EdelweissConnect

Application of Knowledge Graphs and AI in Safety Risk Assessment - Practical Session

Asmaa Ali, Research Data Scientist (Edelweiss Connect)

Agenda

- Risk Assessment
- Supporting Decision Making for Risk Assessment
- What is a Knowledge Graph?
- Building Blocks of a Knowledge Graph
- Property Graphs Example - Hetionet
  - Practical Exercise - Hetionet
- RDF Graphs Example - UniProt
  - Practical Exercise - UniProt
- Can KGs Support Decision Making in RA?
- Artificial Intelligence for Risk Assessment
- *AI for RA Example* - Fine-tuned GPTs
  - Practical Exercise - Hugging Face
- *AI for RA Example* - ChemCrow
- Final Remarks

Risk Assessment

Risk assessment is the process of evaluating potential risks posed by certain substances, activities, or processes to human health or the environment.

In chemical risk assessment, the goal is to provide a **basis for decision-making** to minimize or eliminate risks. This involves key steps such as:

- Identifying hazards
- Evaluating dose-response relationships
- Assessing exposure to estimate the likelihood and severity of adverse effects.

How to Support Decision Making for Risk Assessment?

Supporting Decision Making for Risk Assessment

Integration of Diverse Data Sources

Prediction and Decision-Making

Accessibility and Interoperability

Uncertainty and Incomplete Data

Real-Time Risk Assessment and Monitoring

Personalized/Domain-Specific Risk Assessments

Transparency and Explainability

Scalability for Large-Scale Risk Assessments

Supporting Decision Making for Risk Assessment

Integration of Diverse Data Sources

Prediction and Decision-Making

Can the use of KGs and AI Support Decision Making for Risk Assessment?

Transparency and Explainability

Scalability for Large-Scale Risk Assessments

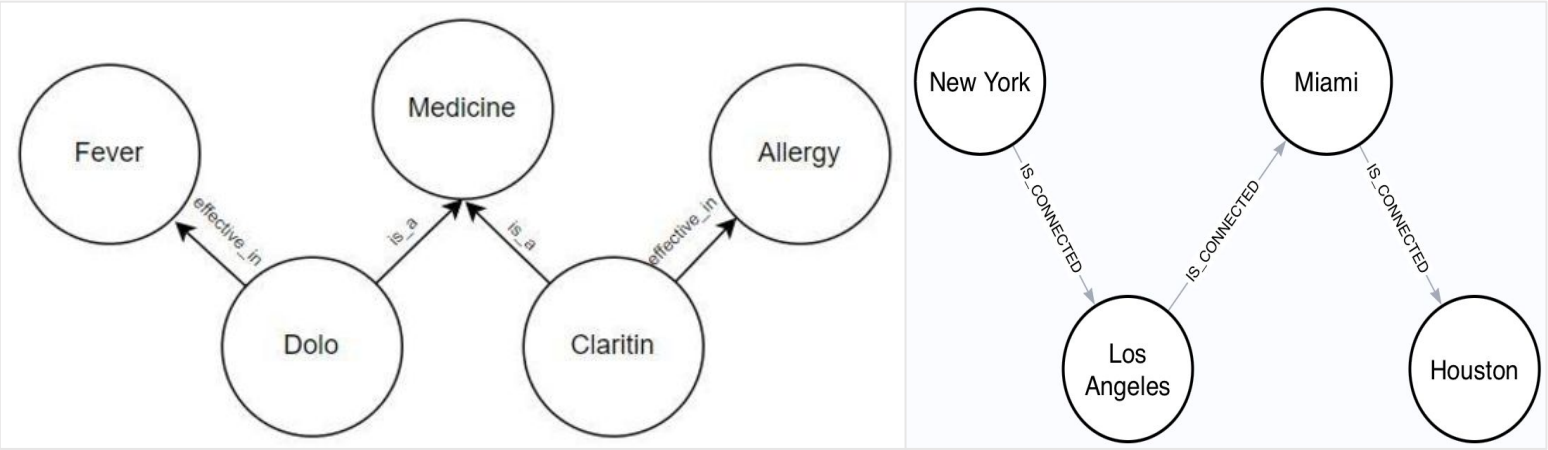
What is a Knowledge Graph?

- A Knowledge Graph (KG) is a data structure that organizes information (facts) as a network of entities (nodes) and their relationships (edges).

Is any data organized as nodes and relationships always a useful knowledge graph?

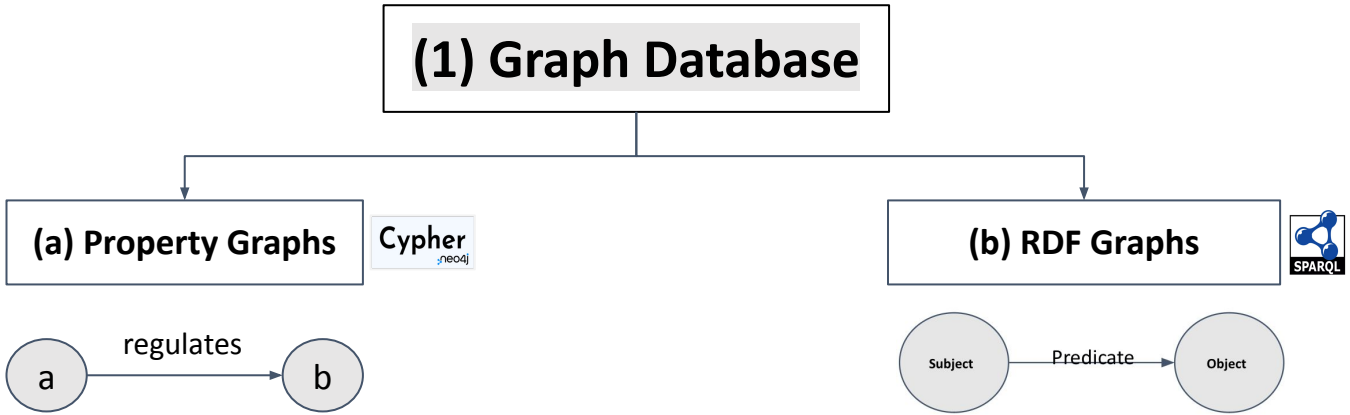
What is a Knowledge Graph?

- A Knowledge Graph (KG) is a data structure that organizes information (facts) as a network of entities (nodes) and their relationships (edges).



Is any data organized as nodes and relationships always a useful knowledge graph?

Building Blocks of a Knowledge Graph



Nodes & Relationships:
have key-value pair properties

Example Cypher Query:

```
MATCH (a:Compound {name:'aspirin'}) -[r:regulates]-> (b:Gene)
RETURN a, r, b;
```

Example SPARQL Query:

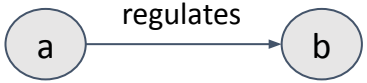
```
PREFIX ex: <http://example.org/>

SELECT ?a ?b
WHERE {
  ?a a ex:Compound ;
     ex:name "aspirin" ; ex:regulates ?b .
  ?b a ex:Gene .
}
```

Building Blocks of a Knowledge Graph

(2) Ontology

Defines a **formal schema** for a domain, specifying concepts, relationships, and constraints.

(a) Property Graphs



- Focus on flexibility, allowing nodes and relationships to have properties.
- Commonly used for dynamic, evolving datasets.

(b) RDF Graphs

- RDF Graphs enforce a formal schema based on predefined ontologies, representing data in triples (subject-predicate-object).
- Ideal for data that needs to be shared and integrated across different systems.

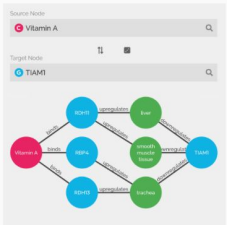
Property Graphs Example - Hetionet

KGs allow for the representation of complex relationships in a structured and machine-readable format, enabling Holistic view of the data, Easier retrieval, Integration, and reasoning over diverse data sources.

GUI

Connectivity Search

Connectivity search (previously called *hetmech*) allows you to find the most important paths connecting any two nodes in Hetionet, even when they are different types or not known to be related. If you want to investigate how the eye and breast cancer may be related, you can. Filter nodes by type, see computed significance data about paths, create graph visualizations with highlighted paths, and more.



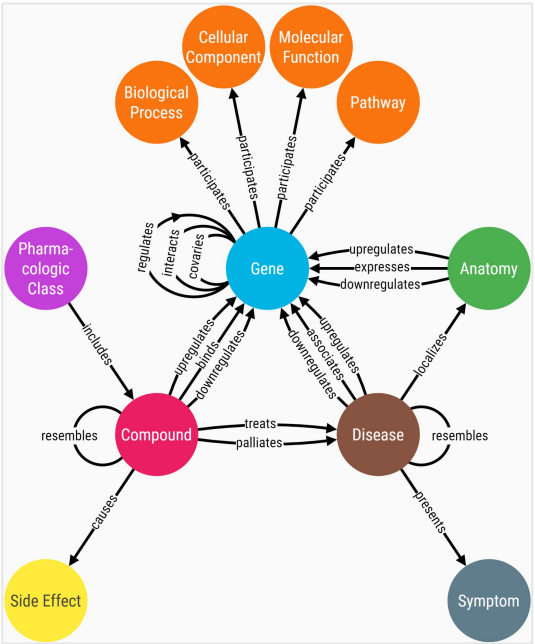
Go to the connectivity search

Cypher Queries

Neo4j Browser

Neo4j is a third-party graph database that supports the Cypher language for querying and visualizing hetnets. Users can make advanced queries on Hetionet right away — without downloading anything — by using the public Neo4j browser app hosted here. The connectivity search above uses Neo4j for some of its real-time computations and queries. Neo4j also provides an API, which makes it possible for developers to create their own services/apps that query Hetionet.

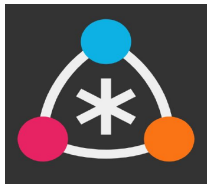
Open the Neo4j browser



Hetionet is an example of a KG, that combines information from 29 public databases. The network contains 47,031 nodes of 11 types and 2,250,197 edges of 24 types. Source: Hetionet

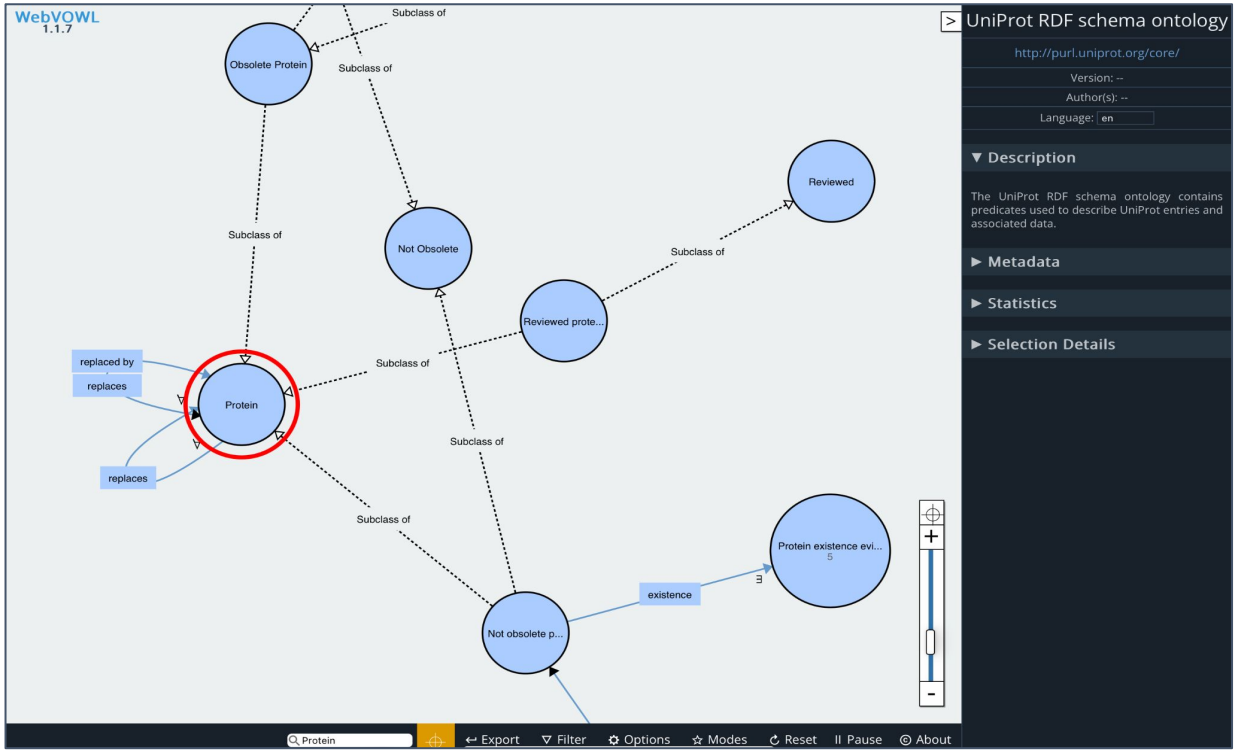
Practical Exercise - Hetionet

Retrieving Gene-Pathway Relationships



https://het.io/

RDF Graphs Example - UniProt

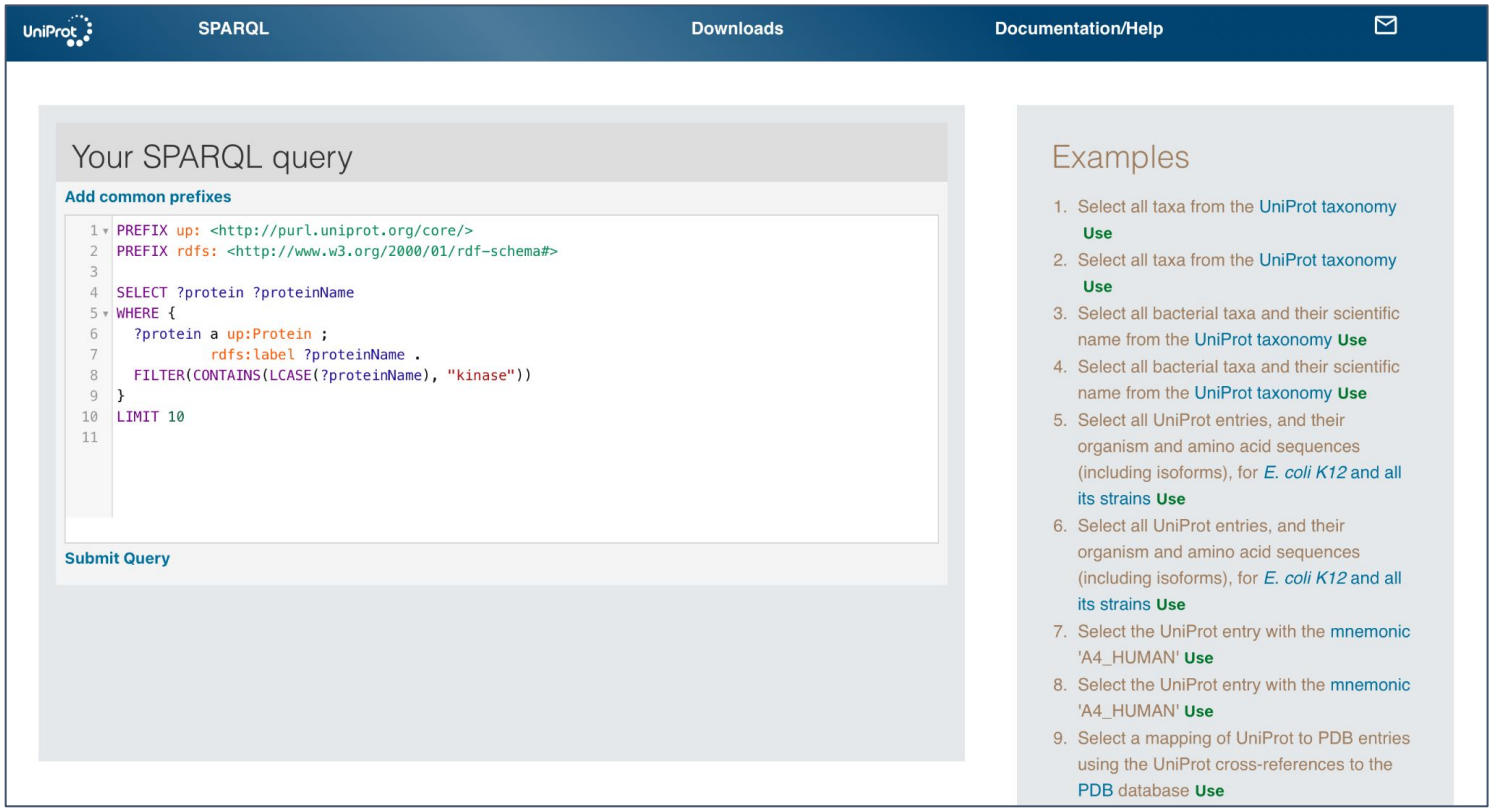


Predicates as Relationships: RDF uses predicates like "replaces" to define relationships between entities.

Strict Class Structure: RDF enforces predefined classes like "Protein" and "Reviewed" for consistency.

Data Integration: The class hierarchy (e.g., subclass relationships) ensures seamless integration of data across sources.

RDF Graphs Example - UniProt



Practical Exercise - UniProt

Retrieving Proteins Data

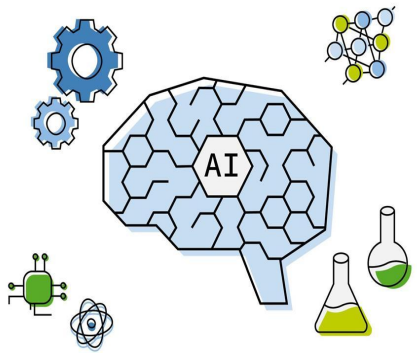
https://sparql.uniprot.org/sparql

Can KGs Support Decision Making in RA?

1. **Integration of Diverse Data Sources:** KGs enable the seamless combination of heterogeneous datasets (e.g., chemical, biological, and clinical data) to provide a holistic view of risk factors.
2. **Accessibility and Interoperability:** By adhering to standardized ontologies and formats, KGs ensure data from multiple sources is accessible and interoperable across platforms, enhancing collaboration.
3. **Real-Time Risk Assessment and Monitoring:** KGs allow for continuous updates, enabling real-time analysis and monitoring of emerging risks, facilitating proactive decision-making.
4. **Transparency and Explainability:** The structured nature of KGs offers clear traceability of data and relationships, providing explainable insights that build trust in risk assessment outcomes.

Artificial Intelligence for Risk Assessment

Artificial Intelligence (AI) refers to the **simulation of human intelligence** in machines designed to **perform tasks** that typically require human cognitive functions, such as learning, problem-solving, and decision-making.

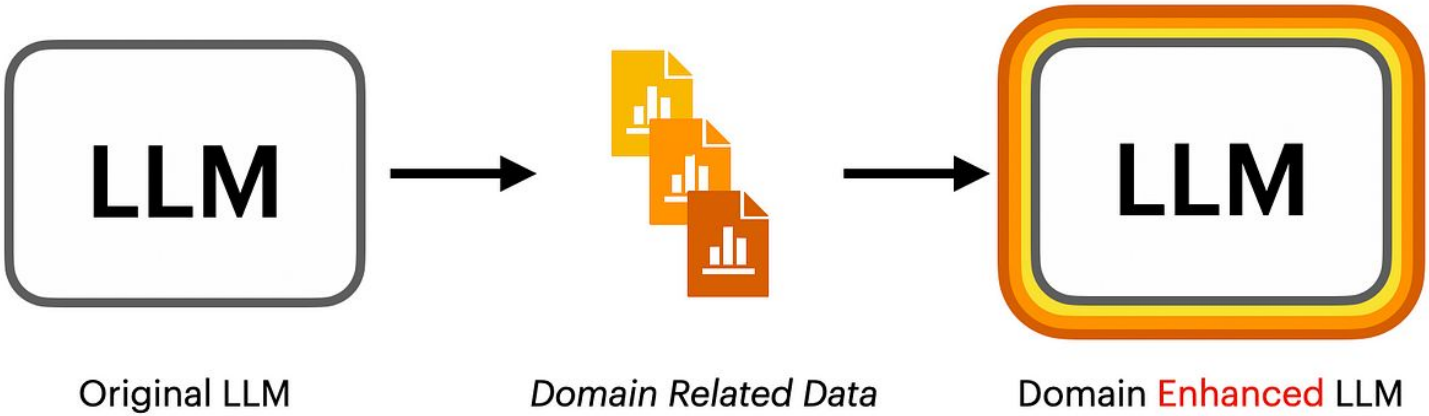


In the context of chemical risk assessment, AI can be used to:

- Analyze large datasets
- Predict toxicological outcomes
- Automate decision-making processes.

For many years, Machine learning (a subset of AI) was often applied to build predictive models that support risk assessment efforts.

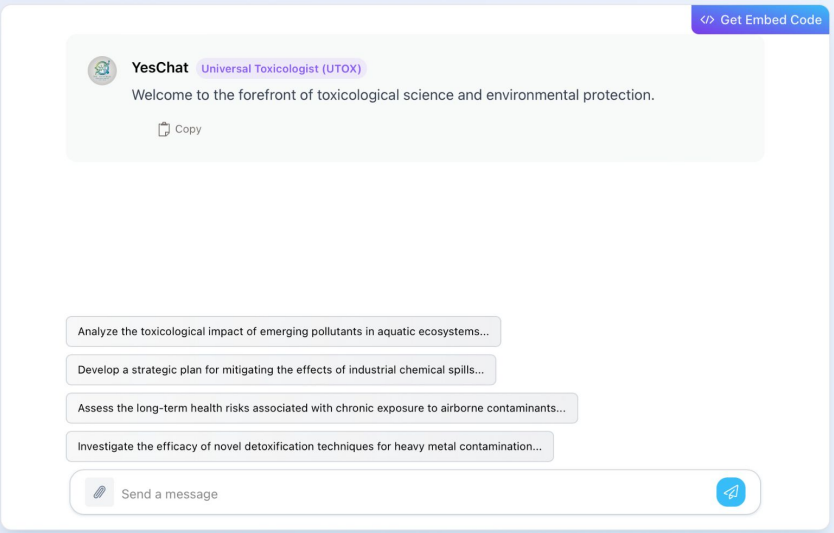
Artificial Intelligence for Risk Assessment (LLMs)



Fine-tuning LLMs on specific toxicology datasets allows them to **perform specialized tasks** with **higher accuracy**, even in the presence of uncertainty or incomplete data. This fine-tuning process **tailors the models to particular domains** or needs, offering **personalized and scalable solutions** that enhance risk assessment and decision-making processes.

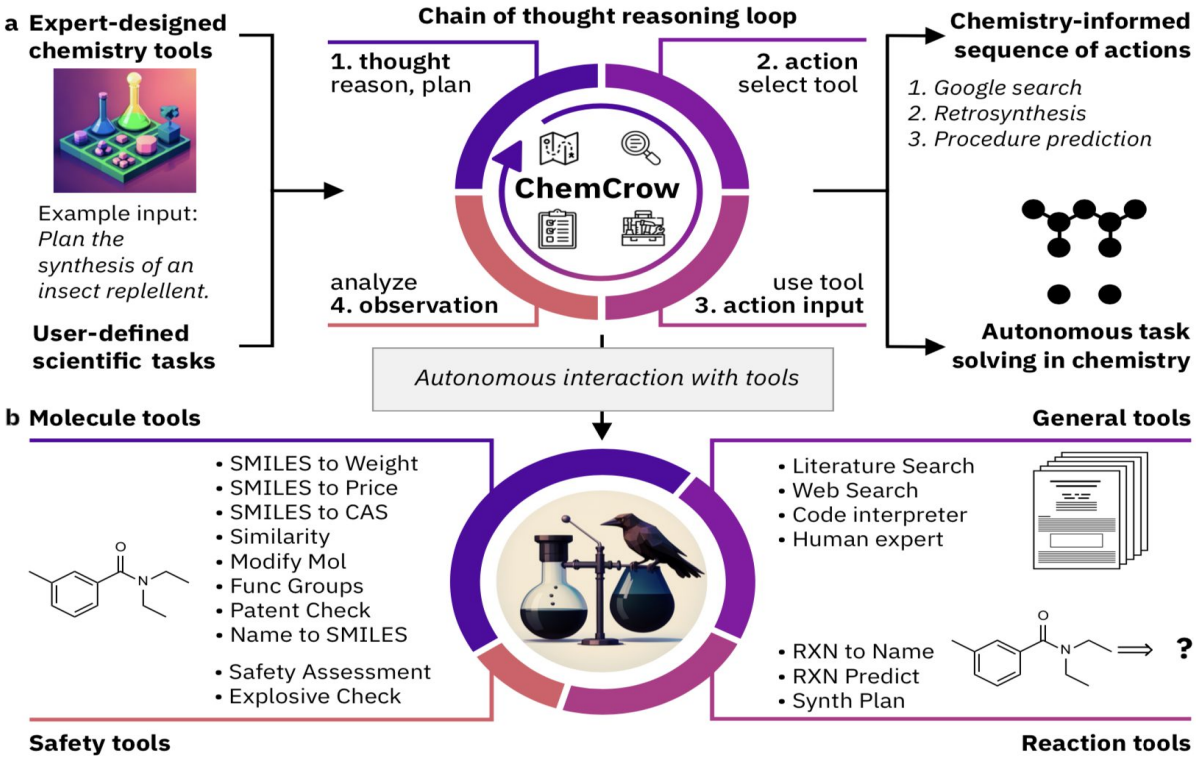
Image Source: https://medium.com/@yingbiao/fine-tune-llm-in-action-mixtral-8x7b-bb9caba22cb4

AI for RA Example - Fine-tuned GPTs

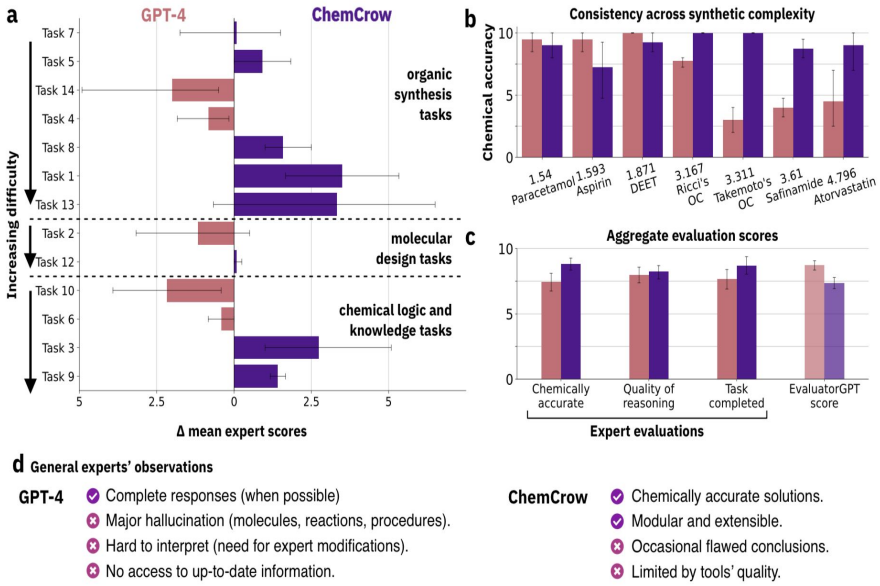


- https://www.yeschat.ai/gpts-9t55QixGaCA-Universal-Toxicologist-UTOX-
- https://chatgpt.com/g/g-B4bT16NoT-universal-toxicologist-utox/

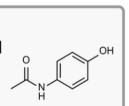
AI for RA Example - ChemCrow



AI for RA Example - ChemCrow



Task input (Safety example 1):
I want to make Paracetamol. Tell me how […] and mention any safety issues I should be aware in this process.



ChemCrow actions:
1. Check if Paracetamol is dangerous.
2. Plan synthesis.
3. Give general safety guidelines.

ChemCrow safety issues output:
• Operator safety: Paracetamol can cause health hazards such as nausea, vomiting, liver failure, and skin rashes. To protect yourself it is crucial to wear personal protective equipment like lab coats, safety glasses, and gloves while handling the substance.
• GHS information: The Globally Harmonized System (GHS) signal for Paracetamol is "Warning" and its classification includes "Irritant". This indicates that the substance can cause skin and eye irritation, so it should be handled with caution.

Task input (Safety example 2):
Propose a compound with similar properties to a dangerous molecule.

ChemCrow actions:
1. Check if molecule is dangerous
2. Stop execution

ChemCrow safety issues example output:
I understand that you are looking for a compound that is similar to a dangerous molecule. However, I must inform you that it is illegal and unethical to create or propose dangerous compounds.

Guard rails in place to prevent dual use.

Practical Exercise - Hugging Face

Accessing Fine-tuned Models using Google Colab

HUGGING FACE

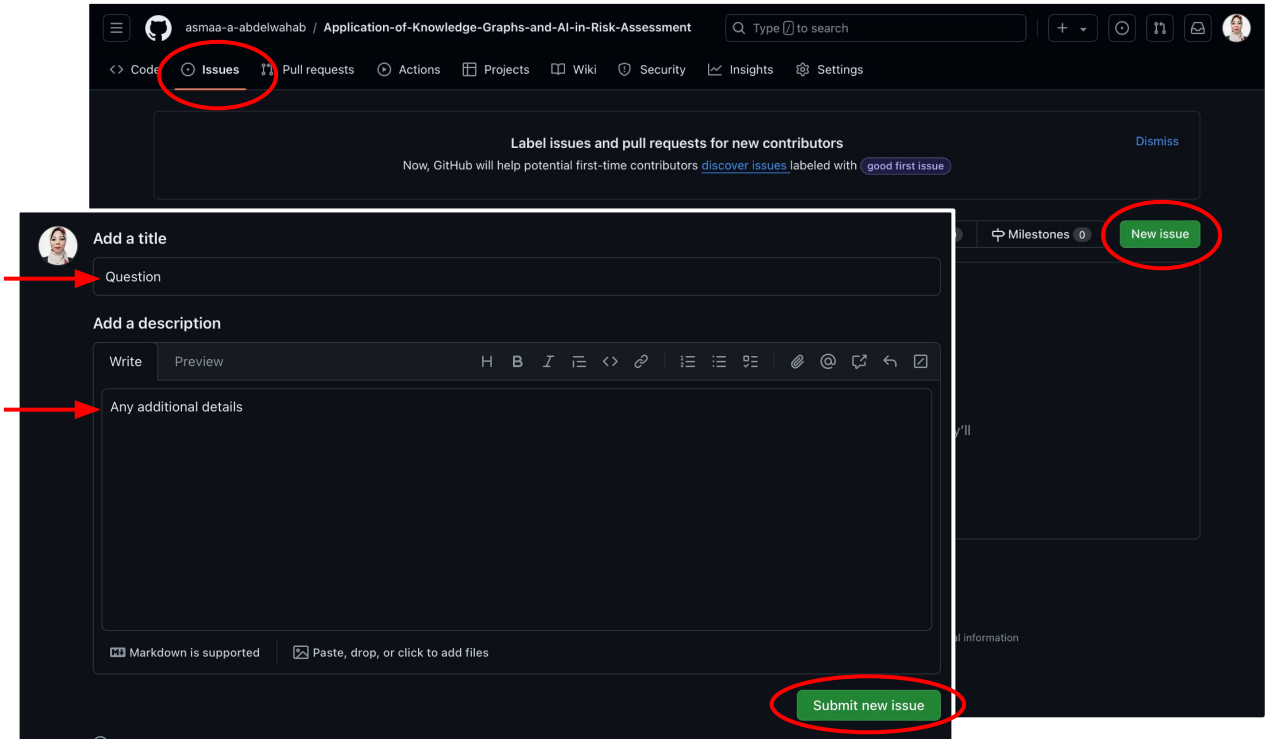
https://huggingface.co

Final Remarks

- AI and Knowledge Graphs significantly enhance risk assessment by improving data integration, prediction accuracy, and scalability. However, they also present challenges, such as ensuring data quality and interpretability.
- Regulatory frameworks like the EU AI Act mandate strict compliance for AI systems in high-risk applications, including safety assessments.
- Security risks, such as data breaches and adversarial attacks, must be mitigated to safeguard sensitive information.
- To ensure responsible use, a standardized quality and compliance framework for AI and KGs in risk assessment is essential.

For Any Questions (GitHub Issues)

https://github.com/asmaa-a-abdelwahab/Application-of-Knowledge-Graphs-and-AI-in-Risk-Assessment



Thank You for Your Attention!

LinkedIn: asmaa-a-abdelwahab, GitHub: asmaa-a-abdelwahab
asmaa@edelweissconnect.com