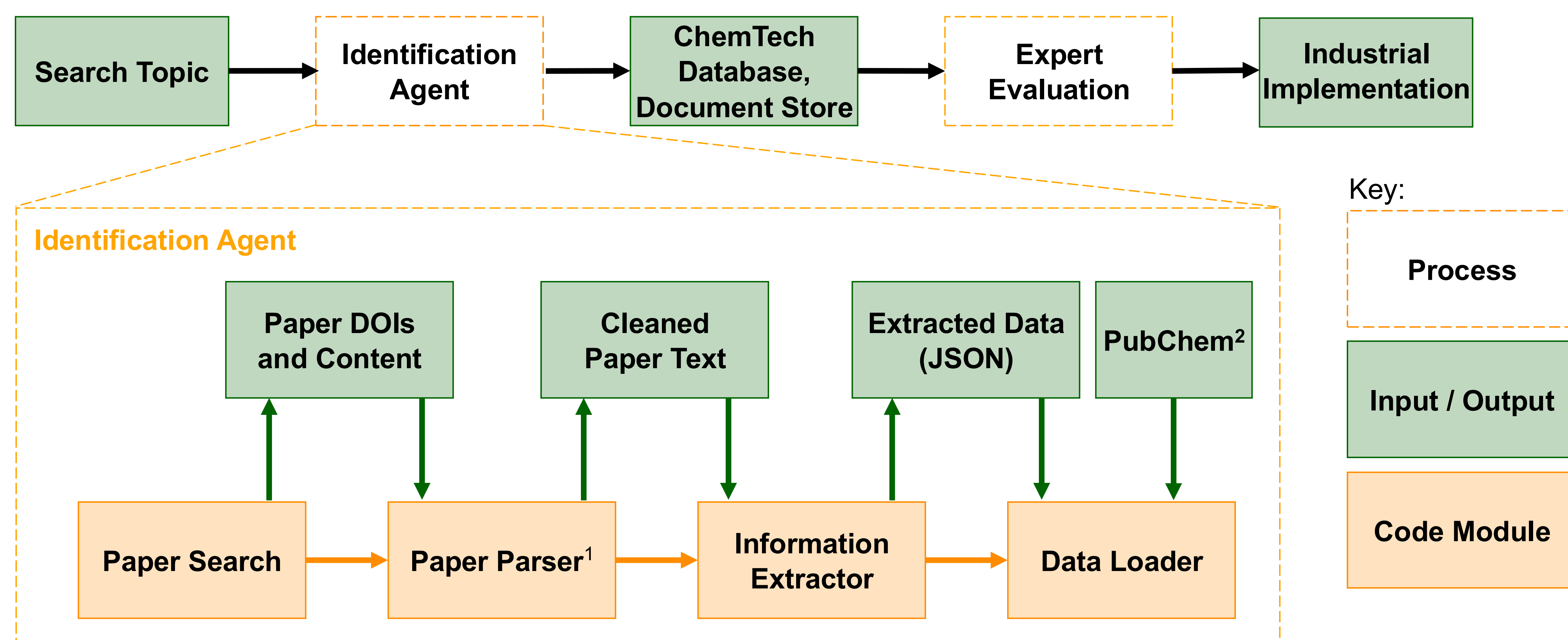


The Innovation – Implementation Gap in Chemical Manufacturing

Novel technologies are constantly being developed across the chemical industry. However, advances are reported across a diverse spread of speciality literature, hindering the conversion of technical advances into actionable insights. **We present a tool to support the evaluation of new technologies in chemical manufacturing, with the goal of informing procurement and manufacturing choices.**

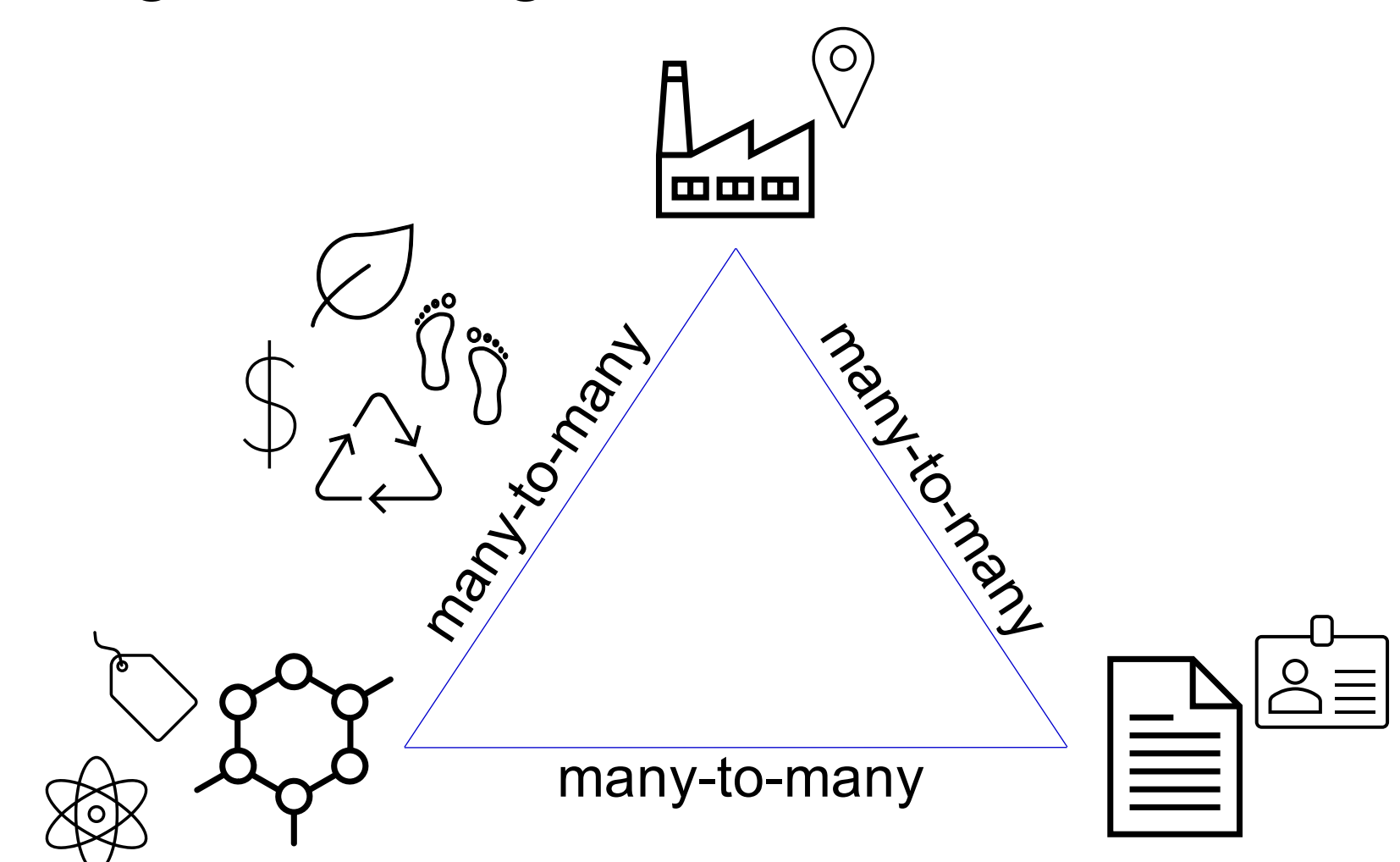


Data Extraction Methodology

- **LLM-powered extraction** of literature to identify novel chemicals and technologies
- Extracts conversion processes annotated with **environmental, chemical, and process data**
- Decoder-only LLM, locally hosted, restricted to structured outputs
- **PubChem²** integration for chemical identifiers
- Structured data saved to ChemTech database; unstructured to document store
- **Future work:** extracted data to be used for evaluation of novel technologies under uncertainty

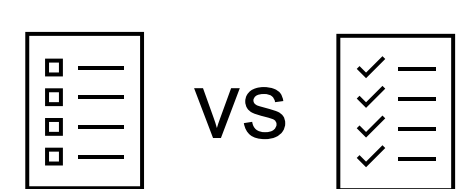
The Result: ChemTech Database

- **PostgreSQL** database storing extracted data on chemical manufacturing technologies in a structured format



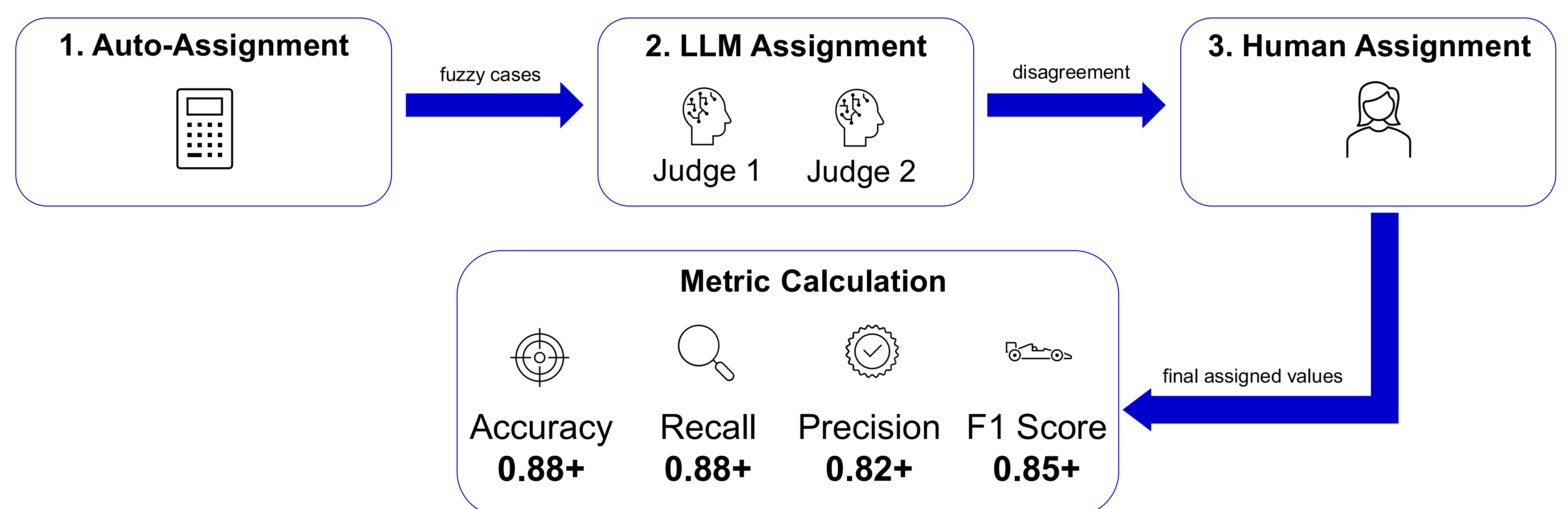
A Hierarchical Framework for the Evaluation of Extracted Data

- Extracted data are compared against a **manually-curated ground truth** and assigned as TP/TN/FP/FN



- Assignment Key:

Ground Truth	Extracted Value	Assignment
A	A	True Positive
null	null	True Negative
A	null	False Negative
null	A	False Positive
A	B	False Positive (Fuzzy)



References

1. Kang, Y. et al. Harnessing Large Language Model to collect and analyze metal-organic framework property dataset. *J. Am. Chem. Soc.* **2025**, *147*(5), 3943-3958. 2. Kim, S. et al. An update on PUG-REST: RESTful interface for programmatic access to PubChem. *Nucleic Acids Res.* **2018**, *46*(W1), W563-W570. doi:10.1093/nar/gky294