香港中文大學
The Chinese University of Hong Kong

Journal Club:
A review of large language models and autonomous agents in chemistry

Huiwen Tan

18 Apr 2025

Cite this: *Chem. Sci.*, 2025, **16**, 2514

A review of large language models and autonomous agents in chemistry

Mayk Caldas Ramos, [ab] Christopher J. Collison [c] and Andrew D. White *[ab]

Large language models (LLMs) have emerged as powerful tools in chemistry, significantly impacting molecule design, property prediction, and synthesis optimization. This review highlights LLM capabilities in these domains and their potential to accelerate scientific discovery through automation. We also review LLM-based autonomous agents: LLMs with a broader set of tools to interact with their surrounding environment. These agents perform diverse tasks such as paper scraping, interfacing with automated laboratories, and synthesis planning. As agents are an emerging topic, we extend the scope of our review of agents beyond chemistry and discuss across any scientific domains. This review covers the recent history, current capabilities, and design of LLMs and autonomous agents, addressing specific challenges, opportunities, and future directions in chemistry. Key challenges include data quality and integration, model interpretability, and the need for standard benchmarks, while future directions point towards more sophisticated multi-modal agents and enhanced collaboration between agents and experimental methods. Due to the quick pace of this field, a repository has been built to keep track of the latest studies: https://github.com/ur-whitelab/LLMs-in-science.

Received 15th June 2024
Accepted 3rd December 2024

DOI: 10.1039/d4sc03921a

rsc.li/chemical-science

LLM vs. Traditional ML

1. **Traditional Machine Learning (ML)**
   - Support Vector Machines (SVM)
   - Random Forests (RF)
2. **Neural Network Architectures in Deep Learning**
   - Recurrent Neural Networks (RNN) – for sequential data (e.g., text, reaction steps)
   - Convolutional Neural Networks (CNN) – for grid-like data (e.g., spectra, electron density maps)
   - Graph Neural Networks (GNN) – for molecular graphs (atoms as nodes, bonds as edges)
   - Transformers – for long-range sequence modeling (e.g., SMILES, protein sequences)

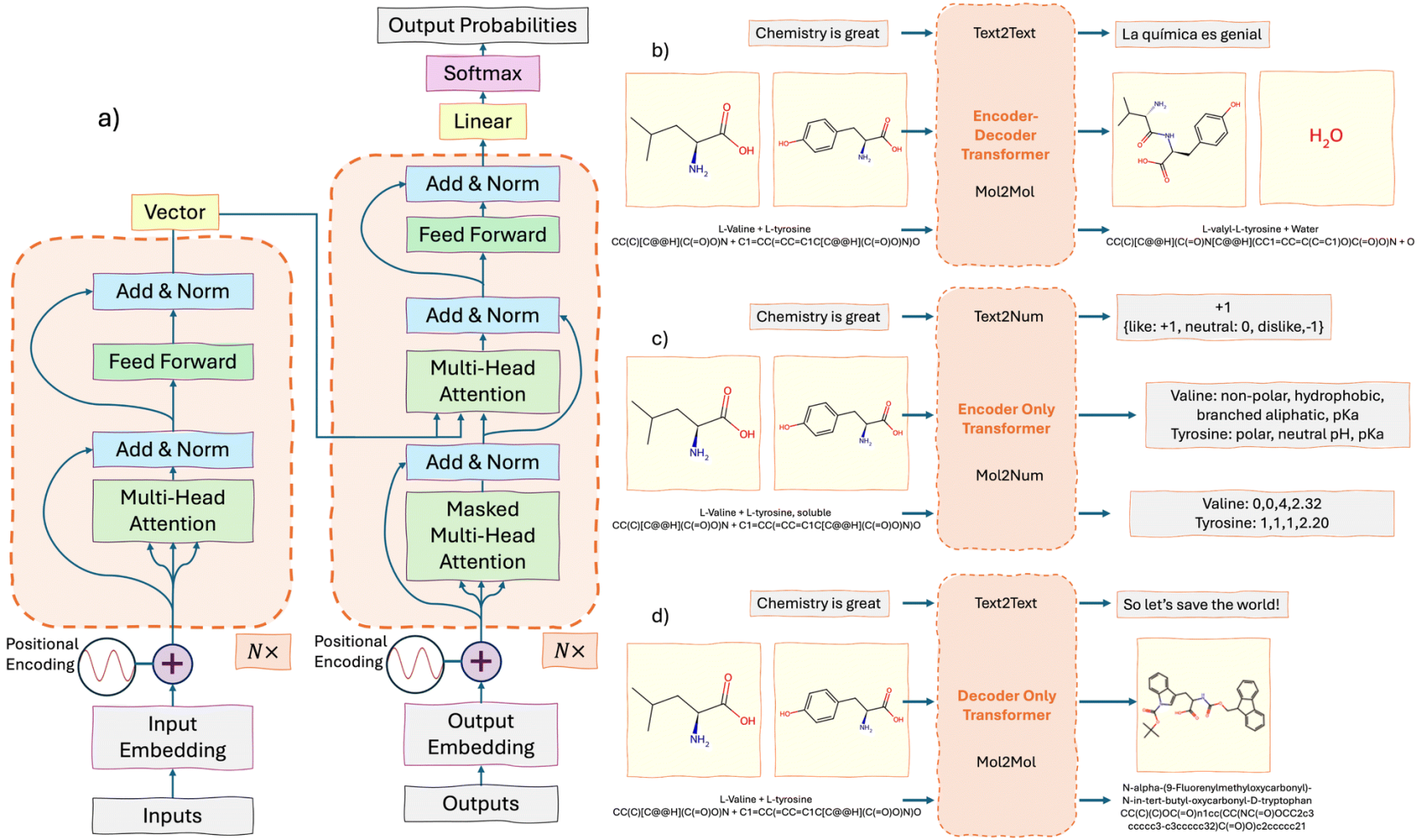
LLM vs. Traditional ML

1. **Traditional Machine Learning (ML)**
   - Support Vector Machines (SVM)
   - Random Forests (RF)
2. **Neural Network Architectures in Deep Learning**
   - Recurrent Neural Networks (RNN) – for sequential data (e.g., text, reaction steps)
   - Convolutional Neural Networks (CNN) – for grid-like data (e.g., spectra, electron density maps)
   - Graph Neural Networks (GNN) – for molecular graphs (atoms as nodes, bonds as edges)
   - Transformers – for long-range sequence modeling (e.g., SMILES, protein sequences)

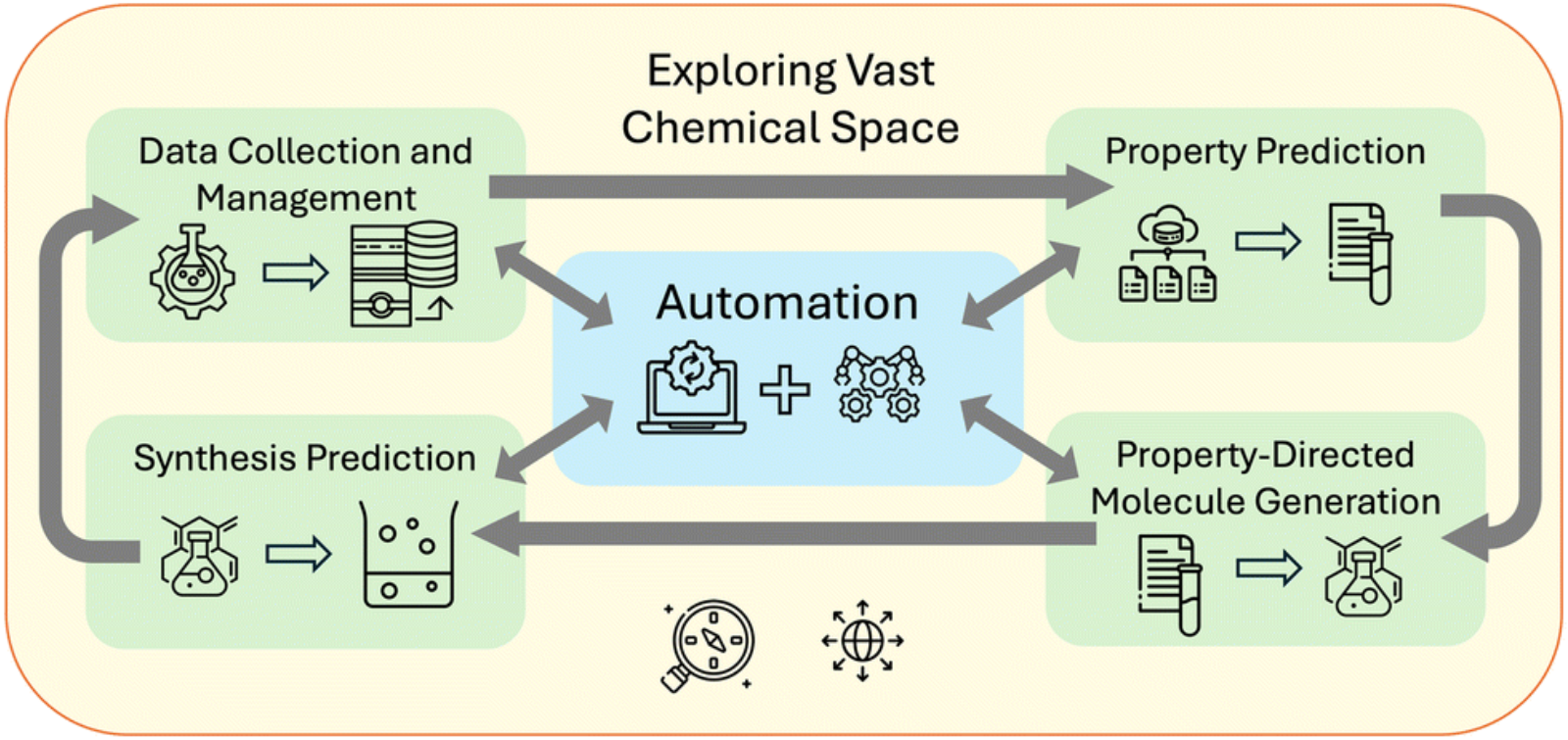
Large Language Models (LLMs)

LLM Types

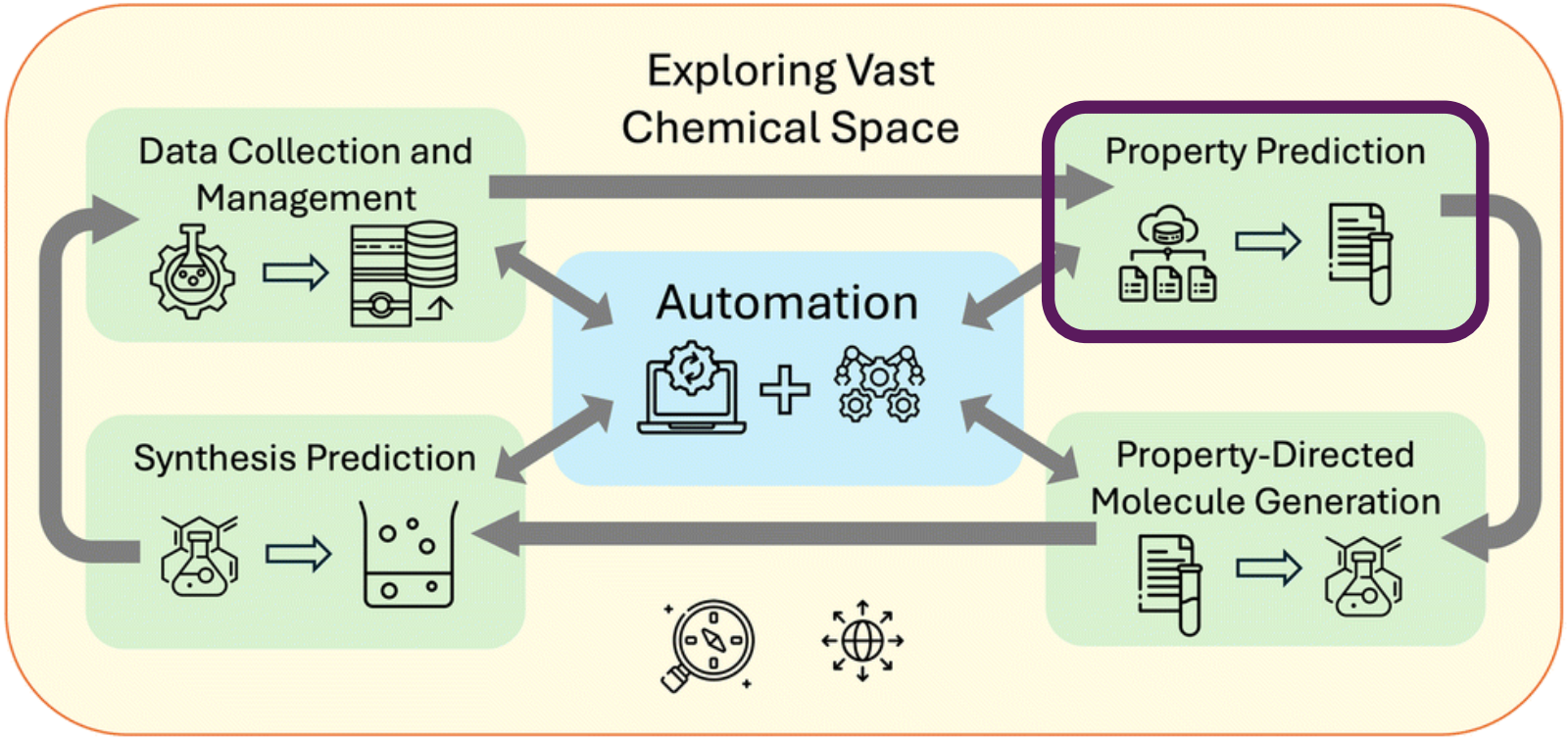
1. **Encoder–decoder models**
   - Bidirectional and Auto-Regressive Transformers (BART)
2. **Encoder-only models**
   - Bidirectional Encoder Representations from Transformers (BERT)
3. **Decoder-only models**
   - Generative Pretrained Transformer (GPT)
4. **Multi-task and multi-modal models**
   - Text-to-Text Transfer Transformer (T5)
   - NExT-GPT



LLM in Chemistry

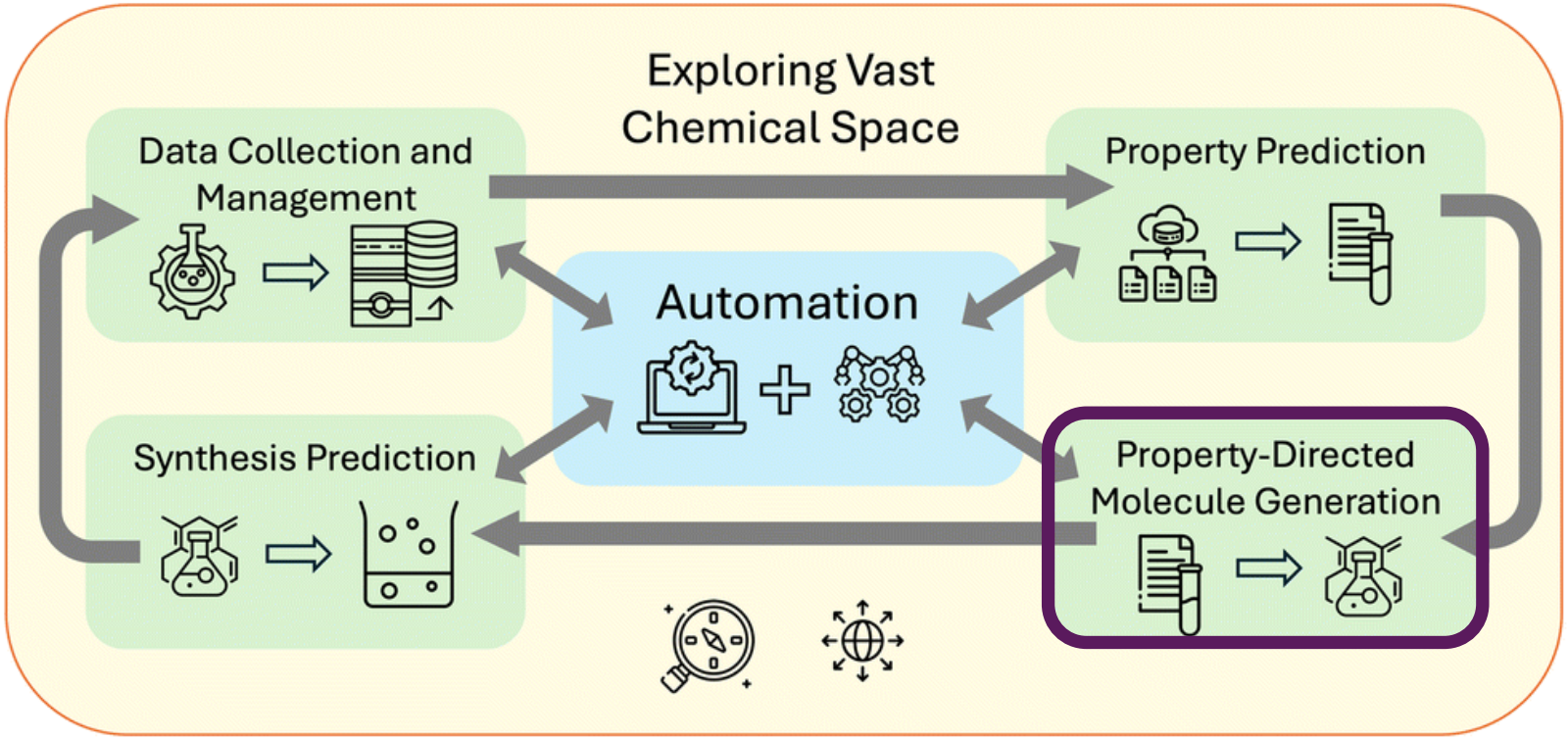


LLM in Chemistry



Encoder-only

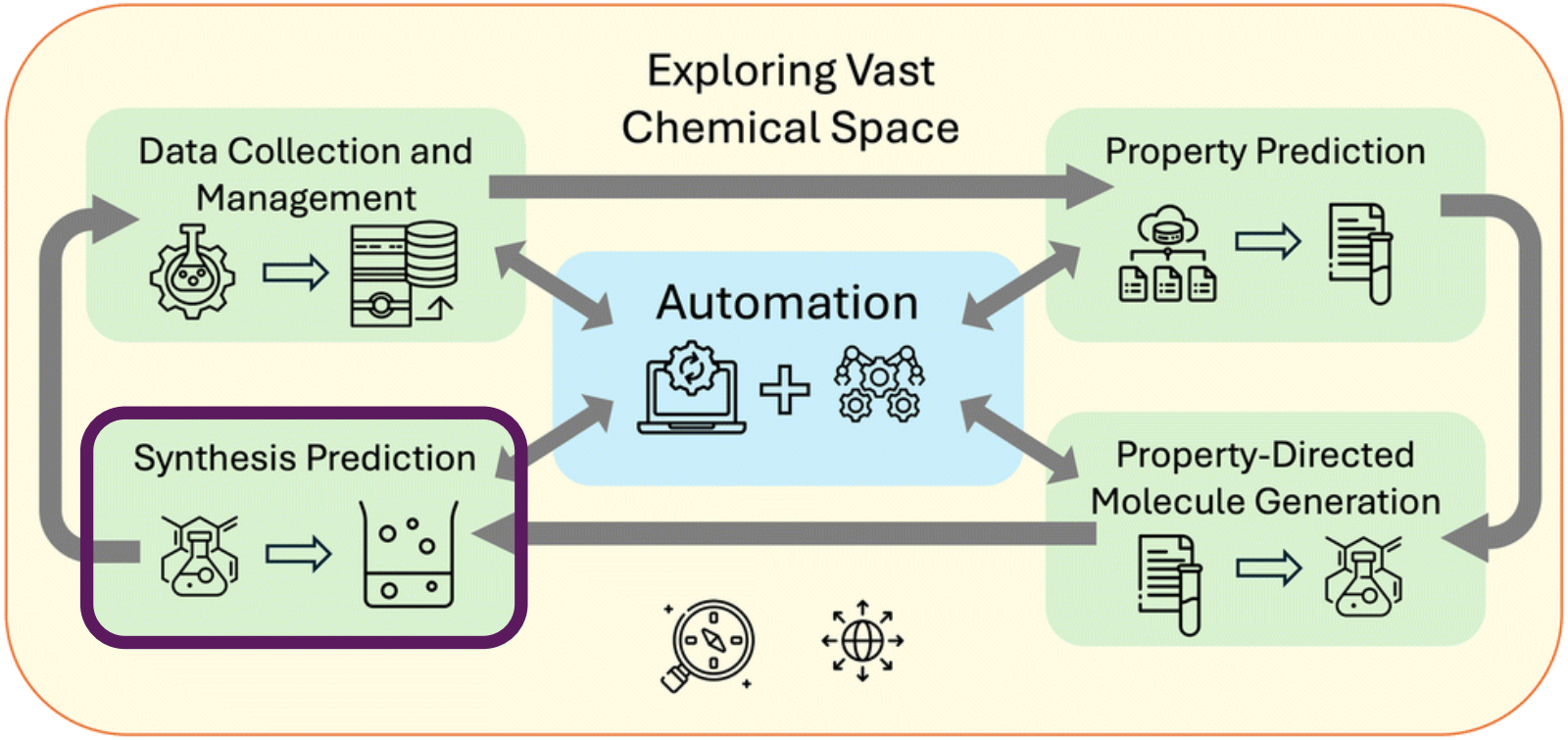
LLM in Chemistry



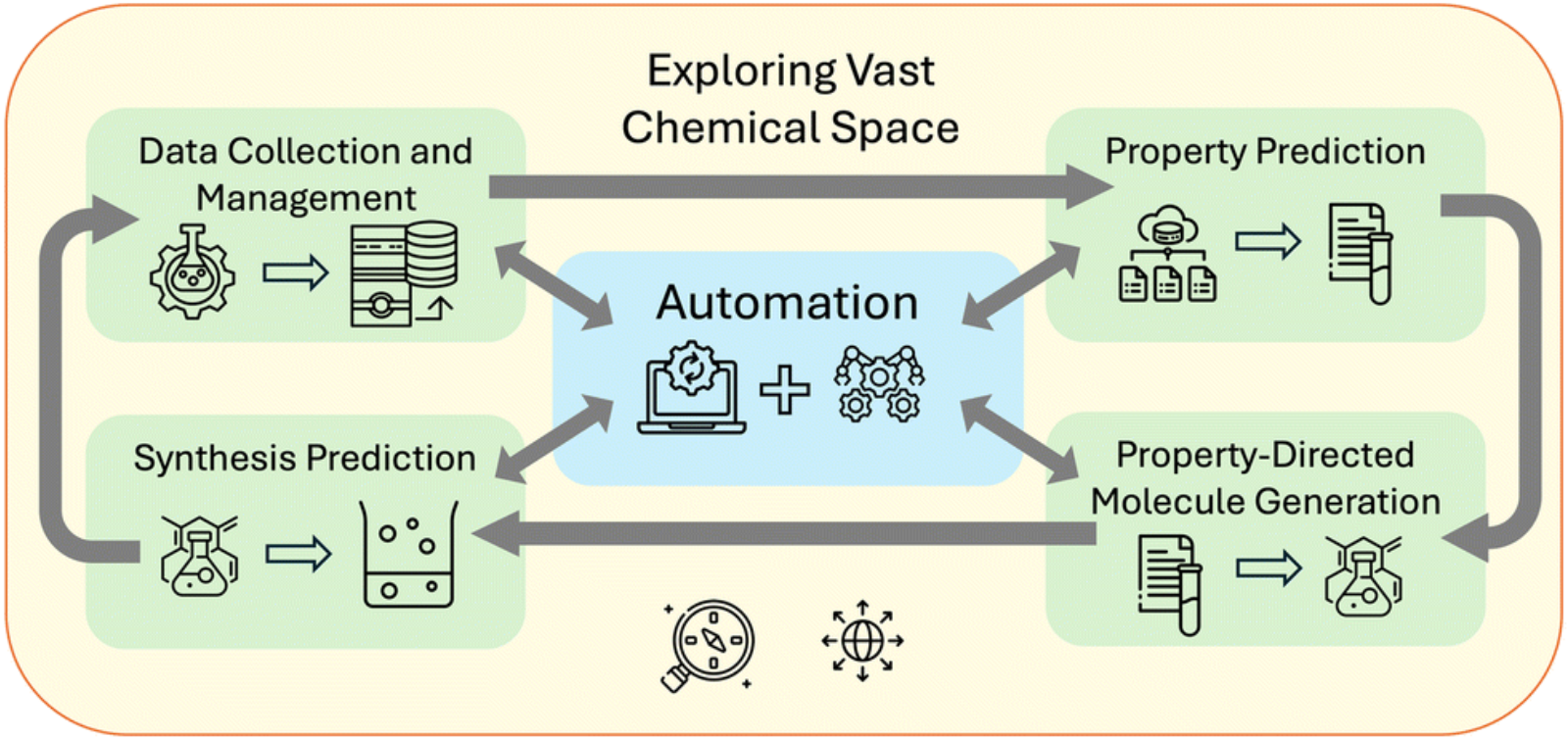
Decoder-only

LLM in Chemistry

Encoder–decoder



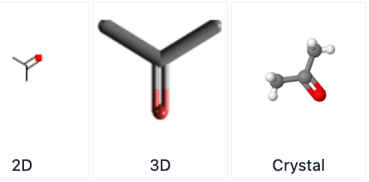
LLM in Chemistry



Mol-LLM: Molecular Representation Based on Strings

PubChem Acetone (Compound)

Acetone

PubChem CID	180
Structure	 2D 3D Crystal
Chemical Safety	 Flammable Irritant Laboratory Chemical Safety Summary (LCSS) Datasheet
Molecular Formula	C_3H_6O CH_3-CO-CH_3
Synonyms	acetone 2-propanone 67-64-1 propanone Dimethyl ketone View More...
Molecular Weight	58.08 g/mol *Computed by PubChem 2.2 (PubChem release 2021.10.14)*

PubChem Acetone (Compound)

2.1 Computed Descriptors

2.1.1 IUPAC Name

propan-2-one

Computed by Lexichem TK 2.7.0 (PubChem release 2021.10.14)

PubChem

2.1.2 InChI

InChI=1S/C3H6O/c1-3(2)4/h1-2H3

Computed by InChI 1.0.6 (PubChem release 2021.10.14)

PubChem

2.1.3 InChIKey

CSCPPACGZOOCGX-UHFFFAOYSA-N

Computed by InChI 1.0.6 (PubChem release 2021.10.14)

PubChem

Simplified Molecular Input Line Entry System

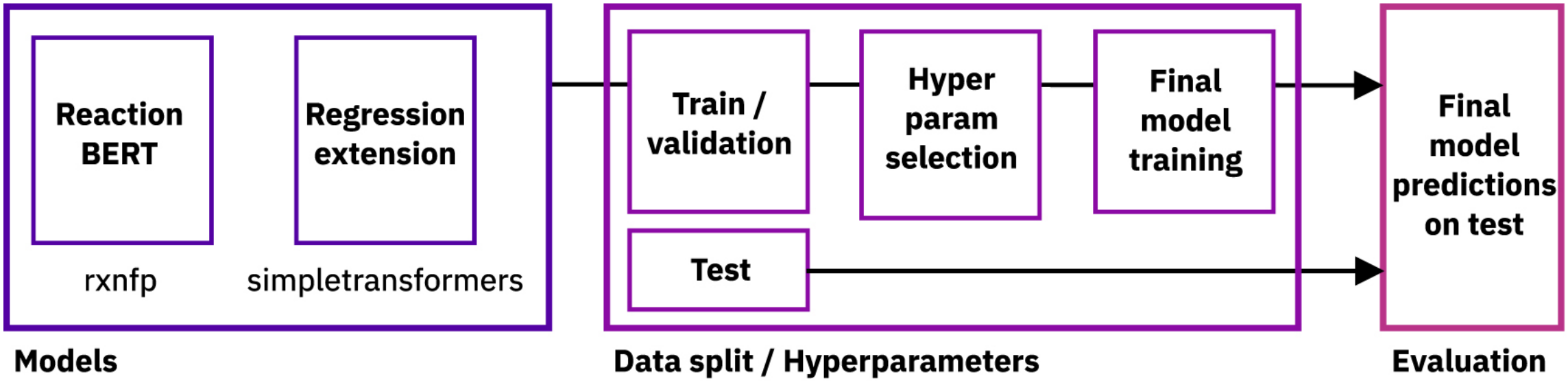
2.1.4 SMILES

CC(=O)C

Computed by OEChem 2.3.0 (PubChem release 2024.12.12)

PubChem

Predicts Reaction Yields Starting from Reaction SMILES



Brc1ccccn1.CCOC(=O)c1cc(C)no1.CN1CCCN2CCCN=C12.COc1ccc(OC)c(P(C(C)(C)C)C(C)(C)C)c1-c1c(C(C)C)cc(C(C)C)cc1C(C)C.Cc1ccc(N)cc1.O=S(=O)(O[Pd]1c2ccccc2-c2ccccc2[NH2]1)C(F)(F)F>>Cc1ccc(Nc2ccccn2)cc1

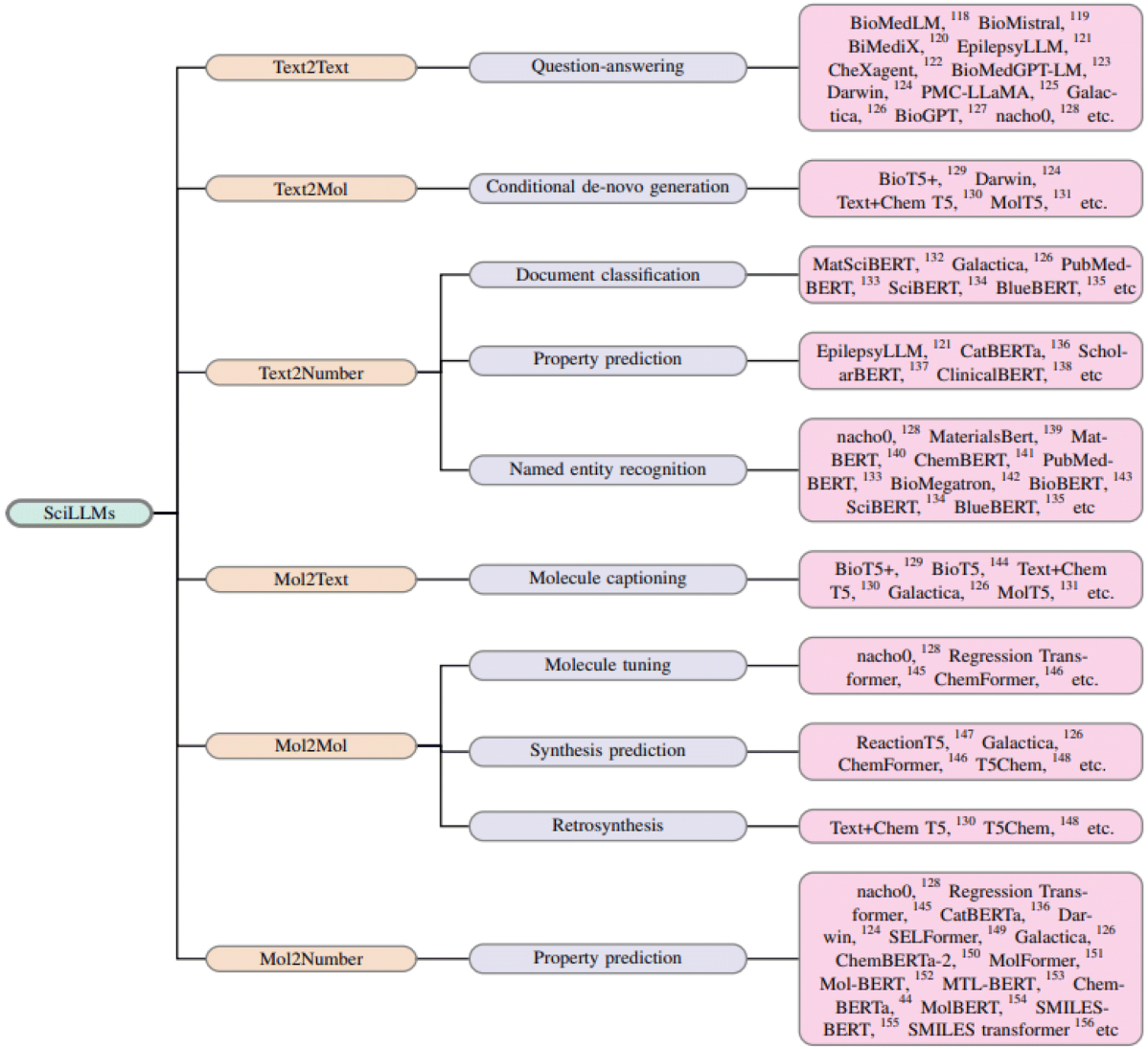
Reaction SMILES

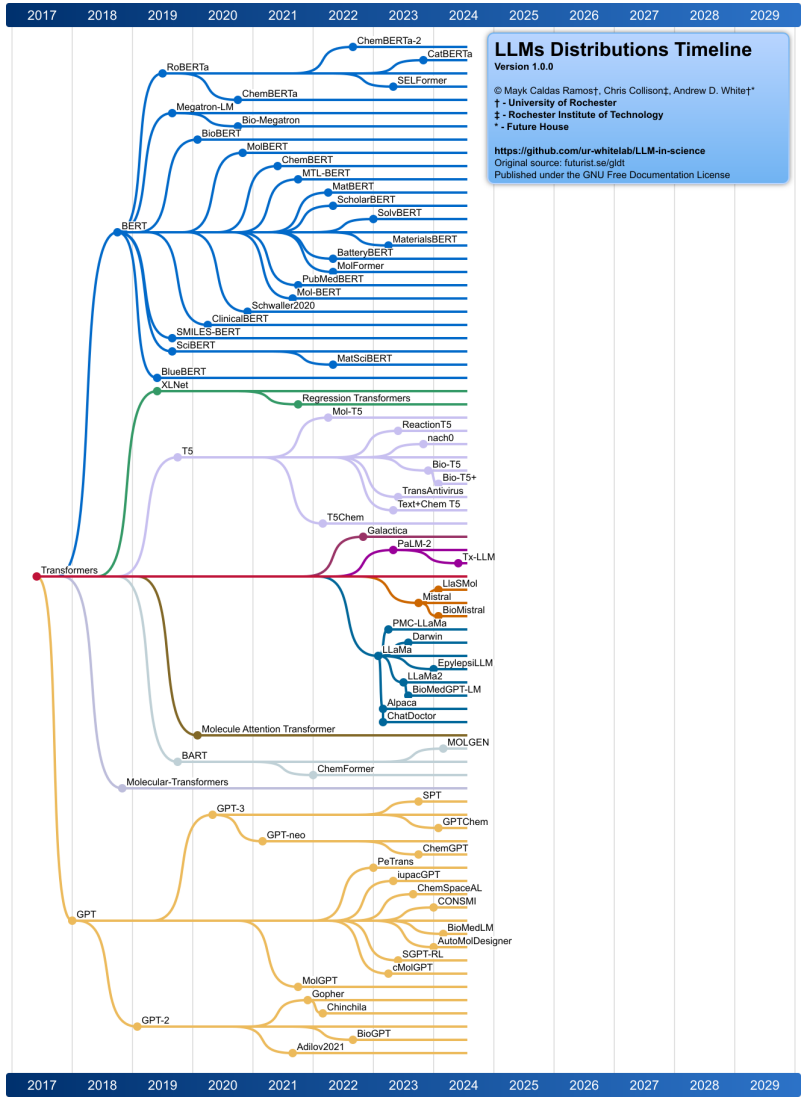
Yield-BERT →

96.6%

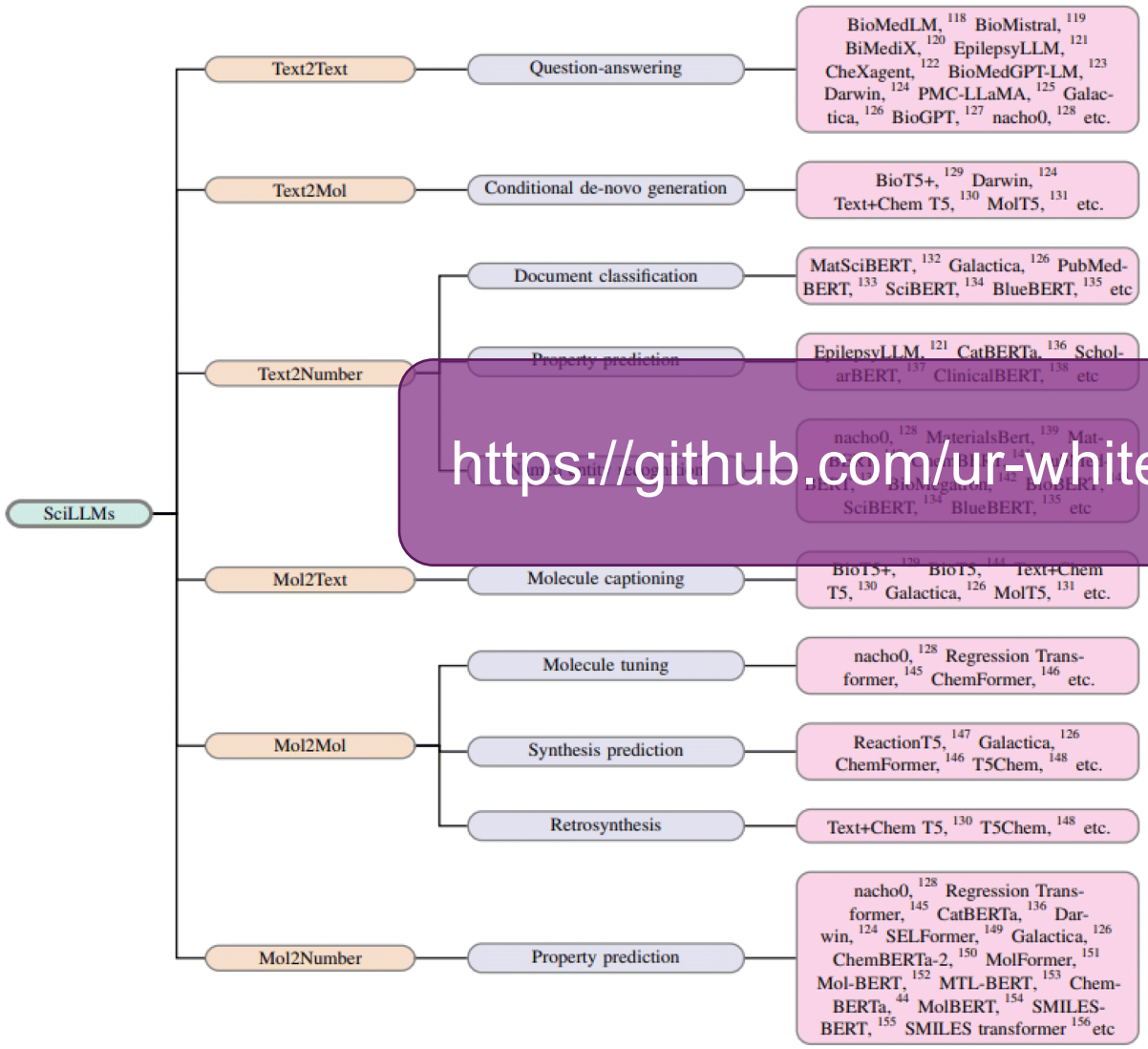
Reaction yield

The Capabilities and Development of LLMs

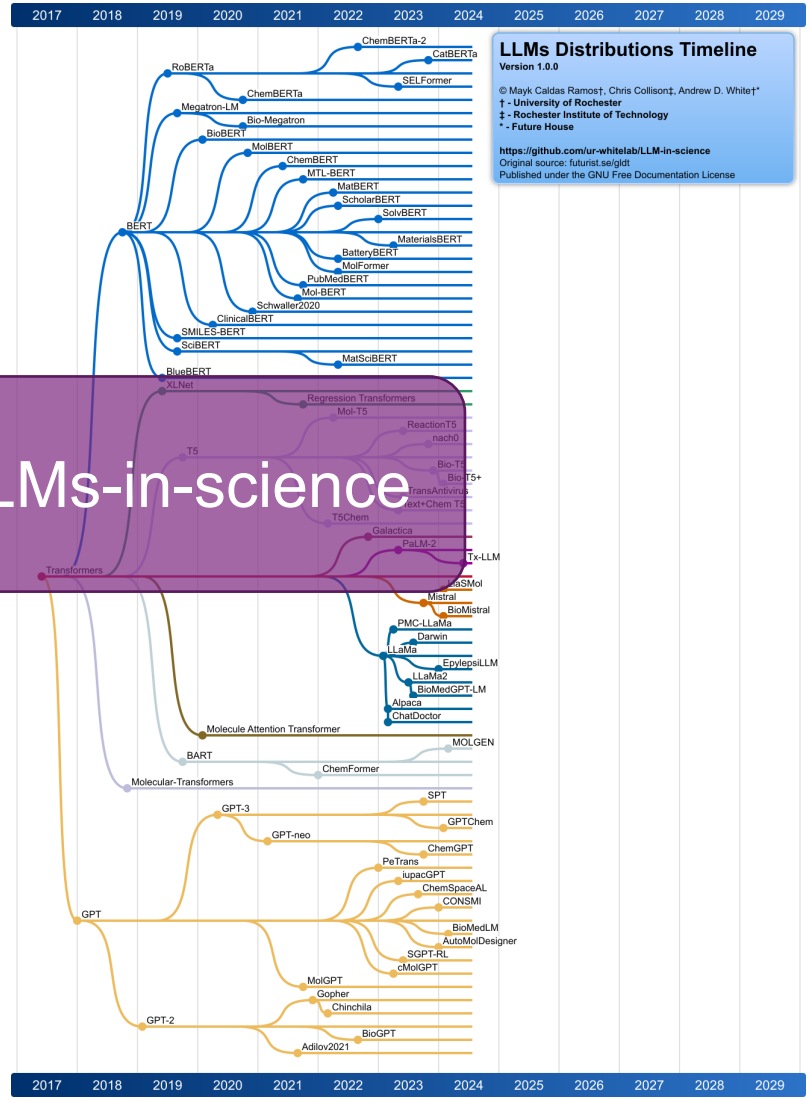




The Capabilities and Development of LLMs

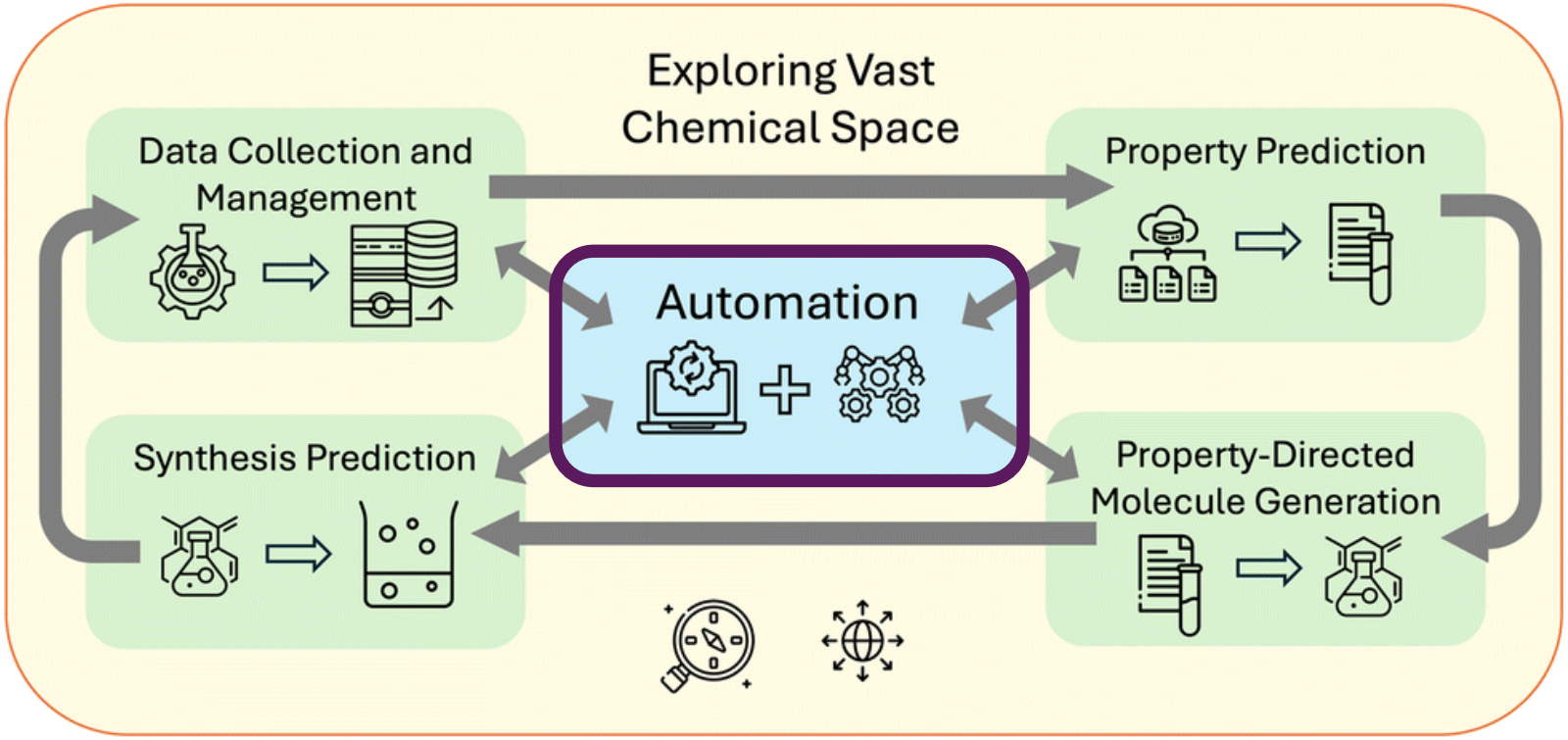


https://github.com/ur-whitelab/LLMs-in-science

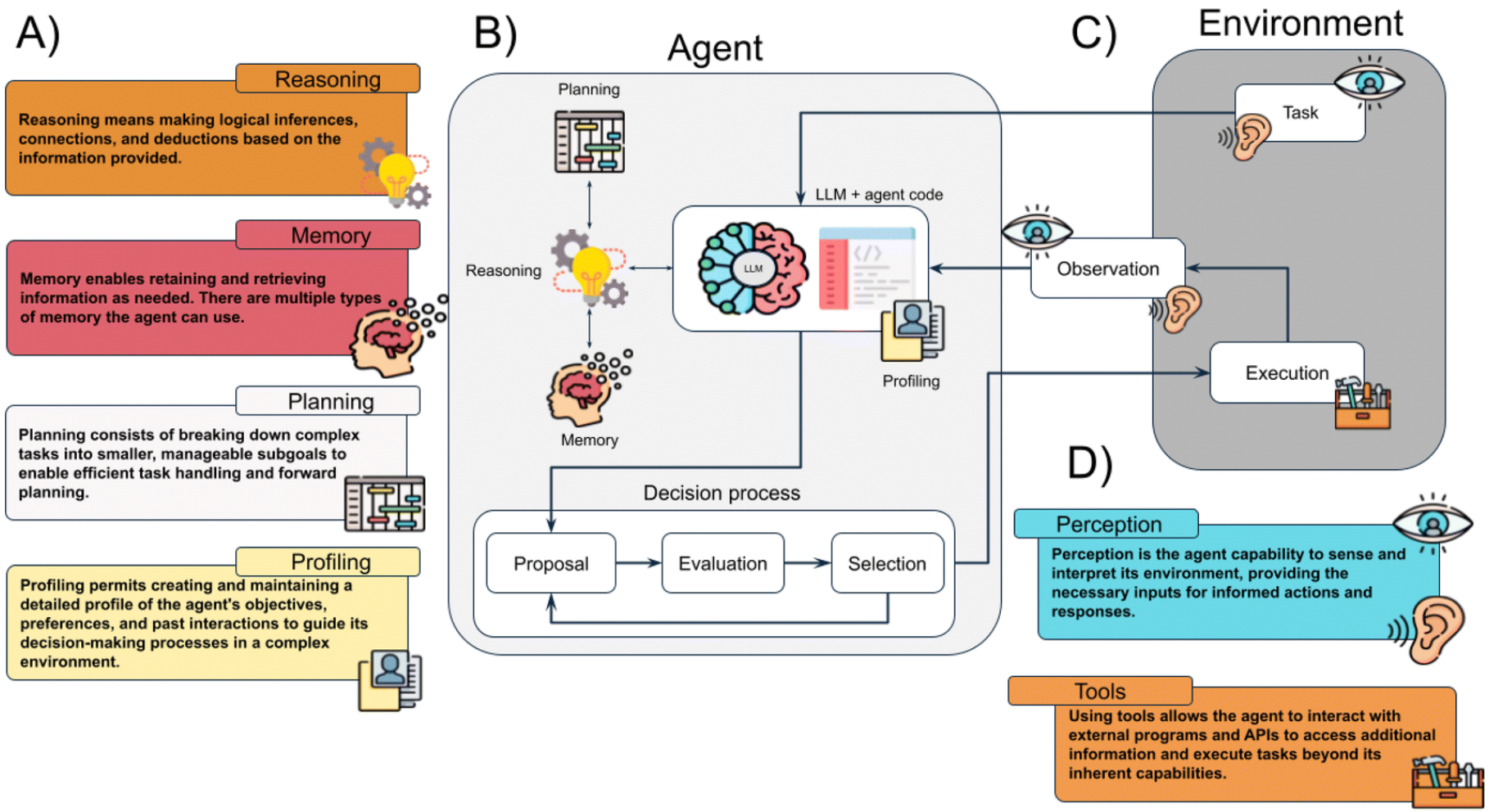


Autonomous Agent

LLM-Agent



LLM-Agent


A)
Reasoning
Reasoning means making logical inferences, connections, and deductions based on the information provided.
Memory
Memory enables retaining and retrieving information as needed. There are multiple types of memory the agent can use.
Planning
Planning consists of breaking down complex tasks into smaller, manageable subgoals to enable efficient task handling and forward planning.
Profiling
Profiling permits creating and maintaining a detailed profile of the agent's objectives, preferences, and past interactions to guide its decision-making processes in a complex environment.
B)
Agent
Planning
Reasoning
Memory
LLM + agent code
LLM
Profiling
Decision process
Proposal
Evaluation
Selection
C)
Environment
Task
Observation
Execution
D)
Perception
Perception is the agent capability to sense and interpret its environment, providing the necessary inputs for informed actions and responses.
Tools
Using tools allows the agent to interact with external programs and APIs to access additional information and execute tasks beyond its inherent capabilities.

Limitations and Challenges

1. **Lack of high-quality and diverse data**

2. **Absence of standardized benchmarks**

3. **Black-box nature and limited interpretability of LLMs**

Thank You