Тепловизионный анализ и эксплуатационные показатели

Сервер в исполнении Tower

Температуры под стресс-тестом (RTX 4090 48GB):

- Ядро GPU: не выше 69 °C (макс. разрешённая температура ядра: 85 °C)
- Память GPU: не выше 78 °C (макс. разрешённая температура памяти: 100–110 °C)

№ Показатели температуры, зафиксированные на тепловизоре, значительно ниже допустимых значений, что является отличным подтверждением эффективности охлаждения.

Совместимость ПО (Container Runtime + TensorRT):

CNN: производительность -3.6%
LLM: производительность +2.5%

Скорость инференса (vLLM):

• Qwen2.5-14B, Deepseek R1 14B: 32 TPS

