

Paper ET09

Navigating the Safe Integration of AI (ChatGPT) into Clinical Trials Programming Workflow

Mykyta Vysotskyi



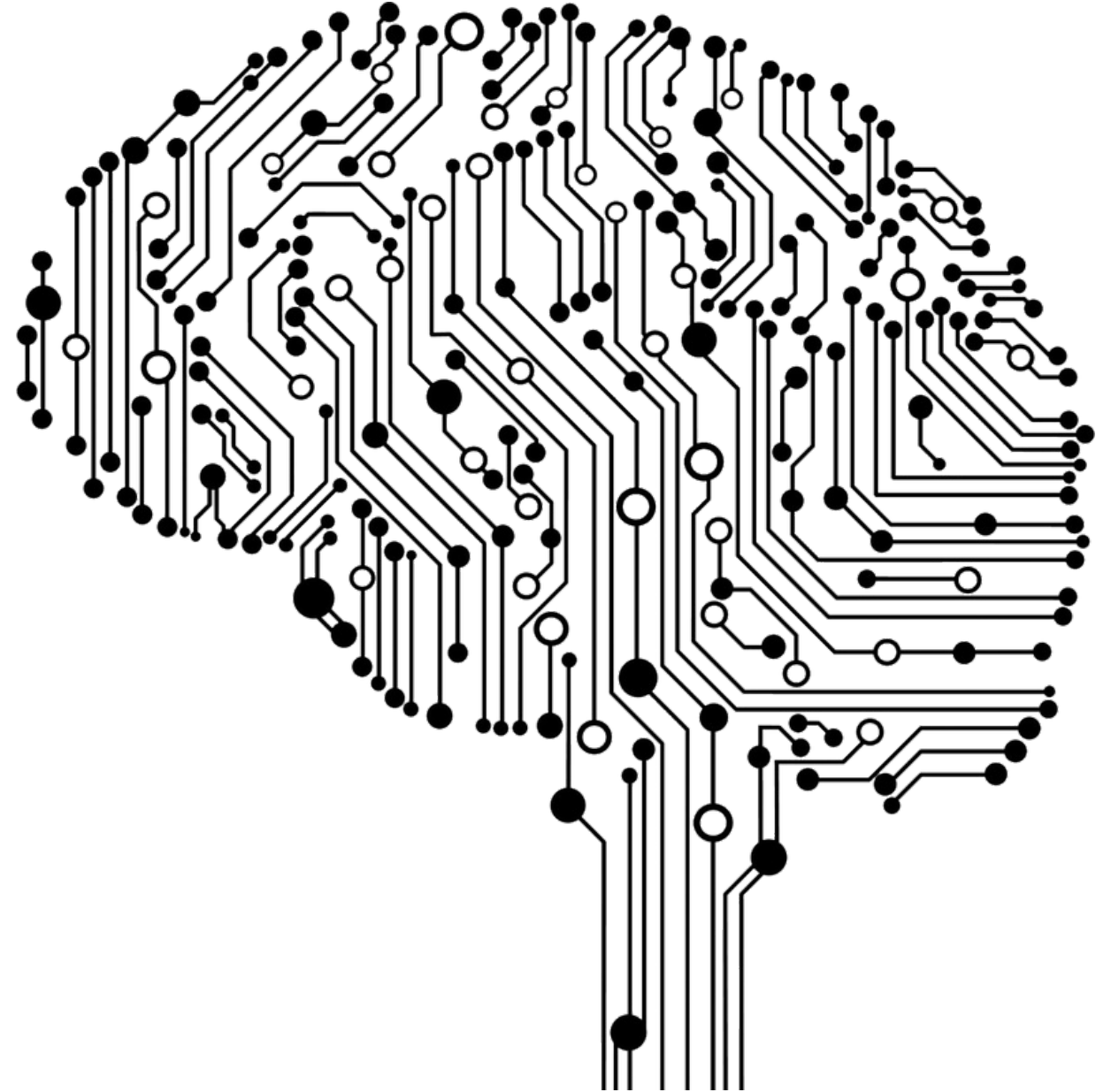
Bethesda, Maryland 2024

AGENDA

- Introduction
- What is ChatGPT 3.5
- How to Use ChatGPT 3.5
- Security and Confidentiality
- Practical Examples
- Conclusions



Introduction

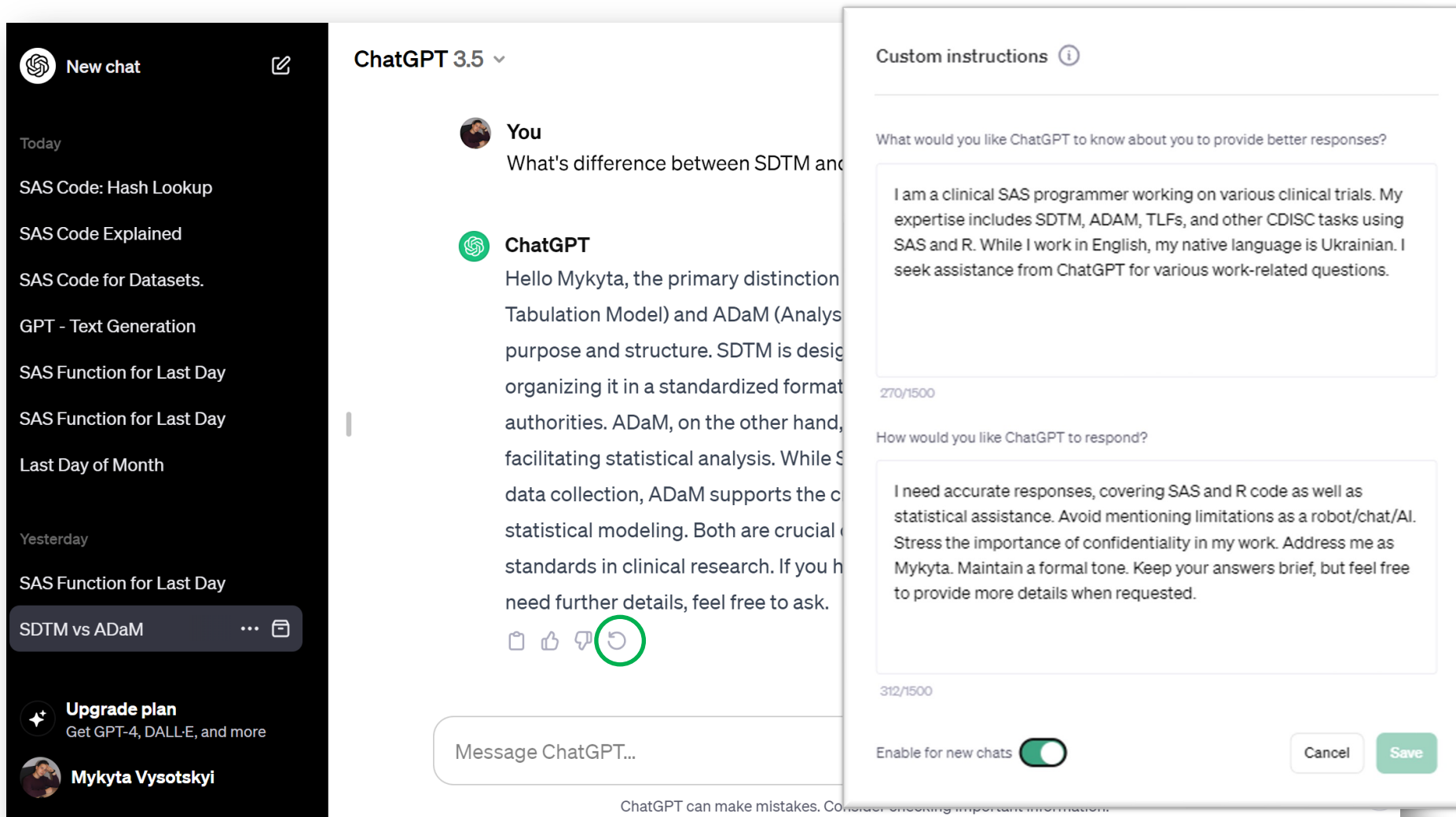


- **ChatGPT 3.5 is an advanced artificial intelligence language model developed by OpenAI, trained on a massive amount of text data from various sources on the internet available up to January 2022.**

What is ChatGPT 3.5

- **It is an iteration of the GPT (Generative Pre-trained Transformer) architecture, specifically designed for natural language understanding and generation tasks.**
- **ChatGPT 3.5 excels in generating human-like text responses, answering questions, completing prompts, and assisting users across a wide range of applications and industries.**

How to Use ChatGPT 3.5



The screenshot displays the ChatGPT 3.5 interface. On the left is a dark sidebar with a 'New chat' button and a list of previous conversations. The main chat area shows a conversation between 'You' and 'ChatGPT'. A 'Custom instructions' modal is open on the right, allowing users to provide context and preferences for their interactions.

ChatGPT 3.5

You
What's difference between SDTM and ADaM?

ChatGPT
Hello Mykyta, the primary distinction between SDTM (Standard Data Tabulation Model) and ADaM (Analysis Data Model) lies in their purpose and structure. SDTM is designed for organizing data in a standardized format for regulatory authorities. ADaM, on the other hand, facilitates statistical analysis. While SDTM focuses on data collection, ADaM supports the process of statistical modeling. Both are crucial standards in clinical research. If you need further details, feel free to ask.

Custom instructions

What would you like ChatGPT to know about you to provide better responses?

I am a clinical SAS programmer working on various clinical trials. My expertise includes SDTM, ADaM, TLFs, and other CDISC tasks using SAS and R. While I work in English, my native language is Ukrainian. I seek assistance from ChatGPT for various work-related questions.

How would you like ChatGPT to respond?

I need accurate responses, covering SAS and R code as well as statistical assistance. Avoid mentioning limitations as a robot/chat/AI. Stress the importance of confidentiality in my work. Address me as Mykyta. Maintain a formal tone. Keep your answers brief, but feel free to provide more details when requested.

Enable for new chats ☒

Cancel Save

- **Keep Patient Data Private:** Don't share any information that could identify patients.
- **Protect Business Confidentiality:** Talk about general topics only. Don't give out details about projects, companies, or drugs.
- **Check ChatGPT's Answers:** Always double-check what ChatGPT tells you, to make sure it's correct.
- **Use for General Advice:** ChatGPT is good for advice on methods and standards, but not for big decisions about research.
- **Be Careful with Files in ChatGPT 4:** Don't upload any files with confidential or sensitive information to ChatGPT 4.

Security and Confidentiality



Practical Example 1: Smart SAS Reference Guide

Pointing out the issue



You



You

Using your code I get the latest date of the NEXT month, when I need the latest day of a month based on input date. Please correct your code



ChatGPT

I apologize for the oversight, Mykyta. You're correct, and I appreciate your patience. Here is the corrected SAS code to obtain the last day of the month based on the input date:

```
sas
data _null_;
  input my_date mmddyy10.;
  last_day = intnx('month', my_date, 0, 'end');
  format last_day date9.;
  put last_day=;
  datalines;
01/15/2024
;
run;
```

By setting the third argument of the INTNX function to 0, we ensure that it returns the last day of the current month. Please test it with your input date, and let me know if you have any further questions.



the INTNX function to find the last day of the
example code:

```
1      OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
68
69      data _null_;
70          input my_date mmddyy10.;
71          last_day = intnx('month', my_date, 0, 'end');
72          format last_day date9.;
73          put last_day=;
74          datalines;
```

last_day=31JAN2024

Statement used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.00 seconds
memory	482.71k
OS Memory	19872.00k

e. The INTNX function is used with the 'month'
t day of the month. The result will be displayed

larification or assistance.

Practical Example 2: Assistance with Unfamiliar Syntax



You

Provide me alternative SAS code using proc sql



ChatGPT

Certainly, Mykyta. Here's an alternative SAS code using `proc sql` to achieve a similar result:

```
sas
Copy code

proc sql;
  create table hash_code as
  select A.*, B.Country
  from capitals as A
  left join work.CountryCapitals as B
  on A.capital = B.Capital;
quit;
```

This `proc sql` code performs a left join between the `capitals` dataset (aliased as A) and the `work.CountryCapitals` dataset (aliased as B) based on the "capital" and "Capital" variables, respectively. The result is stored in a new table named `hash_code`, which includes all columns from the `capitals` dataset along with the "Country" variable from the `work.CountryCapitals` dataset.

This SQL-based approach achieves a similar outcome to the hash object method, combining information from both datasets based on the common key. If you have any specific requirements or questions about this alternative, please let me know.

```
1 data CountryCapitals;
2   length Country $10 Capital $10;
3   input Country $ Capital $;
4   datalines;
5 USA Washington
6 Canada Ottawa
7 France Paris
8 Germany Berlin
9 Japan Tokyo
10 Australia Canberra
11 Brazil Brasilia
12 India NewDelhi
13 Italy Rome
14 SouthKorea Seoul
15 ;
16 run;
17
18 data Capitals;
19   set CountryCapitals;
20   if _N_ <= 5;
21   keep Capital;
22 run;
```

	Country	Capital
1	USA	Washington
2	Canada	Ottawa
3	France	Paris
4	Germany	Berlin
5	Japan	Tokyo
6	Australia	Canberra
7	Brazil	Brasilia
8	India	NewDelhi
9	Italy	Rome
10	SouthKorea	Seoul

Capital
1 Washington
2 Ottawa
3 Paris
4 Berlin
5 Tokyo

```
24 data hash_code;
25   if 0 then delete;
26   if _N_ = 1;
27   declare hash h;
28   h.definekey('Capital');
29   h.definedata('Country');
30   h.defineobj('Country');
31 end;
32 set capitals;
33 rc = h.put(Capital, Country);
34 if rc = 0;
35 run;
```

```
37 proc sql;
38   create table hash_code as
39   select *
40   from capitals
41   left join work.CountryCapitals
42   on capitals.Capital = work.CountryCapitals.Capital;
43 quit;
```

Practical Example 3: Code Transformation from SAS to R

```
/* Calculation of number of subject in treatment group (will be used as denominator for
percentage calculation) */
```

Part 2

```
# Calculation of number of subjects in the treatment group
```

```
s1 = freq out <- s1 %>%
```

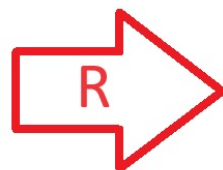
Part 2

AE	subjid	aedecod	aebodsys	trtp
1	001	Event1	SOC1	A
2	001	Event2	SOC2	A
3	001	Event1	SOC1	A
4	004	Event4	SOC4	B
5	005	Event5	SOC5	A
6	006	Event6	SOC1	B
7	007	Event7	SOC2	A
8	008	Event8	SOC3	B
9	009	Event9	SOC4	A
10	010	Event10	SOC5	B
11	011	Event1	SOC1	A
12	012	Event2	SOC2	B
13	012	Event3	SOC3	B
14	014	Event4	SOC4	B
15	015	Event1	SOC1	A
16	015	Event6	SOC1	A
17	017	Event7	SOC2	A
18	018	Event8	SOC3	B
19	019	Event9	SOC4	A
20	020	Event10	SOC5	B

SL	subjid	trtp
1	001	A
2	002	B
3	003	A
4	004	B
5	005	A
6	006	B
7	007	A
8	008	B
9	009	A
10	010	B
11	011	A
12	012	B
13	013	A
14	014	B
15	015	A
16	016	B
17	017	A
18	018	B
19	019	A
20	020	B



	aebodsys	aedecod	A	B
1	SOC1	Event1	3 (30.0)	0 (0.0)
2	SOC1	Event6	1 (10.0)	1 (10.0)
3	SOC2	Event2	1 (10.0)	1 (10.0)
4	SOC2	Event7	2 (20.0)	0 (0.0)
5	SOC3	Event3	0 (0.0)	1 (10.0)
6	SOC3	Event8	0 (0.0)	2 (20.0)
7	SOC4	Event4	0 (0.0)	2 (20.0)
8	SOC4	Event9	2 (20.0)	0 (0.0)
9	SOC5	Event10	0 (0.0)	2 (20.0)
10	SOC5	Event5	1 (10.0)	0 (0.0)



	AEBODSYS	AEDECOD	TRTP_A	TRTP_B
1	SOC1	Event1	3 (33.333)	0 (0.0)
2	SOC1	Event6	1 (11.111)	1 (10.000)
3	SOC2	Event2	1 (11.111)	1 (10.000)
4	SOC2	Event7	2 (22.222)	0 (0.0)
5	SOC4	Event9	2 (22.222)	0 (0.0)
6	SOC5	Event5	1 (11.111)	0 (0.0)
7	SOC3	Event3	0 (0.0)	1 (10.000)
8	SOC3	Event8	0 (0.0)	2 (20.000)
9	SOC4	Event4	0 (0.0)	2 (20.000)
10	SOC5	Event10	0 (0.0)	2 (20.000)



	AEBODSYS	AEDECOD	TRTP_A	TRTP_B
1	SOC1	Event1	3 (30.0)	0 (0.0)
2	SOC1	Event6	1 (10.0)	1 (10.0)
3	SOC2	Event2	1 (10.0)	1 (10.0)
4	SOC2	Event7	2 (20.0)	0 (0.0)
5	SOC4	Event9	2 (20.0)	0 (0.0)
6	SOC5	Event5	1 (10.0)	0 (0.0)
7	SOC3	Event3	0 (0.0)	1 (10.0)
8	SOC3	Event8	0 (0.0)	2 (20.0)
9	SOC4	Event4	0 (0.0)	2 (20.0)
10	SOC5	Event10	0 (0.0)	2 (20.0)

Updated 'header = TRUE'
to 'header = FALSE' in the
'read.table()' function

```
if b = "" then b = " 0 ( 0.0)";
keep AEBODSYS AEDECOD a b;
run;
```

Conclusions

- **ChatGPT 3.5 significantly enhances the capabilities of statistical programmers in clinical trials by providing immediate, free, and effective AI integration into daily tasks.**
- **Despite its potential, it is crucial to use ChatGPT with a high level of confidentiality and adherence to security and ethical standards in clinical research.**
- **ChatGPT does not replace human input but serves as a valuable tool for code generation, syntax explanation, and code transformation between SAS and R, subject to verification and adjustment by the user.**



stions

THANK YOU

Mykyta Vysotskyi
Intego Group LLC
Grzybowska 78, 00-844 Warsaw, Poland
Work Phone: +1 407.512.1006 (ext. 2480)
Email: mykyta.vysotskyi@intego-group.com
Web: www.intego-group.com