
FLIPPED INTERACTION PATTERN, TABLES, INCLUDING NEW INFORMATION AND SOME OTHER TRICKS

BLAŽ ZUPAN, LJUBLJANA, 2024

LECTURE NOTES FOR THE UNIVERSITY OF LJUBLJANA'S COURSE ON
CHATGPT FOR RESEARCHERS



FLIPPED INTERACTION PATTERN

Now that I am writing these notes, I realize how much ground we have covered in only one lecture. Using all these needs practice, and we will defer that to you, the reader, for your daily use of LLMs and to the homework assignments in this course. For the end of this lecture, however, let us revisit the content where we have started today and use LLMs in a role where it, the model, asks as question to gather the data. Here is an example conversion:

I would like you to write a skeleton of a grant proposal. You will ask me questions about my current research work to get the data about it. Questions should be specific and easy to answer. When you have enough information, you will show me the title of the proposal and abstract and provide the outline with one sentence describing the content of each proposal section. Please do not ask me more than seven questions. Now, ask me the first question.

What is the general topic or area of your current research work?

Single-cell data analysis.

Great! Could you briefly describe the specific focus or objective of your research within single-cell data analysis?

I would like to determine the level of synchronicity of the amoeba cells during the development.

Thank you for that information. Could you specify what types of data you are currently collecting or analyzing to study the synchronicity of amoeba cells during development?

Single-cell gene expression data at various stages of development.

Thank you. How are you currently processing or analyzing the single-cell gene expression data in your research?

Using the methods and software I will develop within my project. I need to handle batch effects, dimensionality reduction, data visualization and annotation of visualizations with explanations come from external knowledge bases.

Thank you for the information. How are you planning to validate the level of synchronicity among the amoeba cells based on the processed single-cell gene expression data?

Computationally. We have yet to develop methods to assess and visually depict synchronicity. Our hypothesis is that synchronicity increases through time and during specialization of cells.

Thank you for your responses. Based on the information provided, here is the proposed grant proposal:

Title: Quantifying Synchronicity in Amoeba Cell Development through Computational Analysis of Single-Cell Gene Expression Data

Abstract: This proposal aims to develop novel computational methods to quantify and visualize the level of synchronicity among amoeba cells during developmental stages using single-cell gene expression data. The project will focus on addressing challenges such as batch effects, dimensionality reduction, and integrating external knowledge bases for data interpretation. The hypothesis is that synchronicity increases over time and during cell specialization, which will be validated computationally.

Outline:

1. Introduction

Introduce the significance of studying synchronicity in amoeba cell development.

2. Literature Review

Review current methodologies in single-cell data analysis and existing knowledge about amoeba cell development.

3. Methods

Develop computational methods for handling batch effects in single-cell data.

Implement dimensionality reduction techniques tailored for single-cell gene expression data.

Design algorithms for quantifying synchronicity and visualizing temporal changes.

Integrate external knowledge bases to annotate visualizations and interpret results.

4. Validation

Validate the developed methods using simulated data and benchmark datasets.

Apply the methods to real single-cell gene expression datasets from amoeba cells.

5. Results and Discussion

Present results on the quantification of synchronicity during amoeba cell development.

Discuss findings in relation to the hypothesis and existing literature.

6. Broader Impacts

Discuss potential implications of the research for understanding cellular development and synchronization.

Outline potential applications in biomedical research and beyond.

7. Conclusion

Summarize the significance of the proposed research and future directions.

This skeleton outlines the main sections and content of your grant proposal based on the information provided. If you need further details or adjustments, feel free to let me know!

We may not like the proposed skeleton, but note that this may be just the beginning. We may now request to change it, instruct to refine it, and ask for more data-gathering questions. Or even find a paper like that from the *PLOS Computational Biology* series on “Ten Simple Rules” (on grant writing or structuring our proposal), add it to our prompt, and ask ChatGPT to follow the recommendations.

PERSONAL ASSISTANT PROMPT

In a sense, we have already been introducing this type of pattern in the section above, where we are instructing LLMs to ask us about the information that is needed to fulfill a particular task. Here, we combine a prompt of this type with instructions for a particular pattern that the answers should use. Here is an example prompt:

You are my assistant in helping me to design a scientific presentation and related slides. I want to compose up to 5 slides for a topic and mainly include figures and only some text. For each slide, you will help to define the following: **Title****: <TITLE> ****Subtitle****: <ONE SENTENCE SUBTITLE> ****Figure or text****: <SHORT DESCRIPTION OF FIGURE OR TEXT TO**

INCLUDE> To do this, you'll need to ask me questions about my presentation and research. Ready? Ask one question at a time—no more than seven questions. Propose the slides only after you have finished with the questioning. Now, ask me the first question.

The text enclosed in double asterisk (“**”) marks the part to be printed in bold in the markdown language. Again, Markdown is a lightweight markup language with plain text formatting syntax designed to be converted to HTML and other formats, and is the language that LLMs like ChatGPT’s models are using for textual output that is then rendered in HTML in a chat window. With the text enclosed in “<” and “>” we, somehow, tell LLM that this should be replaced with generated text. Note that this is not algorithmically “encoded” in LLM, but simply results in this kind of marking that was used in the textual documents that LLM have consumed in training. We here hope that our communication will be understood in this way, since else, we need to change the prompt.

Also notice that our prompt finished with “Now, ask me the first question.” This is to let LLM know that we do not want to print all the questions at once, and to reemphasize that we would like to answer one question at a time.

Here is an example of the conversation that resulted from such a initial prompt:

Sure! Let's get started. What is the main topic or focus of your scientific presentation?

Machine learning and visualization

Great! Could you please specify the particular aspect of machine learning and visualization you would like to focus on? For example, are you concentrating on a specific algorithm, application, dataset, or technique?

Point-based visualizations, machine learning techniques for dimensionality reduction, like t-SNE, PCA and MDS

Got it! What is the primary objective or key message you want to convey through your presentation on point-based visualizations and machine learning techniques for dimensionality reduction?

That point-based visualization are a cool way to presenting multivariate data, especially if we can, in some way, explain what particular groups of points actually mean. Emphasis on explainable machine learning and visualizations! :)

Excellent! Which datasets or specific examples do you plan to use to demonstrate these techniques and their effectiveness in your presentation?

Data from biology and molecular biology, like the machine learning zoo data sets and single-cell gene expression data sets

That sounds very interesting! Could you specify any particular results or key findings from your analysis of these datasets that you would like to highlight in your presentation?

Yes. It turns out that when the data is large and there are many data instances, t-SNE nicely uncovers clustering structure. This needs to be explained by other information sources, like data bases on marker genes.

Great! What types of figures or visualizations do you already have, or plan to create, to support these key findings? For example, do you have specific plots, charts, or diagrams in mind?

I do. t-SNE, PCA and MDS visualizations and annotations of t-SNE plots.

Perfect! Could you also share any preferences for the overall design or style of your presentation slides? For instance, do you have a specific color scheme, layout, or template in mind?

No, just plain black and white. Oh, I forgot something: I will include snapshots of Orange Data Mining toolbox to illustrate visual analytics, and should spend a slide on what visual analytics is all about.

Thank you for the additional information! Here's a quick summary of what we have so far:

- 1. Introduction to Machine Learning and Visualization**
- 2. Techniques for Dimensionality Reduction**
- 3. Visual Analytics Concept**
- 4. Examples and Results from Biological Datasets**
- 5. Using Orange Data Mining Toolbox**

I will now propose the slides based on this information:

Slide 1: Introduction to Machine Learning and Visualization

Title: Introduction to Machine Learning and Visualization

Subtitle: An overview of point-based visualizations for multivariate data

Figure or text: Include a brief introduction text and a visual representation of multivariate data, such as a simple scatter plot showing high-dimensional data points.

Slide 2: Techniques for Dimensionality Reduction

Title: Techniques for Dimensionality Reduction

Subtitle: Exploring t-SNE, PCA, and MDS for data visualization

Figure or text: Display side-by-side visualizations of the same dataset using t-SNE, PCA, and MDS, highlighting the clustering structures.

Slide 3: Visual Analytics Concept

Title: What is Visual Analytics?

Subtitle: Combining data analysis with interactive visualizations

Figure or text: Provide a brief explanation of visual analytics and a snapshot of the Orange Data Mining toolbox interface, showing interactive features.

Slide 4: Examples and Results from Biological Datasets

Title: Examples and Results from Biological Datasets

Subtitle: Clustering structures in single-cell gene expression data

Figure or text: Show t-SNE plots of the single-cell gene expression data with annotations explaining clusters using marker genes and additional information sources.

Slide 5: Using Orange Data Mining Toolbox

Title: Using Orange Data Mining Toolbox

Subtitle: A tool for visual analytics and machine learning

Figure or text: Include screenshots of the Orange Data Mining toolbox in action, demonstrating how to create and interpret visualizations.

Would you like to make any adjustments or add any specific details to these slides?

This is a good start. For brainstorming, I mean. Reading through the text we can spot that it would be, for instance, better to show Orange Data Mining with examples of various visualizations, and correct ChatGPT that we would like to merge this slide with the slide two. This conversation stems from brainstorming the author of these lines had with ChatGPT when preparing the slides for “The Joy of Point-Based Visualizations”, a recently given invited talk presenting the techniques mentioned above. The talk was longer and needed more slides, but this could be done with simple prompting, such as

Please expand the number of slides to 10. Also note that I would like to introduce various visualizations by showcasing Orange on different data sets, including the zoo data set,

where I show how similar PCA, MDS, and t-SNE results are, and an example single-cell data set, where it becomes clear that with large data sets with higher number of data instances the resulting visualizations can be very different.

Ok, this is a rather long prompt, but it worked and the resulting proposal for the slides (not shown here) was great. In my presentation, I also needed to show how mathematically very different are the three dimensionality reduction techniques, and asked ChatGPT to render their optimization function.

Please give me the mathematical expression for the optimization function for PCA, MDS and t-SNE. Just math, no explanation.

Certainly! Here are the mathematical expressions for the optimization functions:

PCA (Principal Component Analysis)

$$\arg \max_{\mathbf{w}_k} \mathbf{w}_k^\top \mathbf{S} \mathbf{w}_k \quad \text{subject to} \quad \mathbf{w}_k^\top \mathbf{w}_k = 1$$

where \mathbf{S} is the covariance matrix of the data.

MDS (Multidimensional Scaling)

$$\min_{\mathbf{Y}} \sum_{i < j} (d_{ij} - \|\mathbf{y}_i - \mathbf{y}_j\|)^2$$

where d_{ij} is the distance between points i and j in the original high-dimensional space, and \mathbf{y}_i and \mathbf{y}_j are the corresponding points in the lower-dimensional space.

t-SNE (t-Distributed Stochastic Neighbor Embedding)

$$\min_{\mathbf{Y}} \sum_{i \neq j} P_{ij} \log \frac{P_{ij}}{Q_{ij}}$$

where P_{ij} is the probability that high-dimensional points i and j are neighbors, and Q_{ij} is the probability that their low-dimensional representations are neighbors.

Kind of amazing that this works, right? We have already discussed generation of equations by ChatGPT (see our previous lecture notes) and learned that this is just another sequence of letters, this time in LaTeX, that is then conveniently rendered and typeset in the browser, so, technically, no surprises there. We could go on here with personal assistant prompt, but you got the idea, and this section got long already.

PROGRAMMING WITH CHATGPT

Here, by “programming,” we do not mean conventional computer programming but instead instructing LLM to provide output according to the initial prompt. Consider the following example of prompts issued one by one (ChatGPT’s answer not shown).

Whenever you generate output, turn it into a table with columns of property names and rows with examples. Answer with only a few words or preferably just with numbers.

What kind of city is Ljubljana concerning size, population, and number of theatres? Does it have a subway?

Add Kranj and Koper.

Place cities in rows, and properties in columns (transpose current table).

Compare Maribor, Zagreb and Vienna.

Include all EU capitals.

Here's one of the intermediate answers:

City	Size (km ²)	Population	Number of Theatres	Subway
Maribor	147.5	112,065	5	No
Zagreb	641.4	806,341	20	Yes
Vienna	414.6	1,931,830	50	Yes

Be careful; answers are not necessarily correct: Zagreb has a tram, not a subway. But otherwise, it looks rather impressive, right? Using GPT-4o, when writing this text the latest LLM available through ChatGPT, you can also export the resulting data into an Excel file.

Export the data to Excel file.

About tables: LLMs generate text. The data table, as shown above, is also a sequence of characters and strings. We can copy the table (by clicking on a small copy icon in ChatGPT's interface).

Pasting the copied text to some text editor (your instructor uses Sublime, for instance) reveals the text generated. Or we can simply ask ChatGPT what format it uses for the table:

What is the format that you have used to generate the table? Is this markdown?

Yes, the format I used to generate the table is Markdown. Markdown is a lightweight markup language with plain text formatting syntax, often used for formatting readme files, for writing messages in online discussion forums, and to create rich text using a plain text editor.

Here is the Markdown code for the table:

markdown

Copy code

City	Size (km ²)	Population	Number of Theatres	Subway
Maribor	147.5	112,065	5	No
Zagreb	641.4	806,341	20	Yes
Vienna	414.6	1,931,830	50	Yes

You can use this format to generate tables in any Markdown-supported environment.

You can learn more about Markdown and its simple syntax by asking ChatGPT about it. Here, it suffices to say that even when ChatGPT displays tables, this is rendering of a text in the Markdown syntax, and everything is still in the realm of LLM generating a string of characters. Similarly, for example, ChatGPT renders text in bold (Markdown syntax uses two asterixes and with them embodies the part of the text that should be rendered in **bold**) or can display itemized lists or even nicely set mathematical equations (using LaTeX syntax). Not that we will discuss much about this here, but to know, again: LLMs generate text that is nicely rendered by ChatGPT.

Hey, ChatGPT, show me an example of some complicated equation, no extra explanation, please.

Sure, here is an example of a complicated equation:

$$\sum_{k=1}^{\infty} \left(\int_0^1 \frac{\sin(kx)}{x} dx \right)^2 = \frac{\pi^2}{6} (\ln(2\pi) - \gamma)^2 + \frac{\pi^4}{72}$$

Hm, how was this rendered?

This equation was rendered using LaTeX, a typesetting system that's widely used for mathematical and scientific documents. LaTeX allows for the precise formatting of mathematical notation. Here's how the LaTeX code for the equation looks:

latex

Copy code

```
\[
\sum_{k=1}^{\infty} \left( \int_0^1 \frac{\sin(kx)}{x} dx \right)^2 = \frac{\pi^2}{6} (\ln(2\pi) - \gamma)^2 + \frac{\pi^4}{72}
\]
```

INCLUDING NEW INFORMATION

The large language models we used in our lectures were trained on a fixed data set and are not continuously re-trained with new information. At least not yet.

How did Tadej Pogačar perform in today's race?

I can't provide real-time information as my data only goes up until January 2022. However, you can easily find the latest updates on Tadej Pogačar's performance in today's race by checking sports news websites, cycling forums, or his official social media accounts. If you have any other questions or need information within my knowledge cutoff date, feel free to ask!

This was with GPT3.5. When using a more recent version of LLM, ChatGPT queries a web search engine, includes the search results or perhaps the content of a few resulting web pages in the prompt that is sent to the LLM, and then provides the answer by citing the source. Like:

How did Tadej Pogačar perform in today's race?

Today, Tadej Pogačar delivered an impressive performance in the Giro d'Italia. He maintained his dominance throughout the race, demonstrating his exceptional skill and strategy. Although UAE Team Emirates opted for a more conservative approach during the stage, allowing a breakaway to form, Pogačar remained a strong presence in the peloton (cyclingnews.com) (cyclingnews.com).

Which could be dead wrong as well. And it is, in this case. On the day of the writing of this text (Sunday, May 19, 2024), Pogačar had just won the 15th stage of Giro di Italia in what was the hardest Giro stage in a decade, while the first link provided by ChatGPT takes us to another race, Volta a Catalunya. It may take a while for ChatGPT to master the integration of the current news in its answers. Before it does, note that we can do this manually by including the information we would like ChatGPT to refer to within the prompt:

Consider the following today's news. I will have a few questions about it in my next prompt.

>>Tadej Pogacar secured his fourth victory of this year's race as he breezed to a commanding win on stage 15 of the Giro d'Italia. The queen stage of the Giro posed 222km of tarmac alongside more than 5,000 metres of climb but that did not seem the faze the Slovenian, who blew the field away in convincing fashion from Manerba del Garda to Livigno. Nairo Quintana was finally caught by Pogacar on the steep climb with 2km left, and proceeded to lose a further half a minute by the finish line. Two minutes later, EF Education-EasyPost rider Georg Steinhauser, who was also a solo leader until the penultimate climb, came home to take the final spot on the podium in third. Pogacar said afterwards on Eurosport.com: "Today was a really good route with nice climbs. I'm super happy that we kept it under control. We had to be smart all day. It was a strong breakaway but I gave it my all for the last 10k and I'm super happy to win the queen stage in Livigno, one of my favorite places in Italy."<<

Got it. You can ask me about the news whenever you're ready.

Who won today's race and what was the difference with a second-placed?

Tadej Pogacar won today's race. He finished with a half-minute lead over the second-placed rider, Nairo Quintana.

Who was third?

Georg Steinhauser of EF Education-EasyPost was third, finishing two minutes behind the winner, Tadej Pogacar.

Cool, right? I mean, it's cool that Pogačar won (I am a fan), but it is also cool that we can instruct ChatGPT and, with it, LLM to consider new information and then query about it. Notice that above, we have intentionally formatted the prompt to place delimiters to signify where the text from the articles (which I copied from The Guardian) starts and where it ends. I have used >> and <<, but I could use any other special symbol or mark we are consistent with opening and closing the quotation.

Prompts may provide information that LLM had no access to, such as personal information, information from our organization, your own research, your own research results, or the most recent information that LLM has not been trained on. We can use the prompt to introduce information.

Another example discussed in the class was the analysis of the text from the flight ticket. We took the flight information file (PDF) that is usually supplied by the booking agency, copied the entire text (control-A or similar in a PDF viewer), began with prompt

Following is my flight ticket. Please consider it. I will be asking you about some information contained. Ok?

You have probably already noticed that I am adding an “Ok?” at the end of my introductory prompts. I am doing so to prevent LLM from generating a longer text summarizing my prompt and just answering that it is ready to receive further questions. The questions that followed in the class were of the type

Where am I going?

When will I arrive in Dresden?

What is my connecting time in FRA?

What happens if my flight from Ljubljana is delayed for two hours?

and it was interesting that LLM could answer them all despite quite gibberish text that I have obtained from the PDF.

Note again that this was done with an older version of ChatGPT, using GPT-3.5 as a model. In a newer version of the application, and in most modern LLM interfaces, I could submit the PDF and then ask questions about it. This would involve sending the PDF to the server, where the software on the server (not LLM!) would convert the PDF to text and send it to LLM along with my questions.

This is not too far off from what we did when we copied the text content from the PDF. Also note that LLMs can only handle text, i.e. a string of characters, on the input, and cannot directly handle other data formats, such as PDFs or images.

As the last example on including additional information from the class we have considered the Wikipedia page on John Lennon, copied all the text, found out that for ChatGPT 3.5 it was too long and started a new chat only with the text on his political activism, whose length was within the prescribed limits. The types of prompts, questions, we had, were:

Consider the following information about John Lennon. (we here pasted the John Lennon's text from the political activism part of his Wikipedia page)

What were Lennon's major political beliefs. In one sentence, please.

Which political party would Lennon most likely vote for, Democrats or Republicans?

Why would he not vote for Republicans? In one sentence.

Summarize the information that I have provided in one paragraph.

Summarize the information that I have provided, preserving the information about dates and years.

TEMPLATE PATTERN

Prompt templates are predefined structures or patterns used to guide the language model's text generation. These templates can help in generating more consistent and targeted outputs.

I am going to give you a template for your output. CAPITALIZED WORDS are my placeholders. Fill in my placeholders with your output. Preserve the overall formatting of my template. My template is: **Question:** QUESTION ****Answer:** ANSWER** Do not do anything at this stage. I will give you the data to format in the next prompt. Provided the data, you will create five questions using my template. Ok?****

Sure, I understand! Just let me know when you're ready to proceed with the data.

The data: Julian's Assange's wife has told of her elation that the WikiLeaks founder has been released from Belmarsh prison in London and will soon be a "free man" under a deal in which he will plead guilty to violating US espionage law. Speaking from Australia, where she flew on Sunday to prepare her family's new life, Stella Assange, a human rights lawyer, said she had not told the couple's two young sons, Gabriel and Max, about their father's release after five years in jail for fear of the information leaking. She said: "All I told them

was that there was a big surprise. And, on the morning that we left, I told them where we're heading to the airport, and we got on the plane, and I told them that we were going to visit our family, their cousin, their grandfather and so on. "And they still don't know. We've been very careful, because obviously, no one can stop a five- and a seven-year-old from, you know, shouting it from the rooftops at any given moment. Because of the sensitivity around the judge having to sign off the deal, we've been very careful, just gradually, incrementally telling them information." Assange, 52, was woken at 2am in his cell in Belmarsh on Monday before being handcuffed and transported to Stansted airport to take a chartered jet via Bangkok to the island of Saipan in the Northern Mariana Islands, which is US territory.

Now please create three questions using my template.

Question: Where did Julian Assange travel after his release from Belmarsh prison?

Answer: Julian Assange was transported to Stansted airport in London and then took a chartered jet via Bangkok to the island of Saipan in the Northern Mariana Islands, which is US territory.

Question: Why did Stella Assange refrain from informing her sons about Julian Assange's release immediately?

Answer: Stella Assange refrained from immediately informing her sons about Julian Assange's release due to the sensitivity around the judge signing off on the deal. She was concerned that their young age might lead them to inadvertently disclose the information.

Question: How long had Julian Assange been in jail before his release?

Answer: Julian Assange had been in jail for five years before his release from Belmarsh prison in London.

Let us combine this template with further analysis of the data. Suppose we are interested in people involved in the assassination of archduke Franz Ferdinand in Sarajevo but would like to instruct LLM to provide an output in specific form:

I am providing a template for your output. CAPITALIZED words are placeholders. Preserve the overall formatting. My template:

Bio: <NAME>

****Summary:** <ONE SENTENCE SUMMARY>**

****Born:** <YEAR OF BIRTH>, <PLACE>**

****Major impact:** <ONE SHORT PARAGRAPH OF MAJOR ROLE OF THIS PERSON>**

Extract the names of all the people mentioned in the following text and report on them using the template.

<<<In the summer of 1914, the sovereigns of Europe were woven together by treaties, alliances, as well as secret agreements. The Triple Alliance (1882) encompassed the German Empire, Austria, and Italy.[34]

On 28 June 1914, Archduke Franz Ferdinand of Austria, heir presumptive to Emperor Franz Joseph I of Austria, visited Sarajevo, the capital of the recently annexed Bosnia and Herzegovina. Cvjetko Popović, Gavrilo Princip, Nedeljko Čabrinović, Trifko Grabež, and Vaso Čubrilović (Bosnian Serbs) and Muhamed Mehmedbašić (from the Bosniaks community),[35] from the movement known as Young Bosnia, took up positions along the route taken by the Archduke's motorcade, to assassinate him. Supplied with arms by extremists within the Serbian Black Hand intelligence organisation, they hoped his death would free Bosnia from Austrian rule, although there was little agreement on what would replace it.[36]

Nedeljko Čabrinović threw a grenade at the Archduke's car and injured two of his aides, who were taken to hospital while the convoy carried on. The other assassins were also unsuccessful but, an hour later, as Ferdinand was returning from visiting the injured officers, his car took a wrong turn into a street where Gavrilo Princip was standing. He fired two pistol shots, fatally wounding Ferdinand and his wife Sophie.[37] Although Emperor Franz Joseph was shocked by the incident, political and personal differences meant the two men were not close; allegedly, his first reported comment was "A higher power has re-established the order which I, alas, could not preserve".[38]>>>

Bio: Archduke Franz Ferdinand of Austria

Summary: Archduke Franz Ferdinand of Austria, heir presumptive to Emperor Franz Joseph I, was assassinated in 1914, an event that triggered World War I.

Born: 1863, Graz, Austria

Major impact: Archduke Franz Ferdinand's assassination in Sarajevo by Gavrilo Princip of the Young Bosnia movement on June 28, 1914, sparked a series of events leading to the outbreak of World War I. His death destabilized Europe, setting off a chain reaction of alliances and conflicts that culminated in a global war.

Bio: Gavrilo Princip

Summary: Gavrilo Princip was a Bosnian Serb nationalist who assassinated Archduke Franz Ferdinand of Austria, sparking World War I.

Born: 1894, Obljaj, Bosnia and Herzegovina

Major impact: Gavrilo Princip's assassination of Archduke Franz Ferdinand of Austria on June 28, 1914, was a catalyst for World War I. As a member of the Young Bosnia movement, Princip's actions were driven by a desire for Bosnian independence from Austrian rule, although his act led to widespread and catastrophic consequences.

Bio: Nedeljko Čabrinović

Summary: Nedeljko Čabrinović was one of the assassins involved in the plot to kill Archduke Franz Ferdinand of Austria.

Born: 1895, Sarajevo, Bosnia and Herzegovina

Major impact: Nedeljko Čabrinović attempted to assassinate Archduke Franz Ferdinand of Austria on June 28, 1914, by throwing a grenade at his car. Although unsuccessful in killing the Archduke, Čabrinović's actions contributed to the tensions that led to the outbreak of World War I.

...

The list includes few more people. Amazing, right? In prompt engineering, templates play a crucial role in shaping the way queries or prompts are constructed to elicit specific responses from language models like GPT-3. Templates essentially serve as structured formats or frameworks into which variables or placeholders can be inserted. They can improve standardization and consistency, reproducibility, assist in targeted query generation. Templates can be adapted and scaled across different applications and domains. Once a template structure is defined, it can be reused with minor adjustments for similar tasks or contexts, which saves time and effort in prompt design.

For our final prompt in this lecture, let us use LLM to compose a workout using a template to list each of the exercises:

Create a random strength workout with complementary exercises. I am going to provide a template for your output, where CAPITALIZED-WORDS are placeholders, replaces them with content. This is the template: `###: NAME` `<newline>` `set and timing:` `** REPS @ SETS, ESTIMATED-TIME-IN-MINUTES` `<newline>` `**muscles:` `** MUSCLE GROUPS WORKED` `<newline>` `**difficulty:` `** DIFFICULTY SCALE 1-5` `<newline>` `**notes:` `** FORM NOTES. No itemization.`**

Notice that we have used several tricks. We embedded the text that should be set in bold with `**`; we already know this stems from the markdown syntax. Placeholders are in capital letters. We have used the `"<newline>"` tag to tell LLM where we want line breaks. To avoid the use of bullets, we have added `"no itemization"` at the end of the prompt. The output:

1: Push-ups

set and timing: 15 @ 3, 10 minutes

muscles: Chest, Shoulders, Triceps

difficulty: 3

notes: Maintain a straight body alignment throughout.

2: Dumbbell Shoulder Press

set and timing: 12 @ 3, 12 minutes

muscles: Shoulders, Triceps

difficulty: 4

notes: Keep core engaged to avoid arching the back.

3: Bent-over Dumbbell Rows

set and timing: 12 @ 3, 10 minutes

muscles: Upper Back, Biceps

difficulty: 3

notes: Focus on pulling elbows back and squeezing shoulder blades together.

4: Bodyweight Squats

set and timing: 20 @ 3, 8 minutes

muscles: Quadriceps, Glutes, Hamstrings

difficulty: 2

notes: Keep knees tracking over toes and maintain a neutral spine.

5: Plank

set and timing: 30 seconds @ 3, 6 minutes

muscles: Core, Shoulders, Back

difficulty: 2

notes: Brace your core and avoid sagging hips.

6: Dumbbell Bicep Curls

set and timing: 12 @ 3, 8 minutes

muscles: Biceps, Forearms

difficulty: 2

notes: Control the movement to fully contract the biceps.

7: Glute Bridges

set and timing: 15 @ 3, 8 minutes

muscles: Glutes, Hamstrings, Lower Back

difficulty: 2

notes: Squeeze your glutes at the top of the movement.

Cool-down: Finish with stretching exercises for each major muscle group, holding each stretch for 20-30 seconds.

This workout targets multiple muscle groups and provides a balanced strength training session. Adjust weights and repetitions based on your fitness level and goals.

This is already great, right? Consider AI where together with the text we have a sketch of the exercise, or an instructional video. All this is probably coming, the AI technology of today is still very young. Until then, you can further refine the workout plan using some more prompting and providing constraints to the list:



Group the exercises to the warmup, main part, and cooldown.

Construct the exercise plan to fit within half-an-hour.

Emphasis should be on legs and feet.

Make exercises lighter.

Or anything similar. Try it out!