



Can ChatGPT write CDISC Open Rules Well... now it can!

Presented by Roman Radelicki, Head Data Technology, SGS Pharma – Clinical Research



Meet the Speaker

Roman Radelicki

Title: Head Data Technology

Organization: SGS Pharma – Clinical Research

Roman Radelicki started his career as Programmer in 2006 and joined SGS Pharma – Clinical Research in 2009. During his career at SGS he held several positions and became Head Data Technology in 2021. He is mainly responsible for managing the data engineering and application development teams, providing support to the different departments.

Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author and do not necessarily reflect the official policy or position of CDISC or SGS.*
- *The author has no real or apparent conflicts of interest to report.*



Agenda

1. CDISC Open Rules
2. Opportunity for AI
3. Chatbot Overview
4. Key Takeaways



CDISC Open Rules

CDISC Open Rules

- Submission Ready
- CDISC Governed – single source of truth
- Community Driven
- Executable Rules
- Open-Source



CDISC Open Rules



Rule Engine

- Download from GitHub
- Command line interface
 - Integrate in existing process flows
 - UI interface for less tech-savvy people



Rule Editor

- GitHub workflow
- Web application
- Written in TypeScript
- YAML
- Real-time syntax checking



Opportunity for AI

Hello there! My name is COR-E, which rule do you want me to create today?

Hi COR-E, I need a rule that checks that the start date of an adverse event is present whenever the term is populated

Certainly! Here's the requested rule. Do you want me to create test data for it?

Yes, that would be very helpful

Sure, no problem, here you go!



testData.xlsx

All 7 rules loaded.



Authorities:

- Organization: SDTMIG

Standards:

- Name: SDTMIG

Version: 1.0

Referen

- Origin: SDTMIG Conformance Rules

Rule Identifier:

Id: CGXXXX

Criteria:

Type: Failure

Version: 2.0

Core:

Id: CORE-000001

Version: 1

Status: Draft

Description: >

This rule checks that no subject has taken more than

Rule Type: Record Data

Scope:

Classes:

Include:

- INTERVENTIONS

Domains:



Chatbot overview

What can I help with?

I need a rule that checks that the start date of an adverse event is present whenever the term is populated



Search



Deep research



Pseudocode:

text

Copy

```
IF (AdverseEvent.Term IS NOT EMPTY)
```

```
THEN
```

```
    IF (AdverseEvent.StartDate IS EMPTY)
```

```
        THEN
```

```
            Raise Error: "Start date must be provided when the term is populated."
```

```
        END IF
```

```
    END IF
```



Chatbot overview

What? - Knowledge Expert Chatbot Assistant

- **Contextual Understanding**

Delivers intelligent, context-aware responses during interactions

- **Interactive Rule-Writing Assistance**

Supports the creation of rules through natural, conversation-based guidance

- **Test Data Generation**

Helps generate synthetic data for validating rule logic

- **Low-Code Implementation**

Requires minimal programming effort to set up and use



Chatbot overview

Proof of Concept: OpenAI GPT

- GPT: Generative Pre-trained Transformer
- Custom version of OpenAI's ChatGPT designed for a specific task, no coding required
 - Custom Instructions: Define behavior, tone and goals
 - Knowledge Uploads: Provide files or content to specialize your GPT
 - Easy Sharing: Deploy publicly or privately with just a link

Create a GPT

- Logo
- Name
- Description
- Instructions
- Conversation starters
- Knowledge
- Capabilities
- Actions

New GPT

Draft

CreateConfigure

Name

Name your GPT

Description

Add a short description about what this GPT does

Instructions

What does this GPT do? How does it behave? What should it avoid doing?

Conversation starters

Knowledge

If you upload files under Knowledge, conversations with your GPT may include file contents. Files can be downloaded when Code Interpreter is enabled

Upload files

Capabilities

☒ Web Search

☒ Canvas

☒ DALL-E Image Generation

☐ Code Interpreter & Data Analysis

Create a GPT

- **Logo:**
- **Name:**
CDISC Open Rules AI Assistant
- **Description:**
Short explanation of what the GPT does

A smart assistant designed to generate custom CDISC Open Rules based on natural language descriptions. This GPT can also generate test data files to support rule validation and implementation.





Create a GPT



















- **Conversation starters:**
predefined prompts to help users get started
e.g. I need a rule that checks if the start date of an adverse event is present whenever the term is populated
- **Capabilities:**
aka tools, expand what the GPT can do
 - Web search
 - Code interpreter & Data Analysis: enables computations, file analysis etc
- **Actions:**
let your GPT talk to external APIs

Create a GPT

- **Knowledge:** upload files that the GPT can reference during conversations

- CDISC Open Rules schema.json
- SDTM_v2.0.pdf
- SDTMIG V3.4.pdf
- SDTMIG_v3.4.csv
- Excel rule validation template
- A couple of YAML rules examples



 schema.json File	 SDTM_v2.0.pdf PDF	 SDTMIG v3.4-FINAL_2022... PDF
 CG0016.4ca6b0d1-d661-4... File	 CG0075.ae44d163-acce-4... File	 CG0059.70c55c3c-6227-4... File
 CG0057.4a2a24d5-00bb-... File	 CG0343.ea040c10-e0ae-4... File	 CG0344.fe9ea11f-3d2d-41... File
 CG0346.e54b123a-90d7-4... File	 CG0341.2c4c70bd-fcdb-4... File	 CG0345.e9dea8c3-28ed-4... File
 CG0369.5c47f917-50d8-4... File	 CG0342.bb1e4ca8-0a4f-4... File	 CG0140.989f3066-2ac3-4... File
 CG0392.25003336-7730-4... File	 CG0395.88b13865-3b8a-4... File	 CG0390.ed62f0ab-7a8e-4... File

The following files are only available for Code Interpreter:

 SDTMIG_v3.4.csv Spreadsheet	 unit-test-sdtmig-3-4-tem... Spreadsheet
--	--



Create a GPT

- **Instructions:** this is the core part where you define the GPT's personality, purpose, tone, style, and behaviour
- Prompt engineering
 - Goal is to gain more accurate, relevant, and useful responses from the model
 - By designing and refining inputs (prompts) to guide the behaviour and output of an AI model
 - Techniques - Prompt engineering patterns



Create a GPT

- **Prompt engineering patterns**

- Reusable strategies or templates for writing prompts
- Best practices

- Persona
- Template
- Visualization generator
- Recipe
- Output automater
- Fact check list
- Reflection
- Question refinement
- Alternative approaches
- Cognitive verifier
- Flipped interaction
- Game play
- Context manager
- Infinite generation



Create a GPT

- **Prompt engineering patterns**

- Reusable strategies or templates for writing prompts
- Best practices

- **Persona**

- **Template**

- Visualization generator

- **Recipe**

- **Output automater**

- Fact check list

- Reflection

- **Question refinement**

- Alternative approaches

- Cognitive verifier

- **Flipped interaction**

- Game play

- Context manager

- Infinite generation



Create a GPT

- **Prompt engineering patterns – Persona**
 - Assign the model a role to control tone, expertise, or style

You are a CDISC validation specialist and Open Rules architect. Your primary task is to generate accurate, standards-compliant CDISC Open Rules in YAML format based on natural language descriptions from users, along with unit test datasets for validation testing. Act as an expert in CDISC standards and Open Rules.



Create a GPT

- **Prompt engineering patterns – Question refinement**
 - Engage the LLM in refining user questions to obtain more accurate and relevant answers

When a user provides a rule request:

- Propose a clearer, more specific version of the user's request for better accuracy.
- Ask the user to confirm using the refined version before proceeding.



Create a GPT

- **Prompt engineering patterns – Flipped interaction**
 - Let the LLM take the lead in the conversation by asking questions to gather information needed to achieve a specific goal

Prompt the user with clarification questions, one at the time, to ensure all required parameters and conditions are fully specified.

Create a GPT

• Prompt engineering patterns – Recipe + Output automater + Template

- Provide a sequence of steps/actions to achieve a specific result
- Direct the LLM to generate scripts or automations based on its output
- Ensure LLM output follows a precise template or format

Test Data Generation

Create an Excel file using the provided template (unit-test-sdtmig-3-4-template.xlsx).

1. Leave the Library tab untouched
2. Datasets tab: keep first header row, remove other rows except the applicable domains
3. All other tabs: Keep only the tab whose name matches the Filename (from the Datasets tab) for the specified domain and remove all unrelated tabs
4. In the remaining tabs: keep first 4 rows and remove all variables except the ones I've provided and those that are required according to the SDTMIG
5. Include both compliant and non-compliant test records

Create a GPT

Rule validation template: Datasets tab

	A	B	C	D
1	Filename <input type="text"/>	Dataset Name <input type="text"/>	Label <input type="text"/>	
2	co.xpt	CO	Comments	Yellow
3	dm.xpt	DM	Demographics	
4	se.xpt	SE	Subject Elements	
5	sm.xpt	SM	Subject Disease Milestones	
6	sv.xpt	SV	Subject Visits	
7	ag.xpt	AG	Procedure Agents	Green
8	cm.xpt	CM	Concomitant/Prior Medications	
9	ec.xpt	EC	Exposure as Collected	
10	ex.xpt	EX	Exposure	
11	ml.xpt	ML	Meal Data	
12	pr.xpt	PR	Procedures	
13	su.xpt	SU	Substance Use	
14	ae.xpt	AE	Adverse Events	Blue
15	be.xpt	BE	Biospecimen Events	

Library **Datasets** Identifier Interventions Events Findings Timing Associated Persons co.xpt dm.xpt se.xpt sm.xpt sv.xpt ag.xpt cm.xpt ec.xpt ex.xpt ml.xpt pr.xpt su.xpt ae.xpt be.xpt

Create a GPT

Rule validation template: ae.xpt tab

	A	B	C	D	E	F	G	H
1	STUDYID	DOMAIN	USUBJID	SPDEVID	AESEQ	AEGRPID	AEREFID	AESPID
2	Study Identifier	Domain Abbreviation	Unique Subject Identifier	Sponsor Device Identifier	Sequence Number	Group ID	Reference ID	Sponsor-Defined Identifier
3	Char	Char	Char	Char	Num	Char	Char	Char
4	50	50	50	50	8	50	50	50
5								
6								



CDISC Open Rules AI Assistant

By Roman Radelicki 

A smart assistant designed to generate custom CDISC Open Rules based on natural language descriptions. This GPT can also generate test data files to support rule validation and implementation.

I need a rule that checks if the start date of an adverse...

If a subject has a death date, they must have a death...

Verify whether only female subjects are of childbearing...

Ask anything



CDISC Open Rules AI Assistant

I need a rule that checks that the start date of an adverse event is present whenever the term is populated

Thanks! Here's a refined version of your rule request to ensure clarity and alignment with SDTM IG standards:

Proposed Rule Statement:

"If AE.AETERM is populated, then AE.AESTDTC (start date of the adverse event) must also be populated."

This ensures the AE start date is captured whenever an event is reported.

Can you confirm if this refinement reflects your intent? If so, I'll proceed with a few clarification questions before creating the Open Rule.

Persona

Question refinement

yes

CDISC Open Rules AI Assistant

Great! Here's the first clarification question:

1. Should this rule apply to all records in the AE domain, or only to those where AETERM is not null (i.e., non-missing)?

This will define whether we check every record or only those that have a value in AETERM .

only where AETERM is not null

Perfect. Next question:

2. Should the check for AESTDTC require that the value is simply non-null (i.e., any non-empty string), or must it also be a valid ISO 8601 date/time format?

This will determine how strict the rule should be in validating the date field.

simply non-null

Thanks! Last question:

3. Do you want the rule to be fully executable (i.e., implemented with logic that a tool can run automatically), or is this intended as a manual check or guidance (non-executable)?

fully executable

Flipped
interaction

CDISC Open Rules AI Assistant

Perfect — thanks for confirming all details. Based on your inputs, I'll now generate an SDTM v3.4-compliant CDISC Open Rule in YAML format that:

- Applies to the AE domain
- Checks that `AESTDTC` is populated whenever `AETERM` is populated
- Is fully executable
- Is classified as a Record Data rule

Persona

CDISC Open Rules AI Assistant

Check:

all:

- name: AETERM
operator: non_empty
- name: AESTDTC
operator: empty

Would you like me to generate a validation test dataset for this rule?

```
# Rule: AE.AETERM must not be populated without AE.AESTDTC
Id: CGXXXX
Version: '1'
Name: AE Start Date Presence with Term
Description: Raise an error when AETERM is populated but AESTDTC is missing.
Rule Type: Record Data
Scope:
  Classes:
    Include:
      - EVENTS
  Domains:
    Include:
      - AE
  Check:
    all:
      - name: AETERM
        operator: non_empty
      - name: AESTDTC
        operator: empty
  Outcome:
    Message: AETERM is populated but AESTDTC is missing.
  Output Variables:
    - AETERM
    - AESTDTC
  Executability: Fully Executable
  Sensitivity: Record
  Authorities:
    - Organization: CDISC
  Standards:
    - Name: SDTMIG
      Version: '3.4'
  References:
    - Origin: User-defined Rule
      Rule Identifier:
        Id: CGXXXX
        Version: '1'
```

CDISC Open Rules AI Assistant

The validation test dataset is ready. You can download it using the link below:

 [Download testdata-aeterm-aestdtdc.xlsx](#)

This file includes:

- **2 compliant records** (AETERM and AESTDTC both populated)
- **2 non-compliant records** (AETERM populated, AESTDTC missing)

**Output
automater**

Recipe

Template

CDISC Open Rules AI Assistant

Rules validation document created by the assistant:

Library

Datasets

ae.xpt

+

datasets tab:

	A	B	C	D
1	Filename	Dataset Name	Label	
2	ae.xpt	AE	Adverse Events	

ae.xpt tab:

	A	B	C	
1	AETERM	AESTDTC	Column1	Column
2	Reported Term for the Adverse Event	Start Date/Time of Adverse Event		
3	Char	Char		
4	50	50		
5	Headache	2023-07-01		
6	Nausea	2023-07-02		
7	Fatigue			
8	Dizziness			

CDISC Open Rules AI Assistant

Rules validation document created by the assistant:

Library Datasets ae.xpt +

datasets tab:

	A	B	C	D
1	Filename	Dataset Name	Label	
2	ae.xpt	AE	Adverse Events	

ae.xpt tab:

ChatGPT 5

STUDYID	DOMAIN	USUBJID	AESEQ	AETERM	AESTDTC
<i>Study Identifier</i>	<i>Domain Abbreviation</i>	<i>Unique Subject Identifier</i>	<i>Sequence Number</i>	<i>Reported Term for the Adverse Event</i>	<i>Start Date/Time of Adverse Event</i>
<i>Char</i>	<i>Char</i>	<i>Char</i>	<i>Num</i>	<i>Char</i>	<i>Char</i>
50	50	50	8	50	50
STUDY1	AE	SUBJ001		1 HEADACHE	2023-01-05
STUDY1	AE	SUBJ002		1 NAUSEA	
STUDY1	AE	SUBJ003		1	
STUDY1	AE	SUBJ004		1 DIZZINESS	2023-03-10
STUDY1	AE	SUBJ005		1 FATIGUE	



Key takeaways



Key takeaways

- Chatbot integration is useful
 - Lower technical barrier and makes rules creation more accessible
 - Encourage broader participation
 - Save time and reduce errors
- Not perfect – further refinement needed
 - Rule validation excel template
 - Complex rule
- Future
 - AI creates complete custom rule package based on trial specifications
 - Clinical data platform with real time rule creation and execution



Want to try it out yourself?

<https://try.sgs.com/en/hn/cdisc-ai/>

Thank You!

Contact us

 clinicalresearch@sgs.com

 www.sgs.com/pharma

 sgs.com/linkedinpharma

 Visit us at booth 10

