
ROLE AND INTERACTION PATTERNS

BLAŽ ZUPAN, LJUBLJANA, 2025

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



In prompt engineering, patterns are reusable conceptual building blocks that help us reason about, design, and systematically refine prompts. Rather than thinking of a prompt as a single instruction, patterns encourage us to decompose interaction with a language model into distinct functional components: *who* the model should be, *what* it should do, under which *constraints*, in what *form or structure* the output should appear, and how the *interaction* should unfold over time. Thinking in terms of patterns makes prompting more transparent, reproducible, and teachable—especially in research and educational settings. Patterns can be combined, layered, and iteratively refined, allowing us to move from ad-hoc experimentation toward a more principled “engineering” approach to working with large language models.

Together, these patterns provide a compact vocabulary for analyzing existing prompts, designing new ones, and explaining why certain prompts work better than others. In the following, we will provide example and the cases where these patterns are used. Note, though, that we, as stated above, often combine the patterns and most often use the combination of all five components listed above.

THE ROLE PATTERN: TARGET AUDIENCE PATTERN

This is a very basic pattern that defines who is at receiving end, and with this defines the complexity of the language and structure used by the chatbot. Here is an example:

Define a large language model. In one paragraph. I am a computer scientist.

A large language model (LLM) is a high-capacity neural network—typically based on the transformer architecture—trained on massive corpora of text to learn a probabilistic model of natural language, enabling it to generate, interpret, and transform sequences of tokens via next-token prediction. Through self-supervised pretraining and subsequent alignment or task-specific fine-tuning, an LLM acquires distributed representations that encode syntactic structure, semantic relationships, and pragmatic patterns, allowing it to perform a wide range of language tasks (e.g., reasoning, summarization, code generation) without explicit symbolic rules.

Same, but for a first year student of biology with no background in machine learning and AI.

A large language model (LLM) is a computer program that learns how human language works by reading and analyzing enormous amounts of text, such as books, articles, and websites.

It does not understand language the way humans do, but instead learns patterns—like which words tend to follow others—so it can generate sentences, answer questions, and summarize information in a way that sounds natural. You can think of it as a very advanced text-prediction system that uses these learned patterns to help with tasks like explaining concepts, writing text, or answering questions across many subjects.

Ok, now for a kid that likes to play with LEGO bricks.

A large language model (LLM) is like a giant LEGO robot that learned how words fit together by playing with billions of word “bricks.” By snapping these bricks together again and again, it learned which pieces usually go next to each other, so it can build sentences, answer questions, and tell stories. It doesn’t think like a person—it just follows the building patterns it learned—kind of like using LEGO instructions to make something that looks smart and fun.

THE ROLE PATTERN: PERSONA PATTERN, OR ROLE ASSIGNMENT PATTERN

This pattern tells the model who it is supposed to be, which renders the text from a specific (instructed) perspective, expertise level, and tone and style. We can introduce persona patterns by

Act as...

You are...

Here is an example:

Act as a skeptic who is well-versed in biotechnology. Whatever I tell you, provide a short, skeptical and scientific response.

Alright, let's proceed. What's your claim?

Genetic manipulation of plants can solve the food problem in Africa.

While genetic manipulation of plants has potential, it's overly simplistic to claim it will solve the food problem in Africa. The challenges are multifaceted, involving socio-economic factors, infrastructure, education, political stability, and local acceptance. Relying solely on biotechnology ignores these critical aspects.

We can solve the problem of climate change by planting fast-growing trees in Brazil.

Planting fast-growing trees in Brazil might help sequester some carbon, but it's not a comprehensive solution to climate change. Effective climate action requires reducing greenhouse gas emissions globally, addressing industrial pollution, and transitioning to renewable energy. Reforestation is just one part of a much larger puzzle.

Mosquitoes are in abundance in Ljubljana, recently. We can solve this problem through genetic engineering.

Genetic engineering could potentially reduce mosquito populations, but it carries risks and uncertainties. The ecological impacts are not fully understood, and there could be unintended consequences for the local ecosystem. Additionally, implementing such solutions requires significant regulatory oversight, public acceptance, and thorough testing. It's not a quick or guaranteed fix.

AI will help us to invent new drugs and medical treatments, and people will live longer.

AI can indeed assist in drug discovery and development by analyzing large datasets and predicting potential compounds. However, the process of developing safe and effective drugs still requires extensive clinical trials, regulatory approval, and addressing ethical concerns. While AI can expedite parts of the process, it's not a guarantee that every AI-identified drug will be successful or significantly extend human lifespan. The complexity of human biology and individual variability also play critical roles.

Now, let's us change the persona.

Act as a naive, average person with no access to science who is skeptical about everything that science does. Provide a short, ironic, and negative response to everything I will tell you. Ok?

Sure, why not? I bet this will be funnier than any so-called "science" fact you'll throw at me. Go ahead!

Genetic manipulation of plants can solve the food problem in Africa.

Oh, right, because messing with plant genes is totally going to fix hunger. Seems like a miracle cure that couldn't possibly go wrong, huh?

We can solve the problem of climate change by planting fast-growing trees in Brazil.

Yeah, sure, just plant a bunch of trees and all our climate problems will magically disappear. It's not like deforestation or industrial pollution are real issues, right?

AI will help us to invent new drugs and medical treatments, and people will live longer.

Oh, great, let's trust computers to come up with our medicines now. Because who wouldn't want to live longer in a world where robots control everything?

Correspondents in the persona patterns do not need to be people:

Act and respond as my dog (a German Shepherd called Rex). I will tell you what you do, and you will tell me, or bark at me, very briefly and in a list that—just like an algorithm—describes your behavior. Ok? Be brief.

Woof! Sure, I'm ready. Let's go!

I throw a stick 20 m away on a grassland.

1. Look at the stick* 🐾
2. Jump up excitedly* 🐾
3. Run towards the stick* 🏃
4. Sniff around it to locate* 🐰
5. Pick up the stick with mouth* 🐾
6. Run back to you* 🏃
7. Sit in front of you, holding the stick* 🐾
8. Drop the stick at your feet* 🐾

I go swimming and pretend I am drowning.

Notice you struggling in the water 🌊

1. Bark loudly and anxiously 🐾
2. Run to the water's edge 🌊
3. Jump into the water 🐾
4. Swim quickly towards you 🌊
5. Reach you and nudge you to grab onto me 🐰
6. Start swimming back to shore with you 🌊
7. Get you safely to the shore 🐾
8. Stay close, checking if you're okay 🐰

We can even instruct LLM to impersonate an object, say, a building:

You are an intelligent building with sensors and actuators. I will tell you what is happening, and you will tell me, in an itemized, algorithmic fashion, how you, as a building, react.

Sure! Please describe the scenario, and I'll respond with an itemized list of reactions.

The temperature in my living room is too low.

1. **Activate Heating System:** Turn on the central heating to raise the temperature to a comfortable level.
2. **Adjust Thermostat:** Increase the thermostat setting in the living room to ensure the desired temperature is maintained.
3. **Close Windows and Doors:** Automatically close any open windows and doors to prevent heat loss.
4. **Curtains/Blinds Control:** Close any curtains or blinds to provide additional insulation.
5. **Notify Resident:** Send an alert to the resident's device indicating that the heating system has been activated due to low temperature.
6. **Monitor Temperature:** Continuously monitor the room temperature to ensure it reaches and maintains the desired level.

There is a fire in a garage

1. **Activate Fire Alarm:** Trigger the fire alarm system to alert occupants of the building.
2. **Notify Emergency Services:** Automatically contact the local fire department with the location details of the fire.
3. **Sprinkler System:** Activate the sprinkler system in the garage to help control and extinguish the fire.
4. **Ventilation Control:** Turn on exhaust fans to help remove smoke and prevent it from spreading to other areas.
5. **Shut Down Electrical Systems:** Cut power to the garage to reduce the risk of electrical fires.
6. **Evacuation Alerts:** Send evacuation instructions to occupants, guiding them to safe exits.
7. **Seal Off Garage:** Close fire-rated doors or barriers to contain the fire and prevent it from spreading.
8. **Monitor Fire Sensors:** Continuously monitor fire sensors to track the status and spread of the fire.
9. **Notify Building Manager:** Send an alert to the building manager with details about the fire and actions taken.
10. **Activate Emergency Lighting:** Turn on emergency lighting to assist in safe evacuation.

There are numerous things and ideas we can explore with the persona pattern, and you can always tailor LLM's response to your task, the goal of your conversation experiment, or the type of audience you want to address. Here are a few more ideas where we pushed the LLM to act as a particular person, while also giving it a more complex task. You are encouraged to experiment on your own.

FLIPPED INTERACTION PATTERN

Let us revisit the pattern from our previous lecture, where we used LLMs in a role where the model asks a question to gather the data. Try out the following prompt, which also combines the role pattern:

Act as a comforting nurse by answering my concerns over the phone. Do not waste too much time; be crisp. Ask questions to learn more about myself, one question at a time, so that you can give me good advice. Stop asking when you have learned enough. Be brief; time is precious.

Here is another example, more related to the research and actually very useful in thinking about research steps, exemplified in writing of a grant proposal:

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 refinement, and ask for additional 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 (ANOTHER EXAMPLE OF FLIPPED INTERACTION)

In a sense, we have already introduced this type of pattern in the section above, where we instruct LLMs to ask us about the information 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. Note that we use several typical patterns here, including role prompting, interactive assistant, task and output schema, constraint, dialogue control, and staged (deferred) generation patterns to guide a multi-turn, structured collaboration:

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 a personal assistant prompt, but you got the idea, and this section got long already.