Can LLMs Model the Environmental Impact on SSD?

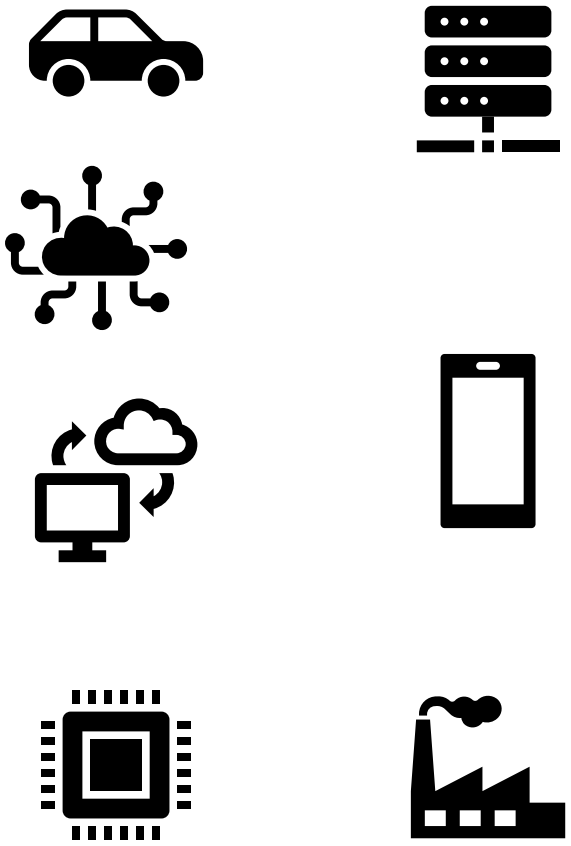
Mayur Akewar[1,*], Gang Quan[1], Sandeep Madireddy[2] and Janki Bhimani[1]

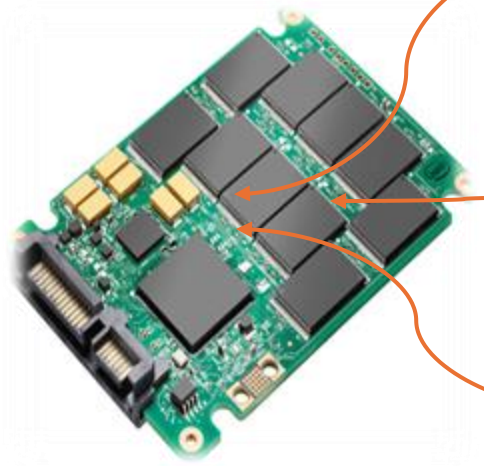
[1]Florida International University [2]Argonne National Laboratory

*makew001@fiu.edu

Operational Environment

Where You Run It, Matters





Temperature

Vibrations

Humidity

From Lab to Reality: SSDs React to Their Environment

[6] https://www.theregister.com/2018/06/22/azure_north_europe_downed_by_pleasant_weather/
[7] https://www.hpcwire.com/2010/01/19/startup_takes_aim_at_performance-killing_vibration_in_datacenter/

Environmental Conditions

Influence on SSD Operation

Increased
Tail Latency [1,2]

Higher
I/O Variability [4]

Bandwidth
Degradation [1,2]

Throughput
Drops [3, 5]

When it gets hot, humid, or shaky, your SSD doesn’t keep up

[1] What Does Vibration Do to Your SSD? [2] Do Temperature and Humidity Exposures Hurt or Benefit Your SSDs? [3] Investigating Power Outage Effects on Reliability of Solid-State Drives [4] SSD Failures in the Field: Symptoms, Causes, and Prediction Models [5] Impact of Environmental Factors on Flash Storage Performance in Autonomous Vehicles

Related Studies

Controlled Environmental Testing

- Costly, small-scale studies (≈ 100 drives) explore limited temperature–humidity scenarios [1, 2]
- Require climate chambers, vibration tables, irradiation facilities → 120h+ per run; many SSDs discarded [1, 2]
- Stress exposure is cumulative: brief peaks cause latent flash wear that skews future results [5]

Data-Driven Modeling

- Analytical formulas & ML built on field SMART logs [4]
- Calibrated to one setting → often break under unseen stressor mixes
- Limited explanatory power: predict what fails, not why

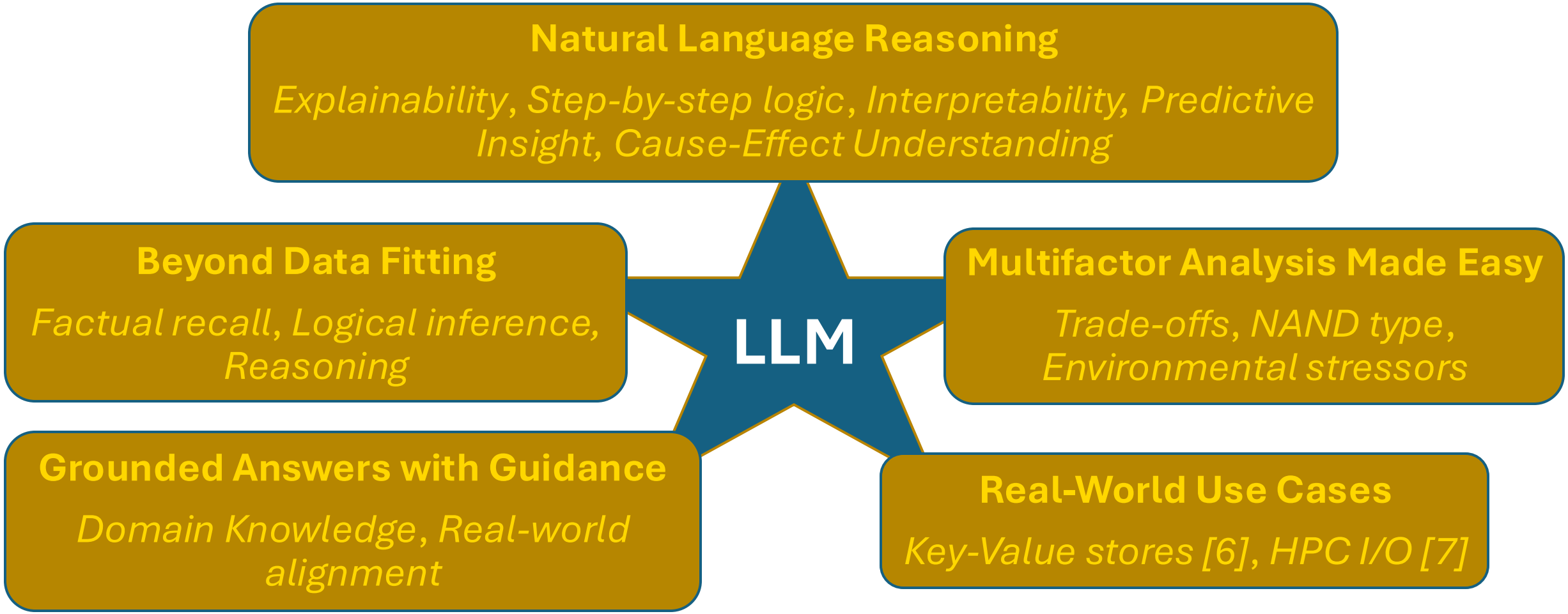
[1] What Does Vibration Do to Your SSD? [2] Do Temperature and Humidity Exposures Hurt or Benefit Your SSDs? [3] Investigating Power Outage Effects on Reliability of Solid-State Drives [4] SSD Failures in the Field: Symptoms, Causes, and Prediction Models [5] Impact of Environmental Factors on Flash Storage Performance in Autonomous Vehicles

Why Is It Challenging to Model Environmental Effects on SSDs?

- Limited Experimental Data
- Historical Exposure Effects (Domino Impact)
- Correlated Environmental Variables (Temp ↔ Humidity)
- NAND Variability (TLC vs. MLC vs. SLC Responses)
- ML & Analytical Models struggle to Generalize (Devices × Environments × Stressors)

Why LLMs?

Using LLMs to Reason About SSDs



A **15°C rise in ambient temperature** can affect the **read latency of an SSD**, but the **impact is typically minimal** under normal operating conditions. Here's a breakdown of what you can expect:

How would a 15°C rise in ambient temperature affect the read latency of an SSD?

🔍 SSD Temperature Sensitivity:

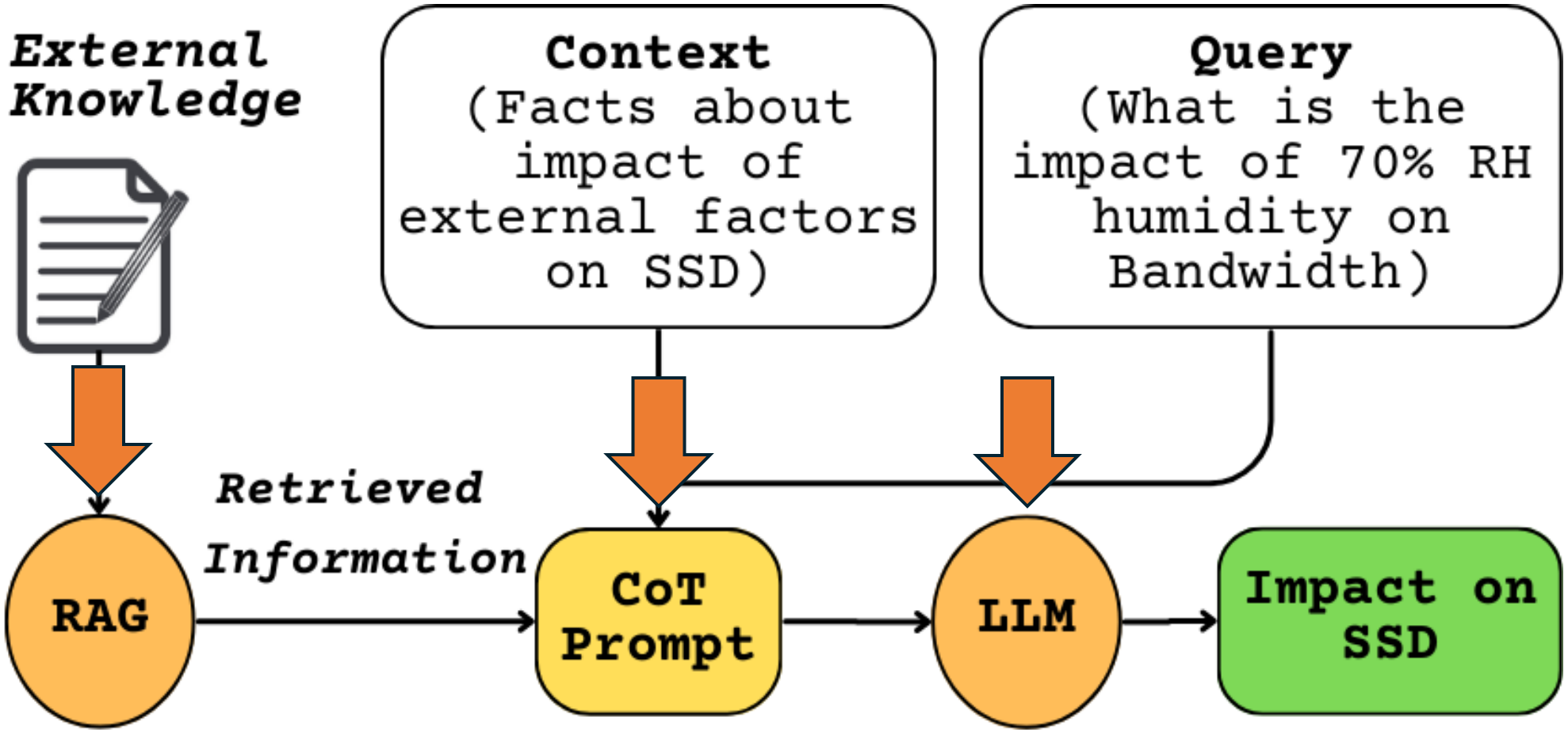
1. **NAND Flash Behavior:**
   - NAND flash memory is slightly temperature-sensitive, but **read operations** are **less affected** than **write or erase** operations.
   - A temperature rise of 15°C is unlikely to cause a dramatic change in **read latency** unless it pushes the SSD past its **optimal thermal range**.
2. **Controller Thermal Throttling:**
   - SSDs have **controllers** that may throttle performance if temperatures approach critical levels (typically ~70–85°C).
   - If the ambient temperature increase causes the **SSD's internal temperature** to exceed its **thermal throttling threshold**, the controller might **slow down all operations**, including reads, to prevent overheating.
   - In such cases, **read latency could increase noticeably**—but only if throttling kicks in.

LLMs Aren't Scientists-Yet

Challenges

- **Prompt Matters**: A vague question leads to a vague answer
- **Context Blindness**: LLMs can misread the situation.
- **Cross-Domain Confusion**: May borrow logic from unrelated domains
- **False Confidence**: Hallucinating unsupported claims

Making LLMs Think Like SSD Experts



Chain-of-Thought Prompting

Teaching LLMs to Think, Not Just Answer

What is Chain-of-Thought (CoT) Prompting?

Guides the LLM to generate intermediate steps before reaching a final answer (mimicking how a domain expert would reason)

Why It Matters for SSD Modeling

Environmental impact, Multi-factor analysis, Complex inference

How We Use CoT Prompting

model explicitly (Think step by step) or implicitly through examples

Benefits Observed

Better accuracy, Explainability, Robust predictions

Chain-of-Thought Prompting

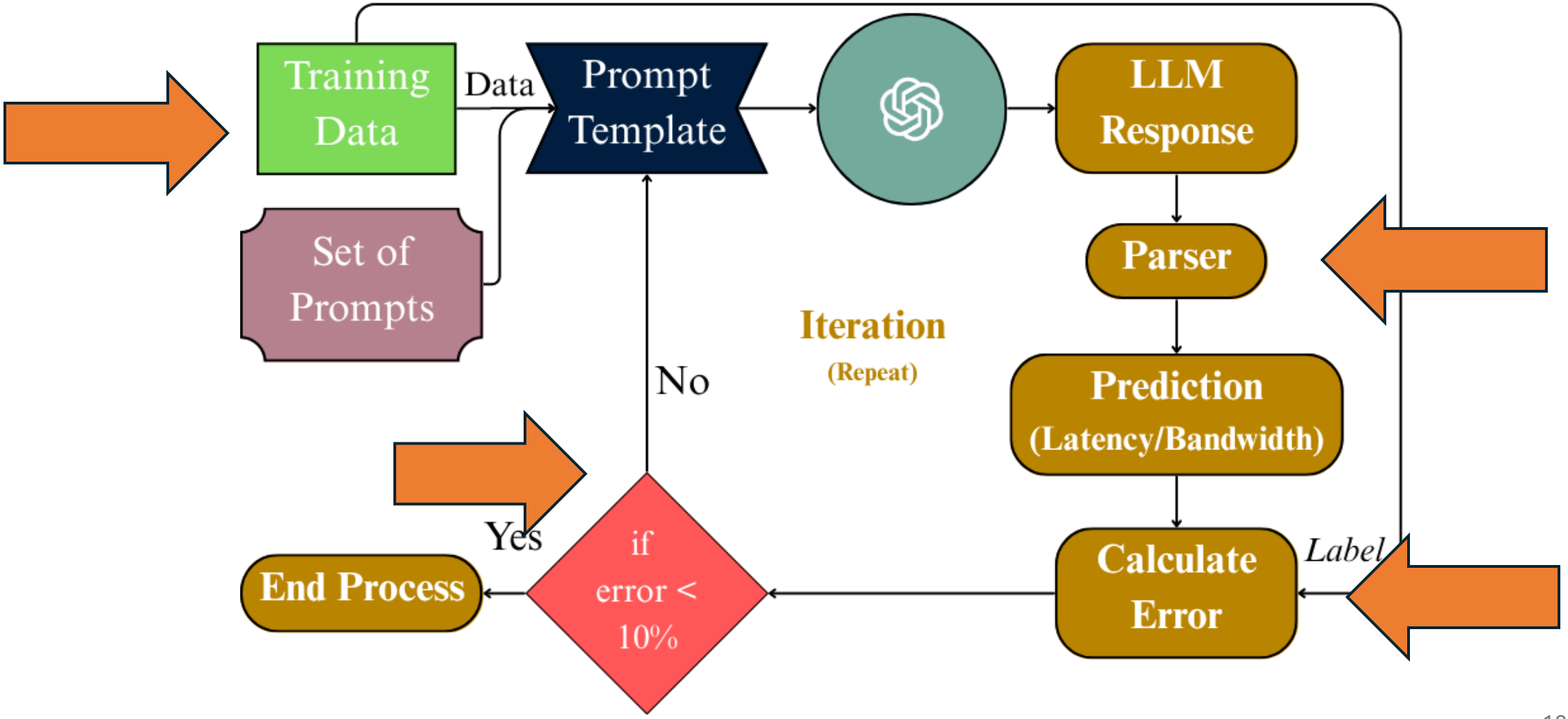
An SSD with TLC NAND runs a workload while the temperature rises from room temperature to $50^{\circ}\mathrm{C}$*, with relative humidity fixed at 50%.*

Think step by step:

① How does increased temperature affect TLC NAND (e.g., electron mobility, latency)? ② Does $50^{\circ}\mathrm{C}$ risk thermal throttling or higher bit error rates in the controller? ③ Does 50% RH impact performance via circuit capacitance or signal delay? ④ Estimate the change in:

(a) Overall performance, (b) Tail Latency (90th–99.99th), (c) Bandwidth, (d) Failure likelihood.

How Effective Is the CoT Prompt?

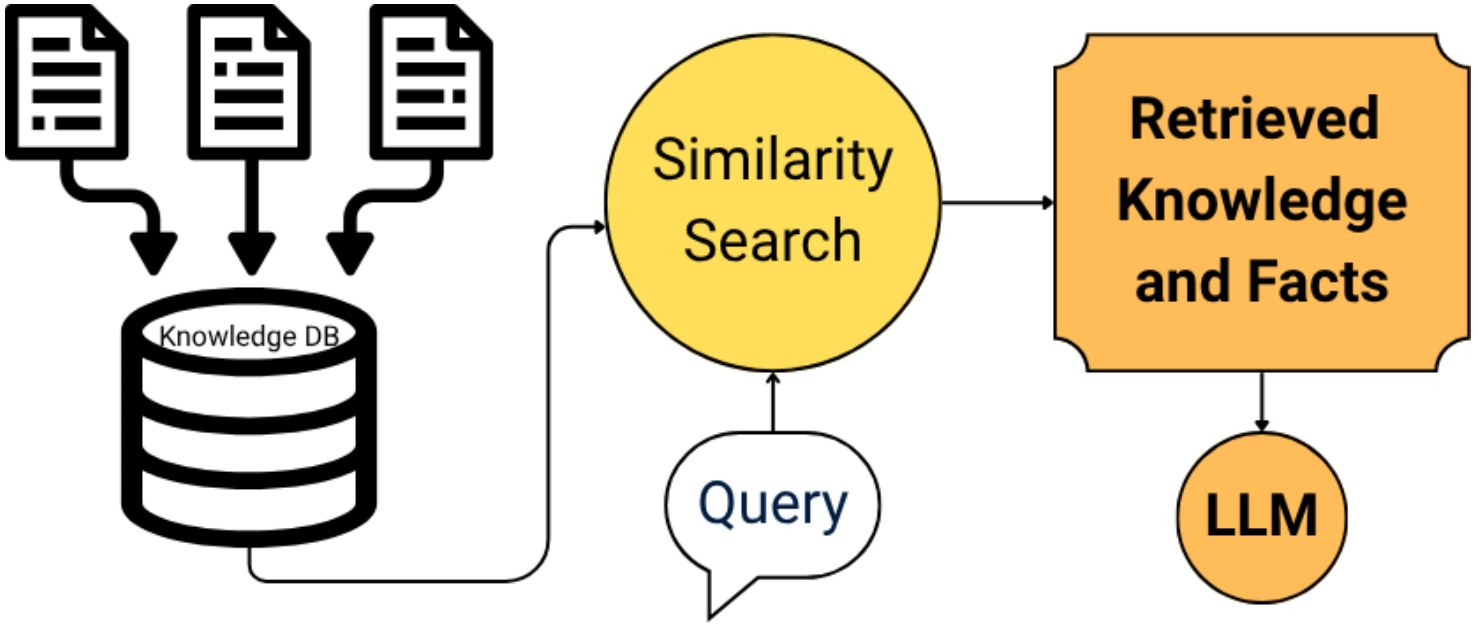


Grounding LLMs with Retrieval-Augmented Generation (RAG)

RAG enhances LLM performance by retrieving relevant knowledge from a curated corpus and injecting it into the prompt

reducing hallucinations and improving trustworthiness

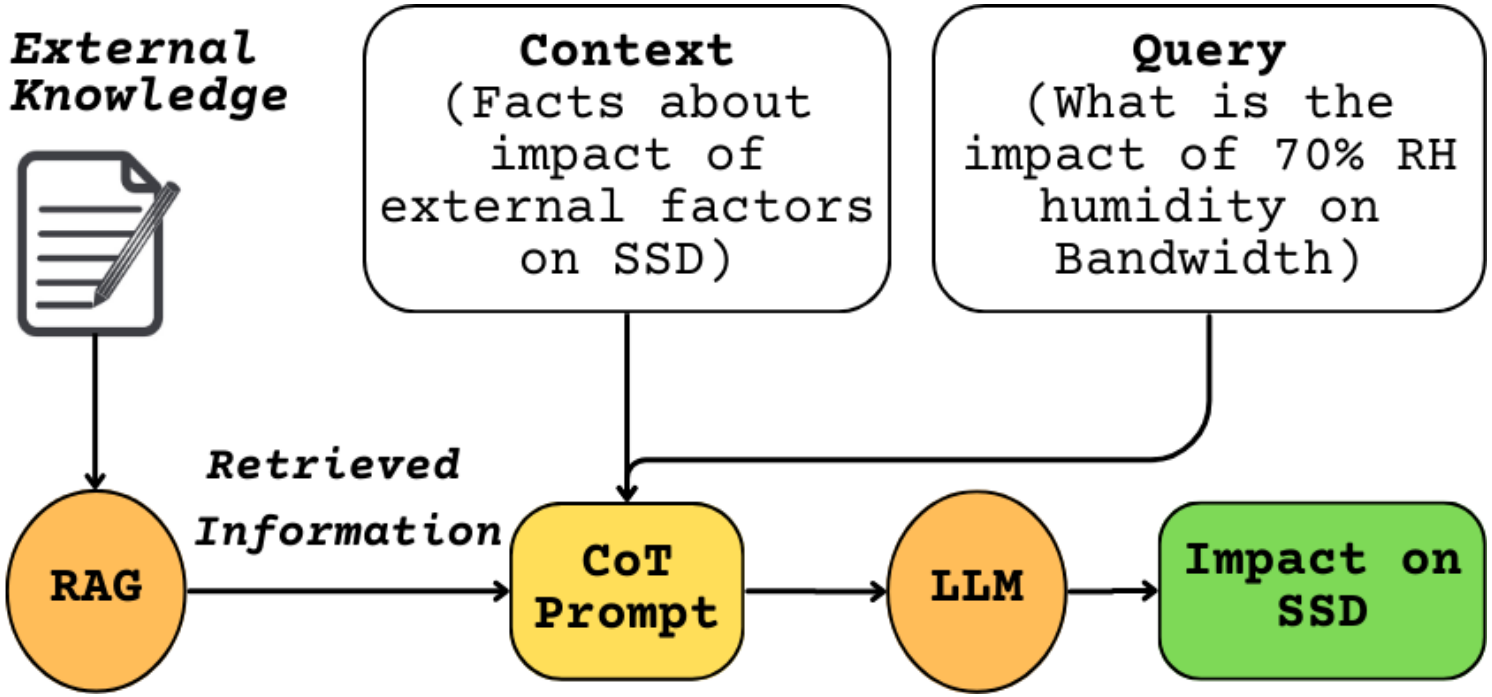
Retrieval-Augmented Generation (RAG)



Why RAG Matters?

- Provides factual support in complex, multifactor scenarios
- Helps the LLM reason based on actual system behavior
- Eliminates the need for large training datasets or retraining the model

Bringing It All Together



By combining CoT Reasoning and RAG, we create an LLM framework that is accurate, interpretable, and grounded in real-world SSD behavior.

Putting the Framework to the Test

Scenario Set

• Temp-humidity swings • Parallel and Vertical vibrations

Data Split

• 20% training • 20% retrieval • 60% test

Prompt Pipeline

a) RAG fetches relevant facts

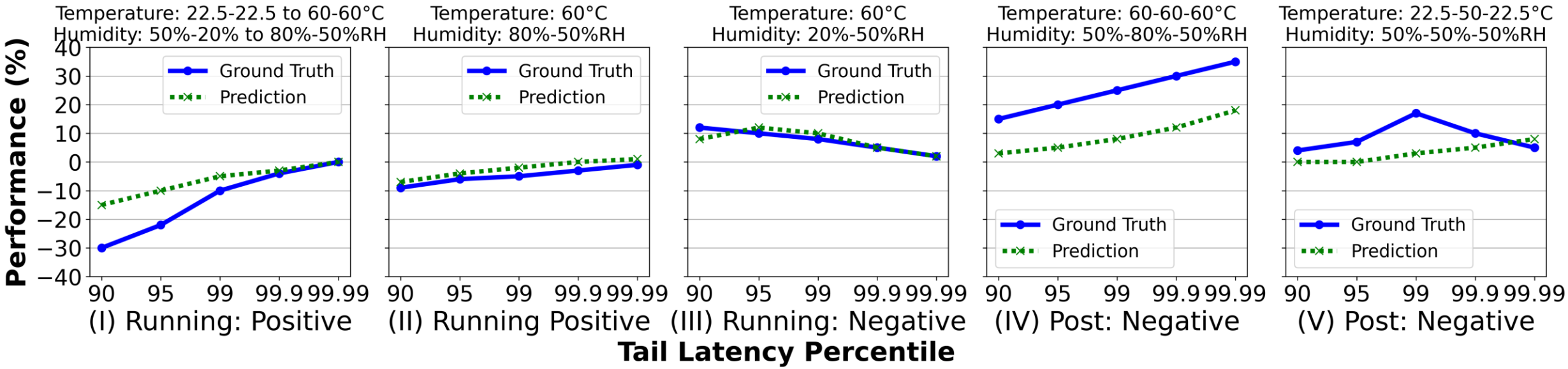
b) Combine with CoT template + query

c) Send to GPT-4o

Validation Loop

• 10 runs per scenario for consistency
• Compare predicted tail-latency & bandwidth to ground truth

Temperature and Humidity Impact



Our results show low RMSE values under the various temperature and humidity settings: about 9% for tail-latency predictions.

Refer to the paper for detailed results and discussion ...

Future Research Directions

Ablation Studies
Isolate RAG vs. CoT roles under varying workloads, applications and stress conditions

Expanded Conditions
Include aging, pressure, radiation in scenario design

Benchmarking Open-Source LLMs
Compare generality, efficiency, and deployment readiness

Summary

- **Problem Statement**
  Can LLMs model the Environmental stressors (temperature, humidity, vibration) impact on SSDs?
- **Key Challenges**
  - Scarcity of comprehensive experimental data
  - Cumulative and interrelated effects of stressors
  - Heterogeneous responses across NAND types (SLC, MLC, TLC) .
- **Proposed LLM-Based Framework**
  - **Chain-of-Thought Prompting** to elicit step-by-step reasoning
  - **Retrieval-Augmented Generation** to ground predictions in empirical results and device specs .
- **Impact**
  - Enables proactive “what-if” SSD health assessments

DaMRL

Argonne NATIONAL LABORATORY

NSF

FIU | FLORIDA INTERNATIONAL UNIVERSITY

Thank you!

https://github.com/Damrl-lab/SSD_LLM

Contact Information

Mayur Akewar

makew001@fiu.edu