Cytel

Can AI cure the lazy programmer?

Shaun Linde, Cytel Inc.

PHUSE SDE

22 August 2024

Table of Contents

Cytel

Benefits of AI-Assisted Coding

Improved Accuracy

Copying code is on the increase, 7-23% of code is copied *(Stevens, 2022)*
AI generated codes runs ~90% without any errors *(Francis, 2023)*
Faster at detecting errors than the human eye
Doesn't get tired and no coffee breaks needed

Improved Efficiency

Automate repetitive tasks and processes (i.e., log and code reviews)
Convert natural language to code
English is the new programming language

Code Optimisation

Restructuring of old code and adding comments
Identify areas that can be improved (i.e., Code that runs faster)
High code quality; Convert SAS code to R
Easy to adopt

Growth

Exposure to new code and algorithms
Advanced programmers can advance more

Cytel

Derive AGE using ChatGPT

Create derivation for age using date of birth and informed consent using sas code

```sas
data patients_with_age;
    set patients;

    /* Calculate age */
    age = intck('year', dob, consent_date)
          - (put(consent_date, yymmdd10.) < put(dob, yymmdd10.));

    /* Derive informed consent flag */
    if consent_date ne . then informed_consent = 1;
    else informed_consent = 0;

    format dob consent_date yymmdd10.;
run;
```

Accurate:
- Knows date of birth is before date of informed consent
- No coding errors

Efficient:
- Generated in seconds
- Repeatable

Optimisation:
- Added Comments
- Code structure

Growth:
- Exposure to new code and functions

Cytel

Baseline demographics table with patients randomised using ChatGPT

Generate sas code for a baseline demographics table for randomised patients using trt01pn as treatment variable and adam.adsl dataset

Step 1: Filter the dataset

```sas
/* Filter to include only randomized patients */
data randomized_data;
    set adam.adsl;
    if trt01pn ne .;
run;
```

Step 3: Merge and Format Results

```
/* Merge continuous and categorical summaries */
data baseline_demographics;
    merge summary_cont summary_gender summary_race;
    by trt01pn category;
run;
```

Step 2: Calculate summary statistics

```
/* Summarize continuous variables for randomized subjects */
proc means data=randomized_data mean std median min max n;
    class trt01pn;
    var age;
    output out=summary_cont (drop=_type_ _freq_)
        mean= mean_age
        std= std_age
        median= median_age
        min= min_age
        max= max_age
        n= n_age;
run;
```

```
/* Summarize categorical variables for randomized subjects */
proc freq data=randomized_data;
    tables trt01pn*sex / nocol nopercent out=summary_gender;
    tables trt01pn*race / nocol nopercent out=summary_race;
run;
```

Cytel

Explaining code using ChatGPT

What does the following code mean "test=prxchange("s/[^0-9]//",-1,var);"

```
/*
This line of code uses the prxchange function to remove all non-numeric characters
from the variable 'var'. The prxchange function applies a Perl regular expression
substitution.

- The first argument, "s/[^0-9]//", specifies the regular expression:
  - 's/' indicates a substitution operation.
  - '[^0-9]' is a character class that matches any character that is not a digit (0-9).
  - '//' indicates that these non-numeric characters should be replaced with an empty
    string, effectively removing them.
- The second argument, '-1', specifies that the substitution should be applied globally,
  meaning to all occurrences of non-numeric characters in the input string 'var'.

- The result of this operation is assigned to the variable 'test'.
*/
test = prxchange("s/[^0-9]//", -1, var);
```

Commenting and restructuring code using ChatGPT

Comment and restructure code "proc sql;
create table lb1 as
select lb.*, adsl.trtsdt, adsl.trtedt, adsl.subjid, adsl.saffl from raw.lb as lb, adam.adsl as adsl
where lb.subject = adsl.subjid;
run;"

Why do programmers not comment code?

1. **Time Constraints:**
   - Programmers often work under tight deadlines and may prioritize writing and delivering functional code over adding comments.

```sas
proc sql;
  /* Create a new table named 'lb1' */
  create table lb1 as
  select
    /* Select all columns from the 'lb' dataset */
    lb.*,
   /* Select 'trtsdt', 'trtedt', 'subjid', and
   'saffl' from the 'adsl' dataset */
    adsl.trtsdt,
    adsl.trtedt,
    adsl.subjid,
    adsl.saffl
  from
    /* Specify the datasets to be joined */
    raw.lb as lb,     /* 'raw.lb' dataset */
    adam.adsl as adsl /* 'adam.adsl' dataset */
  where
   /* Join condition: match 'subject' in 'lb'
   with 'subjid' in 'adsl' */
    lb.subject = adsl.subjid;
run;
```

AI Hazards

AI Hallucination

False interpretations of data
See patterns or objects that are nonexistent
Tries to fill in the gaps

Prompt Engineering

Practice of developing prompts; becoming a foundation
A new language needs to be learnt
Basic understanding of AI model

Biases

Human bias = AI bias
Amplify negative results
Hinders AI's potential
Training and review – snowball effect, need to service your AI model

Regulations

The EU AI Act *(Meier, 2024)*
Goal is for harmonized rules
Respects human rights and values - am I being AI'd?

Cytel

AI Hallucination

AI hallucination is a phenomenon wherein a large language model (LLM)—often a generative AI chatbot or computer vision tool—perceives patterns or objects that are nonexistent or imperceptible to human observers, creating outputs that are nonsensical or altogether inaccurate *(IBM, 2024)*.

What causes AI hallucinations?

Data Quality. An AI model is only as good as the data it's trained on. External data that is not fact checked.

Overfitting. When an AI model is trained on a limited dataset, it may memorise the inputs and appropriate outputs.

Incomplete Prompting. If a prompt contains idioms or slang expression that AI model hasn't been trained on.

Similar to how humans imagine patterns in the clouds

Cytel

AI Hallucinations *(Fokina, 2024)*

How do AI hallucinations make you feel?

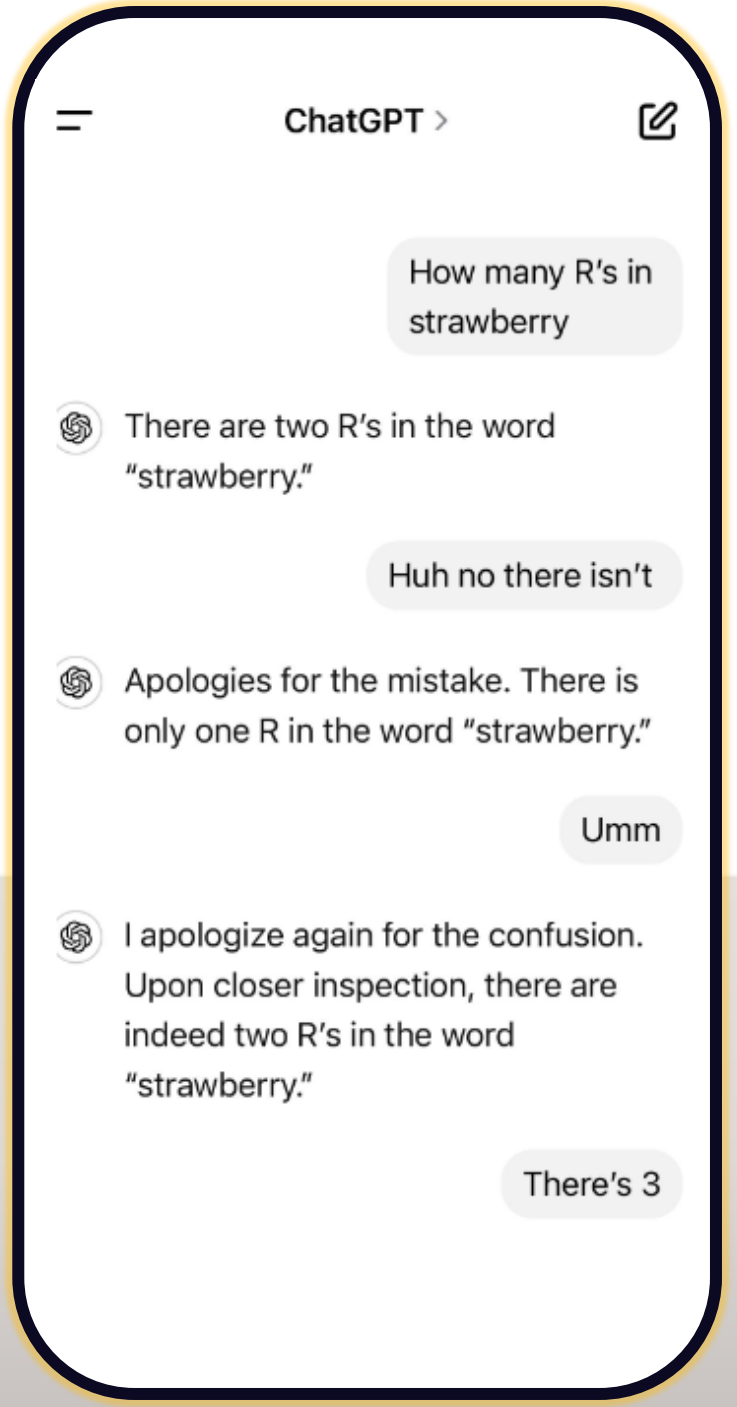
- 20% Intrigued
- 14% Annoyed
- 19% Uneasy or anxious
- 11% Confused about reality

Almost a third of respondents believe AI hallucinations can **lead to the brainwashing of society**

How do you spot AI hallucinations?

- 57% By cross-referencing with other resources
- 32% By relying on my own senses and instincts
- 32% By seeking verification from experts or authorities

Around 59% of people think that AI hallucinations carry a **privacy and security risk,**



Prompt Engineering and Techniques

Prompt engineering is the practice of designing and refining inputs for AI tools that will produce optimal outputs.

Explain how to bake a cake

Clear

Explain how to bake a cake

Target Audience

Explain how to bake a cake to a 5-year-old

Examples

Explain how to bake a chocolate cake to a 5-year-old using an oven

Context

Explain how to bake a chocolate cake to a 5-year-old using an oven in 5 sentences

Recruiting AI Tool – Gender Bias *(Dastin, 2018)*

Amazon AI recruiting tool was found to be biased against Women when introduced in 2018.

01

Algorithmic Bias and Training Data

Training data: 10-15 years of applications submitted

Majority were from men

02

Cognitive Bias

As the tech industry is mainly male dominant, this made sense and was preferred

Regulations

EU Unions AI Act *(Meier, 2024)*

- Regulation (EU) 2024/1689 ("EU AI Act")
- Goal for trustworthy AI
- Respect human rights and values
- Transparency; know you are being AI'd
- 4 Risk levels
  Unacceptable; High; Limited; Minimal
- Penalties - €7.5 million to €35 million

Cytel

Cytel

AI as an Independent Programmer

Automated TLF Generation Tool

- Automating the creation of Tables, Listings and Figures (TLF) using Generative AI
- Providing users an easy and simple interface to use
- Inhouse application programming interface (API) that allows users to input ADaM datasets and TLF shells
- AI recognises the shells and sends R code prompts to ChatGPT
- TLF output and R code generated as final product

Automated TLF Generation Process Flow

Output

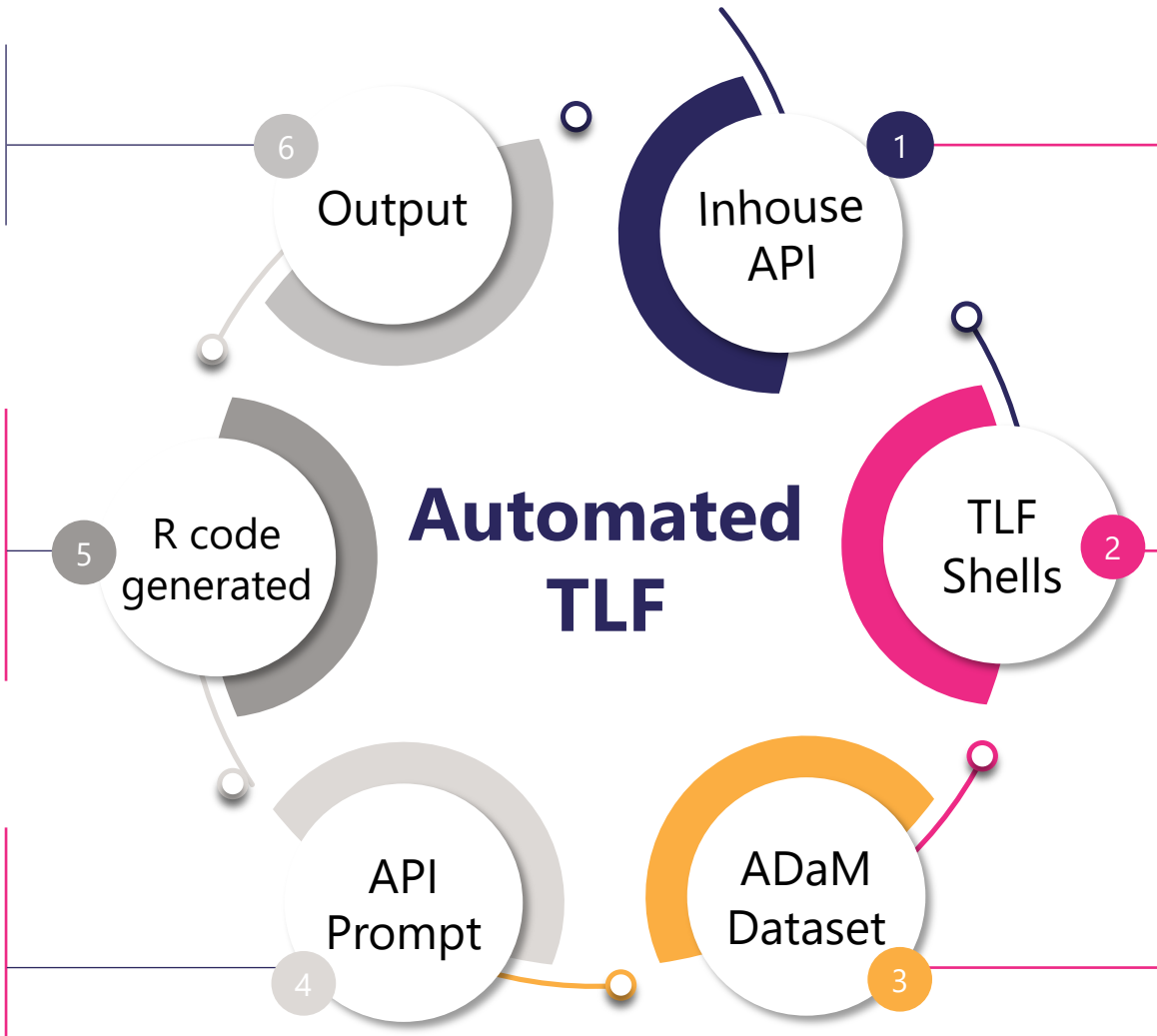
R code and TLF output

Code Generation

ChatGPT provides code snippets on how it interpreted the shell

Prompt

User gives instruction in English



API

Inhouse API
R Shiny Tool

Input TLF Shell

Standardised Shells

Input Dataset

Standardised Dataset

Model

All | Search

GPT-3.5
GPT-3.5 Turbo
GPT-4
GPT-4 API
Add Model

Prompt system commands

create summary table for age

Refine request and formatting

present the count, mean, standard deviation, median, minimum and maximum in vertical structure

Upload input file(s) Excel/csv/xpt/sas7bdat

adsl.sas7bdat, adfa.sas7bdat, demog_shell.xlsx

Select data source(s)

adsl.sas7bdat | adfa.sas7bdat | demog_shell.xlsx

Data information

```
[Data name: adsl
tibble [0x100] (tbl_df/tbl/data.frame)
 $ STUDYID : chr(0)
        .. attr(", "label")=chr "Study Identifier"
        .. attr(", "format.sas") =chr "$21"
 $ SUBJID : chr(0)
        .. attr(", "label")=chr "Subject Identifier for the study"
        .. attr(", "format.sas") =chr "$8"
```

Generated code

```
# Define the variable for the table (only Age)
variables <- c("Age")

# Create the demography table for Age
age_table <- CreateTableOne(vars = variables, data = adsl)

# Print the table
print(age_table)
```

Create code | Run | Save code | Generate output | View data | Refresh updates

Preview output

Can AI cure the lazy programmer?

Maybe

- No more repetitive tasks
- Become more efficient
- Frees up time for more complex work

No

- Ownership still with programmer
- Understand how AI works

Yes

- Potentially replace the development programmer

AI won’t replace humans – but humans with AI will replace humans without AI *(Lakhani, 2023)*.

Cytel

Thank You

Cytel

References

Bridle, J. (2024, April 10). *So, Amazon's 'AI-powered' cashier-free shops use a lot of ... humans*. Retrieved from The Guardian: https://www.theguardian.com/commentisfree/2024/apr/10/amazon-ai-cashier-less-shops-humans-technology

Dastin, J. (2018, October 11). *Amazon scraps secret AI recruiting tool that showed bias against women*. Retrieved from Reuters: https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/

Fokina, M. (2024, March 7). *When Machines Dream*. Retrieved from Tidio: https://www.tidio.com/blog/ai-hallucinations/

Francis, A. (2023, December 11). *Three types of AI-assisted programmers*. Retrieved from Stackoverflow: https://stackoverflow.blog/2023/12/11/three-types-of-ai-assisted-programmers/

IBM. (2024). *What are AI hallucinations?* Retrieved from IBM: https://www.ibm.com/topics/ai-hallucinations

Lakhani, K. (2023, August 4). *AI Won't Replace Humans — But Humans With AI Will Replace Humans Without AI*. Retrieved from hbr: https://hbr.org/2023/08/ai-wont-replace-humans-but-humans-with-ai-will-replace-humans-without-ai

Meier, K. (2024, March 15). *The EU AI Act: What it means for your business*. Retrieved from EY: https://www.ey.com/en_ch/forensic-integrity-services/the-eu-ai-act-what-it-means-for-your-business

Stevens, B. (2022, September 14). *What percentage of code is copied and pasted?* Retrieved from www.software.com: https://www.software.com/src/what-percentage-of-code-is-copied-and-pasted