

BOOSTING SAS PROGRAMMING EFFICIENCY WITH CHATGPT: A CLINICAL TRIALS PERSPECTIVE

CM04 Presenter



ILLIA SKLIAR

Statistical Programmer/Analyst



ABOUT ME

Illia Skliar
Statistical Programmer/Analyst

Based in Kharkiv, Ukraine
Master's Degree in Computer Sciences

3+ years with Intego Group

Analyzing data
Guiding students in SAS programming
Performing music

Introduction



+



Today's Agenda

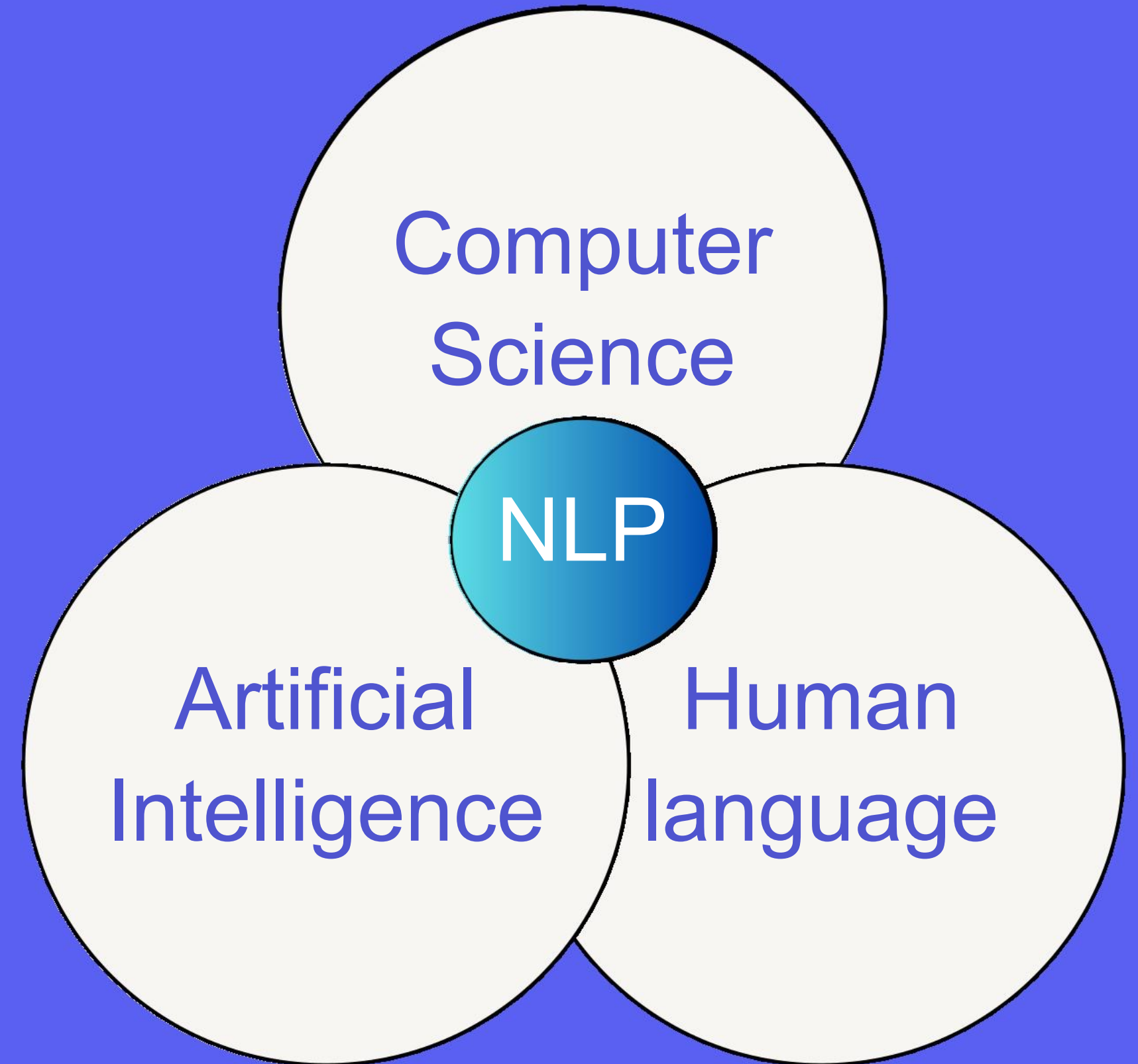


WHAT WE'LL LEARN:

- What is ChatGPT
- What tasks in clinical trials it can resolve
- How to integrate ChatGPT in SAS programming workflow
- What restrictions and limitations we have

UNDERSTANDING CHATGPT

- Key features
- Development History



What is ChatGPT?

This humongous dataset was used to form a deep learning neural network [...] modeled after the human brain—which allowed ChatGPT to learn patterns and relationships in the text data [...] predicting what text should come next in any given sentence.

A prompt and several model outputs are sampled.

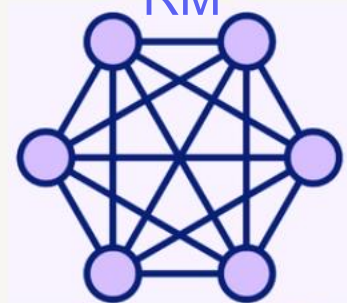
A labeler ranks the outputs from best to worst.

This data is used to train our reward model.


Explain the moon landing to a 6 year old

A B
Explain gravity... Explain war...
C D
Moon is a natural satellite of... People went to the moon...


D > C > A = B


RM
D > C > A = B



INTEGRATION OF CHATGPT IN SAS PROGRAMMING IDE

- Chat Interface within IDE
- Contextual Documentation & Help
- Custom Plugins & Extensions
- Interactive Debugging Assistance



ORGANIZATION

[Personal](#)[Settings](#)[Usage](#)[Rate limits](#)[Members](#)[Billing](#)

USER

[Settings](#)[API keys](#)

API keys

Your secret API keys are listed below. Please note that we do not display your secret API keys again after you generate them.

Do not share your API key with others, or expose it in the browser or other client-side code. In order to protect the security of your account, OpenAI may also automatically disable any API key that we've found has leaked publicly.

NAME	KEY	CREATED	LAST USED	
SAS Key	sk- . . . 9kW3	Aug 25, 2023	Sep 30, 2023	 

+ Create new secret key

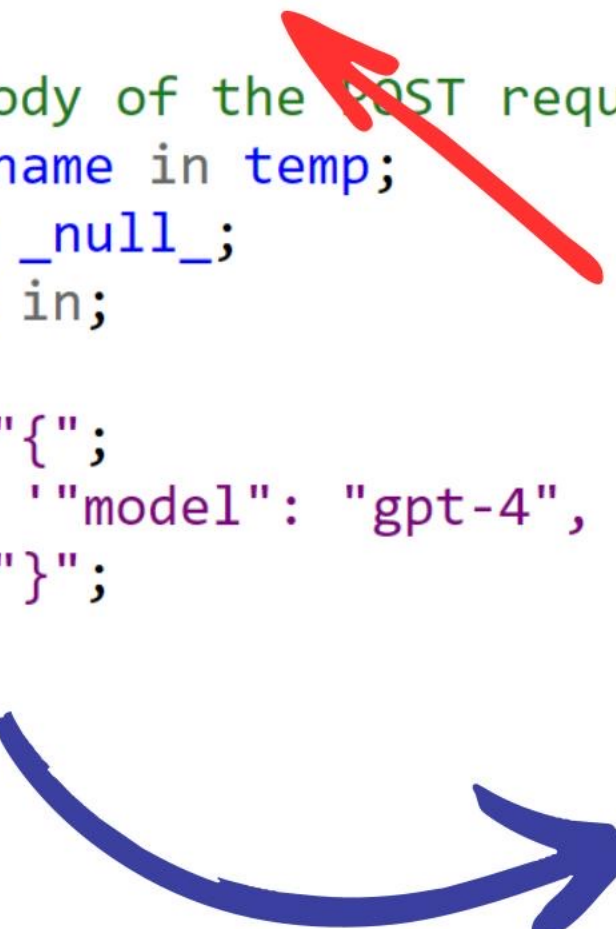
Default organization

If you belong to multiple organizations, this setting controls which organization is used by default when making requests with the API keys above.

Personal

Note: You can also specify which organization to use for each API request. See [Authentication](#) to learn more.

```
1 %macro RunGPTquery(query =);
2
3 %let API_KEY= sk-scZRL3AVhqC9I0Qhw[REDACTED]89kW3;
4 %let QUESTION = %str("%&query.%");
5
6 /* Body of the POST request */
7 filename in temp;
8 data _null_;
9 file in;
10 put;
11 put "{";
12 put '"model": "gpt-4", "messages": [{"role": "user", "content": '"&question }]";
13 put "}";
14 run;
15
16 /* reference that file as IN= parm in PROC HTTP POST */
17 filename resp "%sysfunc(getoption(WORK))/echo.json";
18 proc http
19 method="POST"
20 url="https://api.openai.com/v1/chat/completions"
21 ct="application/json"
22 in=in
23 out=resp;
24 headers "Authorization" = "Bearer &api_key.";
25 run;
26
27 libname response JSON fileref=resp;
28
29
30 proc report data=response.choices_message;
31 column content;
32 define content / display "" style(column)=[cellwidth=6in fontsize=12pt asis=ON];
33 run;
34
35 %mend RunGPTquery;
```



Setup and Call



```
1 %include "/home/u43526594/SAS & ChatGPT/RunGPTquery.sas";  
2  
3 %RunGPTquery(query = SAS code to create boxplot by treatment group and save it to .pdf format);  
4
```

Back-and-Forth to Double-check



```
``SAS
ods graphics on;

ods pdf file="/path_to_your_file/file_name.pdf";

proc sgplot data=YourDataSet;
  vbox variable_name / category=treatment_group;
  title "chart_title";
run;

ods pdf close;
ods graphics off;
```



```
1  /* Example data */
2  data treatment;
3      input TRT $ Value;
4      datalines;
5  A 5
6  A 7
7  A 9
8  B 4
9  B 6
10 B 8
11 C 2
12 C 3
13 C 4
14 ;
15
```

Back-and-Forth to Double-check



```
```SAS
ods graphics on;

ods pdf file="/path_to_your_file/file_name.pdf";

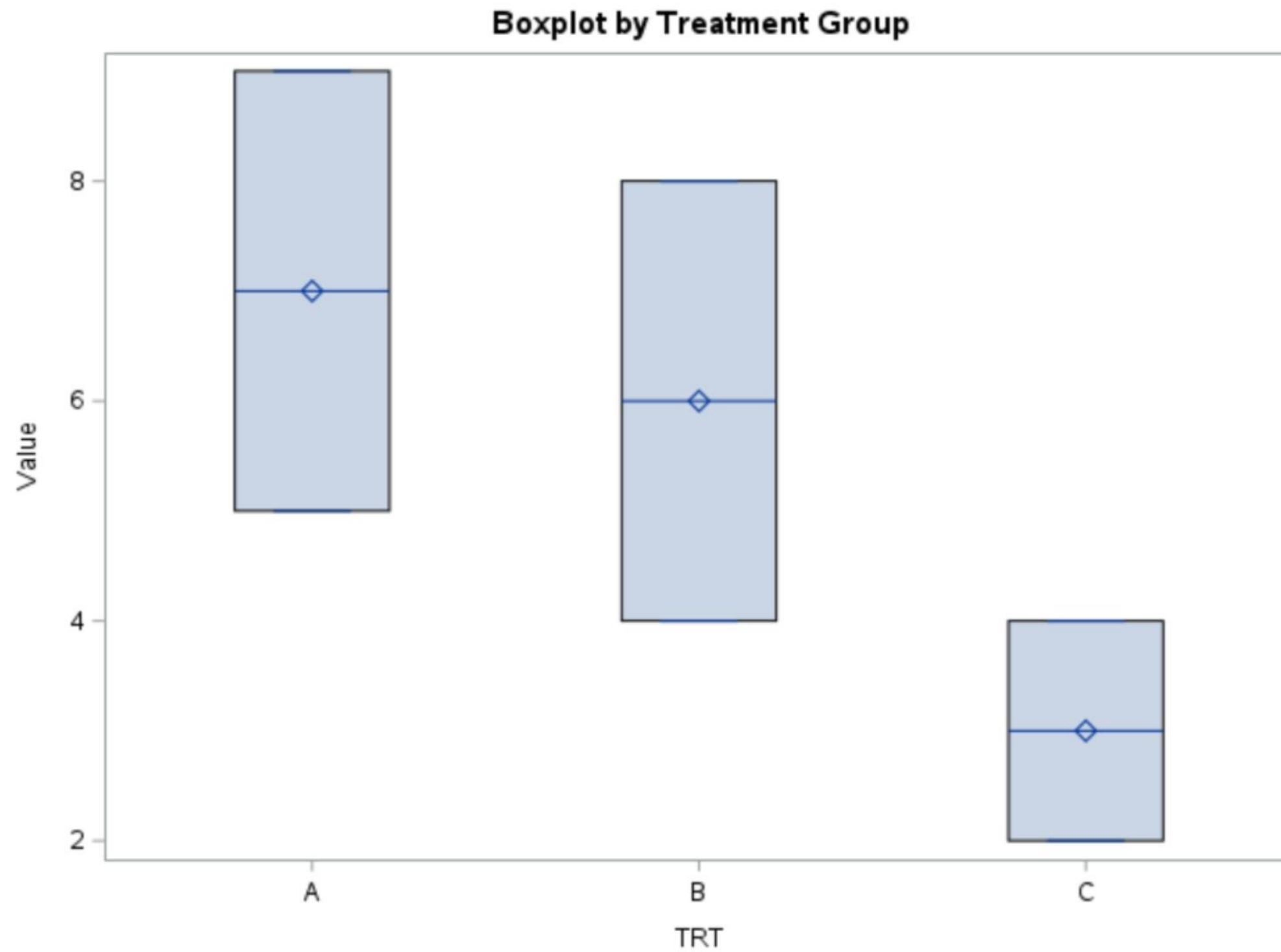
proc sgplot data=YourDataSet;
 vbox variable_name / category=treatment_group;
 title "chart_title";
run;

ods pdf close;
ods graphics off;
```
```



```
16 /*ChatGPT code*/
17 ods graphics on;
18 ods pdf file="/home/u43526594/SAS & ChatGPT/
19
20 proc sgplot data=treatment;
21     vbox Value / category=TRT;
22     title "Boxplot by Treatment Group";
23 run;
24
25 ods pdf close;
26 ods graphics off;
27 /*****
```

The Result

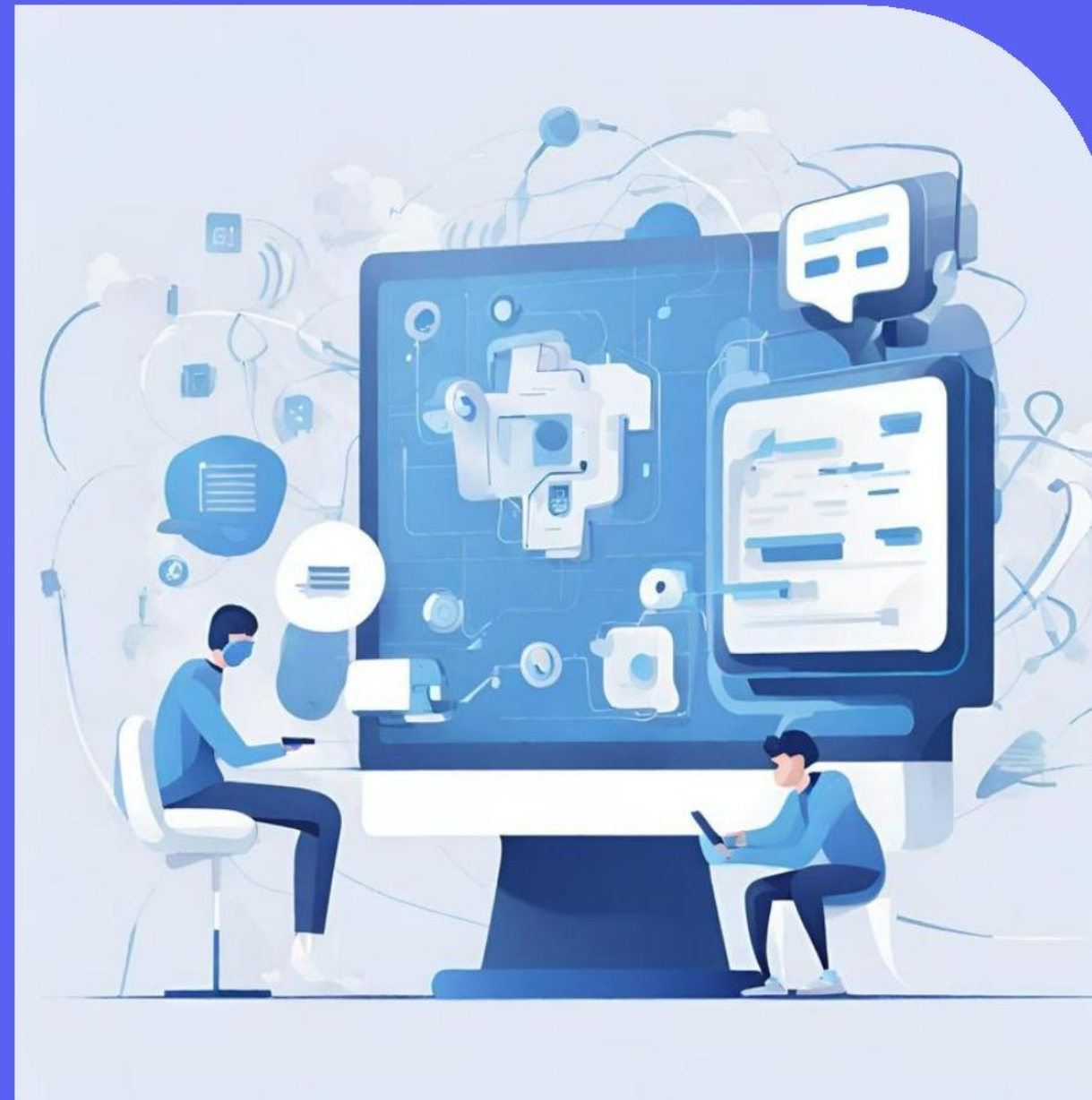


- 📁 SAS & ChatGPT
 - 📄 Check query result.sas
 - 📄 Main.sas
 - 📄 Main.sas~
 - 📄 my_boxplot.pdf ←
 - 📄 RunGPTquery.sas
 - 📄 RunGPTquery.sas~



TASK RESOLUTION WITH CHATGPT

- Create new variables
- Examine data
- Verify compliance with CDISC



```
libname SDTM "/home/u43526594/SDTM"; *MOCK DATA
```

```
proc contents data = SDTM.LB(drop = LBSEQ LBDTC) out = LB_VARS(keep = name label) noprint;
```

```
run;
```

```
data _null_;
```

```
    set LB_VARS end=last;
```

```
    length var_list $2000;
```

```
    retain var_list '';
```

```
    var_list = catx(' ', var_list, name);
```

```
    if last then call symputx('var_list', var_list);
```

```
run;
```

```
%put &var_list.;
```

```
%RunGPTquery(query = check if all the REQUIRED variables according to SDTM IG v.3.3 are present in my LB dataset  
The variables in my LB dataset are as follows: &var_list.);
```

```
libname SDTM "/home/u43526594/SDTM"; *MOCK DATA
```

```
proc contents data = SDTM.LB(drop = LBSEQ LBDTC) out = LB_VARS(keep = name label) noprint;
```

```
run;
```

```
data _null_;
```

```
set LB_VARS end=last;
```

```
length var_list $2000;
```

```
retain var_list '';
```

```
var_list = catx(' ', var_list, name);
```

```
if last then call symputx('var_list', var_list);
```

```
run;
```

```
%put &var_list.;
```

```
%RunGPTquery(query = check if all the REQUIRED variables according to SDTM IG v.3.3 are present in my LB dataset  
The variables in my LB dataset are as follows: &var_list.);
```

THE RESULT



Comparing this list to your dataset, it seems you're missing the LBSEQ variable which is crucial for the order of observations.

Also, LBDTC (Date/Time of Observation/Collection) is a required variable and needs to be included even though you have LBDY



```
47      do,  
48          if input(substr(AEENDTC,6,2),best.)=input(substr(RFXENDT  
49              AENDTF='D';  
50      end;  
51  
52      ASTDY=ASTDT-TRTSDT+(ASTDT>=TRTSDT);  
53      if not missing(AENDT) then AENDY=AENDT-TRTSDT+(AENDT>=TRTSDT);  
54  run;
```



```
47      do,
48          if input(substr(AEENDTC,6,2),best.)=input(substr(RFXENDT
49              AENDTF='D';
50      end;
51
52      ASTDY=ASTDT-TRTSDT+(ASTDT>=TRTSDT);
53      if not missing(AENDT) then AENDY=AENDT-TRTSDT+(AENDT>=TRTSDT);
54  run;

55
56  %let code = %nrstr(ASTDY=ASTDT-TRTSDT+(ASTDT>=TRTSDT));
57  %RunGPTquery(query = Could you create derivation for my specification using this code &code.
58      and check it with IG compliance);
59
```



```

47      do,
48          if input(substr(AEENDTC,6,2),best.)=input(substr(RFXENDT
49              AENDTF='D';
50      end;
51
52      ASTDY=ASTDT-TRTSDT+(ASTDT>=TRTSDT);
53      if not missing(AENDT) then AENDY=AENDT-TRTSDT+(AENDT>=TRTSDT);
54  run;

```



```

55
56 %let code = %nrstr(ASTDY=ASTDT-TRTSDT+(ASTDT>=TRTSDT));
57 %RunGPTquery(query = Could you create derivation for my specification using this code &code.
58     and check it with IG compliance);
59

```



ASTDY is derived based on the input variables ASTDT (start date of abnormal status) and TRTSDT (start date of treatment).

If ASTDT is not missing and TRTSDT is not missing, ASTDY is the difference in days between ASTDT and TRTSDT.

If ASTDT is on or after TRTSDT, a day is added to that difference.

Here is the step-by-step process, and its compliance with CDISC/CDASH standards for clinical trials (IG = Implementation Guide).

1. First, check that ASTDT (start date of abnormal status) and TRTSDT (start date of treatment) are not missing. According to CDISC/CDASH standards, if both are not missing, the calculation is generally compliant.
2. Subtract TRTSDT from ASTDT to get the difference in number of days.
3. Check if ASTDT is on or after TRTSDT. If it is, add one (+) to the difference in days. If it is not, don't add anything.
4. The result is ASTDY, the number of days from the start date of treatment to the start date of abnormal status. If ASTDT is the same as TRTSDT, the result is 0.

In terms of the IG compliance, this calculation is generally compliant with the CDISC/CDASH standards. The calculation is transparent and auditable.

Back-and-Forth Again

```
60 %let rules = %nrstr(  
61 If ASTDT is not missing and TRTSDT is not missing, ASTDY is the difference in days between  
62 Here is the step-by-step process, and its compliance with CDISC/CDASH standards for clinic  
63 1. First, check that ASTDT (start date of abnormal status) and TRTSDT (start date of treat  
64 According to CDISC/CDASH standards, dates and times should be input in ISO 8601 format (YY  
65 The derivation cannot be performed if either of these variables is missing.  
66 2. Subtract TRTSDT from ASTDT to get the difference in number of days.  
67 3. Check if ASTDT is on or after TRTSDT. If it is, add one (+) to the difference in days.  
68 4. The result is ASTDY, the number of days from the start date of treatment to the start d  
69 If ASTDT is the same as or later than TRTSDT, it is counted as an additional day.  
70 );  
71 %put &rules.;  
72  
73 %RunGPTquery(query = create SAS code to derive ASTDY variable using this rules: &rules.);  
74
```

Back-and-Forth Again



Following these instructions, SAS code that meets the CDISC/CDASH standards for clinical trials could look like this:

```
```SAS
DATA new_dataset;
SET old_dataset;
*Check that ASTDT and TRTSDT are not missing; ✓
IF NOT (MISSING(ASTDT) OR MISSING(TRTSDT)) THEN DO;
 *Calculate the difference between ASTDT and TRTSDT in days;
 ASTDY = ASTDT - TRTSDT;

 *Check if ASTDT is on or after TRTSDT. If it is, add one to the difference in days;
 IF ASTDT >= TRTSDT THEN ASTDY = ASTDY + 1; ✓
END;
RUN;
```
```

This code would create a new variable, 'ASTDY', in the 'new_dataset'. If either 'ASTDT' or 'TRTSDT' is missing, 'ASTDY' would remain missing.

Please note that the dates should be in SAS date format, which is the number of days since January 1, 1960. If the date is in other format, you might need to first convert them.

Also, it is good practice to handle missing values with care in SAS. If 'ASTDT' or 'TRTSDT' are missing, the code won't produce any error but it won't create the 'ASTDY' in the new dataset.

Finally, the raw data set is named "old_dataset" here as a placeholder. Please replace it with your actual data set name.

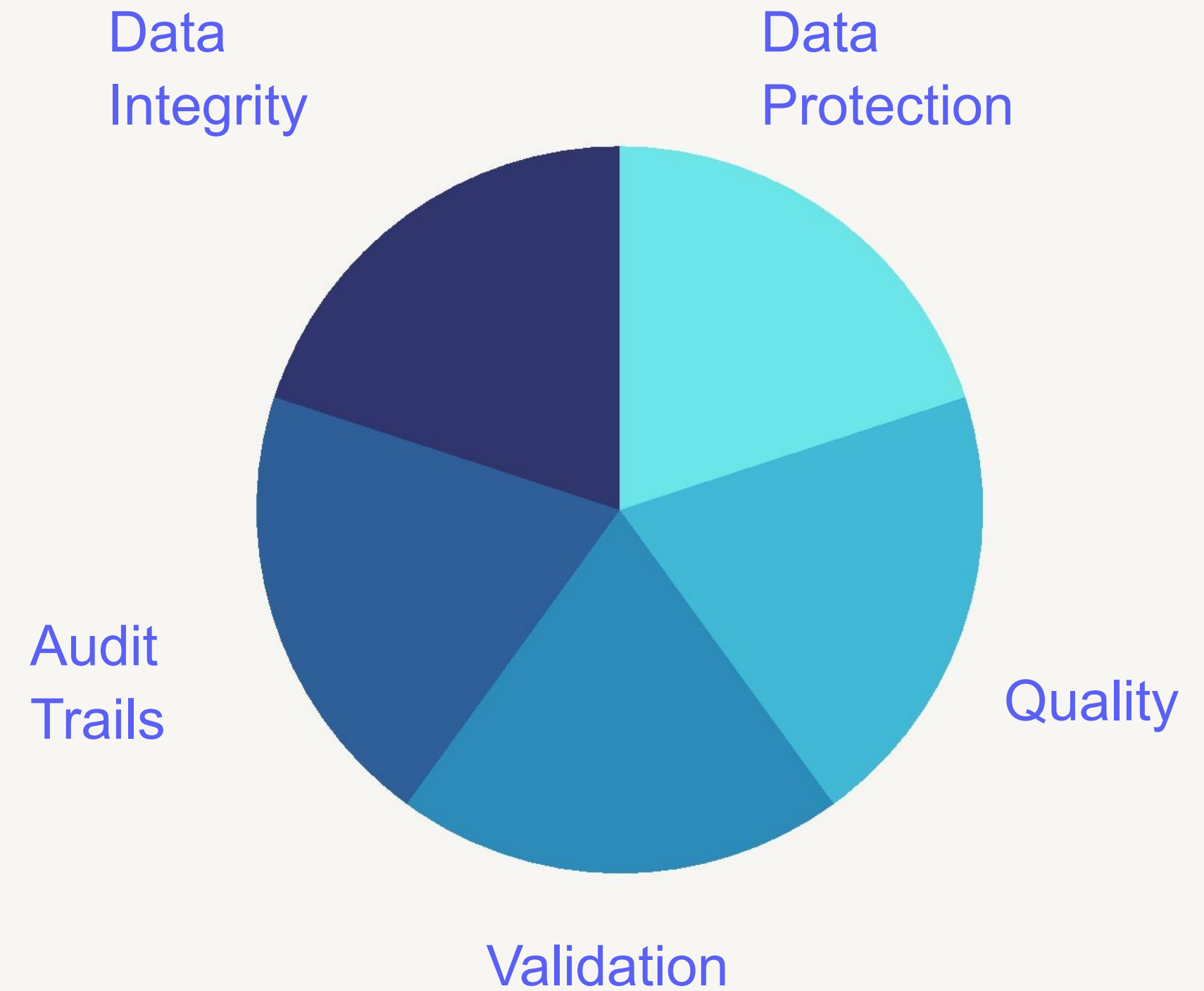
RESTRICTIONS & CONFIDENTIALITY

- Good Clinical Practice, ICH E6(R2)
- The Code of Federal Regulations Title 21
- Other Instruments to Provide Security

“TO PROTECT THE RIGHTS, SAFETY AND WELLBEING OF RESEARCH PARTICIPANTS AND TO SIMPLIFY AND HARMONISE REGULATORY PROCESSES”

Regulatory Framework

CM04



TITLE 21--FOOD AND DRUGS
CHAPTER I--FOOD AND DRUG ADMINISTRATION
DEPARTMENT OF HEALTH AND HUMAN SERVICES
SUBCHAPTER A - GENERAL

PART 50 PROTECTION OF HUMAN SUBJECTS

TITLE 21--FOOD AND DRUGS
CHAPTER I--FOOD AND DRUG ADMINISTRATION
DEPARTMENT OF HEALTH AND HUMAN SERVICES
SUBCHAPTER A - GENERAL

PART 11 ELECTRONIC RECORDS; ELECTRONIC SIGNATURES

TITLE 21--FOOD AND DRUGS
CHAPTER I--FOOD AND DRUG ADMINISTRATION
DEPARTMENT OF HEALTH AND HUMAN SERVICES
SUBCHAPTER A - GENERAL

PART 56 INSTITUTIONAL REVIEW BOARDS

TITLE 21--FOOD AND DRUGS
CHAPTER I--FOOD AND DRUG ADMINISTRATION
DEPARTMENT OF HEALTH AND HUMAN SERVICES
SUBCHAPTER A - GENERAL

PART 54 FINANCIAL DISCLOSURE BY CLINICAL INVESTIGATORS



AI

NLP

GPT

Summary



Efficacy and accuracy of ChatGPT



Profound Effects



Integrating into IDE



Practical Solutions

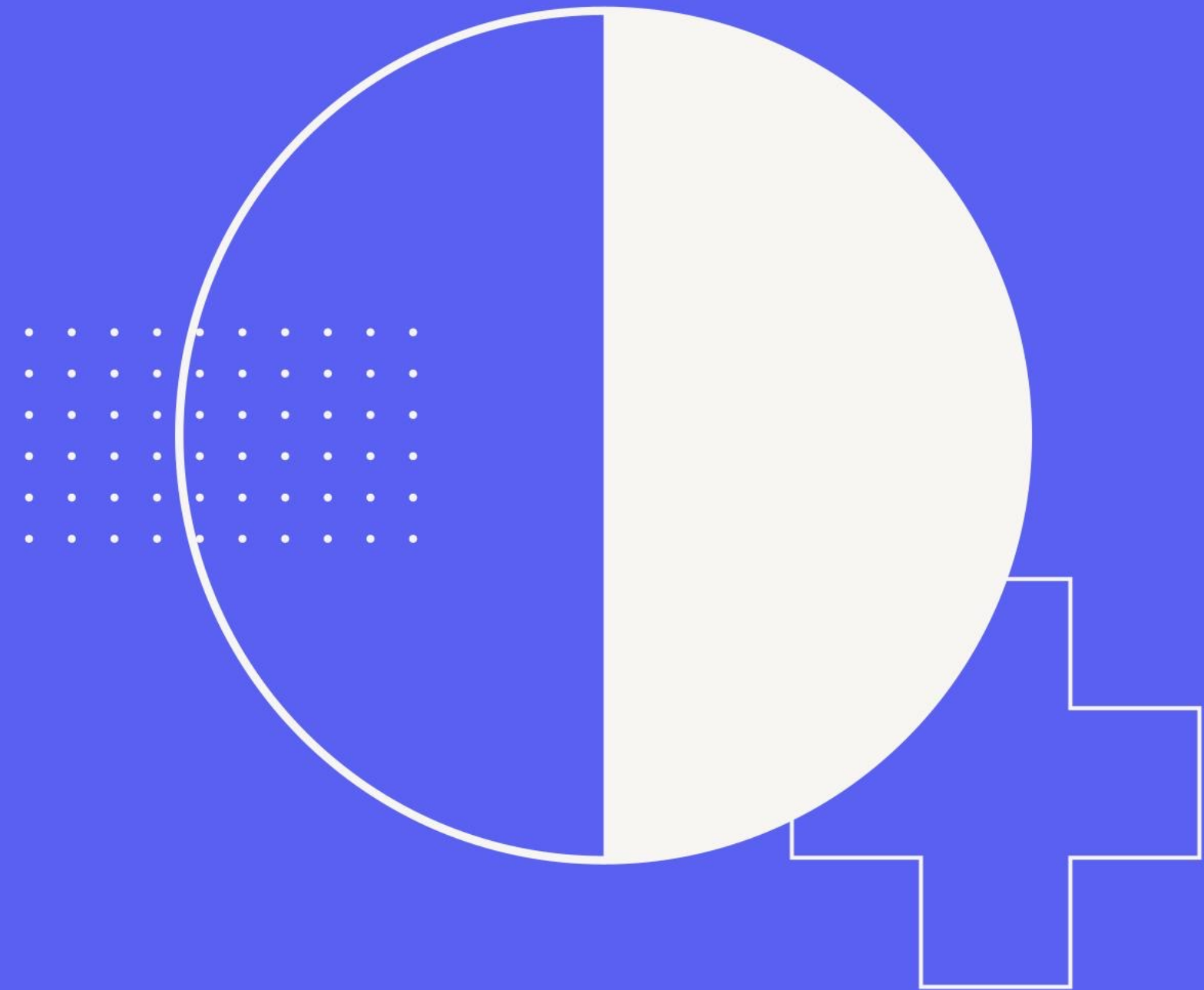


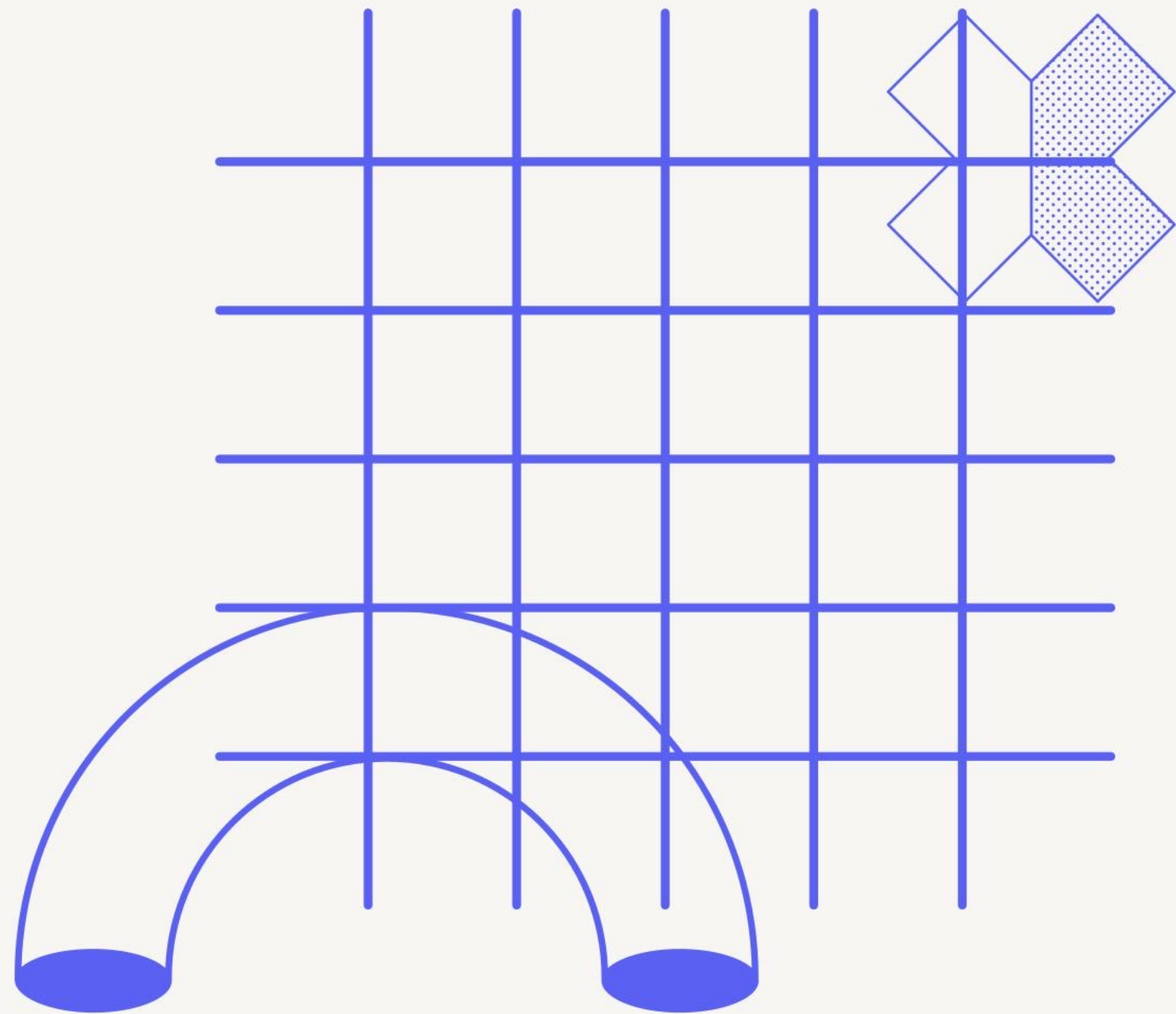
Data Privacy

Q&A

Please remain in our Hall to dive deeper into the world of GPT models

and explore the question:
“Will AI Chatbots Replace Statistical Programmers?”





Thank you!

Please contact me at
Illia.Skliar@intego-group.com
if you have any questions.