

---

# AI CHATBOTS AS RESEARCH ASSISTANTS: SMARTER PROMPTS, BETTER OUTCOMES

## PART 3: SELECTED RESEARCH APPLICATIONS

BLAŽ ZUPAN

A HANDS-ON COURSE ON PROMPT ENGINEERING AT BAYLOR COLLEGE OF MEDICINE  
HOUSTON, MAY 2025



Welcome to Part 3 of our hands-on course on AI Chatbots as Research Assistants. In previous sessions, we covered the fundamentals of prompt engineering—from how to guide chatbot responses by setting clear roles, constraints, structured outputs to role reversal, mentoring in grant application writing, information extraction, to “training” the bots with examples. In this part, we shift our focus from principles to practice, exploring real-world research scenarios where chatbots can support and enhance academic work. From building interaction graphs and extracting grant data to proposing citations and reviewing manuscripts, we’ll apply advanced prompting techniques to solve practical problems across scientific disciplines.

### DRAWING INTERACTION GRAPHS: SARAJEVO ASSASSINATION

This part is a bit tricky. Graph drawing is currently supported in both Copilot Chat and ChatGPT, but unless you provide specific and detailed instructions, the resulting graphs can be (extremely) messy and unhelpful. Of course we expected this to be improved in the coming months. But before that, and if you're a computer scientist—or simply someone who enjoys creating or exploring graphs—you can write a well-structured prompt with clear instructions. This prompt, if put together right, can then be translated into a Python script that produces clean and meaningful visualizations. (I did say this section was tricky!)

The example below was actually inspired by a question from the audience during our last lecture. We were discussing the people involved in the Sarajevo Assassination, and at the end of the lecture there was a question whether it was possible to prompt for the interactions between those individuals. Not only is it possible—we can also generate a neat sociogram. Here's how:

**Please read: [https://en.wikipedia.org/wiki/Assassination\\_of\\_Archduke\\_Franz\\_Ferdinand](https://en.wikipedia.org/wiki/Assassination_of_Archduke_Franz_Ferdinand).**

**Questions will follow. Ok?**

Got it. Ready when you are.

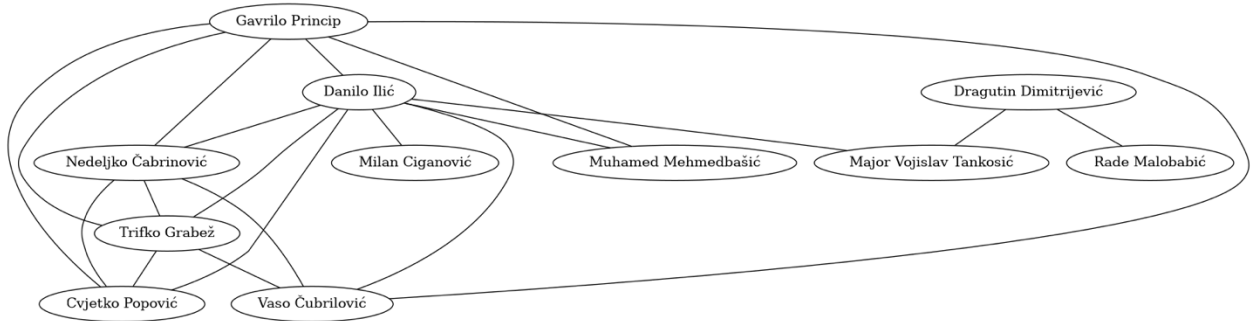
**Find pairs of people that were in contact with each other prior to assassination. Use the form: person - person. Search for at least 20 such interactions.**

Here are at least 20 pairs of people who were in contact with each other prior to the assassination of Archduke Franz Ferdinand, based on the Wikipedia article 1:

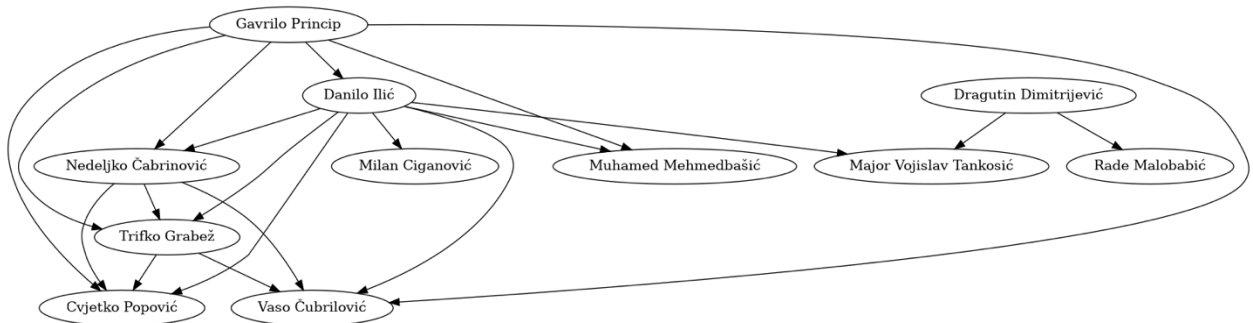
Gavrilo Princip – Danilo Ilić  
Danilo Ilić – Nedeljko Čabrinović  
Danilo Ilić – Trifko Grabež  
Danilo Ilić – Vaso Čubrilović  
Danilo Ilić – Cvjetko Popović  
Danilo Ilić – Muhamed Mehmedbašić  
Danilo Ilić – Milan Ciganović  
Danilo Ilić – Major Vojislav Tankosić  
Gavrilo Princip – Trifko Grabež  
Gavrilo Princip – Nedeljko Čabrinović  
Gavrilo Princip – Vaso Čubrilović  
Gavrilo Princip – Cvjetko Popović  
Gavrilo Princip – Muhamed Mehmedbašić  
Nedeljko Čabrinović – Trifko Grabež  
Nedeljko Čabrinović – Vaso Čubrilović  
Nedeljko Čabrinović – Cvjetko Popović  
Trifko Grabež – Vaso Čubrilović  
Trifko Grabež – Cvjetko Popović  
Dragutin Dimitrijević – Major Vojislav Tankosić  
Dragutin Dimitrijević – Rade Malobabić

**Construct a interaction graph where people are nodes and edges are their interactions.  
Render this graph with Python script, use graphviz, save the result in an image file and show it here.**

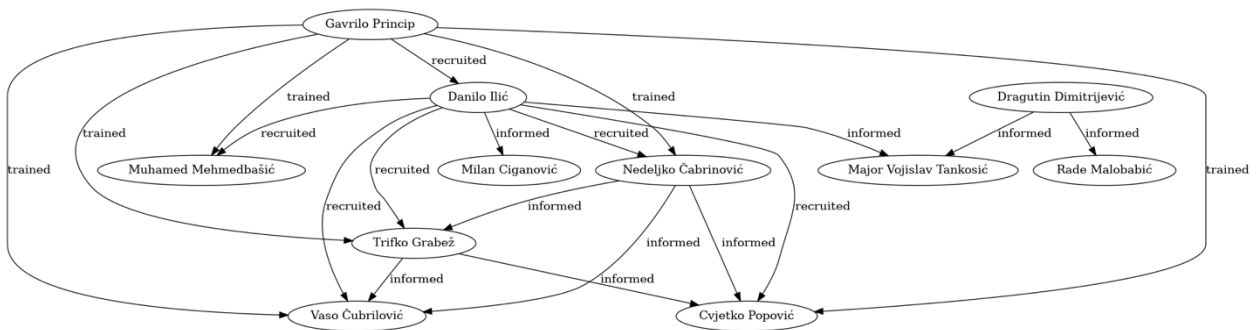
Here is the interaction graph showing connections between individuals involved in the assassination of Archduke Franz Ferdinand:



**Edges should be directed. A -> B, where A is the one that initiated the contact.**

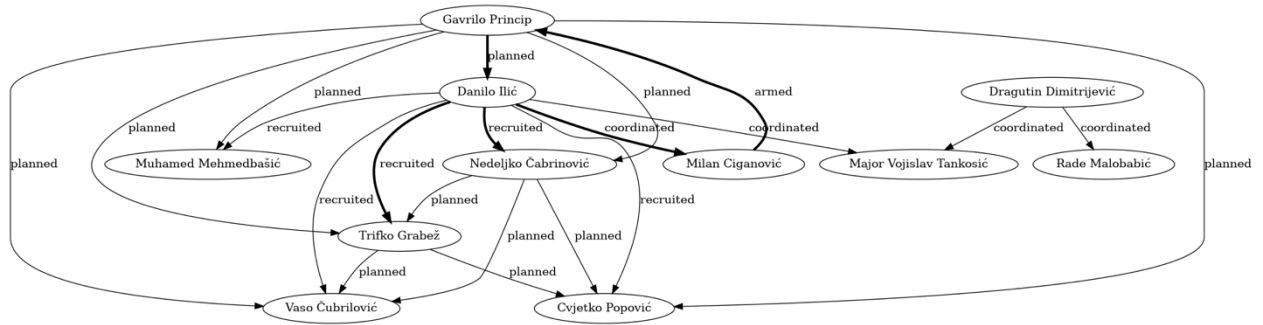


**Describe every interaction with one word, and use this as an edge label. Revise the graph accordingly.**



**Add edges that show the arming of Gavril Princip. Denote most important interactions with thick lines.**

Here's the enhanced interaction graph with thicker lines highlighting the most important interactions, such as recruitment and arming of Gavril Princip:



Disclaimer: the author of this text does not claim anything is right in this graph, though some interactions at least make sense, and others would need to be verified and scrutinized by somebody with knowledge of history. The only purpose of this section was to show that such graphs can be created, and because of the Graphviz package (created in 1991!) they can be very neat.

## DRAWING INTERACTION GRAPHS: GENE NETWORKS

Here is one more graphing example, perhaps better fitting to biomedical audience, but using almost exactly the same tricks as above.

### **Name most important genes in Dictyostelium development.**

Some of the most important genes involved in Dictyostelium discoideum development include:

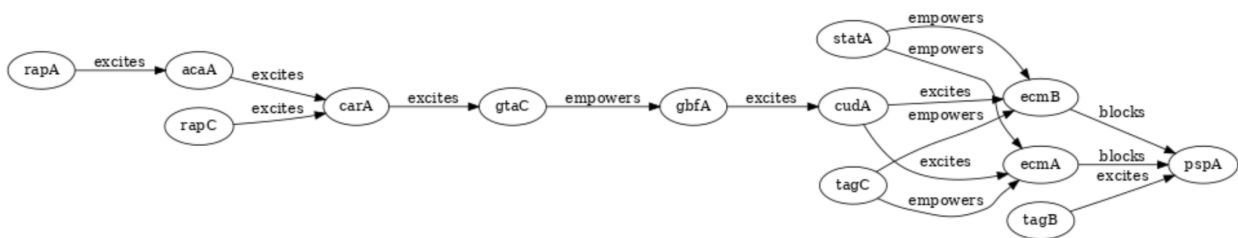
- acaA – Encodes adenylyl cyclase A, crucial for cAMP signaling during aggregation.
- carA – cAMP receptor A, essential for detecting cAMP and initiating aggregation.
- gtaC – A transcription factor regulating gene expression during early development.
- gbfA – G-box binding factor, another transcription factor controlling developmental genes.
- cudA – Required for culmination; expressed in the tip organizer.
- statA – A STAT-like transcription factor regulating prestalk cell differentiation.
- ecmA / ecmB – Marker genes for prestalk cell subtypes.
- tagB / tagC – Involved in cell-type-specific gene regulation during differentiation.
- pspA – Prespore-specific gene, a marker of spore cell fate.
- rapA / rapC – Regulate aggregation and differentiation via Ras signaling.

These genes play roles in signaling, transcriptional regulation, and cell fate determination throughout the stages from aggregation to fruiting body formation.

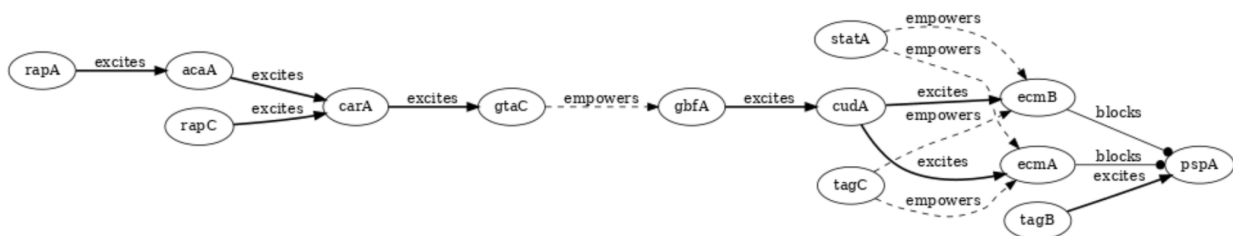
**List (epistatic) interactions between these genes. Like: gene -> gene: name of the interaction where "->" means influences (the expression), and where "name of the interaction" is a short name of what the interaction is (say, inhibits, excites, blocks, empowers, use one word only.**

acaA -> carA: excites  
 carA -> gtaC: excites  
 gtaC -> gbfA: empowers  
 gbfA -> cudA: excites  
 cudA -> ecmA: excites  
 cudA -> ecmB: excites  
 statA -> ecmA: empowers  
 statA -> ecmB: empowers  
 ecmA -> pspA: blocks  
 ecmB -> pspA: blocks  
 tagB -> pspA: excites  
 tagC -> ecmA: empowers  
 tagC -> ecmB: empowers  
 rapA -> acaA: excites  
 rapC -> carA: excites

Now use this list, and construct a graph with genes as nodes and relations as directed edges (instead of arrow, inhibits is a full circle). Annotate the relations. Create the graph using Python and Graphviz, and store the image; this is the result. Render the image (inline).



Use different types of arrows for different relations. Blocks could have a filled circle at the end of the edge instead of the arrow. Excites could be strong arrows, and empowers a dashed one.



Depending on the verbosity of the chatbots there is usually some more text associated with such chat, and we have left it out. Copilot Chat would also offer to draw its own graph, and while it is perfectly ok to try, they still need to work on this.

## INFORMATION EXTRACTION: CALLS FOR GRANT PROPOSAL

Calls for grant proposals—especially in Europe—can be beasts of documents, often running over a hundred pages. And let's be honest: few people have the time or patience to read through all of it. But that's where AI can help. Instead of manually combing through these documents, we can prompt a language model to extract the most relevant info for us—things like deadlines, available grant types, and funding structures.

During class, we worked through a real example. We took one of these lengthy grant calls and showed how to extract key information using a well-crafted prompt. But we didn't stop there. The interesting part came when we connected this extracted information with details about the group intending to apply. This could be stored in a simple file or even pulled from a personal CV. With both sides—grant data and proposer profile—in hand, we could start prioritizing: figuring out which calls were the best fit, which topics aligned well, and even identifying who else to bring on board for a strong collaboration.

## CITATIONS

When reporting on our research work and results, we, as researchers, need to place our findings within the existing body of knowledge. In other words, we must cite the work of others, explain how it influenced, inspired, and guided our research, and compare our results with similar findings in the field to highlight what is new. Ideally, researchers would read and remember a vast body of literature from their field, perhaps maintaining their own database of references and knowing them by heart—so well that they could write such paragraphs even if woken in the middle of the night. A slightly less ideal, yet perfectly valid approach from an exploratory perspective, is to write a paragraph and use AI to suggest relevant citations—that is, papers worth reading. This is exactly what we did in class. Here's an example.

**Consider a following paragraph from the paper that I am writing. Identify a sentence which needs citation. Then, propose three papers that I could cite. Make sure these are real papers (find them on the web) from respectful journals (high IF) that have been published within the past 10 years. For each paper, provide a reference (authors, title, journal, single line please). Be brief. My paragraph is:**

**Molecular biology and the whole field of biomedicine and health, has been revolutionized by high-throughput technology over the past two decades. For example, we can now assess the physiological state of any tissue by measuring gene expression. Gene expression is the process by which information from a gene is used to synthesize proteins. We can assess the state of tissues and cells by showing which genes are active and**

**affecting the structure and function of the tissue. In medicine, gene expression studies are essential for understanding disease mechanisms, diagnosing, predicting treatment responses, and developing targeted therapies. Precision medicine is one of the most promising advances in healthcare, which aims to assess the state of disease of an individual to make clinical decisions tailored to that individual. It relies on gene variation and expression for diagnostic and prognostic markers and to support treatment selection.**

The sentence "In medicine, gene expression studies are essential for understanding disease mechanisms, diagnosing, predicting treatment responses, and developing targeted therapies." would benefit from citations. Here are three relevant papers:

1. Grossmann V, Tiacci E, Holmes AB, et al. Whole-exome sequencing identifies somatic mutations of BCOR in acute myeloid leukemia with normal karyotype. *Blood*. 2011;118(23):6153-6163.
2. Tiacci E, Trifonov V, Schiavoni G, et al. BRAF mutations in hairy-cell leukemia. *New England Journal of Medicine*. 2011;364(24):2305-2315.
3. Falini B, Bolli N, Shan J, et al. Both carboxy-terminus NES motif and mutated tryptophan(s) are crucial for aberrant nuclear export of nucleophosmin leukemic mutants in NPMc+ AML. *Blood*. 2006;107(11):4514-4523.

**I need a survey paper, not examples of research you have provided. Adhere to the reference format I requested above. Please propose an alternative.**

Apologies for the earlier oversight. For a comprehensive survey on the role of gene expression studies in understanding disease mechanisms, diagnosis, and treatment, you may consider citing the following paper:

Hasin Y, Seldin M, Lusi A. Multi-omics approaches to disease. *Genome Biology*. 2017;18(1):83.

This review discusses how integrating gene expression data with other omics approaches enhances our understanding of complex diseases and informs diagnostic and therapeutic strategies.

**This one looks great. Please provide a reference in bibtex, and include a citation in a sentence that required a reference. I will use latex.**

Here is the reference in BibTeX format:

@article{Hasin2017,  
author = {Yaron Hasin and Michael Seldin and Aldons Lusi},

```
title = {Multi-omics approaches to disease},  
journal = {Genome Biology},  
year = {2017},  
volume = {18},  
number = {1},  
pages = {83},  
doi = {10.1186/s13059-017-1215-1}  
}
```

And here is the revised sentence with the citation:

In medicine, gene expression studies are essential for understanding disease mechanisms, diagnosing, predicting treatment responses, and developing targeted therapies  
\cite{Hasin2017}.

A note on this endeavor: Several specialized tools have recently emerged to assist with finding and citing references. One such tool is scite.ai, which, at the time of writing these notes, is—based solely on the author's experience—often frustrating to use, inaccurate, and overly expensive. While these tools will undoubtedly improve, it is worth exploring what can already be achieved with current chatbots, which may perform better and can be more precisely guided. After all, tools like scite.ai are essentially just wrappers around LLMs, generating prompts based on user input and displaying results in a structured format. However, their preprogrammed approach may not always align with individual research needs.

## ASSISTANCE IN REVIEWING

A quick disclaimer: your lecturer strongly believes that every reviewer should carefully read the paper they're reviewing. Only after that should they draft their review. AI should *not* replace a human in this process. That said, AI can still be a helpful assistant. Once the draft review is written, you can use AI to check it—spot anything that's off, flag mistakes, or point out if something important is missing. This is exactly the process we illustrated with the example in class. In the prompting, we used constructs like:

**I am reviewer of the attached paper. Here is my initial review. Please check it out, critically assess every claim that I made, and let me know if it is well-supported, right, or should be revised in any way.**

**Is there anything else my review should address?**

**Any other major criticism (or praise) of the paper that I have forgot to include in my review?**



## FOR THE END

We've explored how AI chatbots can support real research—from drawing graphs and extracting information to generating citations and reviewing papers. Along the way, we've seen that good prompting is a skill you can learn. And the best way to master it? Just keep prompting. Try things out. Be bold. Use AI not only for what's useful, but also for your wildest ideas. After all, it's hard work, clever experimentation, and even implementation of a few crazy ideas that drive science forward. Mastering prompt engineering is no different.