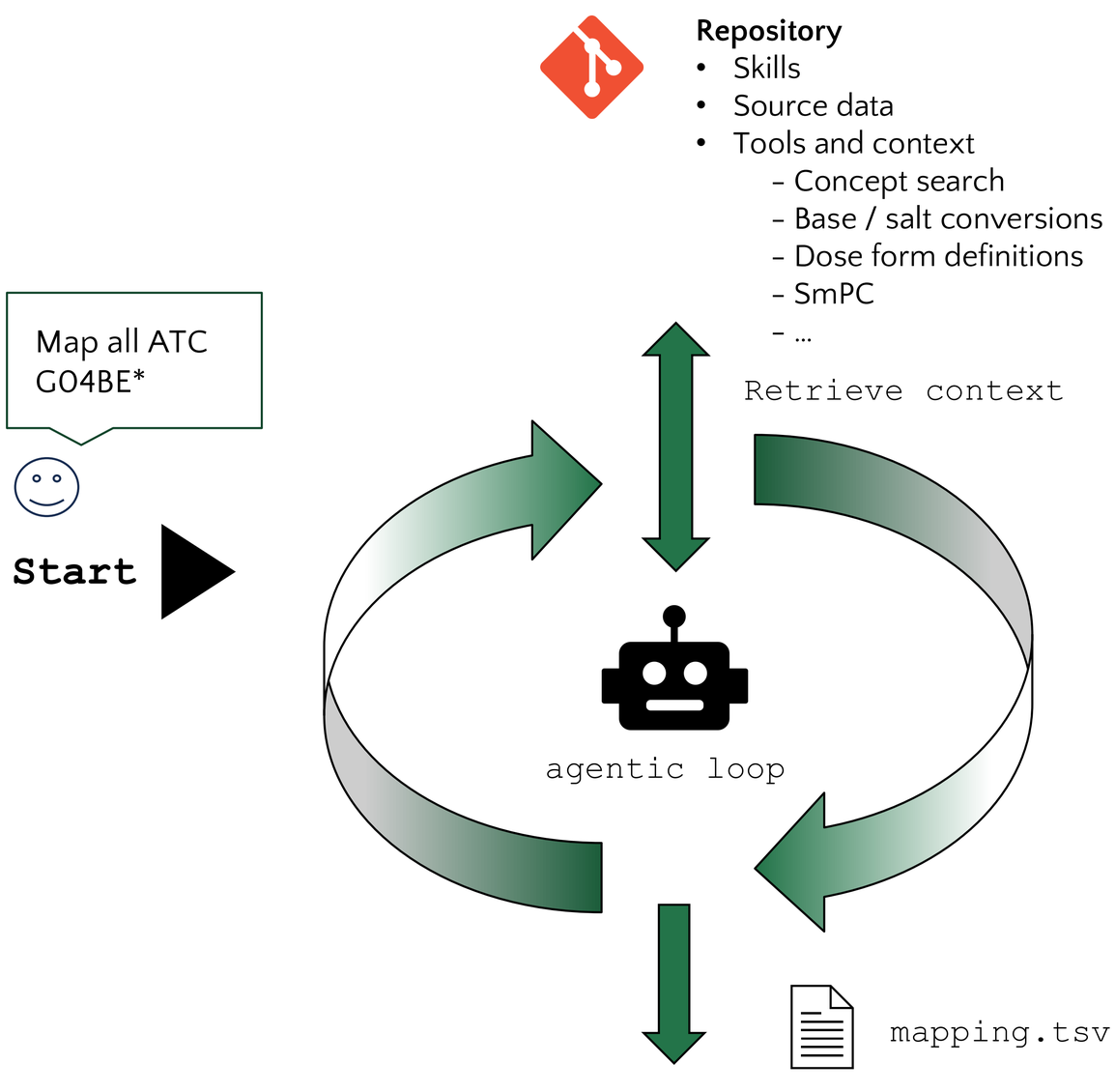
Using coding agents and a skill-based context engineering approach to map EMA authorized drugs to RxNorm

- Despite a need for a centrally maintained and validated mapping of all EMA authorized drugs, until now no publicly available mapping existed.
- Mapping source drug vocabularies to OMOP standardized concepts remains a labor-intensive task.

Common LLM-based mapping systems decompose drug entries into their constituent elements, perform semantic search, and then rerank to obtain the a result. However, RxNorm encodes domain-specific knowledge that cannot always be resolved by search and reranking alone. An LLM does not always know the distinction between various 'injectables', correct base / salt conversions, whether a stated dose is metered or actuated, or the correct biosimilar suffixes. Resolving these cases reliably requires additional context which is not available in a simple drug name string.

The present approach uses a readily available coding agent and its default tools in combination with a curated set of skills, scripts, and relevant information available as plaintext files in a git repository. This allows an agentic LLM to gather context and map drugs with awareness of domain-specific knowledge and conventions. This approach works across languages, formats, coding agents, and source data conventions without requiring significant additional work. It has already been successfully applied to map Spanish and Latvian drugs.



Flow

- User submits mapping task
- Agent reads mapping skill file containing rules and guidelines how to gather context from resources
- Agent accesses source data and optional resources until it has sufficient details to create the mapping
- After the mapping is created the agent runs a validation script with codified rules and fixes issues
- At the end of the loop a validated mapping file is created

Limitations

- LLMs still make mistakes
- Token usage higher than with search-rerank approach
- Formal validation of results is pending

```
…
EU/1/08/476/006     19102192  tadalafil 20 MG Oral Tablet    EXACT                                                  2025-12-04
EU/1/08/476/007     741670    tadalafil Oral Suspension      BROAD     tadalafil 2 MG/ML Oral Suspension [Adcirca]  2026-02-12
…
```

Results:

- 16,575 EMA centrally authorized product presentations mapped
- 31,150 AEMPS (Spain) product presentations mapped
- 35,700 ZVA (Latvia) product presentations mapped
- Repository with skills, scripts, curated context, and mappings publicly available

https://github.com/mi-erasmusmc/ema-authorised-to-rxnorm-mappings

Rowan Parry