



Phenelope – an LLM-assisted concept set builder

Joel Swerdel
Director, Global Epidemiology Organization
Johnson & Johnson
June 9th, 2026



Current Status

- R package released on April 10th
- Repository: <https://github.com/OHDSI/Phenelope/tree/main>
- Documentation: <https://ohdsi.github.io/Phenelope/>



Requirements

- Access to a CDM
 - Used to determine parent child concept relationships
- API access to an LLM
 - We are finding that non-reasoning models may outperform reasoning models
 - We're using OpenAI 4o (which was used in the Mind Meets Machine experiment at the 2025 OHDSI symposium)



Function 1: createConceptSet

- Creates a concept set based on a clinical idea and a seed concept
- Has the LLM examine:
 - All descendants of concept
 - All PHOEBE recommendations for the concept
 - All descendants of accepted PHOEBE recommendations
- Output
 - Dataframe with all the results of the LLM analysis
 - A JSON object with the condensed version (true concept set) of the accepted concepts



Function 1: createConceptSet

- Uses the open-source package ellmer to connect to the LLM
- Example connection

```
llmClient <- ellmer::chat_azure_openai(  
  endpoint = https://{your-resource-name}.openai.azure.com.,  
  model = "gpt-4o", #The deployment id for the model you want to use  
  credentials = "secret_credentials")
```

- Ellmer has some nice features for easier LLM use
 - Set system prompt
 - Force LLM to return responses in the correct format



Function 1: createConceptSet

- **conceptName**
- **originalConceptList**
- **IlmClient**
- **connectionDetails**
- **cdmDatabaseSchema**
- **outputDirectory**
- **excludedConditions**
- **additionalInformation**
- **clinicalContext**



Function 1: createConceptSet - Output

suggestedCondition	conceptId	mainCondition	finalAnswer	rationaleForAnswer
Hypersomnia disorder related to a known organic factor	434891	Somnolence or Sedation	YES	Hypersomnia refers to excessive sleepiness, which is a direct synonym or subtype of somnolence. The phrase also
Persistent hypersomnia	434904	Somnolence or Sedation	YES	Persistent hypersomnia characterizes ongoing excessive sleepiness, which directly relates to somnolence or sedation.
Narcolepsy	436100	Somnolence or Sedation	YES	Narcolepsy is a neurological disorder that includes excessive daytime sleepiness, directly associating it with somnolence
Recurrent hypersomnia	443528	Somnolence or Sedation	YES	Recurrent hypersomnia suggests repetitive occurrences of excessive sleepiness, associating directly with somnolence
Photodocumentation of two or more cecal landmarks to establish a complete examination (Deprecated)	915717	Somnolence or Sedation	NO	The proposed concept relates to a procedural documentation practice and does not associate with somnolence or sedation.
Destruction of intraosseous basivertebral nerve, first two vertebral bodies, including imaging guidance (e.g., fluoroscopy), lumbar/sacrum	953292	Somnolence or Sedation	NO	This procedural concept regarding nerve destruction has no relation to somnolence or sedation.



Function 2: createClinicalDescription

- Creates a clinical description based on a health condition
- Output includes:
 - Overview of condition
 - Presentation
 - Assessment criteria
 - Differential diagnoses
 - Treatment
 - Prognosis



Function 2: createClinicalDescription

createClinicalDescription

- **condition**
- **IlmClient**
- excludedConditions
- outputToWord
- wordFileName



Function 2: createClinicalDescription - Output

Condition: Somnolence or Sedation (Synonyms: Excessive sleepiness, hypersomnia, drowsiness)

Excluded Conditions: None

1. Overview of Somnolence or Sedation:

****Somnolence or Sedation****, commonly referred to as excessive sleepiness or drowsiness, is a temporary or persistent condition in which an individual experiences an unusually strong urge to sleep or is in a state of reduced alertness, often associated with decreased mental and physical activity. This condition can be acute or chronic. Acute somnolence may result from transient factors such as medication side effects or sleep deprivation, and it can recur if the causative factor persists. Chronic somnolence or sedation may be secondary to medical conditions like sleep disorders (e.g., sleep apnea, narcolepsy), mental health conditions, or neurological dysfunction.

No excluded conditions for differentiation.

2. Presentation for Somnolence or Sedation:

- Persistent or intermittent excessive daytime sleepiness
- Difficulty staying awake during typical daily activities
- Low energy levels



Thank You!



Function 1: createConceptSet

ellmer has some nice features for easier LLM use:

```
systemPrompt <- "You are an expert medical doctor specializing ..."
```

```
llmClient$set_system_prompt(systemPrompt)
```

```
text <- llmClient$chat_structured(prompt,  
  type = ellmer::type_array(ellmer::type_object(  
    suggestedCondition = ellmer::type_string(),  
    finalAnswer = ellmer::type_enum(values = c("YES","NO")),  
    rationaleForAnswer = ellmer::type_string()))
```



Functions 1: createConceptSet

Prompt (partial):

General rules:

- 1) Always consider synonyms of MAIN_CONDITION to be absolutely CLINICALLY interchangeable with MAIN_CONDITION in ALL contexts (everyday use, different ontologies, clinical context, etc.).
- 2) DO NOT infer SUGGESTED_CONDITION to be more specific such as by adding a non-specified body part or region or adding extra words for clarity!
- 3) Additional information: ADDITIONAL_INFORMATION4) ALWAYS keep in mind the current clinical context: CLINICAL_CONTEXT - all patients are in this category of subjects.

Use the following specific rules to decide on your answer:

- 1) If the vast majority (> 95%) of people with SUGGESTED_CONDITION, also have MAIN_CONDITION in CLINICAL_CONTEXT, the answer is YES.
- 2) If SUGGESTED_CONDITION is a synonym of MAIN_CONDITION and in the clinical context, CLINICAL_CONTEXT, the answer is YES.
- 3) If SUGGESTED_CONDITION is a subtype of MAIN_CONDITION and in the clinical context, CLINICAL_CONTEXT, the answer is YES.
- 4) If SUGGESTED_CONDITION is a synonym, subtype, or exclusive manifestation of any of EXCLUDED_CONDITIONS (i.e., only happens with EXCLUDED_CONDITIONS), the answer is NO.5) If SUGGESTED_CONDITION is not possible within the clinical context, CLINICAL_CONTEXT, the answer is NO.