

Single Day Event

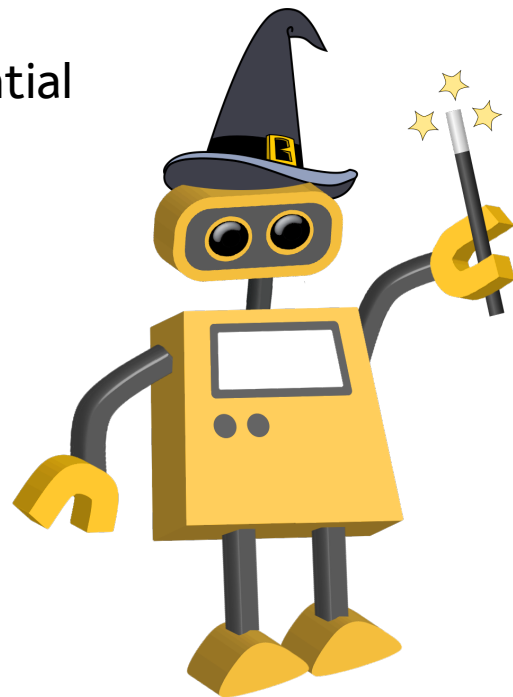
Starting Small: Incorporating ChatGPT into
Pharma Programming One Step at a Time

Jianing Wang, Wang Zhang
MSD



Introduction

- ChatGPT: revolutionizing the AI landscape
- Pharma industry: cautious but exploring AI potential
- ChatGPT as the sidekick to clinical programmers





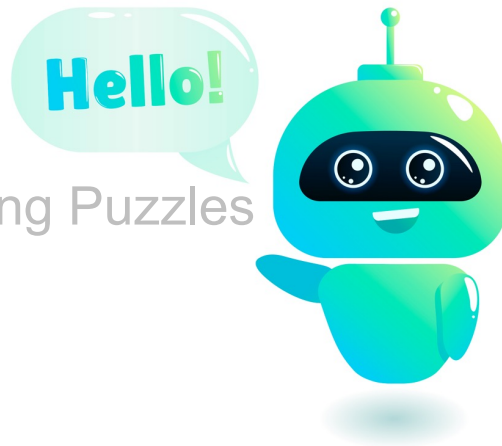
Contents

- I. Evolution of GPT Models
- II. Tips for Crafting Effective Prompts
- III. ChatGPT in Action - Solving Programming Puzzles
- IV. Discussion - The Future, AI and Us



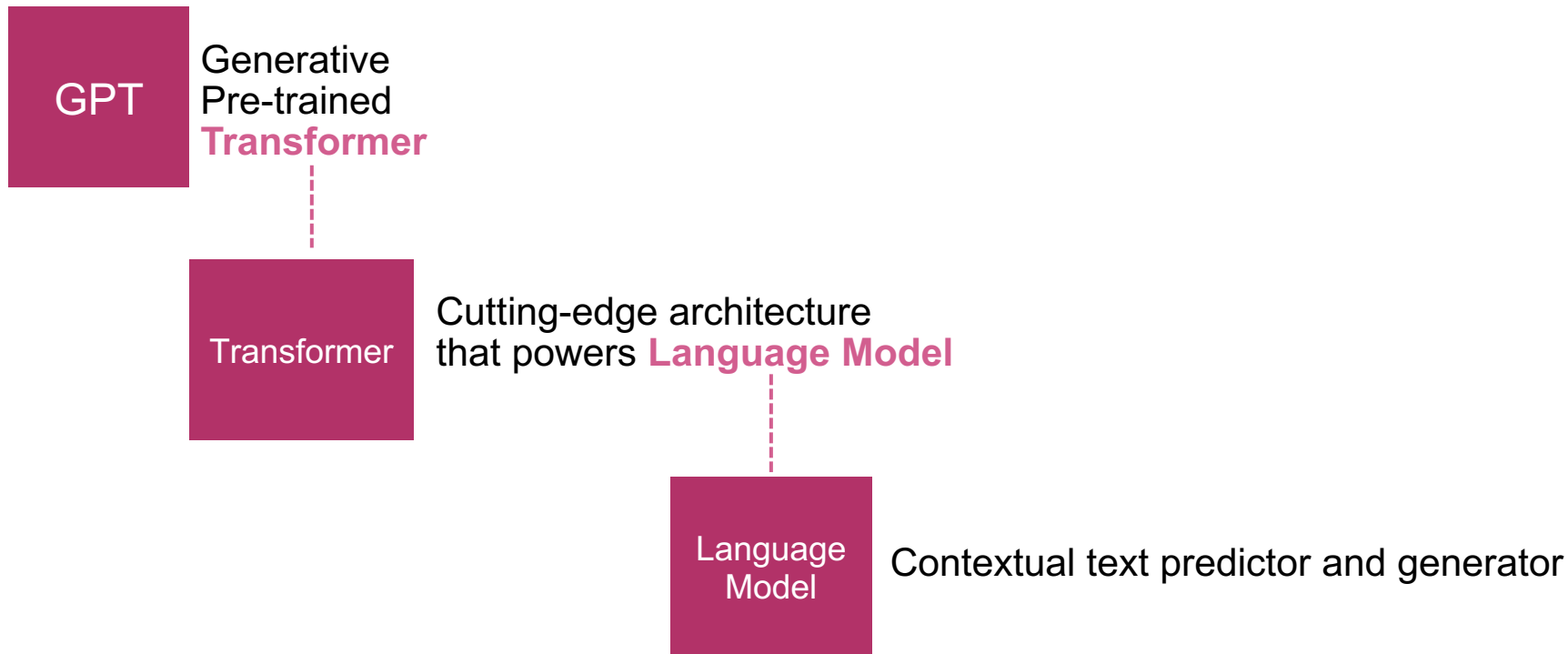
Contents

- I. Evolution of GPT Models
- II. Tips for Crafting Effective Prompts
- III. ChatGPT in Action - Solving Programming Puzzles
- IV. Discussion - The Future, AI and Us



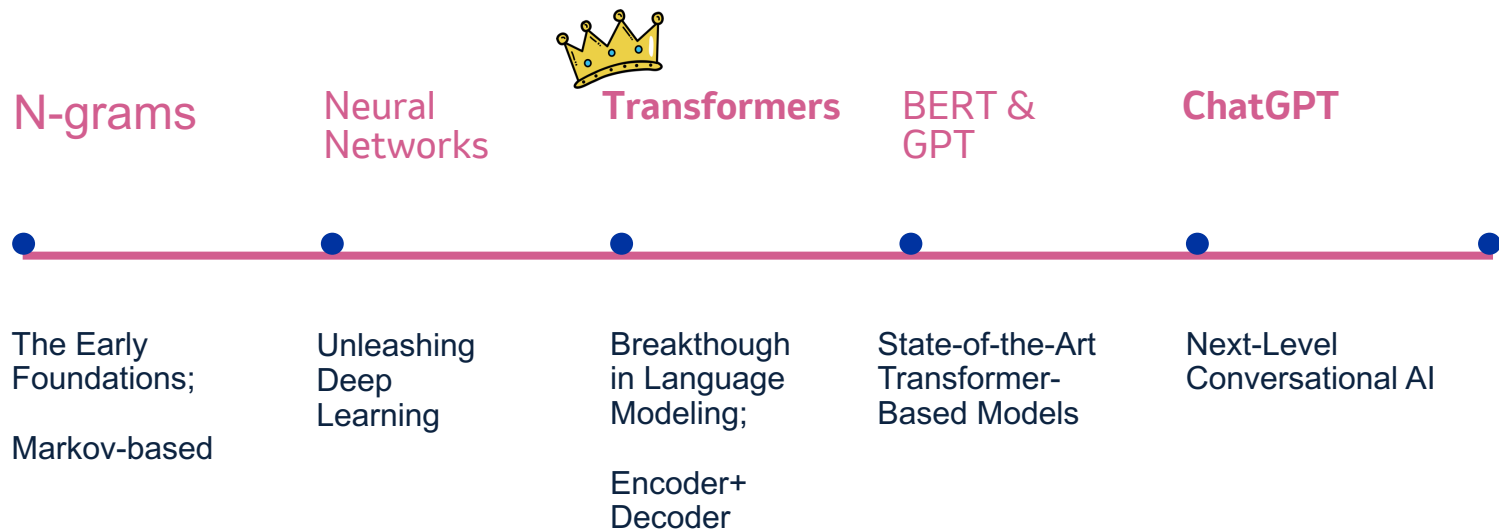


Why “ChatGPT”?





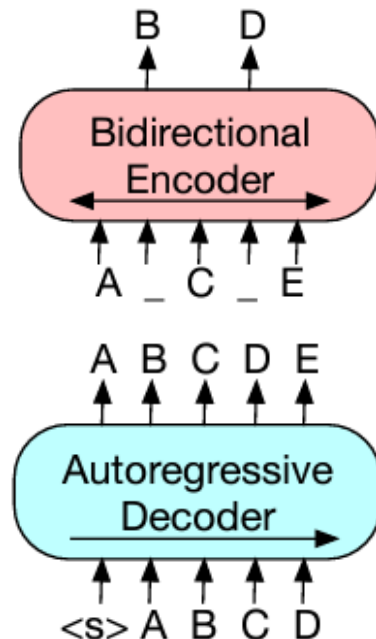
Key Evolution Milestones of Language Models



Protagonists in Transformer family - **BERT** & **GPT**

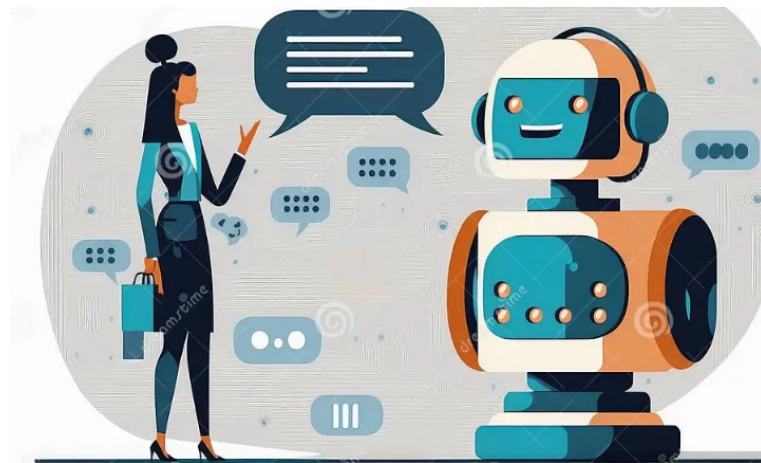
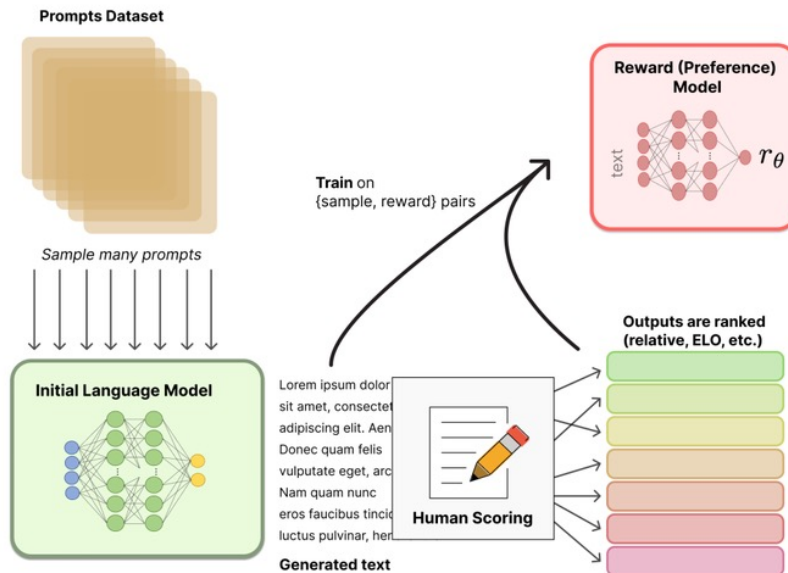
- **BERT** - the Meticulous Detective vs **GPT** - the swift action hero

Model	Architecture	Strengths
BERT	Bidirectional Encoder, No Decoder	1. Excels in understanding context 2. Great at sentiment analysis and entity recognition 3. Requires smaller task-specific training data
GPT	Left-to-right Decoder, No Encoder	1. Trains on large-scale text sources 2. Generates long, coherent sequences of human-like text 3. Exhibits emergent abilities like complex arithmetic and logical reasoning



The Hero of our tale - ChatGPT

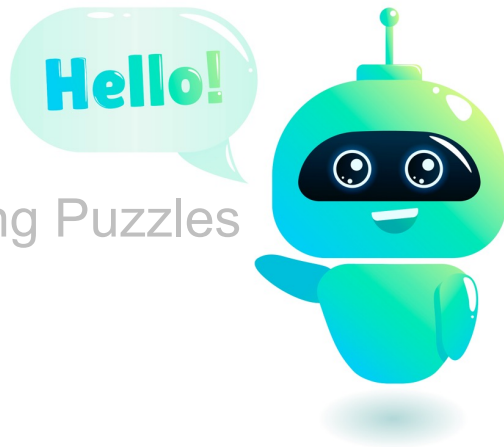
- GPT finetuned with Reinforcement Learning from Human Feedback (RLHF) for a more human-like conversation





Contents

- I. Evolution of GPT Models
- II. Tips for Crafting Effective Prompts**
- III. ChatGPT in Action - Solving Programming Puzzles
- IV. Discussion - The Future, AI and Us





Tips for Crafting Effective Prompts

- Prompt: Instruction to LMs to generate a response/complete a task
- Basic Prompt Framework: Clear Instructions and Relevant Context

Prompt Engineering Guide

Instruction (Required) - a specific task or instruction you want the model to perform

Context (Optional) - can involve sexternal information or additional context that can steer the model to better responses

Input Data (Optional) - is the input or question that we are interested to find a response for

Output Indicator (Optional) - indicates the type or format of the output.



Tips for Crafting Effective Prompts

- Essential Criteria for Prompt Design –
specificity, simplicity, and conciseness
- Effective Strategies:
 - a. Use delimiter (```, """ , ---, <>) to separate the instruction and context
 - b. Assign ChatGPT an identity (e.g., “Act as...”, “You are an expert in ...”) and intended audience
 - c. Avoid Impreciseness; specify all the pertinent details: desired format, length, style, and more. Consider to include examples.



Tips for Crafting Effective Prompts

- Example 1 – Text Sentiment Analysis and Classification



Classify the text into neutral, negative or positive. Text: The weather is too bad. Sentiment:



Text: The weather is too bad. Sentiment: negative



Classify the text into neutral, aa or bb. Text: I think the vacation is wonderful. Sentiment: bb. Text: you are selfish. Sentiment: aa.
Text: The weather is too bad. Sentiment:



Text: The weather is too bad. Sentiment: aa (negative)



Tips for Crafting Effective Prompts

- Advanced techniques : Few-Shot in-context learning

Surprising fact: The role of examples ; sometimes LESS is MORE.



Classify the text into neutral, negative or positive. Text: I think the vacation is wonderful. Sentiment: negative. Text: you are selfish. Sentiment: positive. Text: The weather is too bad. Sentiment:



Negative

Model used: gpt-4; Accumulated tokens: 1790; ResponseTime: 1.05



Classify the text into neutral, negative or positive. Text: I think the vacation is wonderful. Sentiment: aa. Text: you are selfish. Sentiment: bb. Text: The weather is too bad. Sentiment:



Negative

Tips for Crafting Effective Prompts

- Example 2 – Reasoning (arithmetic)

•  Q: The PT took Drug A 4mg/day for 5 consecutive days, then reduced to 2mg/day for 15 consecutive days. Calculate the total drug consumption in first 2 weeks.

 A: $4*5+2*9=38$ mg.

Tips for Crafting Effective Prompts

- Example 2 – Reasoning (arithmetic)

$$4*5+2*9=38 \text{ mg.}$$



The PT took Drug A 4mg/day for 5 consecutive days, then reduced to 2mg/day for 15 consecutive days. Calculate the total drug consumption in first 2 weeks. Do not explain.



56mg

Model used: gpt-4; Accumulated tokens: 699; ResponseTime: 18.04

56 mg



The PT took Drug A 4mg/day for 5 consecutive days, then reduced to 2mg/day for 15 consecutive days. Calculate the total drug consumption in first 2 weeks. Do not explain.



42mg

Model used: gpt-4; Accumulated tokens: 752; ResponseTime: 1.27

42 mg



The PT took Drug A 4mg/day for 5 consecutive days, then reduced to 2mg/day for 15 consecutive days. Calculate the total drug consumption in first 2 weeks. Do not explain.



42mg

42 mg





Tips for Crafting Effective Prompts

- Example 2 – Reasoning (arithmetic)

$$4*5+2*9=38 \text{ mg.}$$



Let's think step by step. The PT took Drug A 4mg/day for 5 consecutive days, then reduced to 2mg/day for 15 consecutive days. Calculate the total drug consumption in first 2 weeks. Do not explain.



38mg

Model used: gpt-4; Accumulated tokens: 1165; ResponseTime: 0.69

38 mg



Let's think step by step. The PT took Drug A 4mg/day for 5 consecutive days, then reduced to 2mg/day for 15 consecutive days. Calculate the total drug consumption in first 2 weeks. Do not explain.



38mg

Model used: gpt-4; Accumulated tokens: 1224; ResponseTime: 1.35

38 mg



Let's think step by step. The PT took Drug A 4mg/day for 5 consecutive days, then reduced to 2mg/day for 15 consecutive days. Calculate the total drug consumption in first 2 weeks. Do not explain.



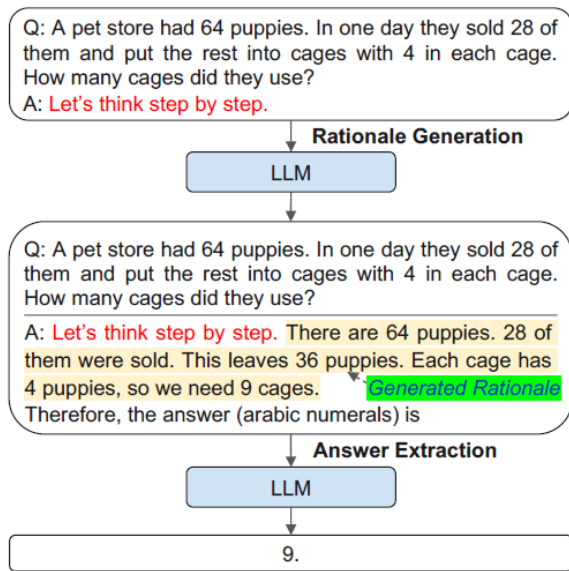
38mg

38 mg

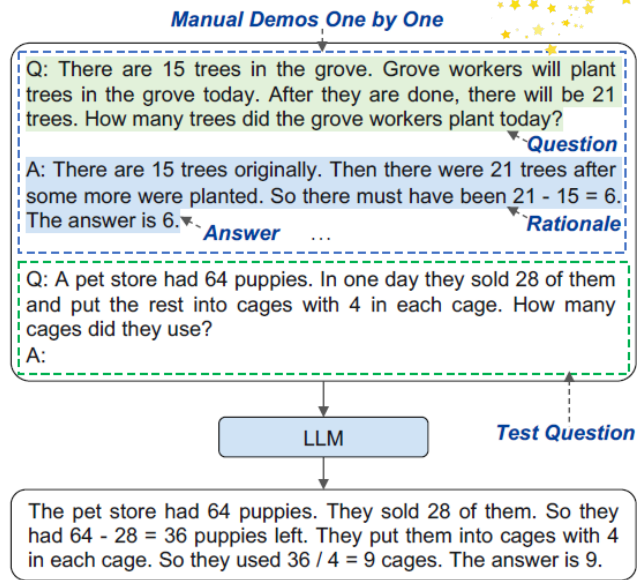


Tips for Crafting Effective Prompts

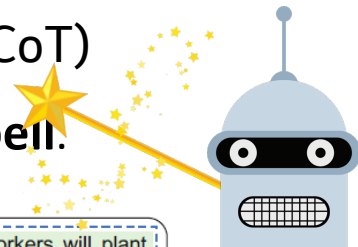
- Advanced techniques : Zero-shot chain of thoughts(CoT)
- Unshackling ChatGPT's Cognitive Potential through **Magic Spell**.



(a) Zero-Shot-CoT



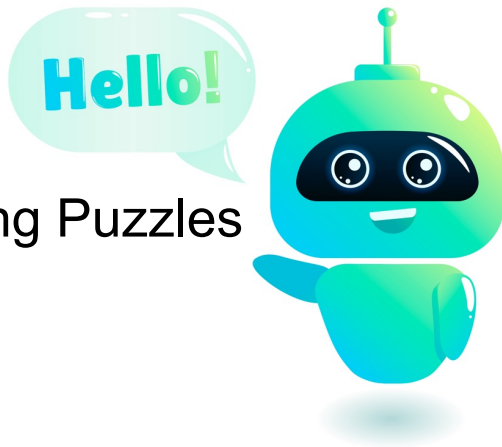
(b) Manual-CoT





Contents

- I. Evolution of GPT Models
- II. Tips for Crafting Effective Prompts
- III. ChatGPT in Action - Solving Programming Puzzles**
- IV. Discussion - The Future, AI and Us





ChatGPT in Action - Solving Programming Puzzles

- Practical Application 1- Information Extraction

Input	
Identify all potential date information from the input text and output them in a Markdown table with two columns: one for the raw text and one for the transformed date value in <u>ddmmmyyyy</u> format. Consider the following guidelines: 1. Correct any typos in the month section. 2.If any day/month/year values are missing but can be inferred from the context, impute those values. For example, if event A happened on 01Aug2020 and 2 weeks later event B happened on a Monday, output both dates: "01Aug2020" and "17Aug2020". 3.Use "xx", "xxx", or "<u>xxxx</u>" to impute missing values when there are no clues to infer the values. "" Patient presented to study site clinic in November,02, 2021. While discussing treatment plan patient informed site of 1st hospitalization from JUL 18 -21 2021, occurring after a surgery in March but cannot recall the specific date. Checking medical records, PT got injured on NORVEMBE-02 in 2002. This was confirmed with CT scan on 12.23 in 2021, and a re-check was performed at about 8 weeks later on a Wednesday in 2022.""	
Output	
Raw text	Transformed date
November,02, 2021	02Nov2021
JUL 18 - 21 2021	18Jul2021-21Jul2021
surgery in March	<u>xxMarxxxx</u>
NORVEMBE-02 in 2002	02Nov2002
CT scan on 12.23 in 2021	23Dec2021
8 weeks later on a Wednesday in 2022	16Feb2022



ChatGPT in Action - Solving Programming Puzzles

- Practical Application 2- Code Writing Assistance

Generate Regex Pattern from Given Text

Input

Create a SAS-compatible regular expression (regex) that can match all possible date formats found in the given text.

Consider the following guidelines:

1. The month part can be provided in full words (e.g., "January", "February"), three-letter abbreviations (e.g., "Jan", "Feb"), or 2-digit format (e.g., "01", "02").
2. The year part can be in either a 2-digit or 4-digit format. For 4-digit years, the first 2 digits should be either 19 or 20.
3. Separators between the day, month, and year are optional and can be one or multiple characters from the set `[ ,;-]` or the word "in".

Here is the text:

"" Patient presented to study site clinic in november,02-2021. While discussing treatment plan patient informed site of 2nd hospitalization from nov 18:2021. Checking medical records, PT got injured on november 2 in 02. This was confirmed with CT scan on 12.23 in 2021. ""

The desired dates that the regex should match:

"" november,02-2021 nov 18:2021 november 2 in 02 12.23 in 2021 ""

Output

```
\b(?:jan(?:uary)?|feb(?:ruary)?|mar(?:ch)?|apr(?:il)?|may|jun(?:e)?|jul(?:y)?|aug(?:ust)?|sep(?:tember)?|oct(?:ober)?|n
ov(?:ember)?|dec(?:ember)?|\d{1,2})[ ,;-]*(?:\d{1,2}[ ,;-]*|\b(?:in)\b[ ,;-]*){1,2}(?:19|20)?\d{2})\b
```



ChatGPT in Action - Solving Programming Puzzles

Practical Application 3- SDTM Mapping

Input

As an SDTM specialist familiar with CDISC SDTM IG 3.2, your task is to perform SDTM mapping for the adverse event page of a CRF. This involves mapping each clinical data collection field on the CRF to the corresponding variables within the CDISC SDTM AE domain. You will be provided with the following information regarding all variables specified in the CDISC SDTM AE domain: variable number, variable name, variable label, and CDISC notes (separated by commas).

""

1.STUDYID, Study Identifier, Unique identifier for a study.

2.DOMAIN, Domain Abbreviation, Two-character abbreviation for the domain.

3.USUBJID, Unique Subject Identifier, Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.

4.AESEQ, Sequence Number, Sequence Number given to ensure uniqueness of subject records within a domain. May be any valid number.

... (list all the variables from SDTM IG 3.2 AE domain)...

""

You are also provided with the CRF page that needs annotation. The input text will have two columns - one for the question number and the other for the question text. Please display the mapping result in a table with three columns: the first two columns will be the same as what provided (question number and question text), and the third column should contain the most appropriate SDTM AE variables to map to.

Note:

1.If it is not possible to find a suitable SDTM AE variable for mapping, map it to "SUPPAE" and format it as "QVAL when QNAM = 'AEXXXXXX' in SUPPAE". Replace "XXXXXX" with a 3-6 letter abbreviation of the question.

2.Ensure that column 3 matches either a variable name from CDISC SDTM AE domain (e.g., "AEOUT") or the exact format "QVAL when QNAM = 'AEXXXXXX' in SUPPAE".

3.Each CDISC SDTM AE variable should only be mapped to one CRF item, and column 3 should only contain one SDTM AE variable with the highest possibility for matching.

4.Present your answer in a Markdown table format.

""

1.Sequence Number

2.AE Term (For lab AE use the term 'increased' or 'decreased')

3.LLT

4.PT

5.SOC

6.Location of Reaction (For injection site AE only)

7.Maximum Size (For injection site AE only)

... (list all the fields in this legacy study's CRF AE page)...

""



ChatGPT in Action - Solving Programming Puzzles

- Practical Application 3- SDTM Mapping

Output (appending the original annotation in column 4 as for comparison)			
CRF Question Number	CRF Question Text	SDTM AE Variable	Original Annotation
1	Sequence Number	AESEQ	AESPID
2	AE Term (For lab AE use the term 'increased' or 'decreased')	AETERM	AETERM
3	LLT	AELLT	AELLT
4	PT	AEDECOD	AEDECOD
5	SOC	AESOC	AEBODSYS
6	Location of <u>Reaction</u> (For injection site AE only)	AELOC	AELOC
7	Maximum <u>Size</u> (For injection site AE only)	QVAL when QNAM = 'AEMAXSZ' in SUPPAE	QVAL when QNAM = 'AESIZE' in SUPPAE
8	Onset <u>Date</u> (Lab collection date if lab AE)	AESTDTC	AESTDTC
9	Outcome of AE	AEOUT	AEOUT
10	Intensity	AESEV	AESEV
11	Action taken with Study Medication	AEACN	AEACN
12	Relationship to Study Medication	AEREL	AEREL
13	Serious Criteria (check all that apply)	AESER	AESER
14	Serious Criteria of Inpatient hospitalization or prolongation of hospitalization	AESHOSP	AESHOSP
15	Serious Criteria of Cancer	AESCAN	AESCAN
16	Serious Criteria of Immediately life threatening	AESLIFE	AESLIFE

ChatGPT in Action - Solving Programming Puzzles

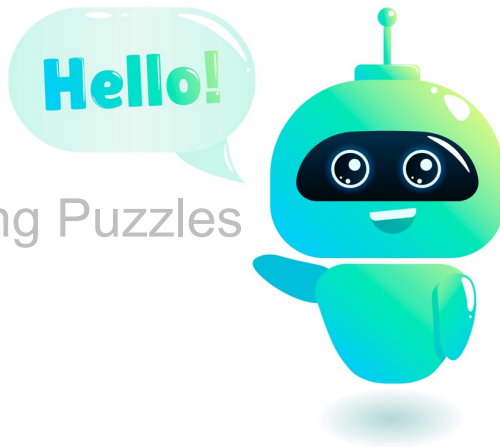
Practical Application 3- SDTM Mapping

17	Serious Criteria of Persistent or significant disability/incapacity	AESDISAB	AESDISAB
18	Serious Criteria of Congenital anomaly/birth defect	AESCONG	AESCONG
19	Serious Criteria of Overdose	AESOD	AESOD
20	Serious Criteria of <u>Other</u> important medical event	AESMIE	AESMIE
21	Serious Criteria of Death	AESDTH	AESDTH
22	Is the AE an Event of Clinical Interest?	QVAL when QNAM = 'CLINEV' in SUPPAE	QVAL when QNAM = 'AECLINT' in SUPPAE
23	Did the event improve or disappear after stopping study medication? (<u>Dechallenge</u>)	QVAL when QNAM = 'AEDEC' in SUPPAE	QVAL when QNAM = 'AESDCH' in SUPPAE
24	Did the event improve or disappear after stopping study medication? (<u>Rechallenge</u>)	QVAL when QNAM = 'AEREC' in SUPPAE	QVAL when QNAM = 'AESRCH' in SUPPAE
25	Relationship of event to any Other Suspect Therapy?	AERELNST	QVAL when QNAM = 'AERELNST' in SUPPAE
26	Briefly describe the presentation of the event including Signs and Symptoms	QVAL when QNAM = 'AESYM' in SUPPAE	QVAL when QNAM = 'AESS1' in SUPPAE
27	Describe the clinical course of the event	QVAL when QNAM = 'AEOUTCOM' in SUPPAE	QVAL when QNAM = 'AECLCRS1' in SUPPAE
28	Course of treatment for event	QVAL when QNAM = 'AETRAT' in SUPPAE	QVAL when QNAM = 'AECRSTRT' in SUPPAE
29	Evidence supporting <u>recovery</u> (return to baseline or description of new baseline)	QVAL when QNAM = 'AERECOV' in SUPPAE	QVAL when QNAM = 'AEOCEVID' in SUPPAE
30	Hospital Discharge Date	QVAL when QNAM = 'AEHOSDTC' in SUPPAE	QVAL when QNAM = 'AEHODDTC' in SUPPAE
31	List Discharge diagnosis(es)	QVAL when QNAM = 'AEDISDX' in SUPPAE	QVAL when QNAM = 'AEHODDGS' in SUPPAE
32	Provide follow-up to subject's clinical course	QVAL when QNAM = 'AEFUP' in SUPPAE	QVAL when QNAM = 'AECRSFU' in SUPPAE



Contents

- I. Evolution of GPT Models
- II. Tips for Crafting Effective Prompts
- III. ChatGPT in Action - Solving Programming Puzzles
- IV. Discussion - The Future, AI and Us





Discussion - The Future, AI and Us

Will AI take your job?

Unreplaceable:

- Human trait -Empathy, Intuition.
- Human expertise – Review and supervise

Best practices for Implementing Generative AI

- Perform repetitive tasks, unlock programmers' full potential on creative and strategic problem-solving work

