AI-assisted LCA Database Development: The TianGong Experience

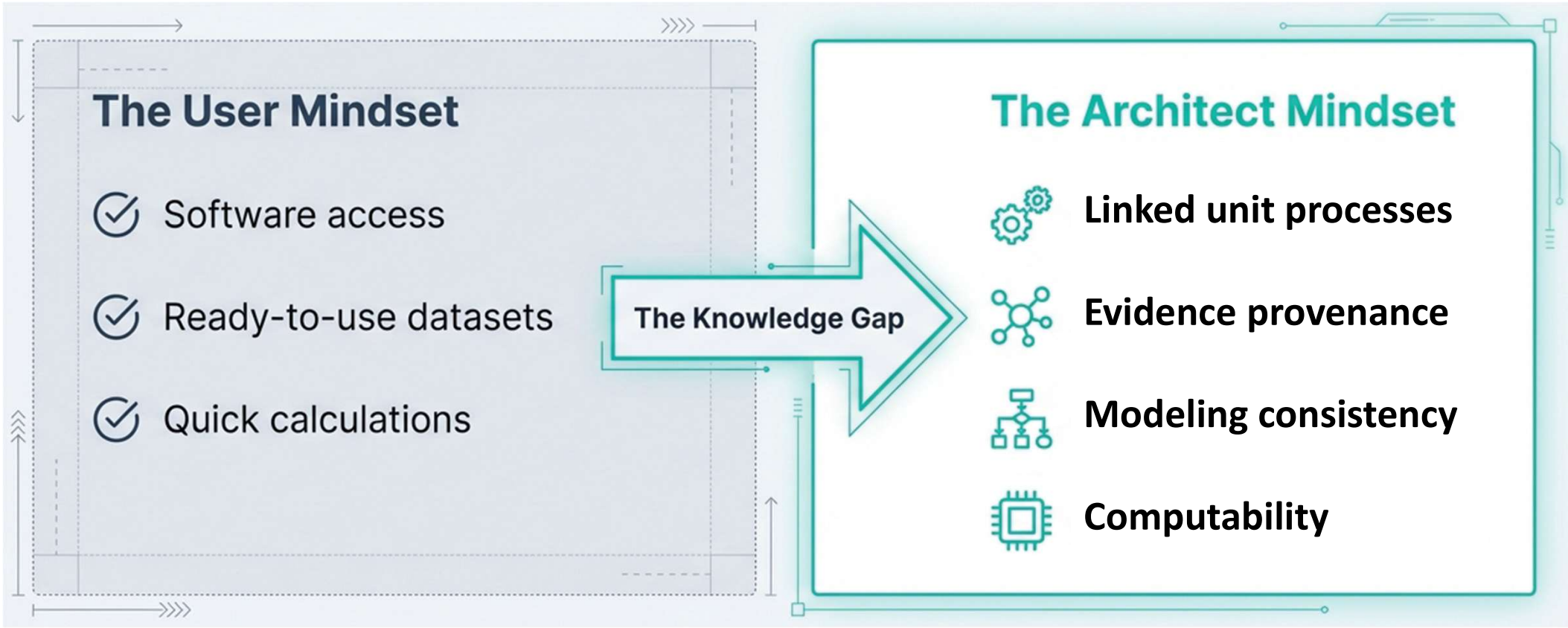
Ming XU
Tsinghua University
xu-ming@tsinghua.edu.cn
https://www.tiangong.earth

碳足迹产业技术创新联盟
Carbon Footprint Alliance

TianGong

Few LCA Database Users Ask What Makes One Good

Using a database is not the same as understanding it



LCA Database ≠ Factor Tables / Excel Summary
LCA Database = Connected Graph of Unit processes + evidence

Inputs/Outputs List (Exchanges)

inputs, outputs, emissions

	electricity/电力	(kWh)
	coal/煤炭	(kg)
	water/水	(m^3)
	transport/运输	(t*km)
CO_2	CO_2/二氧化碳	(kg)
SO_2	SO_2/二氧化硫	(kg)

Not just numbers: each item has unit, flow name, classification

(not just numbers)

Unit Process Dataset

A Core data

reference flow

inputs/outputs matrix

B Metadata

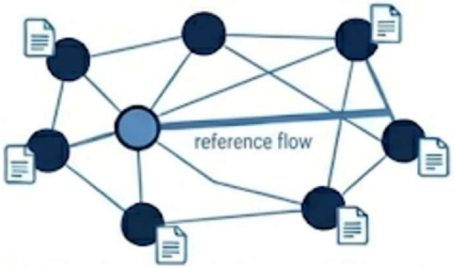
geography, time, technology
uncertainty, representativeness
consistent identifiers, units, classifications

C Evidence documents (Important!)

statistical yearbooks, corporate reports, papers, methodology notes, review records

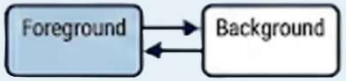
Unit Process Network (connected graph)

unit processes linked by reference flows



Product System Model

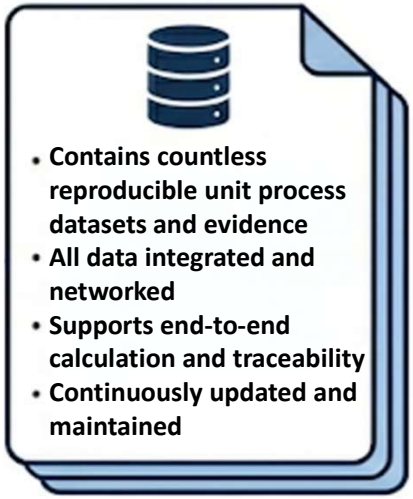
(Product system model)



End-to-end computable

Complete LCA Database

(LCA database)

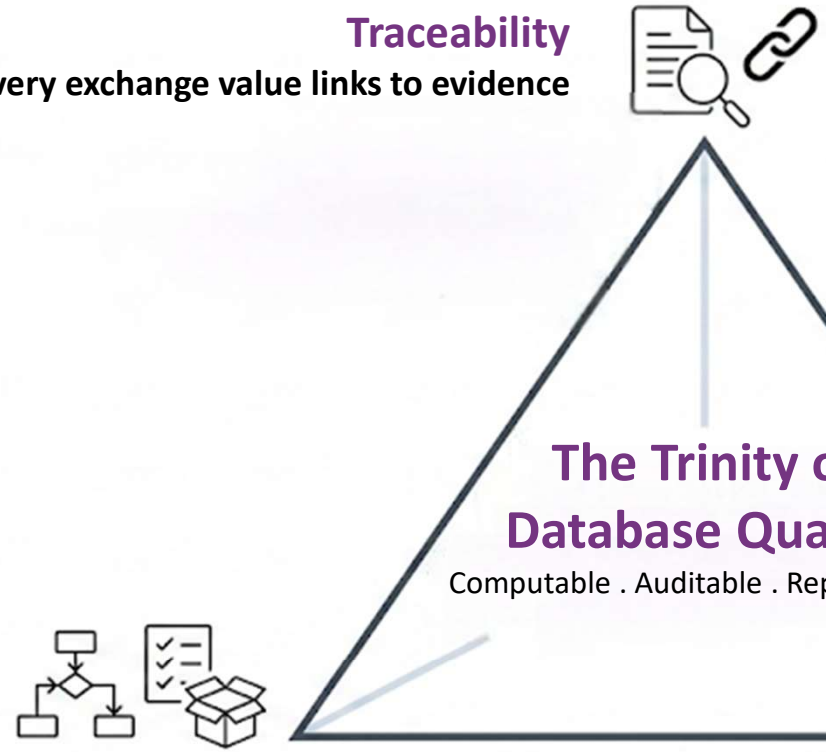


- **Contains countless reproducible unit process datasets and evidence**
- **All data integrated and networked**
- **Supports end-to-end calculation and traceability**
- **Continuously updated and maintained**

Complete LCA Database = A Collection of Countless Reproducible LCA Study Reports

A database is a library of reproducible studies.

What Defines a "High-Quality" LCA Database?



Traceability

Every exchange value links to evidence

The Trinity of Database Quality

Computable . Auditable . Reproducible

Transparency

Methods & decisions are **explicit** and **auditable**

Comprehensiveness

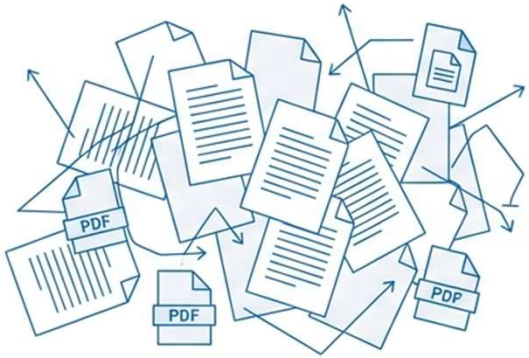
Not **'more data'**, but **completeness**

The network is both connected and sufficiently covered

Nice-to-have： Interoperability / Standard Alignment

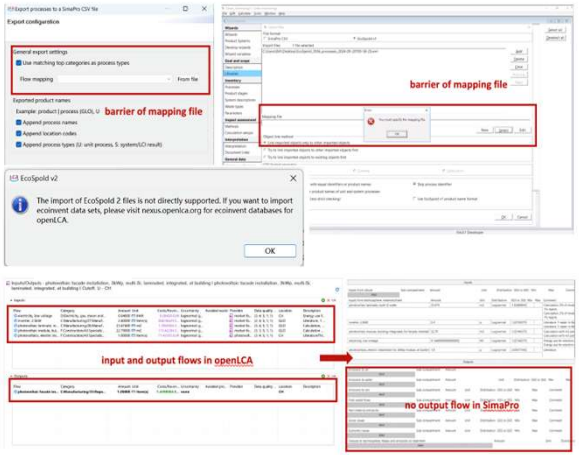
The Status of LCA Data: "Data-Rich but System-Poor"

Fragmentation



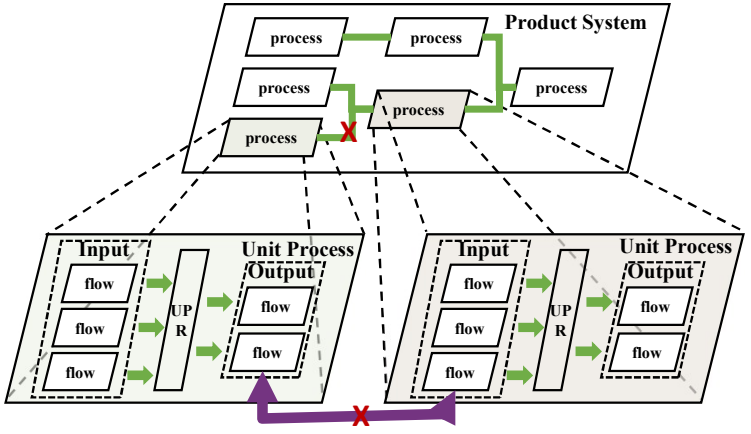
Scattered across 10,000+ sources
- Data silos prevent calculation
- Inconsistent units and boundaries

Segregation



Different data formats
Difficult in data conversion

Inconsistency



Different UUID of elementary flow
Cannot get LCIA result from different database sources

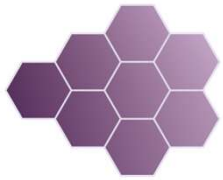
The TianGong Initiative: Addressing LCA data challenges, in China

A community-driven Initiative led by Tsinghua University with 60+ investigators and 200+ researchers to develop and maintain an open, transparent LCA database in China: TianGong Database.

The **Tiangong Kaiwu**, or The Exploitation of the Works of Nature, published in 1637, is an encyclopedia describing the **agricultural and manufacturing processes** in China.

https://www.tiangong.earth

University-Industry Partnership: Carbon Footprint Alliance (CFA)

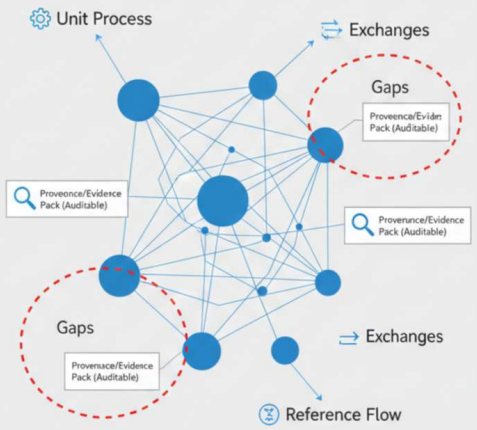


https://www.carbonfootprint.network/

The TianGong Strategic Roadmap: Three Steps

Step 1

Build the Computable Foundation



computable baseline graph + evidence

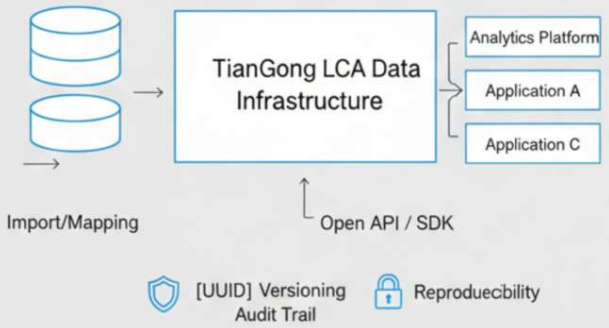
Step 2

Scale with Production Toolchain

scalable pipeline + quality gates + review workflow

Step 3

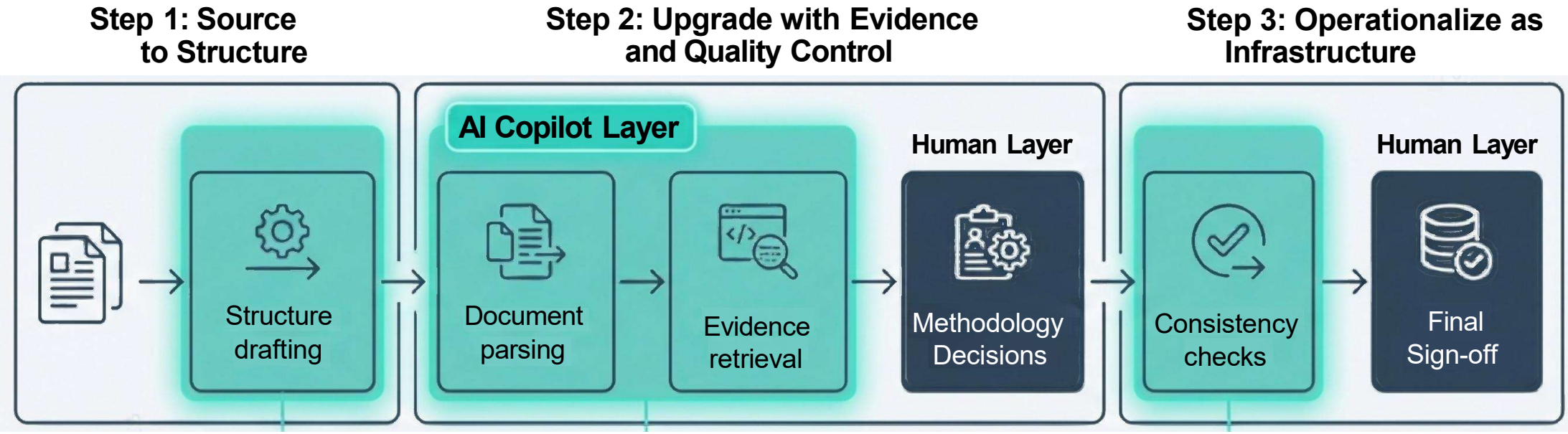
Operate as Infrastructure



governance + interoperability layer

Connectivity → Quality → Sustainability

Where AI Fits in the Workflow

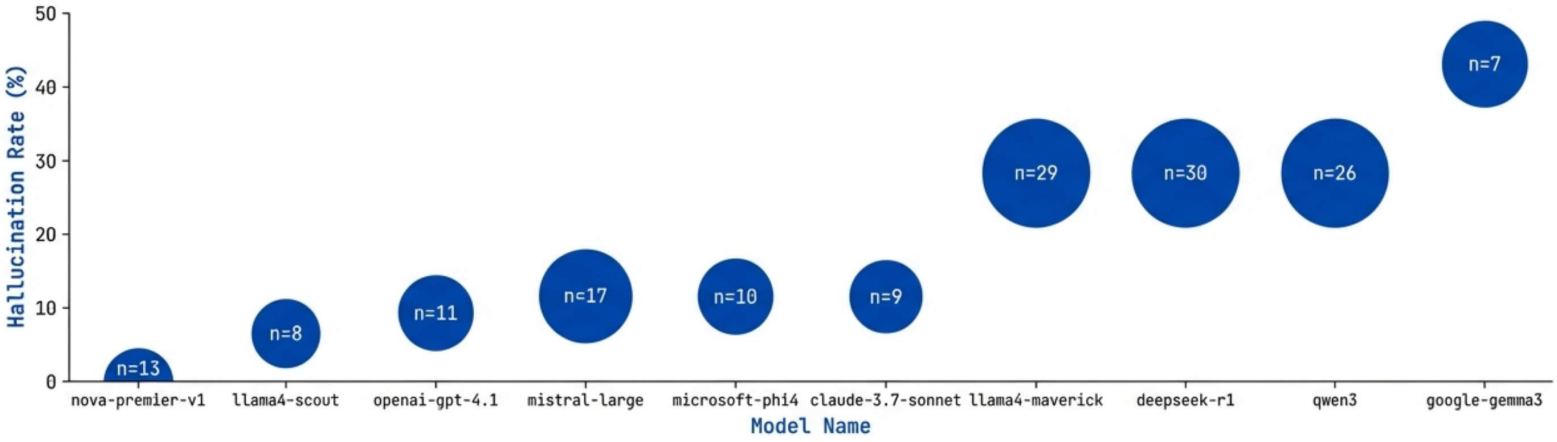


AI is embedded as a copilot layer within a controlled workflow, not as a substitute for methodological judgment.

Bad context creates bad answers—from AI and from humans

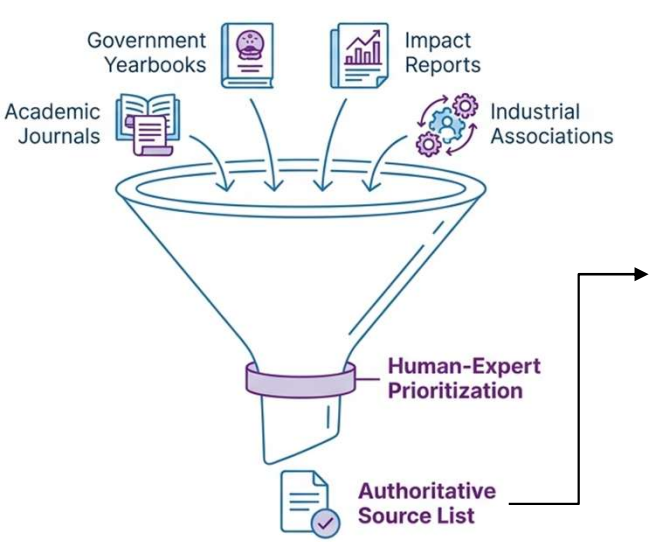
LCA is not binary. Decisions on Allocation (Economic vs Mass) and System Boundaries are subjective and context-dependent.

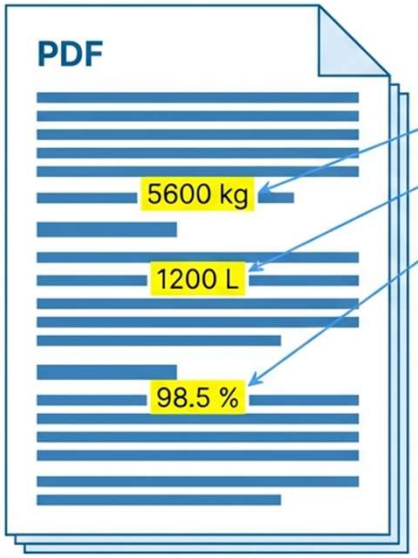
When context is weak, both AI and humans can fabricate. The issue is not AI itself—it's how we use it.



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

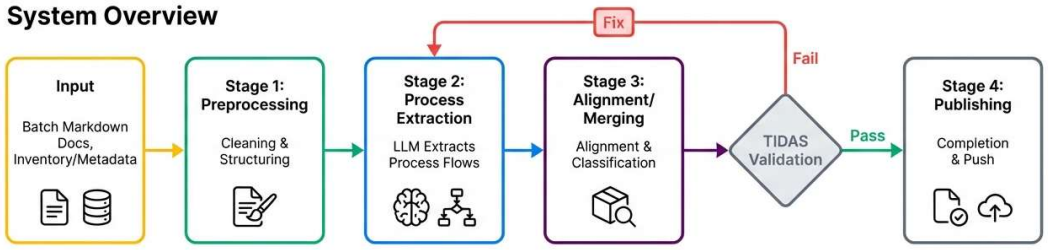
Sourcing and AI-Assisted Extraction

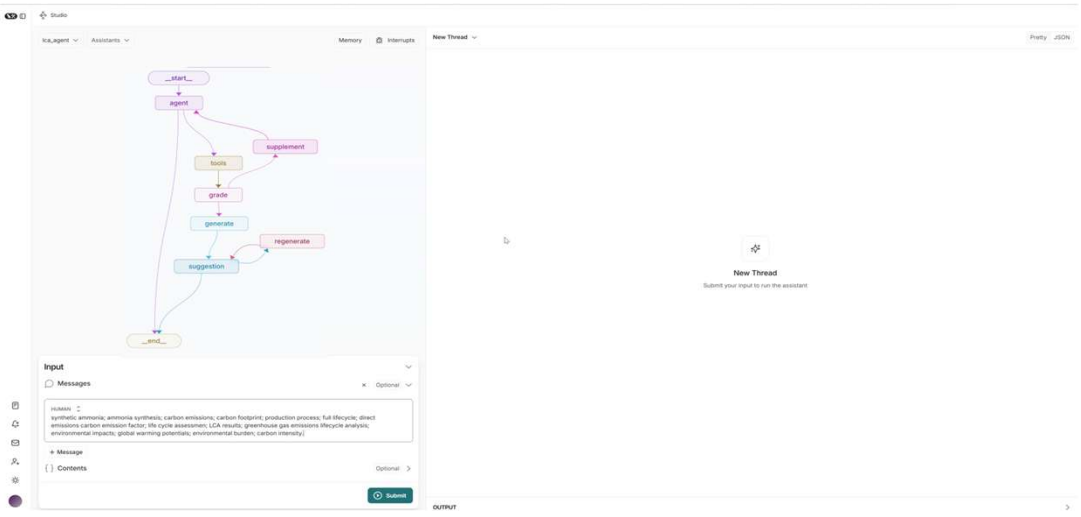




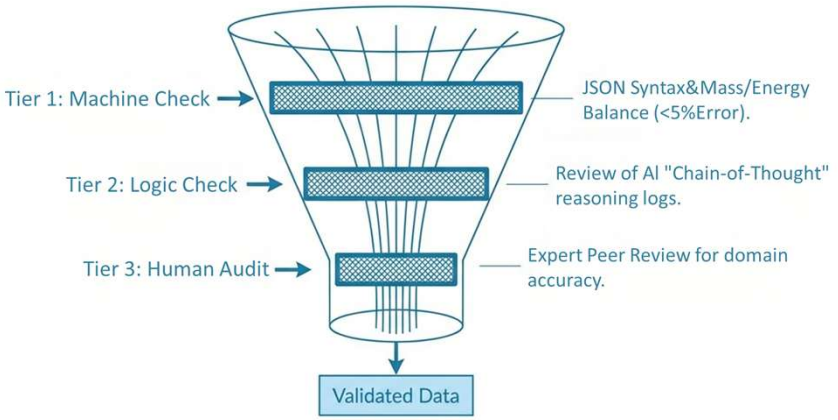
Flow Name	Value	Unit	Source UUID
Water Consumption	5600	kg	PDF-DOC-A1
Fuel Usage	1200	L	PDF-DOC-A2
Efficiency Rate	98.5	%	PDF-DOC-A3

Locating, NOT Generating.

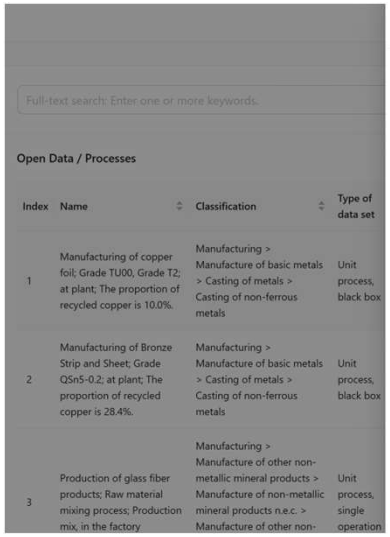


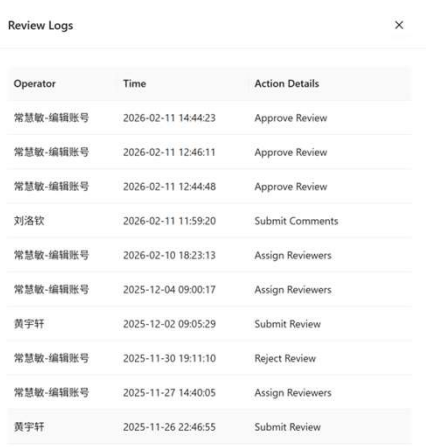


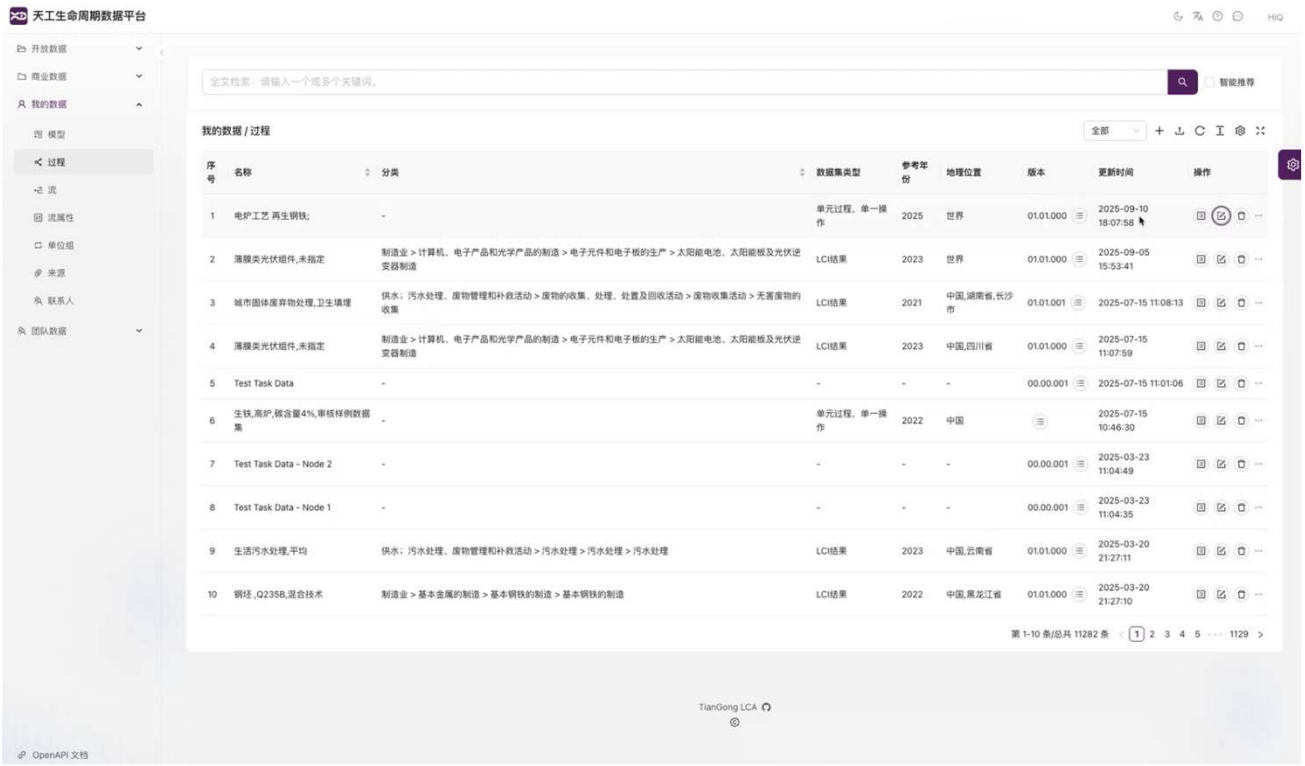
Multi-Tier Validation Protocol



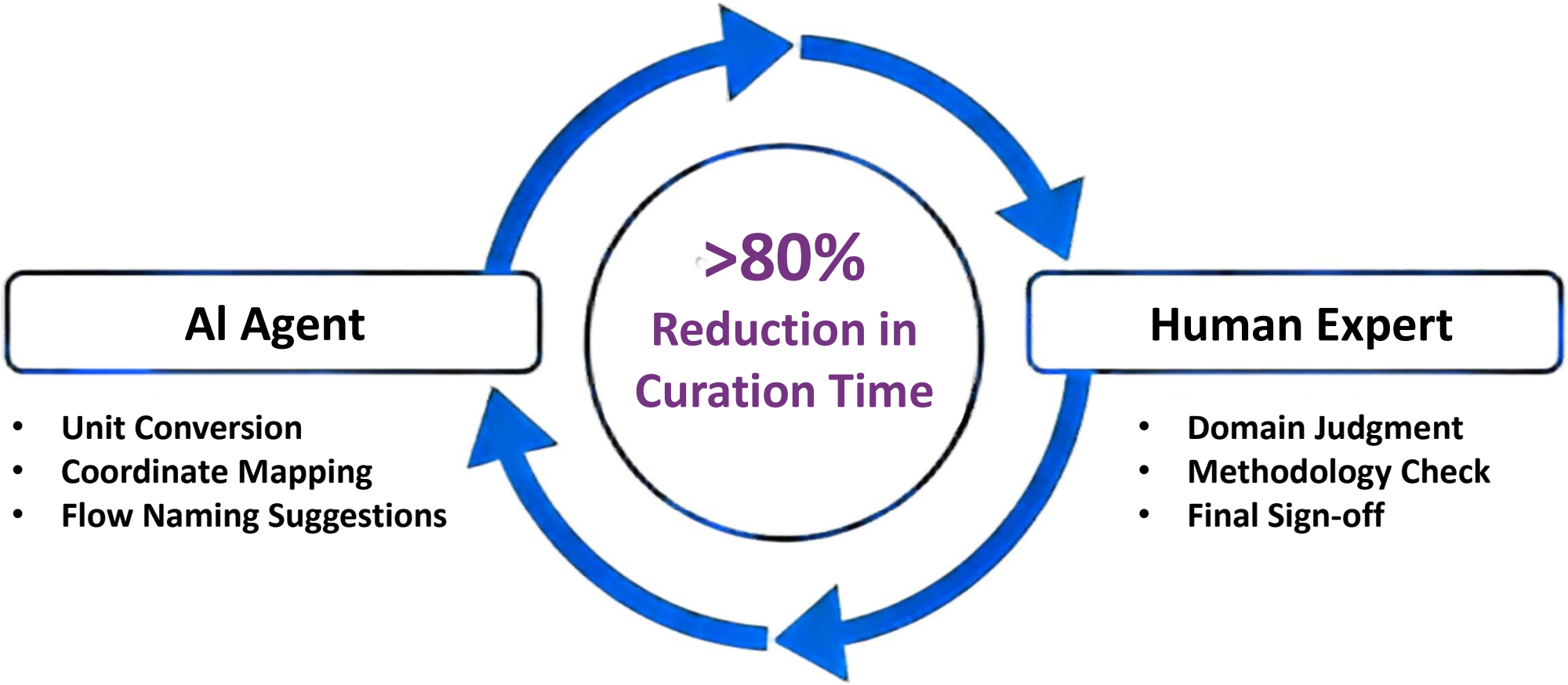
➢ **Smart validation and automatic correction of integrated data content significantly enhance the accuracy, completeness, and consistency of data content.**



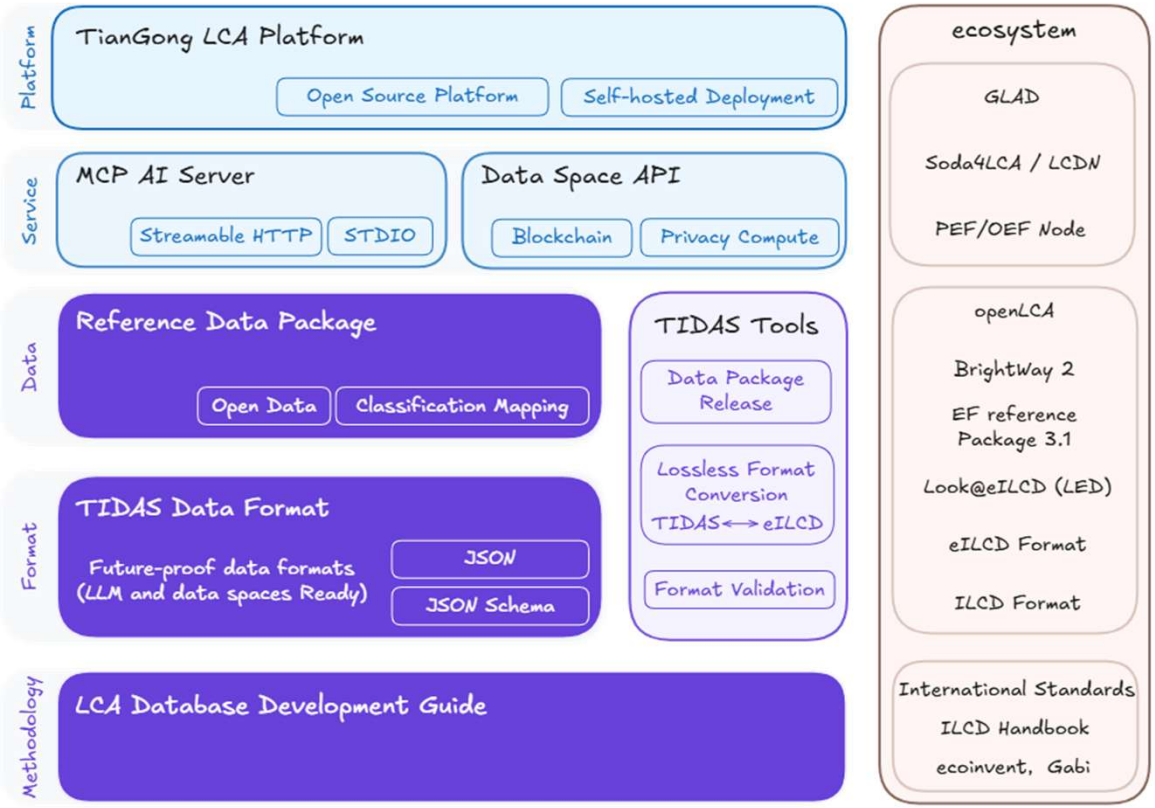




Human-in-the-Loop Collaboration



TianGong Data System - TIDAS



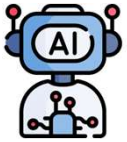
- **Open Source**
  - Globally open-source, free of use
- **Ready to Use**
  - Data, toolkits, and platform, available out of the box
- **International Compatibility**
  - Lossless conversion and interoperability with international data systems such as ILCD

https://tidas.tiangong.earth

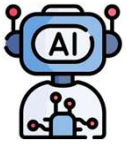
An AI Agent helping us for everything

INPUT
(User Command)

Hi, Jarvis, find input and output data for xxx unit process.



AI PROCESSING
(Agent)

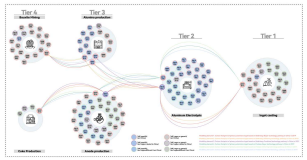


Search Data

Bill of materials

Suppliers

Modelling



Processing...

95% Complete

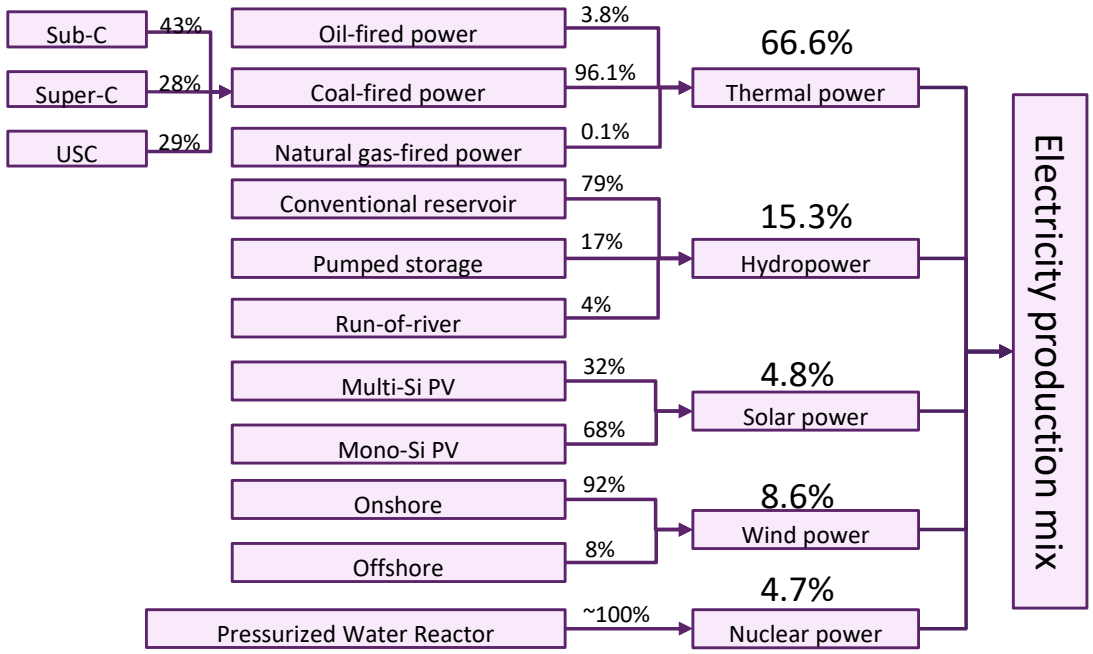
OUTPUT
(Instant Result)

AI as curator: retrieval + structuring + checks, not inventing inventories.

Example:
TianGong Results vs EF Node Chinese Electricity dataset (GaBi)

Electricity grid mix, consumption mix (TianGong)

Year 2022



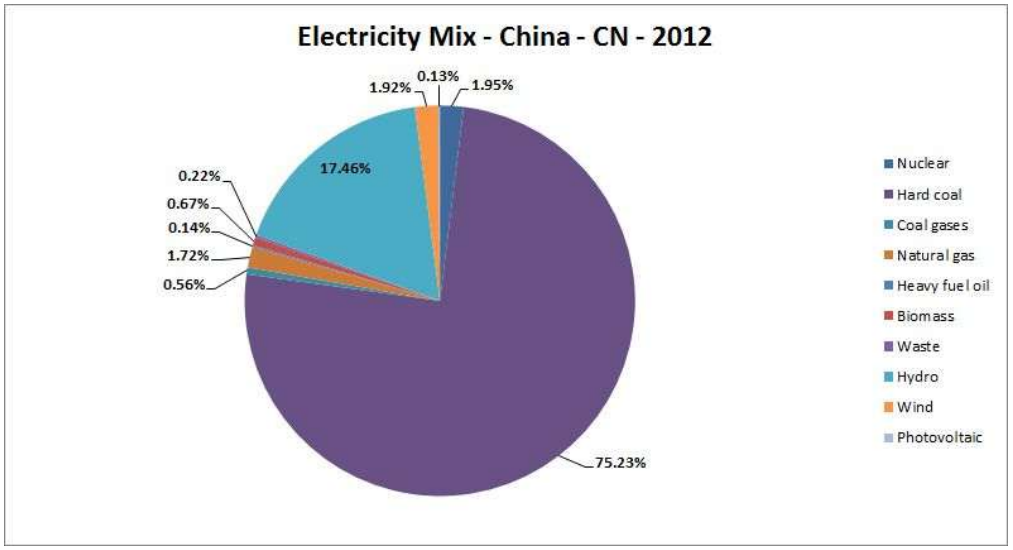
35-330 KV electricity consumption mix, 0.6294 kg CO2/kwh

1-35 KV electricity consumption mix, 0.6314 kg CO2/kwh

<1 KV electricity consumption mix, 0.6654 kg CO2/kwh

Sphera LCDN Node

Year 2012



1-60kV electricity consumption mix, 0.9419 kg CO2/kwh

Thanks

Ming XU
Tsinghua University
xu-ming@tsinghua.edu.cn
https://www.tiangong.earth

碳足迹产业技术创新联盟
Carbon Footprint Alliance

TianGong