

Angewandte Generative KI in den (digitalen) Geisteswissenschaften

Einführung in das Prompt Engineering

<http://chpollin.github.io/GM-DH>

Christopher Pollin

<https://chpollin.github.io>

Zentrum für Informationsmodellierung, Graz

<https://informationsmodellierung.uni-graz.at/de/>

Digital Humanities Craft OG

www.dhcraft.org

TEI Modeller - Named Entities. ChatGPT September 25 Version. October 5, 2023.

<https://chat.openai.com/share/89449f80-1f46-4c38-82b3-8125d791308d>

TEI Modeller - Handwritten correspondence. ChatGPT September 25 Version. October 5, 2023.

<https://chat.openai.com/share/38b65c7c-dcc7-427b-a6fa-28558d98fae2>

“TEI Modeler” by Christopher Pollin [GPT-4]

Take a deep breath and think step by step. I'm delving into [*Type of Historical Text: e.g., manuscripts, letters, legal documents*] and am confronted with modeling challenges. I aim to make the best use of the P5: Guidelines for Electronic Text Encoding and Interchange (TEI). Can you, as an expert in TEI textual modeling, systematically guide me through this? For each modeling challenge I describe, accompanied by phenomena details and a sample text, I'd appreciate a comprehensive exploration of ways to model the phenomena in TEI. It's essential to weigh different modeling approaches like granularity, simplicity, and flexibility. Dive deep and provide explanations tailored for someone well-versed in advanced TEI, and if there's ever a moment of doubt about the evolving TEI standards post-2021, please request specific parts of the current TEI Guidelines to be pasted for verification. List out the pros and cons of each method, and showcase the example texts as markdown code blocks with TEI XML. Ignore verbosity and complexity as a cons. Additionally, identify any pitfalls to sidestep during TEI modeling. The TEI must always be valid and the XML well-formed. Never use elements that do not exist in the TEI.

Text phenomena to be modeled: [*phenomena*]

Excerpt from text source

...

[*sample text*]

...

Grundlagen-der-Informatik-Assignment-Solver - Wahrheitstabelle. ChatGPT September 25 Version. October 5, 2023. <https://chat.openai.com/share/92483933-777c-4e28-9a62-1814ca7af1f7>

Grundlagen-der-Informatik-Assignment-Solver - RegEx. ChatGPT September 25 Version. October 5, 2023. <https://chat.openai.com/share/0dc7905-a0eb-4ac3-833f-3e96cbab9b43>

“Grundlagen-der-Informatik-Assignment-Solver” by Christopher Pollin [GPT-4]

Take a deep breath and think step by step. As an expert in didactics and computer science, especially in areas like Data Information Knowledge, Information Theory, bits & bytes and character encoding, Boolean algebra and propositional logic, Sets and predicate logic, Formal languages, regular expressions and automata, Internet and Web, Turing machines, programming (C), data structures, and assembler, I'd like a detailed definition followed by the solution to the following input:

...

[word, term, phrase, paragraph or exercise related to introduction to computer science]

...

If necessary, use the Wolfram Alpha plugin to help solve problems. In addition, offer a simple analogy or metaphor for easier understanding. Always ask the user to give you a syntax example for Wolfram Alpha

Only give an answer if you are sure that your answer is correct. If not, ask for further information.

Klarheit im Begriffsschubel



Midjourney: ai, two male programmer with laptops holding hands in front of a dense jungle, footpath in the jungle, landscape, cinematic lighting, Geof Darrow style
--weird 50 --chaos 10 --quality 0.25 --ar 16:9

Was sind Large Language Models (LLM)?

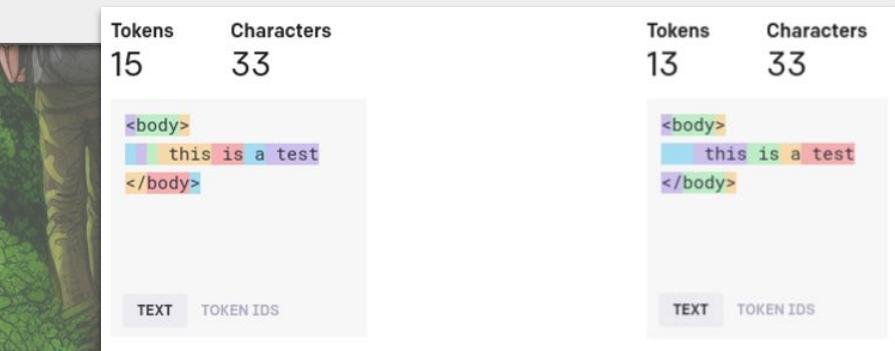
Auf Deep-Learning-Techniken basierende Transformer-Modelle, die mit riesigen Datenmengen in natürlicher Sprache trainiert und mit menschlichen Feedbackschleifen (RLHF) verfeinert wurden, um eine “*Reasoning Maschine*” zu schaffen. (GPT-3.5, GPT-4, BERT, LLAMA 2)

Token & Embedding

“A helpful rule of thumb is that one token generally corresponds to ~4 characters of text for common English text [...] (so 100 tokens ~ = 75 words)”.

<https://platform.openai.com/tokenizer>

Token werden als Zahlenfolge in einem multi-dimensionalen Vektorraum abgebildet. Die Abbildung im Vektorraum (= Embedding) repräsentiert die “Bedeutung” des Tokens.



Erkläre mir, was Projektmanagement ist!

Projektmanagement ist die Kunst der Planung, Organisierung von Ressourcen, um bestimmte Ziele innerhalb eines Budgets zu erreichen. Es ist ein systematischer Ansatz, der zu einem Ergebnis führt, normalerweise innerhalb des Umfangs, den es vorgibt. Projektmanagement findet Anwendung in vielen verschiedenen Bereichen wie dem Gesundheitswesen, Energie und vielen mehr.

Kernkomponenten des Projektmanagements

ChatGPT

... ist ein von OpenAI entwickeltes KI-basiertes (fine-tuned GPT) Sprachmodell und Chatbot, das in der Lage ist, menschenähnlichen Text auf der Basis von Kontext und früheren Unterhaltungen zu erzeugen.

Prompt

... ist ein Text in natürlicher Sprache, der die Aufgabe beschreibt, die eine KI ausführen soll = Input für eine KI.

Prompt Engineering

... ist der Prozess des Entwerfens, Verfeinerns und Optimierens von Prompts, um die Absicht der User*innen effektiv an ein LLM zu kommunizieren.



Erkläre mir, was Projektmanagement ist!



Projektmanagement ist die Kunst der Planung von Ressourcen, um bestimmte Ziele innerhalb von Budgets zu erreichen. Es ist ein systematisches Ergebnis, normalerweise innerhalb des Umlands.

Projektmanagement findet Anwendung in vielen verschiedenen Branchen wie IT, Bauwesen, Gesundheitswesen, Energie und vielen mehr.

Kernkomponenten des Projektmanagements:

1. **Initiierung:** Die Phase, in der das Projekt definiert und genehmigt wird.



You are an expert project manager and scientist. What are the most important things to consider when planning my research project?

My project is about developing best practice for research data acquisition workflows. The project team consists of 5 people.

Go through your analysis step by step. Be very detailed and focus on the project management aspect.

Give me the results in a markdown table.



Prompt Engineering is ... Context, Context, Context

Now, let's summarize these points in a Markdown table.

Category	Key Consideration	Description
Objectives	Clarity of Purpose	Define the key aims and questions of the project.
	Success Metrics	Establish KPIs to measure success.
Scope	Inclusions and Exclusions	Specify what is and isn't part of the project.
	Boundaries	Define the limits in terms of data sources, methodologies, etc.
Team & Roles	Skill Mapping	Map team skills to project needs.
	Role Assignment	Assign roles and responsibilities.
Timeline	Milestones	Identify key milestones and deadlines.
	Work Breakdown Structure (WBS)	Allocate tasks across timeline and team members.
Resources	Budget	Outline financial resources and allocation.
	Tools and Software	Choose the necessary software and tools.
Risk Management	Identification	Identify potential risks.
	Mitigation Plans	Create contingency plans for identified risks.
Communication & Documentation	Communication Channels	Establish communication methods and frequency.
	Documentation	Decide on documentation and storage methods.

Grundlagen des Prompt Engineering - Prinzipien und Techniken

Hauptbestandteile eines Prompts

Prompt Engineering Prinzipien

Anwendungsbeispiele

Einsatzbereiche von GPT

Zusammenfassung	“Extract the main findings from the following text”
Klassifizierung	“Classify the following text”
Übersetzung	“Translate the following text to <language>”
Textgenerierung	“Write a <type> about <topic>.”
Beantwortung von Fragen	“Explain <topic> to me”
Feedback	“Give me (critical) feedback on the following text”.
Verbesserung	“Improve the following text code”.

... es geht immer besser - Prompt Engineering 1/2

Zusammenfassung

“Extract the key findings from the following text”

Verbesserte Version:

“Summarize the key points from the following text **in three sentences.**”

Klassifizierung

“Classify the following text”

Verbesserte Version:

“Classify the following text **into one of the following categories: Business, Technology, Entertainment.**”

Übersetzung

“Translate the following text to <language>”

Verbesserte Version:

“Translate the following text to <language>, **maintaining the tone and context.**”

Textgenerierung

“Write a <type> (e.g., essay, article) about <topic>.”

Verbesserte Version:

“Compose a <type> about <topic>, **emulating the writing style of <style/author>.**”

... es geht immer besser - Prompt Engineering 2/2

Beantwortung von Fragen

Verbesserte Version:

“Explain <topic> to me”

“Explain <topic> to me **in layman's terms, focusing on its practical applications.**”

“Explain <topic> to me like I'm 10 years old (and show me why it's cool or useful)”

Feedback

Verbesserte Version:

“Give me (critical) feedback on the following text”.

“**Provide constructive criticism on the following text, highlighting areas for improvement.**”

Verbesserung

Verbesserte Version:

“Improve the following text|code”.

“**Optimize the following text|code for readability and efficiency.**”

Prompt Engineering Principles 1

Spezifität und Klarheit: Prompts sollten spezifisch und eindeutig sein. Vage oder mehrdeutige Vorgaben können zu ungenauen oder unerwünschten Ergebnissen führen. Es ist wichtig, genau zu definieren, was vom Modell erwartet wird und welche Aktionen es ausführen soll.

Zeit zum “Nachdenken” geben: Große Sprachmodelle wie GPT-4 benötigen Zeit, um Informationen zu verarbeiten und auf Eingabeaufforderungen zu reagieren. Es ist wichtig, dem Modell genügend Zeit zum “Nachdenken” zu geben, bevor eine Antwort erwartet wird.

Kontext und Beispiele verwenden: Kontext und Beispiele können helfen, die Qualität der Antworten zu verbessern. Sie können dem Modell helfen, die Anforderungen der Frage besser zu verstehen und relevante und genaue Antworten zu geben.

Iterativer Ansatz: Die Entwicklung von Prompts ist oft ein iterativer Prozess. Es kann mehrere Versuche und Anpassungen erfordern, um den “perfekten” Prompt zu finden. Es ist wichtig, offen für Änderungen und stets bereit zu sein, den Prompt auf der Grundlage der erhaltenen Antworten zu verfeinern und anzupassen.

Prompt Engineering Principles 2

Verständnis der Fähigkeiten von GPT: Die Fähigkeiten des Modells liegen darin, zusammenzufassen, Informationen abzuleiten, Daten in verschiedene Formate umzuwandeln und Ideen zu entwickeln, wie auch Texte erweitern.

Explizite Einschränkungen verwenden: Manchmal kann es nützlich sein, explizite Einschränkungen in den Prompt einzufügen, um die Ausgabe zu steuern: Länge der Antwort, Detailgrad, Format oder Struktur der Ausgabe sind wichtig.

Vermeidung von Überlastung: Zu komplexe oder zu viele Aufgaben auf einmal können für das Modell problematisch sein und zu ungenauen oder unvollständigen Antworten führen. Oft ist es besser, komplexe Aufgaben in kleinere, leichter zu verarbeitende Teile zu zerlegen.

Proposal - Funding Call 2022: Interoperability and reusability of DH data and tools

Titel: Sharing the CROWN - Establishing a Workflow from Collection Data to Linked Research Data

The Imperial Crown of the Holy Roman Empire is one of the most important symbols of European history. Today it is part of the collections of the Kunsthistorisches Museum in Vienna (KHM). In the course of the so-called CROWN project initiated by the KHM, a comprehensive analysis of the Imperial Crown is being carried out. For this purpose, all components like gems, pearls, plates, etc. will be analysed from a scientific, historical and art-historical point of view. This project, which is currently running until 2024, is thus a truly interdisciplinary endeavour. Research data resulting from the application of highly sophisticated analytical techniques to study manufacturing techniques and materials used, is recorded using The Museum System (TMS). TMS is a widely used software solution designed for museums. It provides a relational database that can be used for inventorying, documenting, and managing collections.

The CROWN project is faced with an in-depth and complex analysis of a single object. It must fulfil the scientific mandate of a world leading museum and has to be based on the already established TMS for data management. Therefore, it is necessary to implement newly developed workflows in the CROWN project using TMS. Concerning data processing this is a flagship project within the KHM Museumsverbund. It is no longer "usual" collection





Extract the main findings from the following text and translate them to german:

Sharing the CROWN - Establishing a Workflow from Collection Data to Linked Re

The Imperial Crown of the Holy Roman Empire is one of the most important symbols of European history. Today it is part of the collections of the Kunsthistorisches Museum Vienna (KHM). In the course of the so-called CROWN project initiated by the KHM, a comprehensive analysis of the Imperial Crown is being carried out. For this purpose, components like gems, pearls, plates, etc. will be analysed from a scientific, historical and art-historical point of view. This project, which is currently running until 2024, is an interdisciplinary endeavour. Research data resulting from the application of highly sophisticated analytical techniques to study manufacturing techniques and materials is recorded using The Museum System (TMS). TMS is a widely used software specifically designed for museums. It provides a relational database that can be used for inventarizing, documenting, and managing collections.

“Main Findings Translated to English: ... “

Warum wird zuerst auf Englisch zusammengefasst?

...

Trennung von Anweisung und zu verarbeitenden Text.

ChatGPT-4 Conversations

Main Findings Translated to German:

1. **Interdisziplinäre Analyse:** Das CROWN-Projekt des Kunsthistorischen Museums in Wien zielt darauf ab, eine umfassende, interdisziplinäre Analyse der Reichskrone des Heiligen Römischen Reiches durchzuführen.
2. **Datenmanagement:** Das Projekt verwendet The Museum System (TMS), eine weit verbreitete Software für Museen, für das Datenmanagement. Es muss hochspezifische Forschungsdaten verarbeiten, die über die üblichen Sammlungsdaten hinausgehen.
3. **Zusammenarbeit:** Aufgrund des Fehlens spezifischer Fachkenntnisse im Museum wird das Projekt in Zusammenarbeit mit dem Institut Zentrum für Informationsmodellierung, Universität Graz, durchgeführt.
4. **FAIR-Kriterien:** Das Ziel ist die Erstellung eines Workflows, der Daten in FAIR (Findable, Accessible, Interoperable, Reusable) Forschungsdaten umwandeln kann.
5. **Semantischer Webansatz:** Das Projekt wird einen semantischen Webansatz verwenden, um die Daten formal darzustellen, basierend auf einer domänenpezifischen Ontologie, die von CIDOC-CRM und Basic Formal Ontology abgeleitet ist.
6. **Arbeitspakete:** Das Projekt beinhaltet Aufgaben wie die Modellierung und Formalisierung einer Domäne als Ontologie, die Umwandlung von Daten in ein Linked Open Data und FAIR-Datensatz, die semantische Anreicherung von Daten und die Implementierung eines Proof-of-Concept-Prototyps.
7. **Dokumentation und Wiederverwendbarkeit:** Eines der wesentlichen Ziele ist die ordnungsgemäße Dokumentation der Erkenntnisse und Methoden, um die Wiederverwendbarkeit des Workflows in anderen Museumsystemen sicherzustellen.



Explain the key findings to me like I'm 10 years old and show me why it's cool or useful.

Context Window
("Aufmerksamkeitsspanne") sind
32.000 Token!

Komplexe Themen im Stil eines
10-Jährigen erklären, um Sachverhalte
besser verstehen zu können (aber
auch Verlust von Informationsdichte, da
komplexe Sachverhalte mit dem
Vokabular eines 10-Jährigen nicht
erklärt werden können)



Sure, let's break down the big ideas from that long text into something easier to understand, and I'll tell you why each one is super cool!

Key Findings Explained for a 10-Year-Old

1. **Studying a Super Cool Crown:** Imagine a really ancient and fancy crown that was worn by emperors a long time ago. A museum in Vienna is studying this crown super closely. They're looking at every gem, pearl, and piece of metal to learn everything they can about it!
 - **Why It's Cool:** It's like being a detective but for history! You get to find out all the secrets this crown holds.
2. **Special Computer Program:** They use a computer program made just for museums to keep track of everything they find out about the crown.
 - **Why It's Cool:** Imagine having a super-organized digital treasure chest where you keep all the cool facts and discoveries. You'd never forget anything!
3. **Team-Up with Brainiacs:** The museum is working with some really smart people from a university because they need special knowledge to study the crown properly.
 - **Why It's Cool:** It's like when superheroes team up to defeat a big villain. More brainpower equals more discoveries!
4. **Making Data Friendly:** They want to make sure all the information they collect can be easily found, read, and used by other people in the future.
 - **Why It's Cool:** Imagine you made a really detailed treasure map and shared it, so others can also find cool stuff easily. Sharing is caring!
5. **Data Gets Its Own Language:** They're using a special way to describe everything they find so computers can understand it too.
 - **Why It's Cool:** It's like teaching your computer a secret language that helps it understand complicated things about the crown. Like teaching your pet dog tricks but way more advanced!



Provide constructive criticism on the initial text, highlighting areas for improvement.

T

Certainly, the initial text is quite detailed and informative, but there are several aspects that could be improved to make it more effective and accessible to a wider audience. Here's my constructive criticism:

Language and Terminology

1. **Technical Jargon:** The text is filled with technical terms like "domain-specific ontology," "FAIR criteria," and "Semantic Web approach." While these terms are appropriate for a specialized audience, they could be confusing for a general reader.
 - **Improvement:** Consider adding a glossary or footnotes to explain these terms in simpler language. Alternatively, provide examples to make the concepts easier to understand.

Structure and Flow

1. **Long Paragraphs:** Some of the paragraphs are quite long, making it difficult to absorb the information.
 - **Improvement:** Break down long paragraphs into shorter ones, focusing each on a single idea for better readability.
2. **Lack of Subheadings:** The text could benefit from more subheadings to guide the reader through the various sections.
 - **Improvement:** Use subheadings to highlight the main points or stages of the project, such as "Objectives," "Collaboration," "Methodology," and so on.



Revise the original text and incorporate the feedback.

T

Certainly, I'll revise the original text to make it more reader-friendly while retaining its informative content. I'll incorporate subheadings, simplify technical terms, and make the text more engaging.

“This project is like a history detective mission ... “

Warum ist der Stil nicht wissenschaftlich, wie beim Antragstext zu Beginn?



Überarbeite den ursprünglichen Textes und führe die Einarbeitung des Feedbacks durch

Deutscher Prompt → Deutsche Response
Englischer Prompt → Englische Response

Introduction: Why the Imperial Crown is a Big Deal

The Imperial Crown of the Holy Roman Empire isn't just a pretty piece of jewelry; it's a powerful symbol of European history. Right now, it's housed at the Kunsthistorisches Museum in Vienna (KHM). Our project, called CROWN, aims to study this crown from every angle—its gems, pearls, plates, and more. This project is like a history detective mission, blending science, history, and art!

Objectives: What We Aim to Do

The goal of CROWN is simple but ambitious:

- Analyze the Imperial Crown in a w
art.

T

Natürlich, ich werde den ursprünglichen Text unter Berücksichtigung des Feedbacks überarbeiten, um ihn lesbarer und zugänglicher zu gestalten.

Einführung: Warum die Reichskrone eine große Sache ist

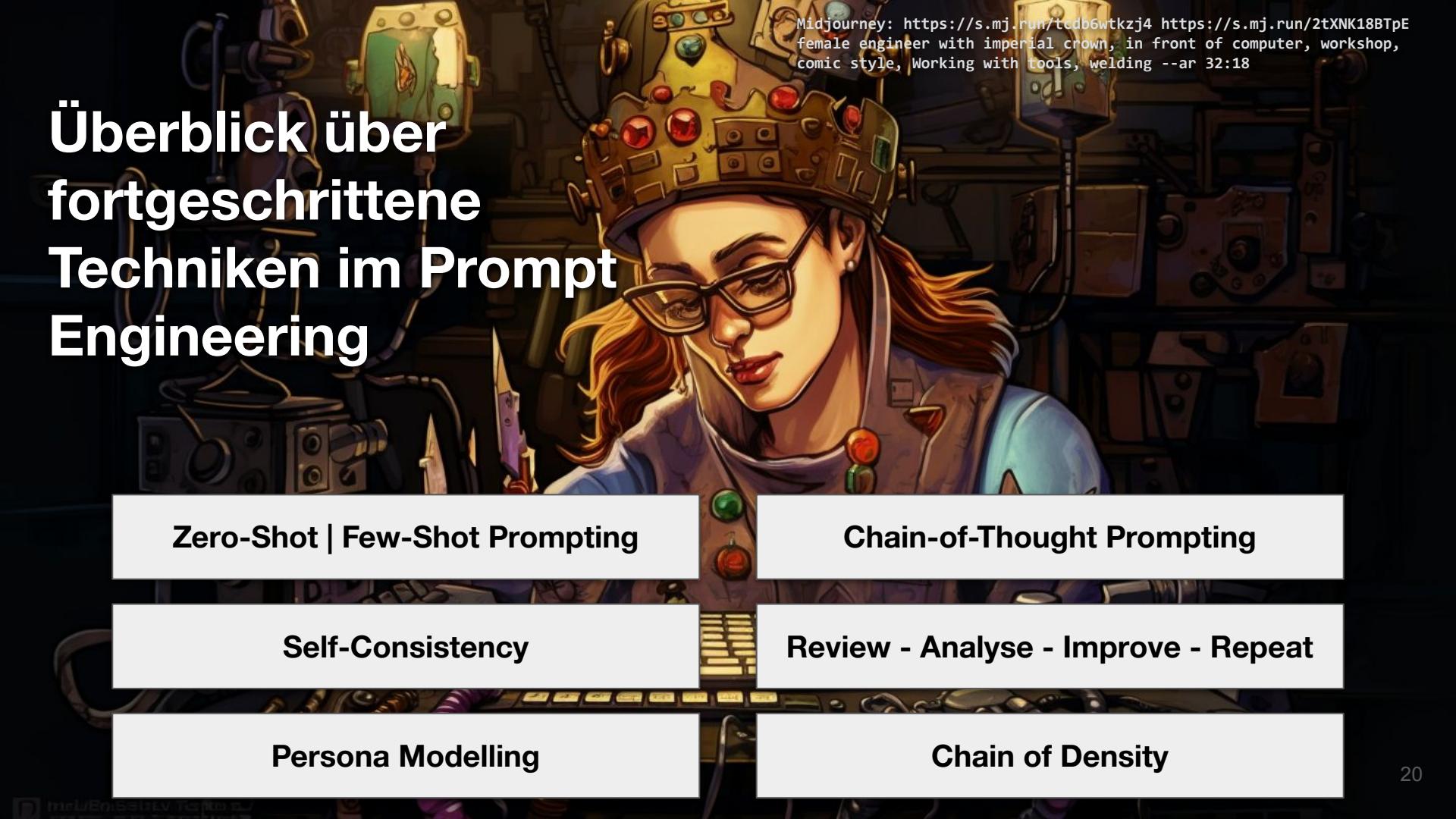
Die Reichskrone des Heiligen Römischen Reiches ist nicht nur ein hübsches Schmuckstück; es ist ein mächtiges Symbol der europäischen Geschichte. Sie befindet sich derzeit im Kunsthistorischen Museum in Wien (KHM). Unser Projekt, genannt CROWN, hat das Ziel, diese Krone von allen Seiten zu untersuchen—ihre Edelsteine, Perlen, Platten und mehr. Dieses Projekt ist wie eine detektivische Mission, die Wissenschaft, Geschichte und Kunst miteinander verbindet!

ChatGPT-4 und seine Besonderheiten (die man wissen muss!)

- **GPT-4 weiß und versteht nichts:** Ermittelt wird immer nur das nächstbeste Token. GPT-4 muss zuerst die Key Findings in Englisch in die *Conversation* schreiben, damit es dann gut übersetzt werden kann. GPT-4 ist jedoch in der Lage, Reasoning über Text durchzuführen.
- **Der Prompt beeinflusst den Output:** Wenn der *Prompt* auf Deutsch ist, wird die Ausgabe (höchstwahrscheinlich) auch auf Deutsch sein, es sei denn, es wird ausdrücklich anders angegeben.
- **GPT-4 hat eine begrenzte “Aufmerksamkeitsspanne”:** GPT-4 hat ein maximales Token-Limit von 32.000 Token für sowohl die Eingabeaufforderung als auch die Antwort. Wenn die Unterhaltung dieses Limit überschreitet, geht Information verloren.
- **“Kontextverschmutzung”:** Der Kontext oder Stil, in dem etwas gefragt wird, kann zukünftige Antworten (innerhalb einer *Conversation*) beeinflussen. Wenn man beispielsweise GPT-4 anweist, es solle so antworten, als würde es mit einem 10-Jährigen sprechen, kann dieser Kontext auf nachfolgende Anfragen übertragen werden.

ChatGPT-4 und seine Besonderheiten (die man wissen muss!)

- **GPT-4 kann nicht zählen oder rechnen und nur bedingt Logik.**
- **Es ist hilfreich, die Antworten einfach neu generieren zu lassen.**



Überblick über fortgeschrittene Techniken im Prompt Engineering

Midjourney: <https://s.mj.run/tcdb6wtkj4> <https://s.mj.run/2tXNK18BTpE>
female engineer with imperial crown, in front of computer, workshop, comic style, Working with tools, welding --ar 32:18

Zero-Shot | Few-Shot Prompting

Chain-of-Thought Prompting

Self-Consistency

Review - Analyse - Improve - Repeat

Persona Modelling

Chain of Density

“Wie man etwas sagt, ist sehr wichtig!”

Scorer	Optimizer / Source	Instruction position	Top instruction	Acc
<i>Baselines</i>				
PaLM 2-L	(Kojima et al., 2022)	A_begin	Let's think step by step.	71.8
PaLM 2-L	(Zhou et al., 2022b)	A_begin	Let's work this out in a step by step way to be sure we have the right answer.	58.8
PaLM 2-L		A_begin	Let's solve the problem.	60.8
PaLM 2-L		A_begin	(empty string)	34.0
text-bison	(Kojima et al., 2022)	Q_begin	Let's think step by step.	64.4
text-bison	(Zhou et al., 2022b)	Q_begin	Let's work this out in a step by step way to be sure we have the right answer.	65.6
text-bison		Q_begin	Let's solve the problem.	59.1
text-bison		Q_begin	(empty string)	56.8
<i>Ours</i>				
PaLM 2-L	PaLM 2-L-IT	A_begin	Take a deep breath and work on this problem step-by-step.	80.2
PaLM 2-L	PaLM 2-L	A_begin	Break this down.	79.9
PaLM 2-L	gpt-3.5-turbo	A_begin	A little bit of arithmetic and a logical approach will help us quickly arrive at the solution to this problem.	78.5
PaLM 2-L	gpt-4	A_begin	Let's combine our numerical command and clear thinking to quickly and accurately decipher the answer.	74.5
text-bison	PaLM 2-L-IT	Q_begin	Let's work together to solve math word problems! First, we will read and discuss the problem together to make sure we understand it. Then, we will work together to find the solution. I will give you hints and help you work through the problem if you get stuck.	64.4
text-bison	text-bison	Q_end	Let's work through this problem step-by-step:	68.5
text-bison	gpt-3.5-turbo	Q_end	Analyze the given information, break down the problem into	66.5

Prompting, Prompting, Prompting

Zero-Shot Prompting bezieht sich auf die Fähigkeit eines maschinellen Lernmodells, eine Aufgabe ohne vorherige spezifische Beispiele oder Anweisungen für diese Aufgabe auszuführen. Bei dieser Methode verlässt sich das Modell auf sein vorheriges Training und die allgemeinen Anweisungen im Prompt, um eine Antwort oder Lösung zu generieren.

Few-Shot Prompting Few-Shot Prompting ist eine Technik, bei der einem maschinellen Lernmodell einige Beispiele (oftmals als "Shots" bezeichnet) für eine bestimmte Aufgabe vorgelegt werden, um dem Modell einen Kontext oder eine Anleitung für die Aufgabe zu geben. Diese Beispiele helfen dem Modell, die gewünschte Aufgabe besser zu verstehen und korrekt auszuführen

Prompt Engineering Guide. <https://www.promptingguide.ai/>

Min, Sewon, Xinxi Lyu, Ari Holtzman, Mikel Artetxe, Mike Lewis, Hannaneh Hajishirzi, und Luke Zettlemoyer. „Rethinking the Role of Demonstrations: What Makes In-Context Learning Work?“ arXiv, 20. Oktober 2022. <https://doi.org/10.48550/arXiv.2202.12837>.

Zero-Shot & Few-Shot Prompting Beispiele

1. Zero-Shot Prompting:

Eingabe: "Konvertiere folgende bibliographische Daten in csv-Format: Autor: J.K. Rowling, Titel: Harry Potter und der Stein der Weisen, Verlag: Carlsen, Jahr: 1997"

Ausgabe:

mathematica

 Copy code

Autor, Titel, Verlag, Jahr

J.K. Rowling, Harry Potter und der Stein der Weisen, Carlsen, 1997

2. Few-Shot Prompting:

Beispiele (Shots):

- Autor: Mark Twain, Titel: Tom Sawyer, Verlag: ABC, Jahr: 1876 -> Mark Twain, Tom Sawyer, ABC, 1876
- Autor: Jane Austen, Titel: Stolz und Vorurteil, Verlag: XYZ, Jahr: 1813 -> Jane Austen, Stolz und Vorurteil, XYZ, 1813

Eingabe nach den Beispielen: "Autor: J.K. Rowling, Titel: Harry Potter und der Stein der Weisen, Verlag: Carlsen, Jahr: 1997"

Ausgabe:

mathematica

 Copy code

J.K. Rowling, Harry Potter und der Stein der Weisen, Carlsen, 1997

Prompting, Prompting, Prompting

Chain-of-Thought Prompting (CoT) bezeichnet eine Technik des Prompt Engineering, bei der komplexe Fragestellungen in mehrere logische oder sequentielle Schritte unterteilt werden, um eine tiefere oder genauere Antwort von einem Sprachmodell zu erhalten. Durch das Hinzufügen von Zwischenschritten oder -überlegungen in den Eingabeaufforderungen wird das Modell dazu angeleitet, den Denkprozess sichtbar zu machen und so komplexe Aufgaben zu lösen, die typischerweise mehrschichtige Überlegungen oder mehrstufiges Reasoning erfordern. CoT Prompting kann besonders nützlich sein, um komplexe Probleme zu zerlegen, Entscheidungsprozesse zu klären oder das Verständnis eines Modells für bestimmte Konzepte zu verbessern.

Self-Consistency ist eine fortgeschrittene Technik im Bereich des Prompt Engineering, die darauf abzielt, die Standard-Dekodierungsmethode in Chain-of-Thought (CoT) Prompting zu verbessern. Statt sich auf einen einzigen, oftmals gierigen Dekodierungspfad zu verlassen, generiert Self-Consistency multiple und vielfältige Denkpfade durch Few-Shot CoT. Anschließend wird die am konsistentesten Antwort aus diesen Generierungen ausgewählt. Dieser Ansatz wurde vorgeschlagen, um die Leistung von CoT Prompting bei Aufgaben zu steigern, die arithmetisches und allgemeines Denken erfordern.

Chain-of-Thought Prompting

Standard Prompting	Chain-of-Thought Prompting
<p>Model Input</p> <p>Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?</p> <p>A: The answer is 11.</p> <p>Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?</p>	<p>Model Input</p> <p>Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?</p> <p>A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.</p> <p>Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?</p>
<p>Model Output</p> <p>A: The answer is 27. X</p>	<p>Model Output</p> <p>A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9. ✓</p>

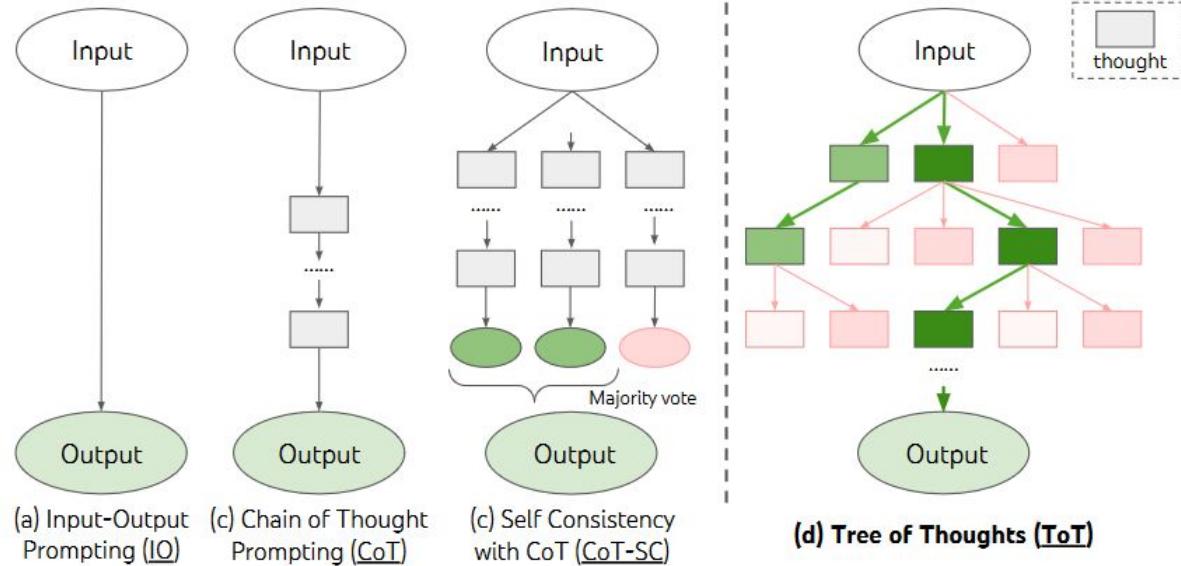


Figure 1: Schematic illustrating various approaches to problem solving with LLMs. Each rectangle box represents a *thought*, which is a coherent language sequence that serves as an intermediate step toward problem solving. See concrete examples of how thoughts are generated, evaluated, and searched in Figures 2|4|6.

Yao, Shunyu, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik Narasimhan. 'Tree of Thoughts: Deliberate Problem Solving with Large Language Models'. arXiv, 17 May 2023. <https://doi.org/10.48550/arXiv.2305.10601>.

Chain-of-Thought Prompting & Outline



A young researcher wants to submit a research proposal for a project called "Sharing the CROWN - Establishing a Workflow from Collection Data to Linked Research Data". He has a basic idea but is not sure how to structure and present the proposal. What steps should he take when writing the research proposal?

Take a deep breath and work on this problem step-by-step.

Research Proposal Writing Guide. ChatGPT August 3 Version.

<https://chat.openai.com/share/15dba3f6-b767-4c3c-8cfa-3a3c06f76ed2>

Reviewer & Literatur & Review - Analyse - Improve - Repeat

- AskPDF
- Science
- Scholar Assist

Request for Research Paper.
ChatGPT August 3 Version.
<https://chat.openai.com/share/4f7271fd-28b2-4ed0-81f7-11e2131253a3>

You have a draft research proposal that needs to be improved. You will play the role of 3 reviewers and write critical and constructive reviews on the proposal.

Reviewer 1
Professor of Digital Humanities
Expert in data modelling and semantic web, CIDOC-CRM, Fair Data

Reviewer 2
Professor of Computer Science
Expert in Semantic Web, workflows, graph databases and knowledge graphs

Reviewer 3
Digital historian
Expert in imperial crowns, digitization

Each reviewer makes 5 arguments about the literature. Only use literature from the literature and base the arguments on the literature and base the arguments on the literature.

Each argument focuses on a different aspect of the research proposal, being creative and constructive.

Research Proposal:
Sharing the CROWN - Establishing a Semantic Web Integration for Cultural Objects

Reviewer 1: Professor of Digital Humanities
Expert in data modelling and semantic web, CIDOC-CRM, Fair Data

1. Data Modelling and Semantic Web Integration

- **Literature:** "Semantic Web Technologies for Intelligent Engineering Applications" (2022) by Stefan Biffl et al.
- **Argument:** The proposal does not adequately address the challenges of integrating heterogeneous data sources, especially when considering the vast complexities of museum collections. Biffl et al. (2022) emphasize the importance of semantic integration techniques in addressing these challenges. The proposed methodology should incorporate advanced semantic integration techniques to ensure quality.

2. CIDOC-CRM Implementation

- **Literature:** "The CIDOC Conceptual Reference Model for Cultural Objects Sharing" (2017) by Stephen Stead.
- **Argument:** While the proposal mentions collecting data from various sources, it lacks a clear strategy for implementing the CIDOC-CRM standard. The implementation plan for the proposal should detail a comprehensive plan to address the intricacies of CIDOC-CRM and its potential for alignment with global standards.

Analyse the reviews step by step and then improve the research proposal.

Check if the urls of the Sources are correct.

Certainly! Let's break down the reviews step by step, analyze the feedback, and then incorporate the suggestions to improve the research proposal. Additionally, I'll verify the URLs provided in the sources.

Step-by-Step Analysis of the Reviews:

1. Reviewer 1: Professor of Digital Humanities

- **Feedback:** Emphasized the importance of semantic integration techniques, CIDOC-CRM implementation, adherence to FAIR data principles, advanced data enrichment techniques, and structured stakeholder engagement.
- **Improvement Strategy:**
 - Incorporate advanced semantic integration techniques.
 - Detail a comprehensive plan for CIDOC-CRM implementation.
 - Explicitly address adherence to FAIR principles.
 - Delve deeper into data enrichment techniques.
 - Outline a structured plan for continuous stakeholder engagement.

Persona Modelling

- **Identität:** Name, Beruf, Hintergrund, Rolle im Gesprächskontext
- **Merkmale:** Persönlichkeitsmerkmale, Kommunikationsstil, Überzeugungen, Werte
- **Wissen und Erfahrung:** Spezifische Wissensgebiete, Fachkenntnisse
- **Erlebnisse:** Relevante Erfahrungen, Errungenschaften, Herausforderungen
- **Motivationen:** Ziele, Wünsche, Treibende Kräfte
- **Emotionale und beziehungsbezogene Aspekte:** Emotionaler Zustand, Grad der Empathie, Art und Weise des Beziehungsaufbaus und -pflege
- **Kontext:** Spezifische Umgebung oder Situation

Persona Modelling

As Christopher Pollin, a dedicated Ph.D. candidate in Digital Humanities, I have a deep-rooted passion that stems from my upbringing in Upper Styria. My journey has led me to fervently believe in and advocate for the synthesis of semantic web technologies and generative AI in the study and visualization of historical financial documents. Given my background and the academic audience present, I would like to provide a pragmatic and scholarly overview of the myriad challenges and opportunities in our field. I'll cover both broad topics and specific methodologies, drawing on my experience in building semantic web platforms, my knowledge of areas such as TEI, RDF, and ontologies, and reflecting the academic rigor and practical insights I've gained.

1. **Identification and Digitization of Historical Financial Documents:** [2 months]
 - Collaborate with international archives, private collections, and libraries to identify relevant financial documents related to the crown.
 - Use high-resolution scanners to digitize these documents ensuring their longevity and ease of access.
2. **Transcription and Annotation of Financial Documents:** [3 months]
 - Engage specialists to transcribe the ancient scripts, considering various historical financial terminologies.
 - Annotate the transcriptions with metadata such as date, place, involved parties, type of transaction, and amounts.
3. **Integration of Financial Data into TMS:** [2 months]
 - Develop a template in TMS tailored to capture the unique details present in historical financial documents.
 - Populate TMS with the transcribed and annotated data.
4. **Semantic Modelling of Financial Data:** [3 months]
 - Extend the domain-specific ontology to accommodate financial terminologies and concepts.
 - Define relationships between financial data and existing object data within the ontology, e.g., correlating a specific gem's acquisition cost with its physical and historical attributes.
5. **Mapping and Transformation of Financial Data to RDF:** [2 months]
 - Use the established workflow to convert the financial data from CSV to RDF.
 - Ensure that relationships between the crown's components and their associated financial data are preserved in the RDF structures.
6. **Semantic Enrichment of Financial Data:** [2 months]
 - Employ the reconciliation processes to semantically enrich the financial data. This might involve linking historical currencies to their modern equivalents or identifying historical merchants and linking them to their respective places of operation via Wikidata.
7. **Visualization & Analysis of Financial Data:** [3 months]
 - Extend the web-based prototype to incorporate visualizations specific to financial data such as time-series graphs showcasing the fluctuation of the crown's value, heatmaps indicating regions of intense economic activity related to the crown, and pie-charts detailing expenditure breakdown.
 - Introduce analytical tools to deduce insights from the financial data, such as the economic significance of the crown during specific periods or regions.
8. **Documentation & Workflow Extension:** [1 month]
 - Update the project documentation to include methodologies specific to the handling of historical financial documents.
 - Ensure that the enriched workflow can be applicable to other similar projects aiming to integrate historical financial data into their systems.

Chain of Density Prompting

<https://arxiv.org/pdf/2309.04269.pdf>

Generate concise article summaries. ChatGPT September 25 Version
October 5, 2023

<https://chat.openai.com/share/5cea3f37-c8a4-4466-bacd-0819a37bda0b>

You will generate increasingly concise entity-dense summaries of the above article.

Repeat the following 2 steps 5 times:

Step 1: Identify 1-3 informative entities (delimited) from the article which are missing from the previously generated summary.

Step 2: Write a new denser summary of identical length which covers every entity and detail from the previous summary plus the missing entities.

A missing entity is:

Relevant: to the main stories.

Specific: descriptive yet concise (5 words or fewer).

Novel: not in the previous summary.

Faithful: present in the article.

Anywhere: located in the article.

Guidelines:

The first summary should be long (4-5 sentences, ~80 words), yet highly non-specific, containing little information beyond the entities marked as missing. Use overly verbose language and fillers (e.g., “this article discusses”) to reach ~80 words. Make every word count. Rewrite the previous summary to improve flow and make space for additional entities. Make space with fusion, compression, and removal of uninformative phrases like “the article discusses”. The summaries should become highly dense and concise, yet self-contained, e.g., easily understood without the article.

Missing entities can appear anywhere in the new summary. Never drop entities from the previous summary. If space cannot be made, add fewer new entities. Remember: Use the exact same number of words for each summary.”



Prompt Engineering Goddess

Midjourney: female programming shaman with imperial crown, in front of computer, altar, finding perfection, ascending to divinity, by Alejandro Jodorowsky, comic style --weird 10 --ar 16:9

Ressourcen

- AI Explained, <https://www.youtube.com/@ai-explained>
- Matthew Berman, https://www.youtube.com/@matthew_berman
- Yannic Kilcher, <https://www.youtube.com/@YannicKilcher>
- ChatGPT Prompt Engineering for Developers, <https://learn.deeplearning.ai/chatgpt-prompt-eng>
- Höller, Johannes and Warmuth, Christian. OpenHPI. ChatGPT: Was bedeutet generative KI für unsere Gesellschaft? <https://open.hpi.de/courses/kizukunft2023/items/3jMwelGPssDCNQw63XXsn>
- The AI Advantage. <https://www.youtube.com/@aiadvantage>
- Google. Maschinelles Lernen von Google. <https://developers.google.com/machine-learning?hl=de>
- freeCodeCamp.org. Development with Large Language Models Tutorial – OpenAI, Langchain, Agents, Chroma. <https://youtu.be/xZDB1naRUIk?si=tK24Bf7q73YHskyx>
- Ekin, Sabit. „Prompt Engineering For ChatGPT: A Quick Guide To Techniques, Tips, And Best Practices“. TechRxiv, 29. April 2023. <https://doi.org/10.36227/techrxiv.22683919.v1>.
- Møller, Anders Giovanni, Jacob Aarup Dalsgaard, Arianna Pera, und Luca Maria Aiello. „Is a prompt and a few samples all you need? Using GPT-4 for data augmentation in low-resource classification tasks“. arXiv, 26. April 2023. <https://doi.org/10.48550/arXiv.2304.13861>.
- Yang, Chengrun, Xuezhi Wang, Yifeng Lu, Hanxiao Liu, Quoc V. Le, Denny Zhou, und Xinyun Chen. „Large Language Models as Optimizers“. arXiv, 6. September 2023. <https://doi.org/10.48550/arXiv.2309.03409>.
- Yao, Shunyu, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, und Karthik Narasimhan. „Tree of Thoughts: Deliberate Problem Solving with Large Language Models“. arXiv, 17. Mai 2023. <https://doi.org/10.48550/arXiv.2305.10601>.
- Zhang, Muru, Ofir Press, William Merrill, Alisa Liu, und Noah A. Smith. „How Language Model Hallucinations Can Snowball“. arXiv, 22. Mai 2023. <https://doi.org/10.48550/arXiv.2305.13534>.
- <https://discord.com/invite/openai>
- Prompt Engineering Guide. <https://www.promptingguide.ai/>

Examples: “TEI Modeler” by Christopher Pollin [GPT-4]

Transaction (bit tricky)

<https://chat.openai.com/share/888cbf01-32b8-4160-8785-f83584f9562b>

Critical Apparatus

<https://chat.openai.com/share/1fe978ba-c27d-4693-8419-541fca8356f3>

Page break, connecting digital facsimiles and columns

<https://chat.openai.com/share/2c218a9b-9b6d-4e84-9e7a-91d69d8f5573>