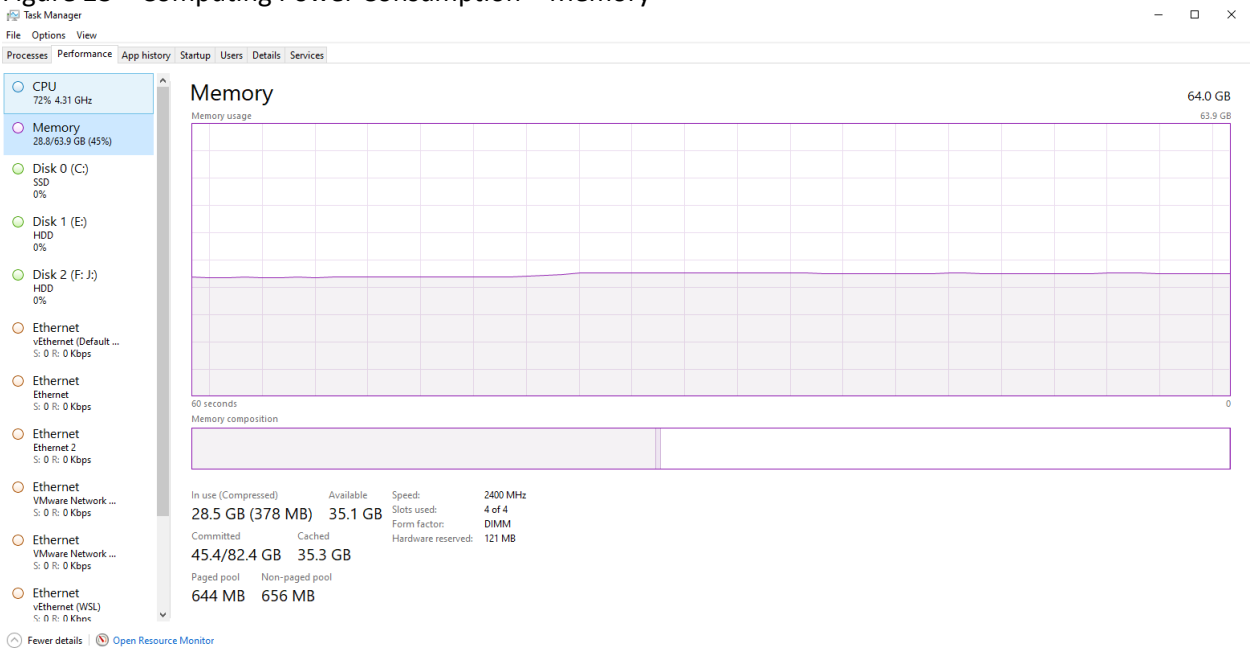


Figure 24 – Computing Power Consumption – GPU 0 Display

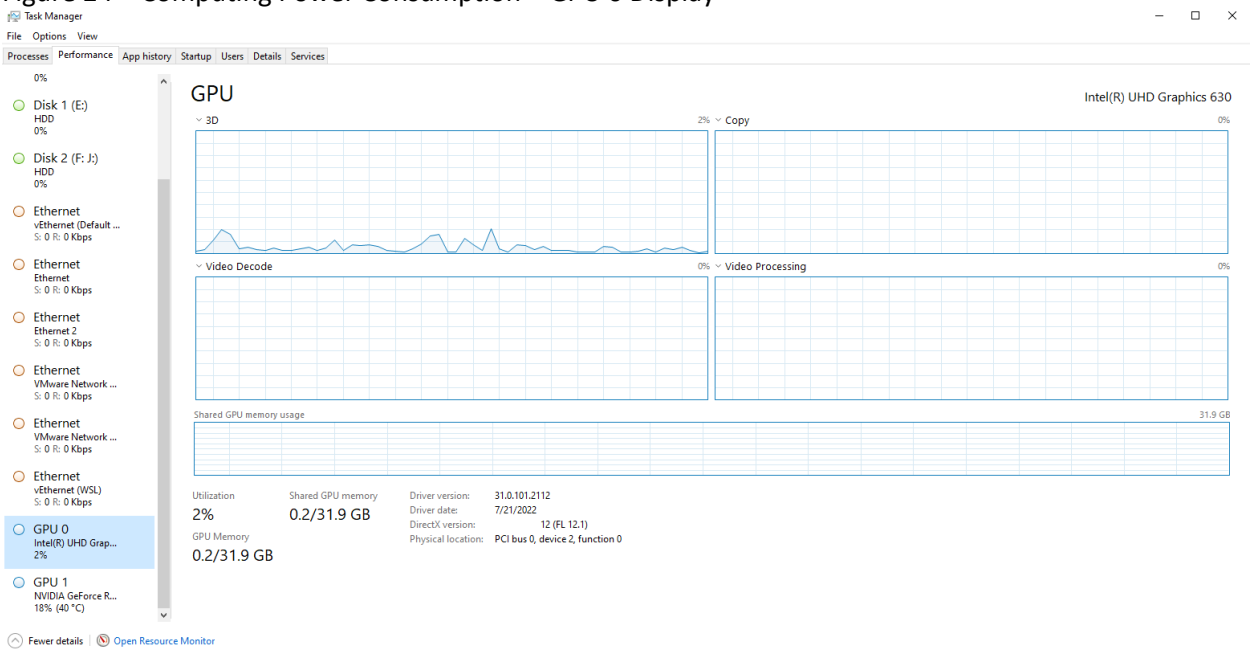
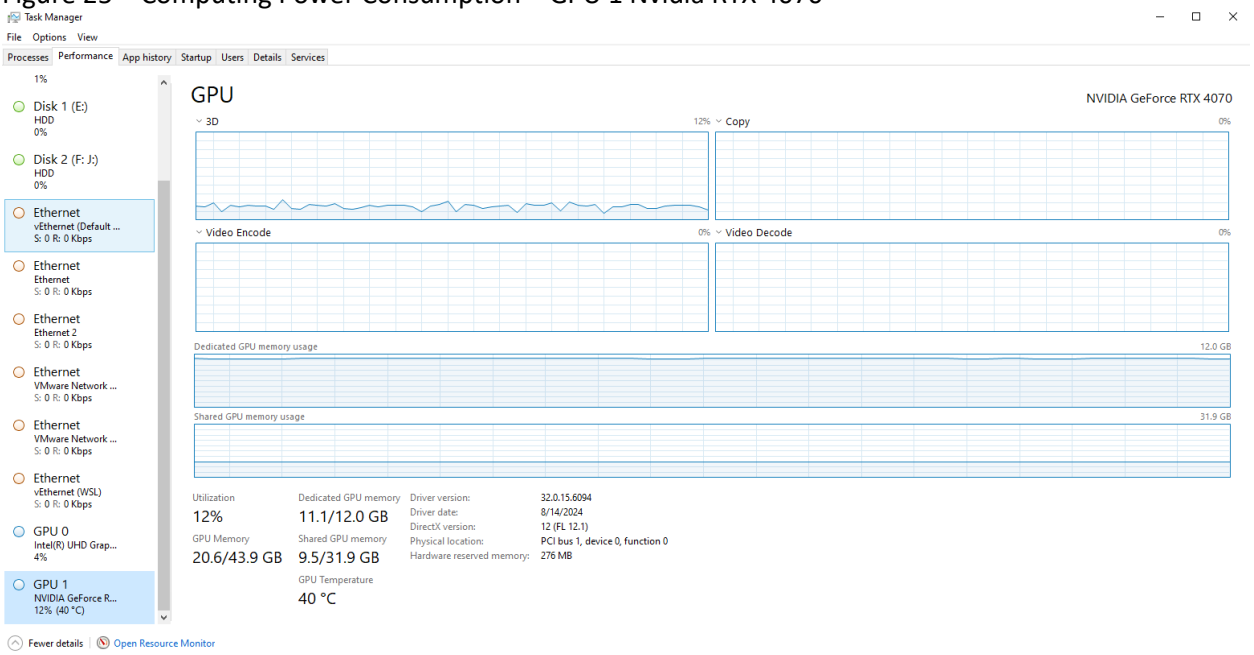


Figure 25 – Computing Power Consumption – GPU 1 Nvidia RTX 4070



Test 2 – Image

DeepSeek R1 is not capable of processing images.

Observations

For Text

- **Memory Usage:** Gemma3 is lower than DeepSeek
- **GPU Usage:** Gemma3 is lower than DeepSeek
- **CPU Usage:** Gemma3 is lower than DeepSeek
- **Disk I/O:** Both Gemma3 and DeepSeek are low
- **Processing Time:** Gemma3 took 8 minutes, DeepSeek took 9 minutes, to complete their response respectively
- **Output:** Gemma3: **840 words**, including Code Interpreter Scorecard and Output; DeepSeek: **345 words**

For Images

- **Processing Time:** Gemma3 took 20 minutes; DeepSeek: Not capable of processing images
- **Output:** Gemma3: Detailed; DeepSeek: Not capable of processing images

Revelation

This simple test with Gemma3 and with DeepSeek R1 for a comparison revealed “currency of model training” challenge, by which it means a model must be trained with current data to work well or the model would produce incorrect responses. A gap always exists between the point of time when a model was trained with then available data and the point of time when the model is used. The greater the gap, the less accurate the model would be.

For example, in our first test, the text question posted to Gemma3 and DeepSeek R1 was identical, which read: “**Please assess Trump’s first six months performance in his second term**.”

Both models treated the question as a hypothetic analysis request, whereas it was supposed to be a fact analysis because when the question was posted on July 11, 2025, Trump was already in office of his second term for more than six months.

Why did both Gemma3 and DeepSeek miss it? Because both models were trained with the data before Trump’s second term started.

Figure 26 – Gemma3

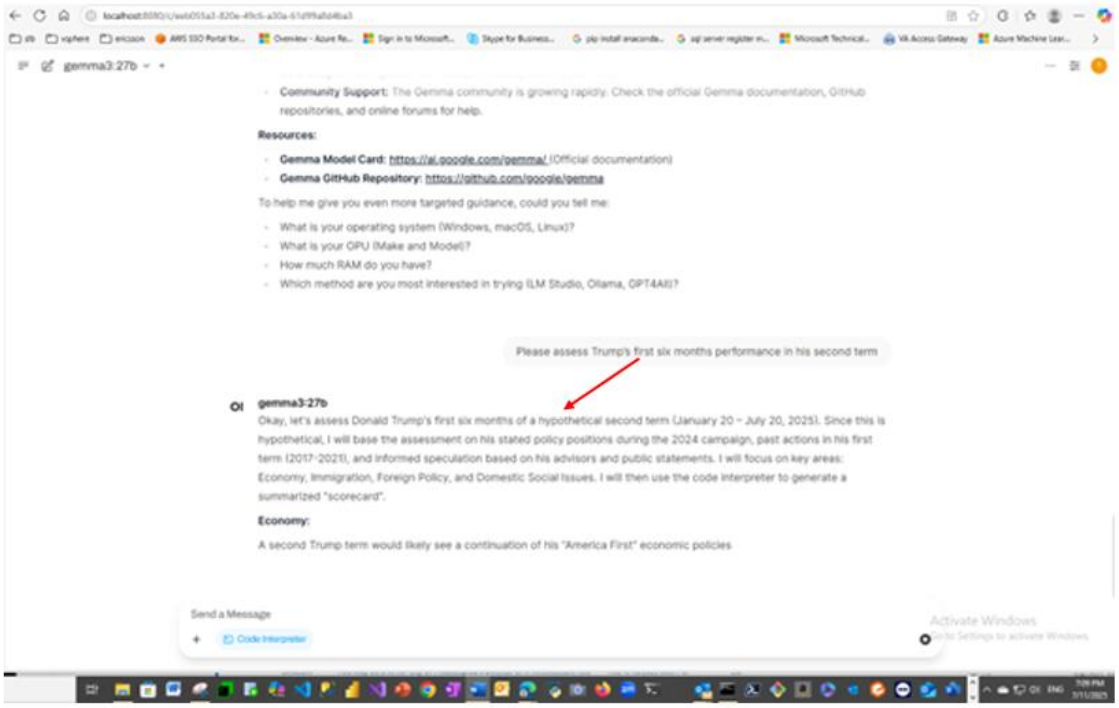
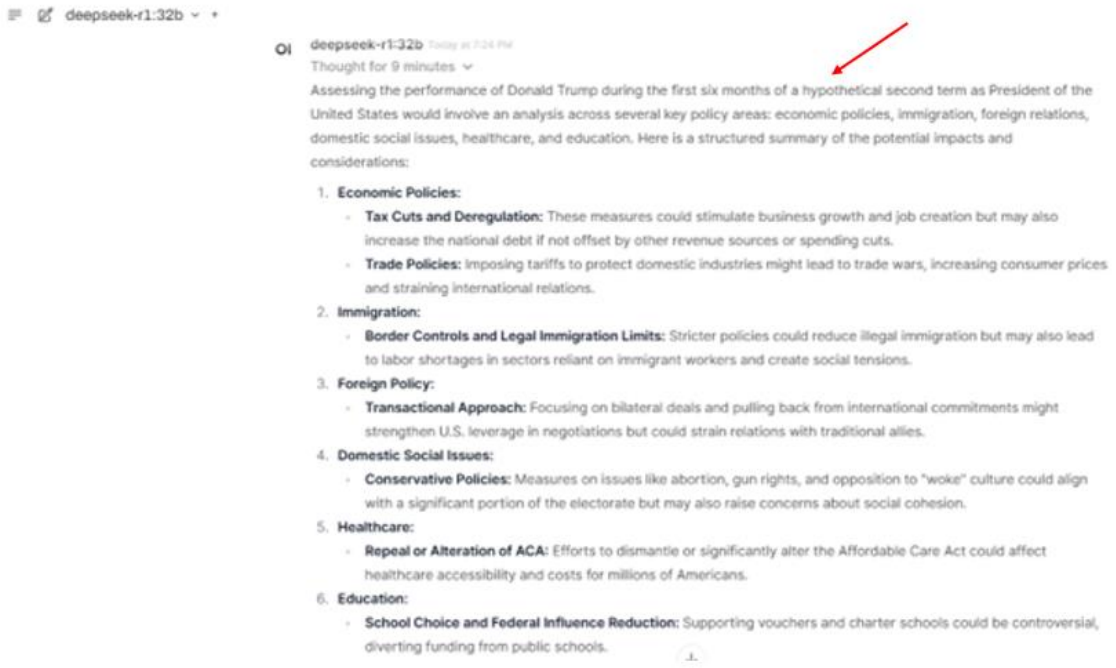


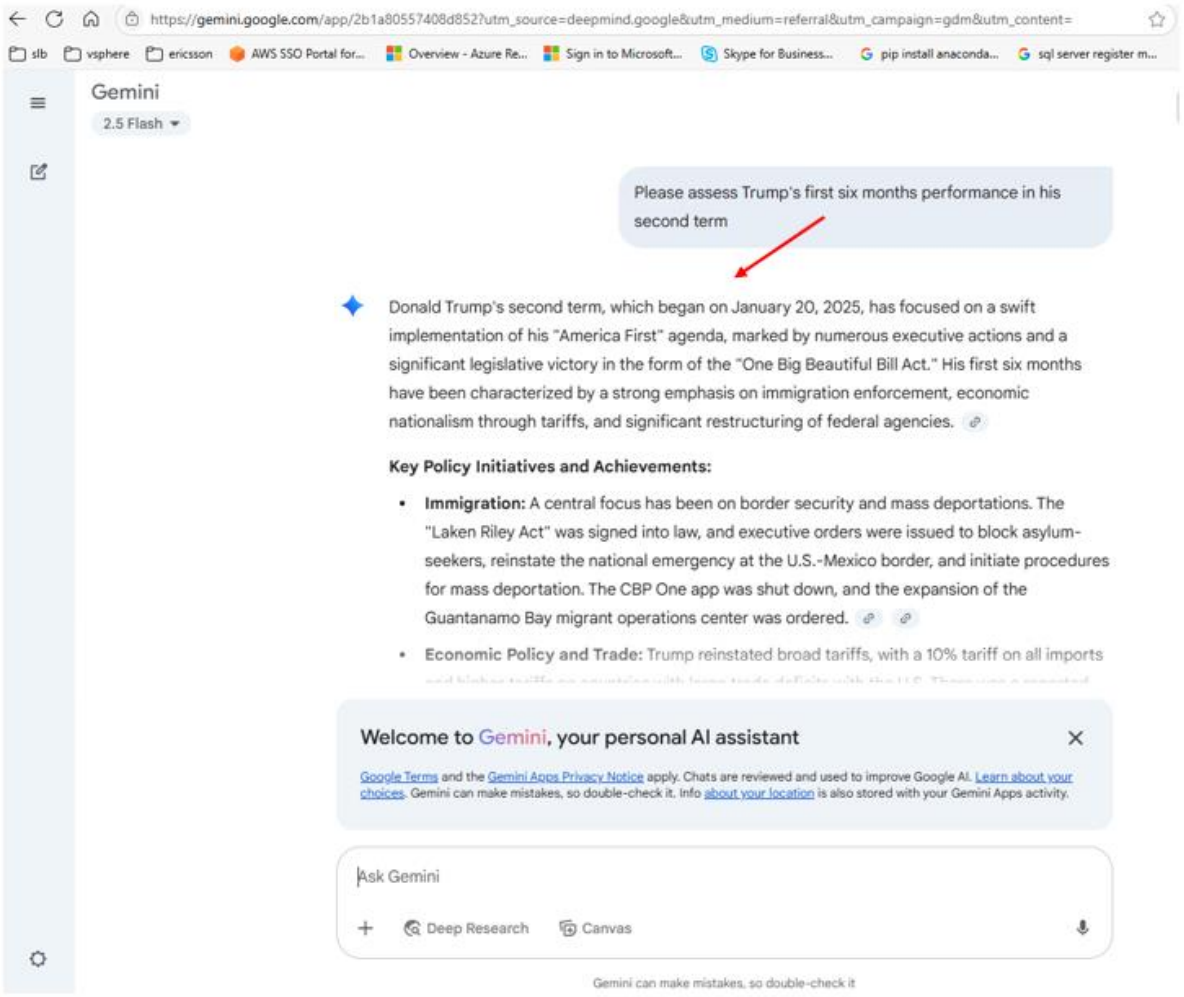
Figure 27 – DeepSeek



How about their online counterparts? Do they do better?

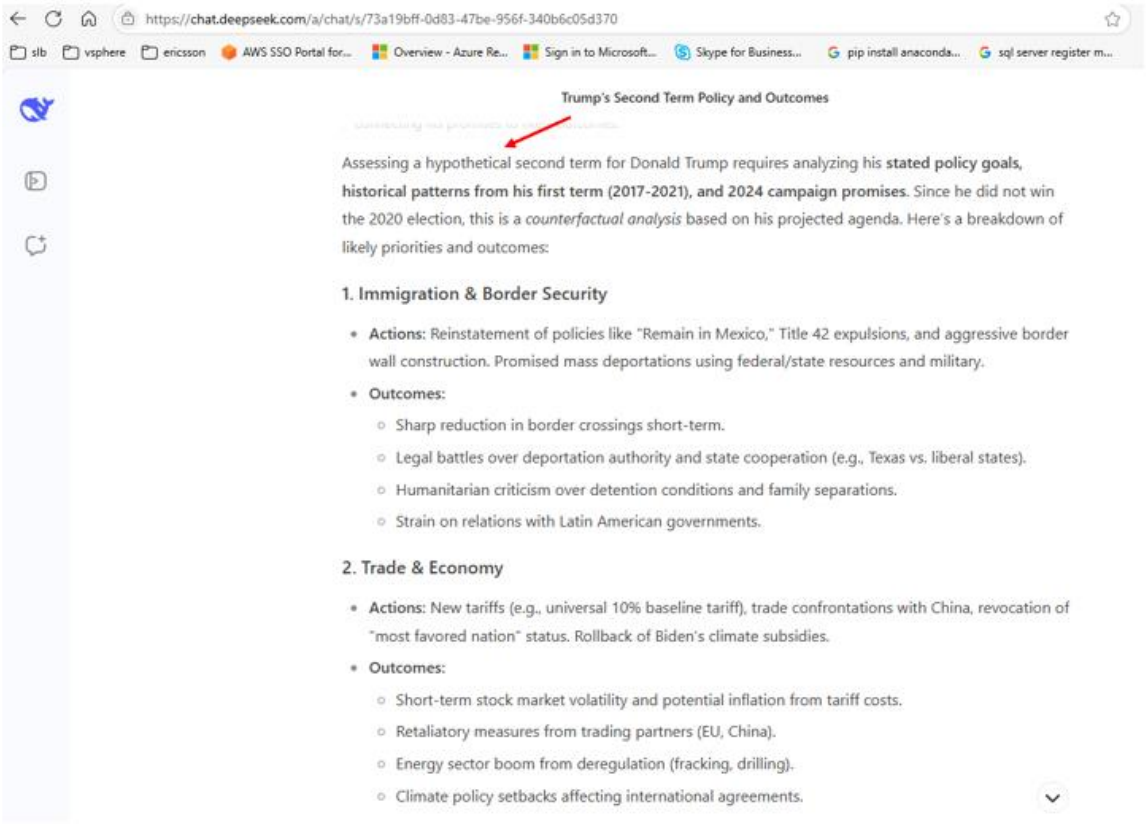
Google Gemini got it right. See Figure 28 below

Figure 28 – Gemini



DeepSeek online chat still thinks it is a hypothetical analysis request. See Figure 29 below.

Figure 29 – DeepSeek online chat



Apparently, DeepSeek has not retrained its model with new data. The model becomes stale.

Summary

A performance comparison test was run on a Windows 10 PC equipped with a low-end Nvidia RTX-4070 GPU card. Two Opensource LLM models were selected for the test, one is Gemma3, the other, DeepSeek R1. Both models come in different sizes. Small-size ones are chosen to run the test, and they are Gemma3:27b and DeepSeek-r1:32b, the most comparable small-size models.

The test run showed that Gemma3 is ahead of DeepSeek R1 considerably. Gemma performs better, demands less computing power, and, most importantly, it is a multi-modal model capable of not only processing languages but also images, whereas DeepSeek R1 has stayed static with no apparent updates since January 2025.

Revision History
Created on July 12, 2025
Revised on July 13, 2025