FROM THEORETICAL SCALE TO PRECISION ENGINEERING IN SUSTAINABILITY DATA

Sangwon SUH, PhD

Xinghua Chair Professor
School of Environment, Tsinghua University
TianGong Initiative

Agenda

1. About TianGong
2. From automation to quality
3. Evaluating quality
4. Concluding remarks

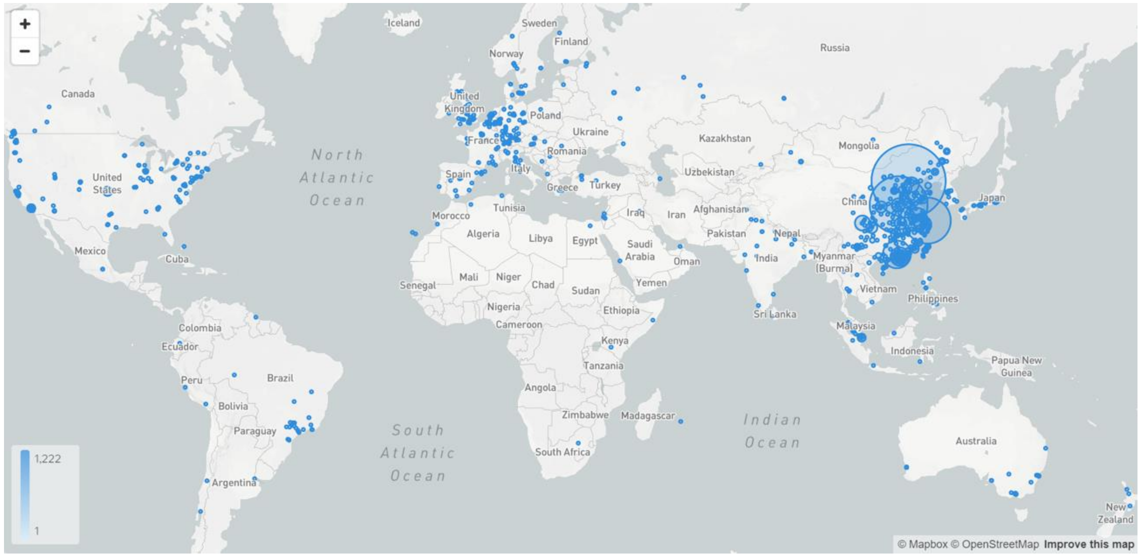
3 What is TianGong?

- A community-driven, AI-powered initiative to develop and maintain open, transparent LCA background database.
- Led by Tsinghua University with 43 investigators and 200+ researchers.

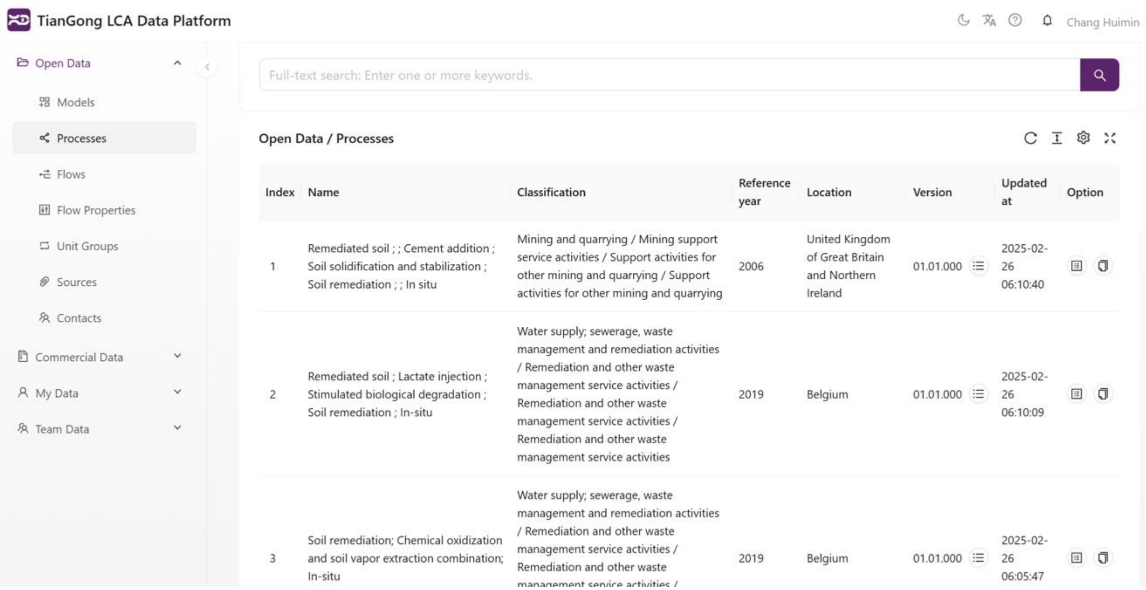
Sample stats

Database	ecoinvent	GaBi	USLCI	TianGong Database
Unit process	~15,000	~17,000	~600	4,315
Geographical coverage	Primarily Europe	Primarily Europe	US	China
Time to develop	~20 years	~20 years	~10 years	1 year

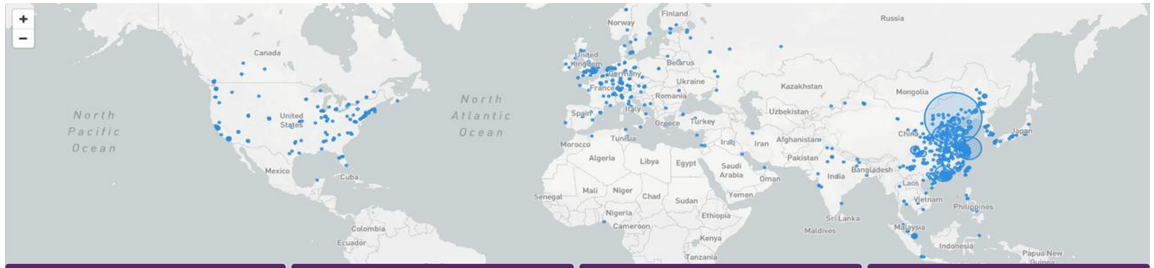
User locations



4 TianGong Database (currently, based largely on the literature)



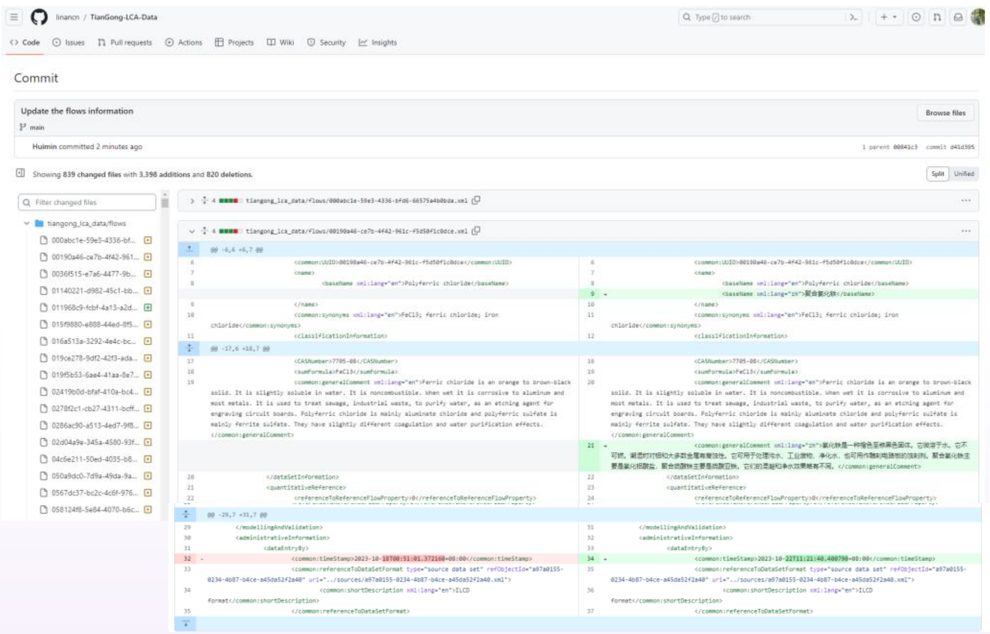
Index	Name	Classification	Reference year	Location	Version	Updated at	Option
1	Remediated soil ; ; Cement addition ; Soil solidification and stabilization ; Soil remediation ; ; In situ	Mining and quarrying / Mining support service activities / Support activities for other mining and quarrying / Support activities for other mining and quarrying	2006	United Kingdom of Great Britain and Northern Ireland	01.01.000	2025-02-26 06:10:40	
2	Remediated soil ; Lactate injection ; Stimulated biological degradation ; Soil remediation ; In-situ	Water supply; sewerage, waste management and remediation activities / Remediation and other waste management service activities / Remediation and other waste management service activities / Remediation and other waste management service activities	2019	Belgium	01.01.000	2025-02-26 06:10:09	
3	Soil remediation; Chemical oxidization and soil vapor extraction combination; In-situ	Water supply; sewerage, waste management and remediation activities / Remediation and other waste management service activities / Remediation and other waste management service activities /	2019	Belgium	01.01.000	2025-02-26 06:05:47	



Visitors	Regions	Visits	Downloads
150,000+	88	2,300,000+	21,000

https://lca.tiangong.earth/

Data sources	Unit processes	Flows
1,188	4,315	96,518



Transparent, and traceable

5 TianGong Database

Deployment Options

SaaS / Online: immediate access and collaboration

Private Deployment: Docker-based local deployment for enterprise security

Developer Tools

MCP: Local/Remote servers for AI Agents

(https://github.com/linancn/tiangong-lca-mcp)

Python SDK: Data structures and validation for LCA workflows ***(tidas-tools)***

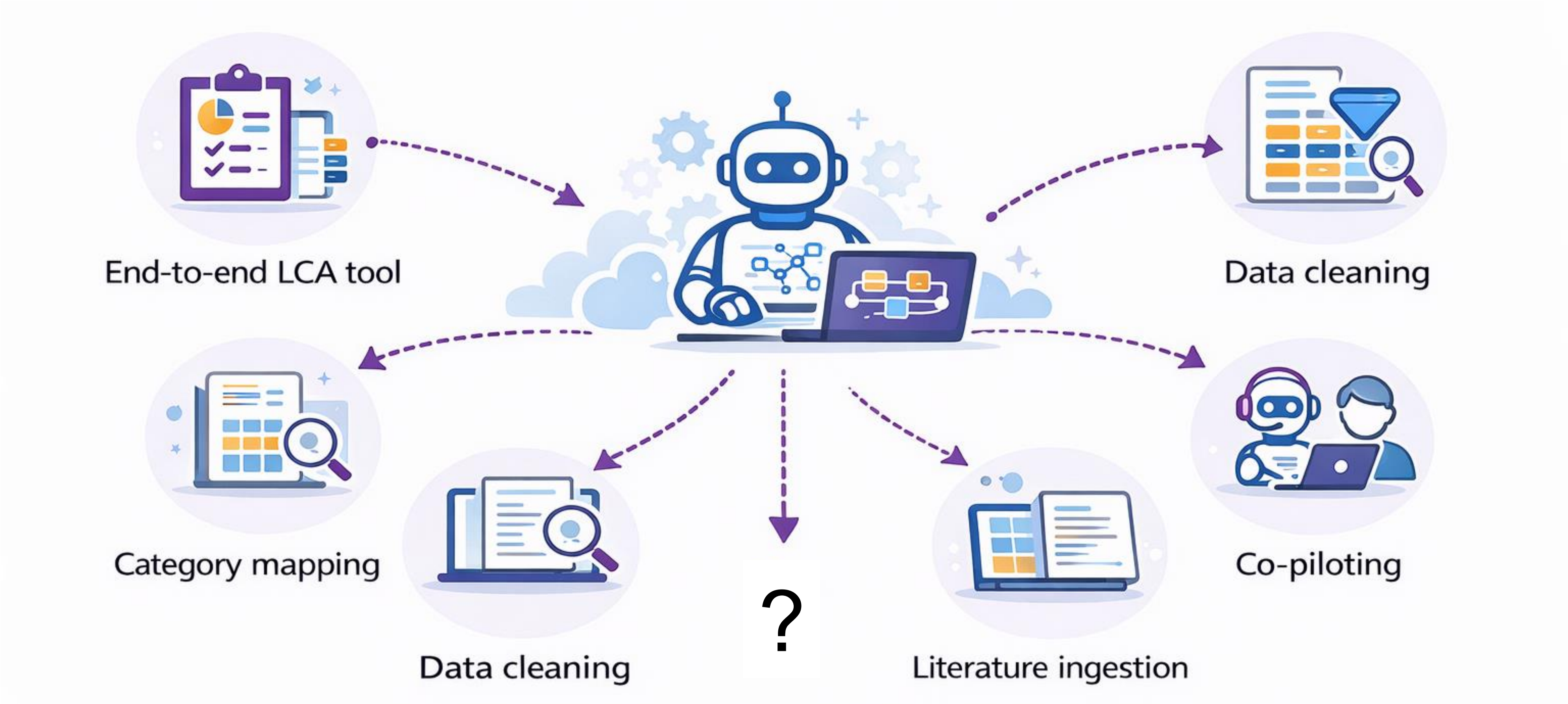
Embedding model: Trained on LCA datasets and process semantics, improving data understanding

Open LCA database. Flexible deployment. Ready for AI.

https://lca.tiangong.earth/

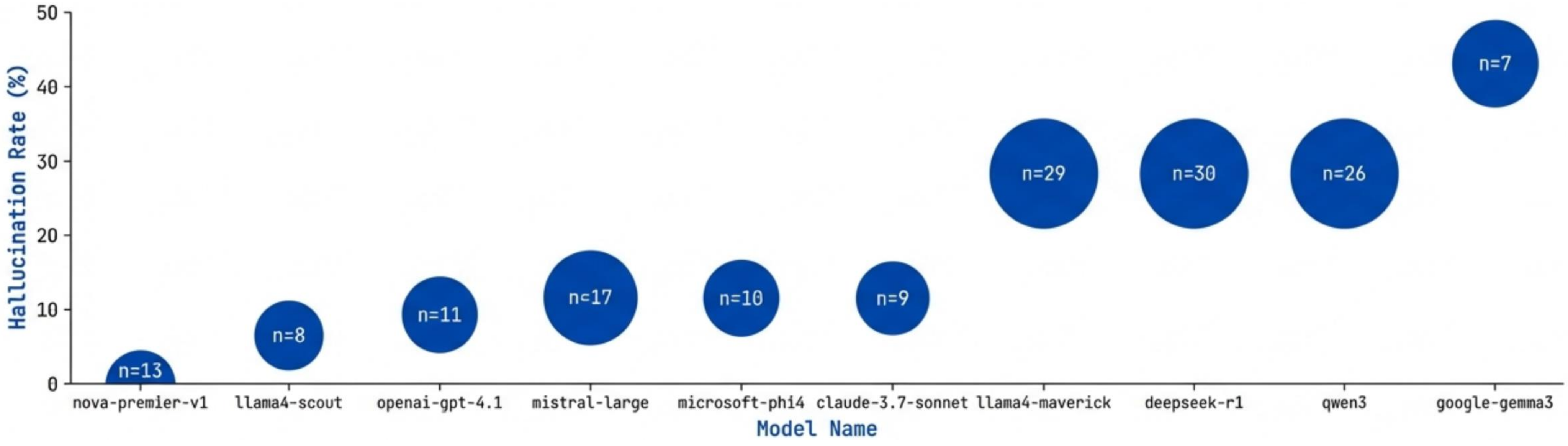
FROM AUTOMATION TO QUALITY

7 Proliferation of AI in LCA



8 Reality check

- **37.2%** of responses contained **incorrect or misleading information**
- **Up to 40% hallucination on citation** for some models

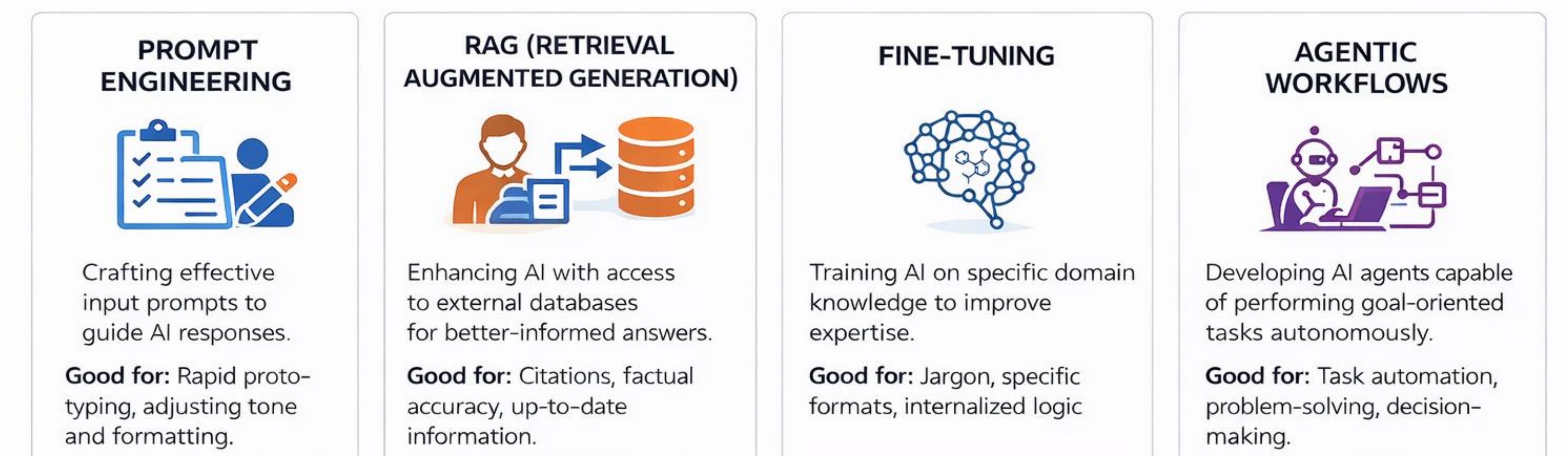


Source: Donaldson, A., Balaji, B., Oriekezie, C., Kumar, M. & Patouillard, L. An Expert-grounded benchmark of General Purpose LLMs in LCA. 2025. https://arxiv.org/abs/2510.19886

Quality concerns: 2025 ACLCA challenge session feedback

- "**AI needs deep training by subject matter experts** before being trusted to select appropriate technologies."
- "AI needs more **stringent, specific rubrics** compared to human expert judgment."
- "At least initially, **human-inspection by sector-expert** on technological representativeness is needed."
- "AI should not both perform and validate allocation; needs **clear documentation for human-verification**."
- "Currently don't trust AI fully for **source credibility assessment**; would require **human checks** initially"
- "Common conventions (e.g., medium voltage preference) can be used to flag **unconventional AI choices for human review**."

Approaches for improving the performance of AI



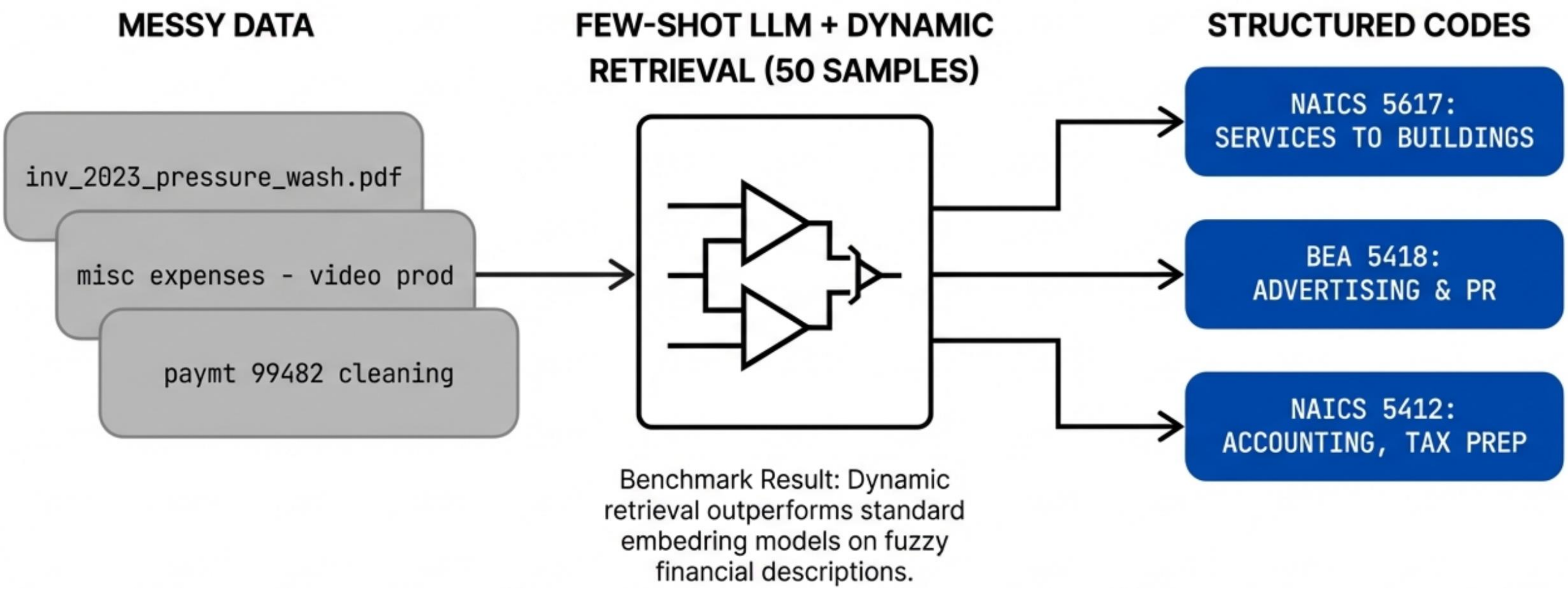
Giving **detailed instructions** on homework assignments

Giving the **references info to be cited** in their homework

Giving an **intensive review on a specific subject** before the exam

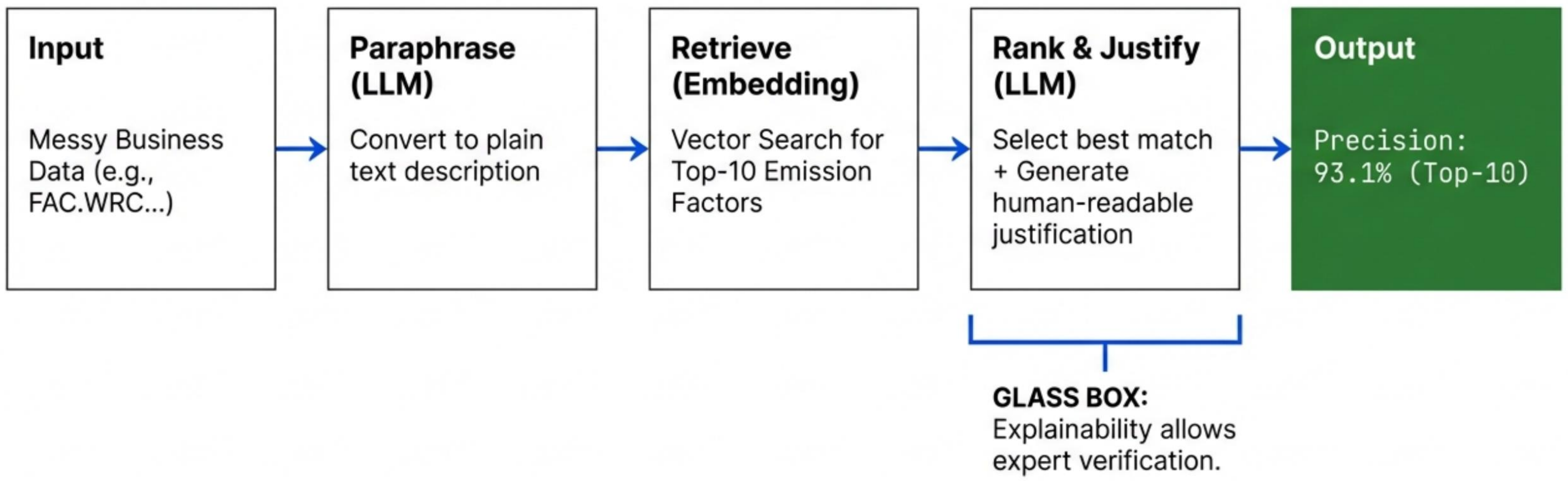
Allowing students to **work together** on a homework assignment

ATLAS for mapping (prompt engineering w/ RAG)



Source: Watershed (https://neurips.cc/virtual/2024/100600)

PARAKEET (prompt engineering w/ paraphrasing + RAG)



Source: Amazon (https://www.amazon.science/publications/parakeet-emission-factor-recommendation-for-carbon-footprinting-with-generative-ai)

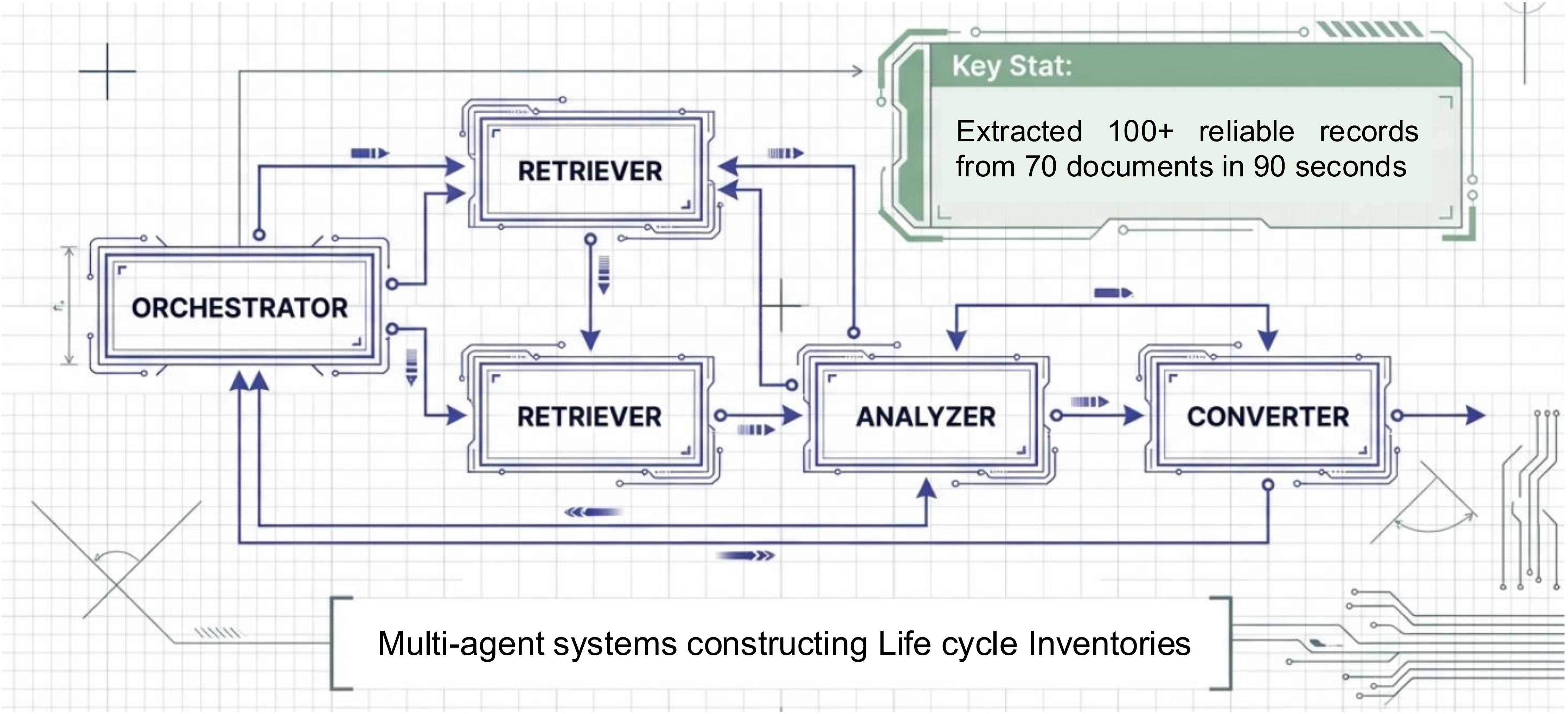
Fine Tunning using domain-specific knowledge: TianGong

Redacted (manuscript under preparation)

LCA domain-specific embedding outperformed raw or generic embedding models at all cutoff-rank values (*K*)

Source: TianGong: manuscript under preparation

Agentic workflow: High- LCA Data Extraction at Scale



Source: TianGong from Tsinghua University **(see Jinliang's presentation later today)**

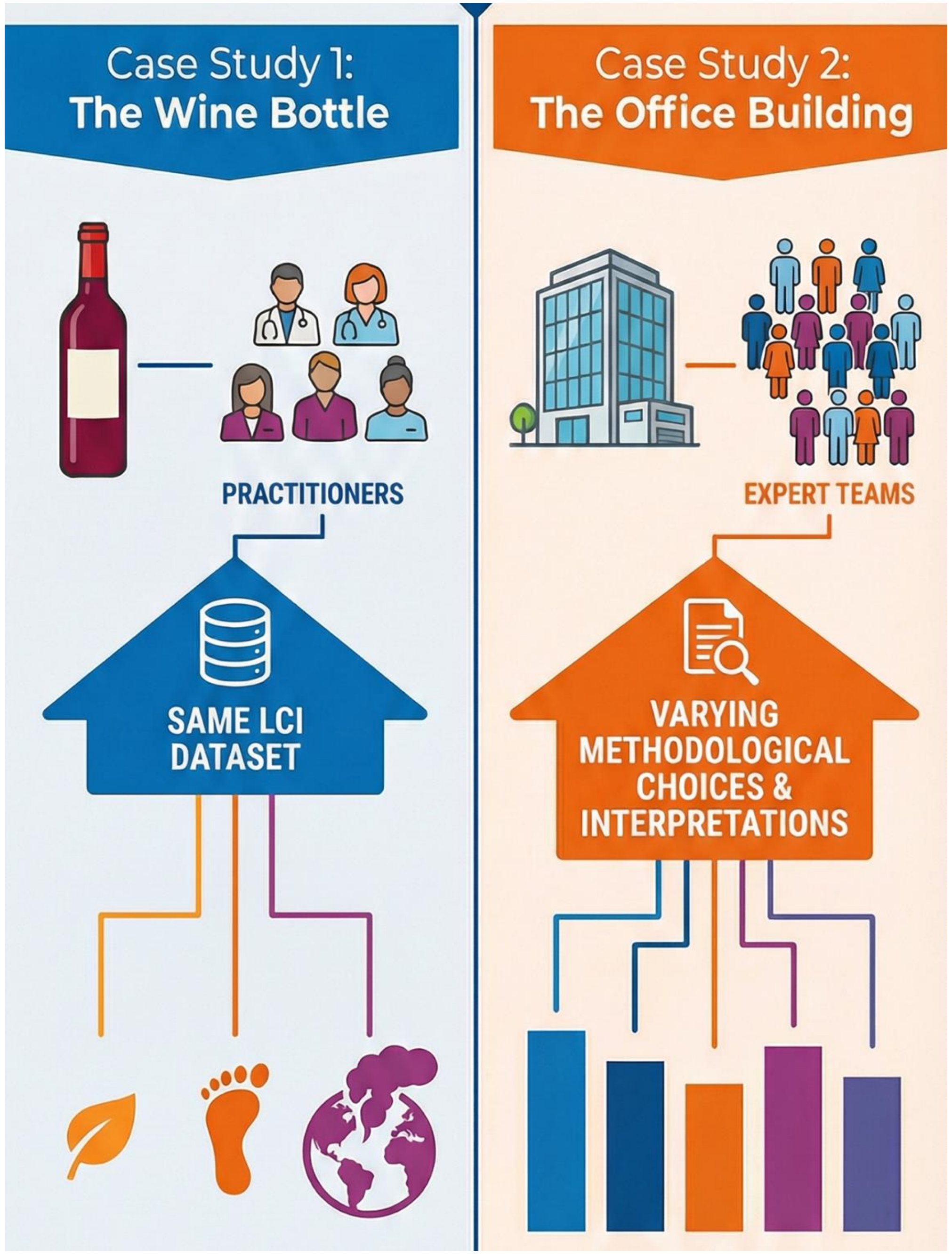
EVALUATING QUALITY

Challenges

- **The "Ground Truth" Paradox:** High-quality labeled LCA datasets are scarce, and experts often don’t agree with each other.

The "Ground Truth" Paradox

- **Practitioner Effect (Literature):**
  - Independent practitioners produce **+-20% difference** on a simple product w/ same LCI dataset (Scrucca et al., 2020).
  - Round-robin study shows a **X 7 variation** (10 v.s. 71 kg CO_2-eq m^{-2} yr^{-1}) on the same building with different (process) LCI datasets.
- AI cannot simply "predict" the expert's consensus, because often there is none.



Challenges

- **The "Ground Truth" Paradox:** High-quality labeled LCA datasets are scarce, and experts often don’t agree with each other.
- **Goal & Scope Dependency:** Quality is not absolute; it depends on the study's context.

The "Ground Truth" Paradox: from ACLCA 2025

LCA is not binary. Decisions on **Allocation (Economic vs Mass)** and **System Boundaries** are **subjective** and **context-dependent**.

Challenges

- **The "Ground Truth" Paradox:** Not only that high-quality labeled LCA datasets are scarce, but also that experts often don’t agree with each other.
- **Goal & Scope Dependency:** Quality is not absolute; it depends on the study's context.
- **Hallucination:** Studies show that LLMs are rather lazy; they would gladly make things up rather than laboring to find the truth.
- **Traceability in Agentic Swarms:** Tracing the root cause of an error is becoming increasingly complex as multiple agents interact with each other.

How to navigate?

- **The "Ground Truth" Paradox:** the goal is to set the boundaries for acceptable options (and their quality hierarchies, if any).
  - For procedural quality: synthetic Q&As based on smart chunking + human validation

Example: synthetic Q&A pair

Category 1: Goal & Scope Definition

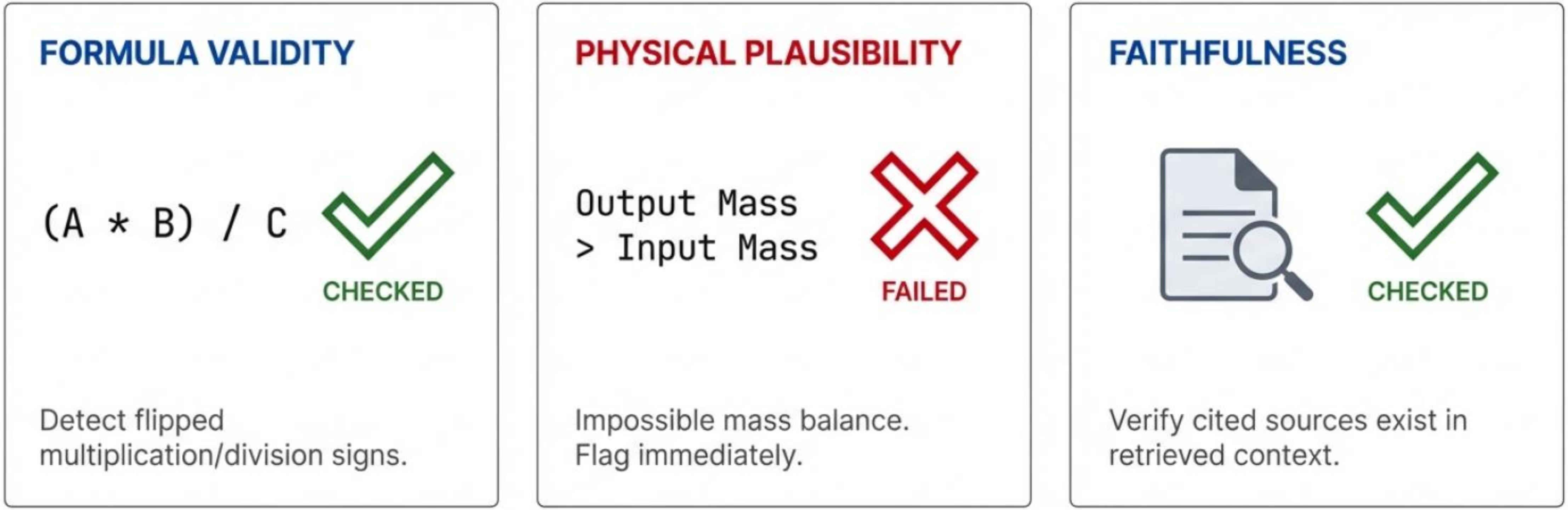
Testing the AI's ability to detect ambiguity and enforce definition requirements.

- **Example: Functional Unit Sufficiency**
  - **Scenario/Prompt:** "A study compares two beverage packaging options (Glass vs. PET) and reports results in 'kg CO2e per bottle produced.' Is this functional unit sufficient for a comparative assertion intended for public disclosure?"
  - **Reference Anchor:** ISO 14044, Clause 5.2.3 (Functional Unit Requirements).
  - **Expected "Gold" Answer: No.** The functional unit is insufficient.
- **Required Reasoning:**
  1. Per bottle" describes a reference flow, not a functional unit.
  2. It fails to define the **performance characteristics** (e.g., volume delivered: 500ml) and **duration/quality** (e.g., shelf life or carbonation retention).
  3. Without equating the service provided, the comparison is invalid under ISO standards.
- **Critical Failure Flag:** If the AI answers "Yes" or proceeds to compare the materials without flagging the ambiguity,.

How to navigate?

- **The "Ground Truth" Paradox:** the goal is to set the boundaries for acceptable options (and their quality hierarchies, if any).
  - For procedural quality: synthetic Q&As based on smart chunking + human validation
  - For empirical validity: property eval + key ground truth ranges (e.g., electricity)

Empirical validity: Property Evals as the First Line of Defense



Source: Parikh and Dumit (2025). *A practical framework for LLM system evaluations for multi-step processes*, Watershed (https://watershed.com/blog/a-practical-framework-for-llm-system-evaluations-for-multi-step-processes)

How to navigate?

- **The "Ground Truth" Paradox:** the goal is to set the boundaries for acceptable options (and their quality hierarchies, if any).
  - For procedural quality: synthetic Q&As based on smart chunking + human validation
  - For empirical validity: property eval + key ground truth ranges (e.g., electricity)
- **Goal & Scope Dependency:** Machine-readable G&S template

Context-awareness: machine-readable G&S template

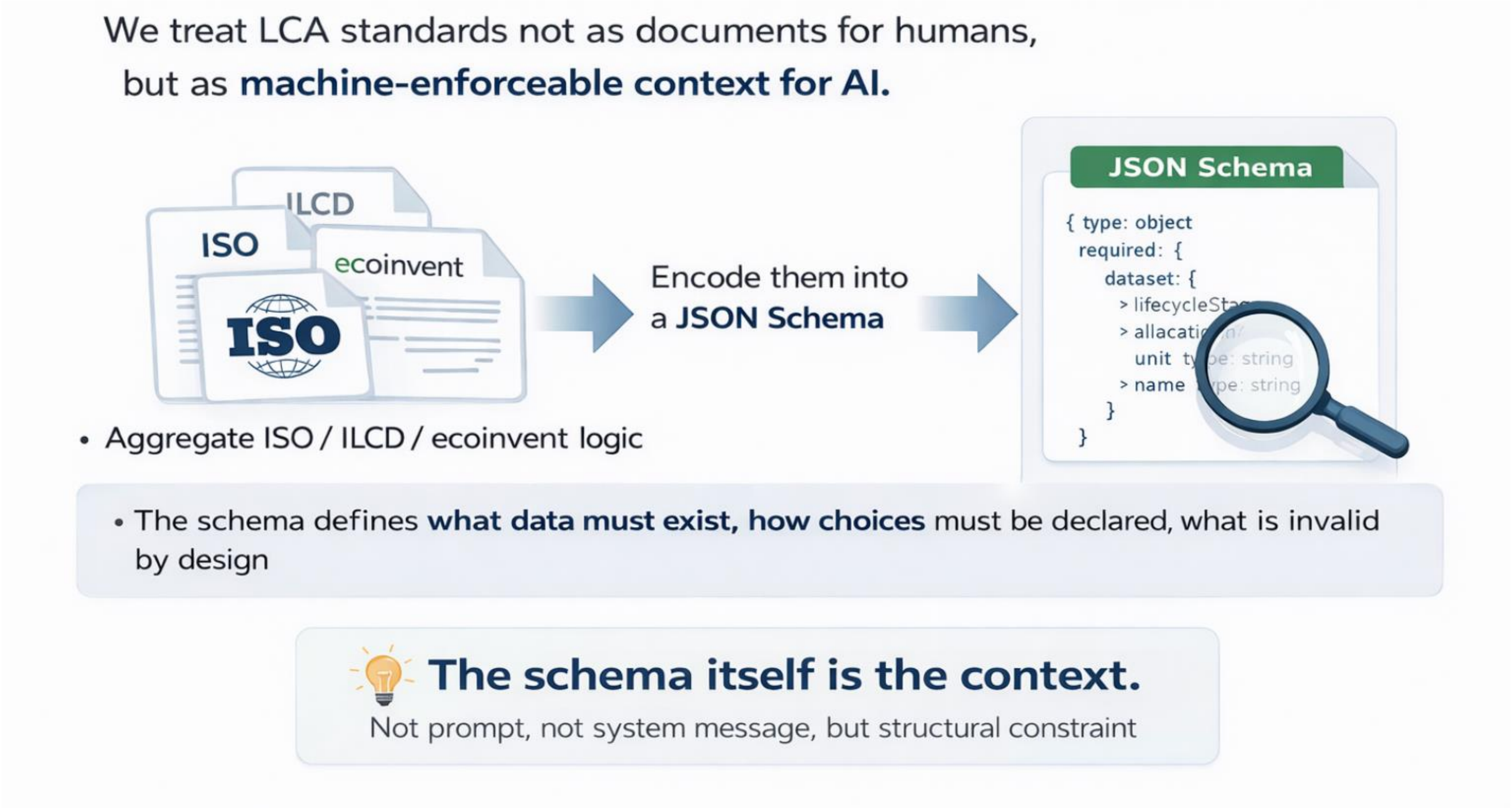
- Addresses "it depends" problem by constraining the choice of method and data based on the context.

```
{
  "GOAL_DEFINITION": {
    "intended_application": "comparative_assertion",
    "intended_audience": "public_consumer",
    "comparative_assertion_public": true,
    "regional_context": "US-NA",
    "methodology_trigger": "ISO_14044_Clause_5.3"
  }
}
```

TRIGGER: Must enforce sensitivity analysis & critical review.

CONSTRAINT: Triggers specific PCRs and Grid Mixes.

Example: TIDAS—procedural quality via schema-as-context



Source: *https://tidas.tiangong.earth/en/docs/intro*

How to navigate?

- **The "Ground Truth" Paradox:** the goal is to set the boundaries for acceptable options (and their quality hierarchies, if any).
  - For procedural quality: synthetic Q&As based on smart chunking + human validation
  - For empirical validity: property eval + key ground truth ranges (e.g., electricity)
- **Goal & Scope Dependency:** Machine-readable G&S template
- **Hallucination & Faithfulness:** Hallucination check—*does the reference exist*?
- **Traceability in Agentic Swarms:** Minimum documentation requirement including the documentation of key “reasoning” or “chain-of-thought”.

Standardizing Minimum Documentation Requirement

A. **System Provenance:** Model version, exact prompt log, and timestamp to trace hallucinations.

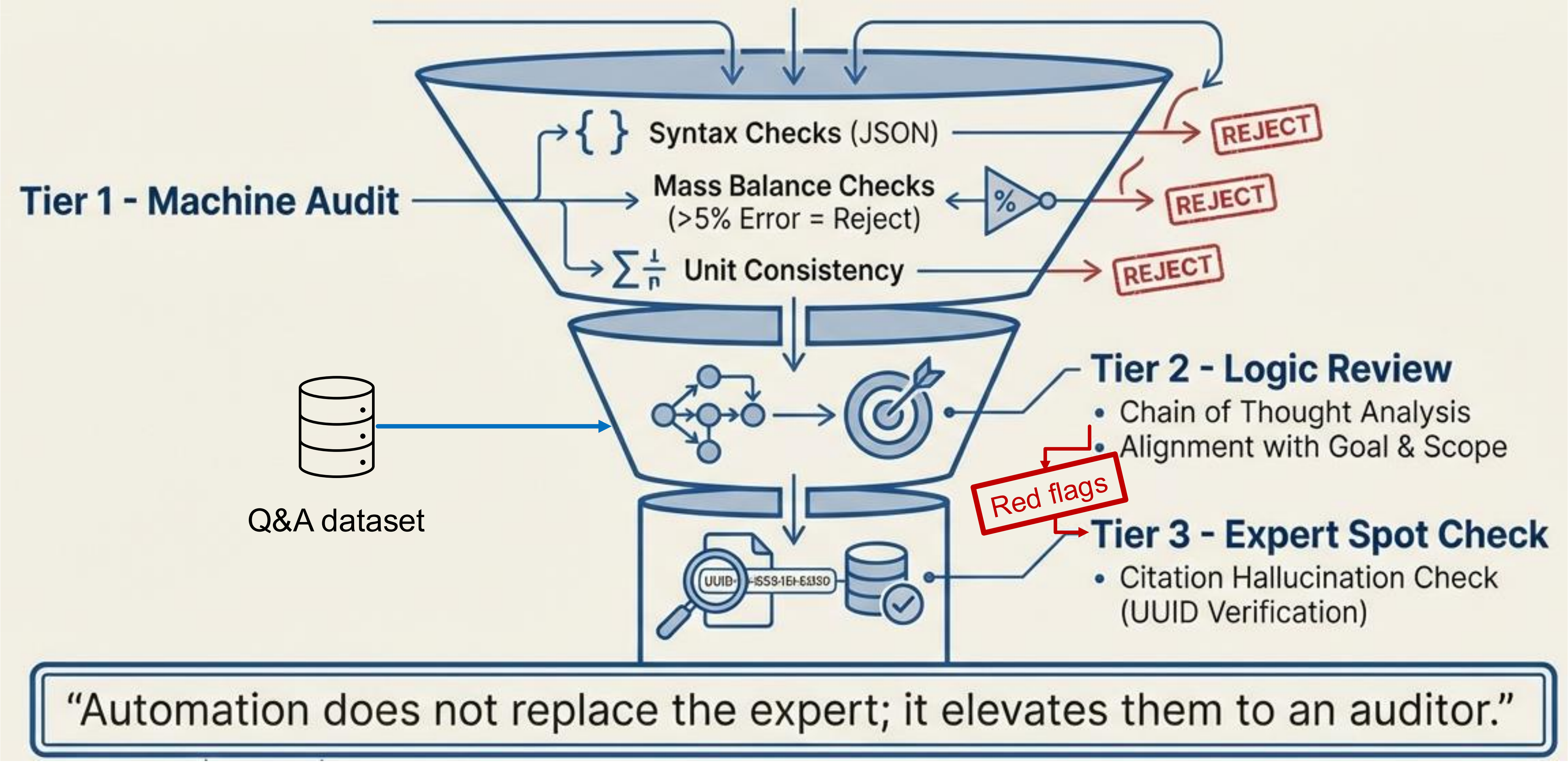
B. **The "Context Key" (Goal & Scope)**.

C. **The ”Chain of Thought" (Logic):** A log of decision hierarchies (e.g., *“Attempted System Expansion → Failed → Economic Allocation*)

D. **Data Integrity:** Secondary data must link to unique **UUIDs/DOIs** (check citation hallucination)

E. **Automated audit trails:** Generated mass/energy balance checks and unit consistency flags.

Overall Evaluation Protocol: a tiered approach



CONCLUDING REMARKS

Concluding remarks: Toward Trustworthy AI for LCA

- **Automation to precision:** The field is moving from *automation* with questionable quality to *precision engineering*.
- **Solving "It Depends" problem is crucial:** a choice should be evaluated against machine readable goal & scope context.
- **Scale Eval:** Synthetic Q&As + Property Eval + selective numerical validation would be the first line of defense.
- **Traceability and Documentation:** Unified provenance models and standardized documentation requirement.
- **Human-in-the-loop:** The goal is to making human audits more effective at scale.

“As LCA grows more intelligent and scalable, so must the way we evaluate it.”

Sangwon SUH, PhD

Xinghua Chair Professor
School of Environment, Tsinghua University
TianGong Initiative