

# Angewandte Generative KI in den (digitalen) Geisteswissenschaften

Digitale Edition und TEI-Erzeugung am Beispiel von  
Briefen des Hugo Schuchardt Archivs

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

Christopher Pollin

<https://chpollin.github.io>

Zentrum für Informationsmodellierung, Graz

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

Digital Humanities Craft OG

[www.dhcraft.org](http://www.dhcraft.org)

# News: Open Source Models Are Coming



## OpenLLaMA: An Open Reproduction of LLaMA

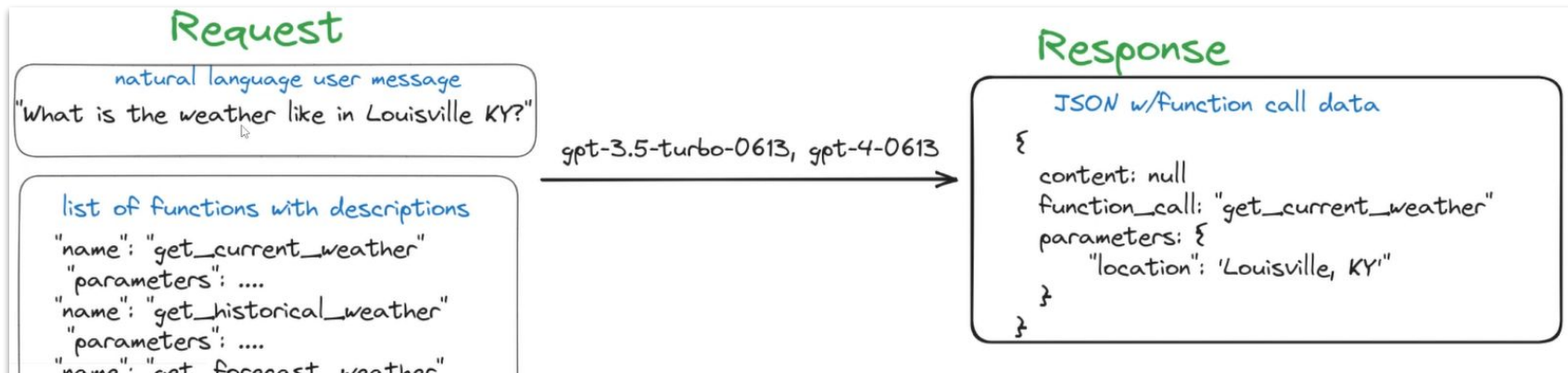
**TL;DR:** we are releasing our public preview of OpenLLaMA, a permissively licensed open source reproduction of Meta AI's LLaMA 7B and 13B trained on the RedPajama dataset. Our model weights can serve as the drop in replacement of LLaMA in existing implementations. We also provide a smaller 3B variant of LLaMA model.

In this repo, we present a permissively licensed open source reproduction of Meta AI's [LLaMA](#) large language model. We are releasing 3B, 7B and 17B models trained on 1T tokens. We provide PyTorch and JAX weights of pre-trained OpenLLaMA models, as well as evaluation results and comparison against the original LLaMA models. Stay tuned for our updates.

- **Insane New AI Model - ORCA - That Finally Beats GPT-4.** AI Revolution, <https://www.youtube.com/watch?v=gyeiL1JZnXc>
- **Orca: The Model Few Saw Coming.** AI Explained. [https://www.youtube.com/watch?v=Dt\\_UNg7Mchg](https://www.youtube.com/watch?v=Dt_UNg7Mchg)
- **EW "Orca" 🐋 Open-Source Model Surprised Everyone.** Matthew Berman. <https://www.youtube.com/watch?v=Kol6G7oWYvM>
- Mukherjee, Subhabrata, Arindam Mitra, Ganesh Jawahar, Sahaj Agarwal, Hamid Palangi, and Ahmed Awadallah. 'Orca: Progressive Learning from Complex Explanation Traces of GPT-4'. arXiv, 5 June 2023. <https://doi.org/10.48550/arXiv.2306.02707>.
- **OpenLLaMA: An Open Reproduction of LLaMA.** GitHub. [https://github.com/openlm-research/open\\_llama](https://github.com/openlm-research/open_llama)
- <https://huggingface.co/decapoda-research/llama-13b-hf>
- Und viele andere! Z. B. <https://open-assistant.io/de>

Achtung bei den Lizenzen!

# News: GPT-4 Function Calling

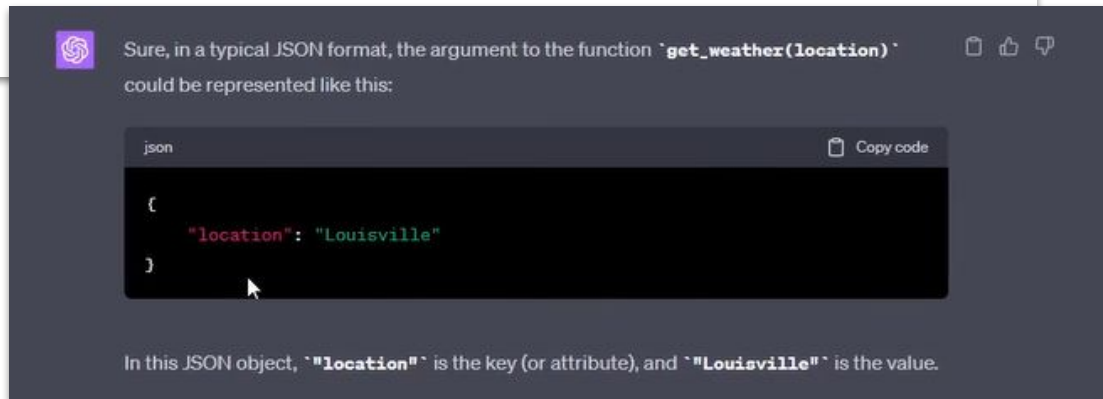


GPT-4 Function Calling Complete Guide | NEW

ChaptGPT Models | Openai. Nick Franck.

[https://www.youtube.com/watch?v=N0pWog\\_6dlg](https://www.youtube.com/watch?v=N0pWog_6dlg)

<https://openai.com/blog/function-calling-and-other-api-updates>



# Ressourcen

## Conversations:

- **Digitale Edition und TEI Erzeugung am Beispiel von Briefen des Hugo Schuchardt Archivs 1. GPT-4, ChatGPT May 24 Version. 08.06.2023.** <https://chat.openai.com/share/561def82-00bf-44da-93c6-cea141a4ce5b>.
- **Digitale Edition und TEI Erzeugung am Beispiel von Briefen des Hugo Schuchardt Archivs 2. GPT-4, ChatGPT May 24 Version. 08.06.2023.** <https://chat.openai.com/share/5931cbc2-5561-4099-a8fc-0f86fe6a40a9>
- Gou, Zhibin, Zhihong Shao, Yeyun Gong, Yelong Shen, Yujiu Yang, Nan Duan, and Weizhu Chen. **‘CRITIC: Large Language Models Can Self-Correct with Tool-Interactive Critiquing’**. arXiv, 19 May 2023. <https://doi.org/10.48550/arXiv.2305.11738>.
- Zhang, Muru, Ofir Press, William Merrill, Alisa Liu, and Noah A. Smith. **‘How Language Model Hallucinations Can Snowball’**. arXiv, 22 May 2023. <https://doi.org/10.48550/arXiv.2305.13534>.
- **Why AI Is Incredibly Smart and Shockingly Stupid** | Yejin Choi | TED. <https://www.youtube.com/watch?v=SvBR0OGT5VI>.
- **Orca: The Model Few Saw Coming.** AI Explained. [https://www.youtube.com/watch?v=Dt\\_UNg7Mchg](https://www.youtube.com/watch?v=Dt_UNg7Mchg)
- Mukherjee, Subhabrata, Arindam Mitra, Ganesh Jawahar, Sahaj Agarwal, Hamid Palangi, and Ahmed Awadallah. **‘Orca: Progressive Learning from Complex Explanation Traces of GPT-4’**. arXiv, 5 June 2023. <https://doi.org/10.48550/arXiv.2306.02707>.

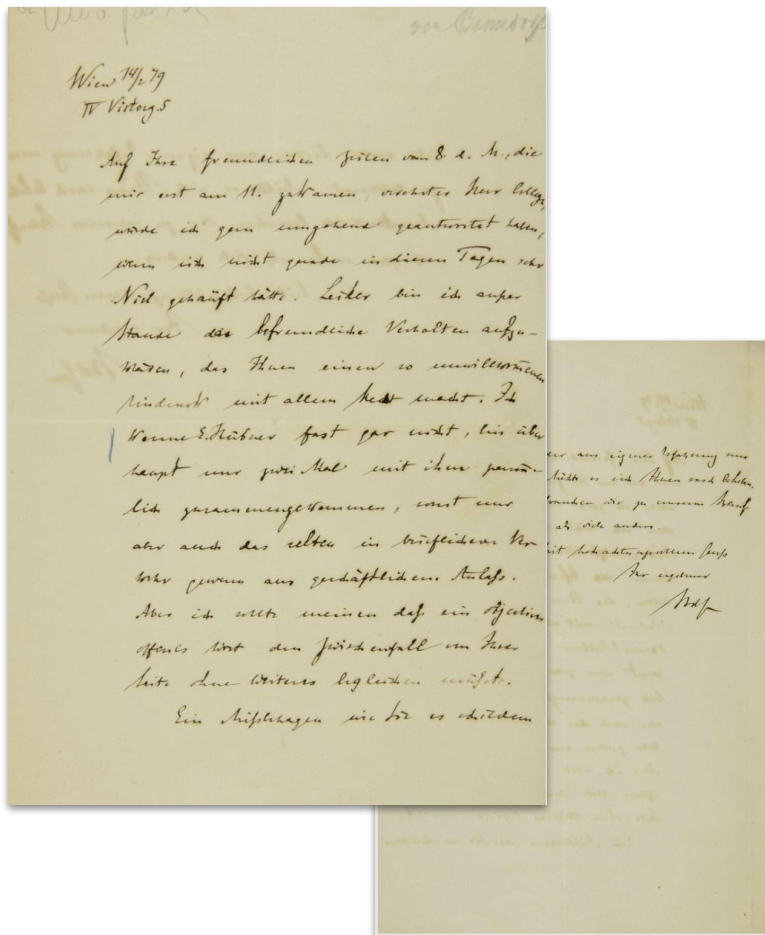
# Ressourcen

## YouTube und weitere Tutorials

- AI Explained, <https://www.youtube.com/@ai-explained->
- Matthew Berman, [https://www.youtube.com/@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>
- Hötter, Johannes and Warmuth, Christian. OpenHPI. ChatGPT: Was bedeutet generative KI für unsere Gesellschaft? <https://open.hpi.de/courses/kizukunft2023/items/3jMwelGWPssDCNQw63XXsn>

## Paper

- Baktash, Jawid Ahmad, and Mursal Dawodi. 2023. '**Gpt-4: A Review on Advancements and Opportunities in Natural Language Processing**'. arXiv. <https://doi.org/10.48550/arXiv.2305.03195>.
- Møller, Anders Giovanni, Jacob Aarup Dalsgaard, Arianna Pera, and Luca Maria Aiello. 2023. '**Is a Prompt and a Few Samples All You Need? Using GPT-4 for Data Augmentation in Low-Resource Classification Tasks**'. arXiv. <https://doi.org/10.48550/arXiv.2304.13861>.
- Mukherjee, Subhabrata, Arindam Mitra, Ganesh Jawahar, Sahaj Agarwal, Hamid Palangi, and Ahmed Awadallah. 2023. '**Orca: Progressive Learning from Complex Explanation Traces of GPT-4**'. arXiv. <https://doi.org/10.48550/arXiv.2306.02707>.
- Roberts, Jonathan, Timo Lüddecke, Sowmen Das, Kai Han, and Samuel Albanie. 2023. '**GPT4GEO: How a Language Model Sees the World's Geography**'. arXiv. <http://arxiv.org/abs/2306.00020>.
- Watkins, Ryan. '**Guidance for Researchers and Peer-Reviewers on the Ethical Use of Large Language Models (LLM) in Scientific Research Workflows**'. OSF Preprints, 21 April 2023. <https://doi.org/10.31219/osf.io/6uh8p>.
- 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://arxiv.org/abs/2305.10601>.



<sup>1</sup>Wien 14/2 79

IV Victorg. 5<sup>2</sup>

Auf Ihre freundlichen Zeilen vom 8. d. M., die mir erst am 11. zukamen, verehrter Herr College, würde ich gern umgehend geantwortet haben, wenn sich nicht gerade in diesen Tagen sehr Viel gehäuft hätte. Leider bin ich ausser Stande das befremdliche Verhalten aufzuklären, das Ihnen einen so unwillkommenen Eindruck mit allem Recht macht. Ich kenne [E. Hübner](#)<sup>3</sup> fast gar nicht, bin überhaupt nur zwei Mal mit ihm persönlich zusammengekommen, sonst nur aber auch das selten in brieflichem Verkehr gewesen aus geschäftlichem Anlass.<sup>4</sup> Aber ich sollte meinen dass ein objectives offenes Wort den Zwischenfall von Ihrer Seite ohne Weiteres begleichen müsste.

Ein Missbehagen wie Sie es schildern <sup>[2]</sup> ist mir leider aus eigener Erfahrung nur zu bekannt. Möchte es sich Ihnen rasch beheben. Munterkeit brauchen wir zu unserm Beruf so gar mehr als viele andere.

Mit hochachtungsvollem Gruss

Ihr ergebener

[Bdf](#)



1Wien 14/2 79  
IV Victorg. 5<sup>2</sup>

Auf Ihre freundlichen Zeilen vom 8. d. M., die mir zukamen, verehrter Herr College, würde ich gern geantwortet haben, wenn sich nicht gerade in die gehäuft hätte. Leider bin ich ausser Stande das b Verhalten aufzuklären, das Ihnen einen so unwillk mit allem Recht macht. Ich ke überhaupt nur zwei Mal mit ihm sonst nur aber auch das selte geschäftlichem Anlass.<sup>4</sup> Aber offenes Wort den Zwischenfall begleichen müsste.

Ein Missbehagen wie Sie es s Erfahrung nur zu bekannt. Mö Munterkeit brauchen wir zu unserm Beruf so gar andere.

Mit hochachtungsvollem Gruss

Ihr ergebener

Bdf

GPT

```
<text>
<body>
<div subtype="original" type="letter" xml:id="L.7711" xml:lang="de">
  <pb facs="#IMG.1" n="1"/>
  <dateline> <note n="1" place="foot" type="editorial">Antwortbrief auf <ref
    target="https://gams.uni-graz.at/o:hsa.letter.7710">01-HSFB</ref>.</note>Wien
    14/2 79</lb/> IV Victorg. 5<note n="2" place="foot" type="editorial">Benndorf
    wohnte seit seinem Wechsel von Prag nach Wien in der Victorgasse 5 in
    Wien-Wieden, einem Stadtteil, die bei Universitätsprofessoren sehr beliebt war.
    In dieser Gegend hatte schon sein Vorgänger am Wiener Lehrstuhl für Klassische
    Archäologie, <persName ref="https://gams.uni-graz.at/o:hsa.persons#P.4100"
    >Alexander Conze</persName>, gewohnt, und auch sein Kollege am
    Archäologisch-epigraphischen Seminar, <persName
    ref="https://gams.uni-graz.at/o:hsa.persons#P.1771">Otto Hirschfeld</persName>,
    hatte in Wieden sein Zuhause gefunden.</note> </dateline>
    chen Zeilen vom 8. d. M., die mir erst am 11. zukamen, verehrter
    rde ich gern umgehend geantwortet haben, wenn sich nicht gerade
    sehr Viel gehäuft hätte. Leider bin ich ausser Stande das
    halten aufzuklären, das Ihnen einen so unwillkommenen Eindruck
    macht. Ich kenne <persName
    s.uni-graz.at/o:hsa.persons#P.1805">E. Hübner</persName> <note
    t" type="editorial"> <persName
    s.uni-graz.at/o:hsa.persons#P.1805">Emil Hübner</persName>
    ssischer Philologe und Epigraphiker, seit 1863 Professor für
    ps://gams.uni-graz.at/o:hsa.subjects#S.4435">klassische
    an der <ref
    gams.uni-graz.at/o:hsa.subjects#S.4296">
    ns-Universität Berlin</ref>. </note> fast gar nicht, bin
    ei Mal mit ihm persönlich zusammengekommen, sonst nur aber auch
    ieflichem Verkehr gewesen aus geschäftlichem Anlass.<note n="4"
    es="editorial">Mit <persName
    s.uni-graz.at/o:hsa.persons#P.1805">Emil Hübner</persName> stand
    Benndorf von 1868 bis 1895 in brieflicher Korrespondenz (16 Schreiben), vgl.
    <ref target="http://data.onb.ac.at/rec/AC14105379">Österreichische
    Nationalbibliothek, Sammlung von Handschriften und alten Drucken, Autogr.
    646/36</ref>.</note> Aber ich sollte meinen dass ein objectives offenes Wort den
    Zwischenfall von Ihrer Seite ohne Weiteres begleichen müsste.</p>
    <p>Ein Missbehagen wie Sie es schildern <pb facs="#IMG.2" n="2"/> ist mir leider aus
    eigener Erfahrung nur zu bekannt. Möchte es sich Ihnen rasch beheben. Munterkeit
    brauchen wir zu unserm Beruf so gar mehr als viele andere.</p>
    <closer> <salute>Mit hochachtungsvollem Gruss</salute> <lb/> <salute>Ihr
    ergebener</salute> <lb/> <signed> <persName
    ref="https://gams.uni-graz.at/o:hsa.persons#P.1127">Bdf</persName> </signed>
    </closer>
  </div>
</body>
```

[1]

Graz 6. 6. '20

Lieber Freund,

Verzeihe daß ich erst jetzt Deinen lieben Brief vom 2. April beantworte. Ich bin alt, krank, matt, und über die Augen habe auch ich zu klagen; und besonders leide ich alljährlich nach Beginn der Gewitterperiode an dauerndem Kopfdruk.

Die Angelegenheit, wegen deren ich Dir zuletzt schrieb, war schon damals eine vollendete Tatsache, und es bleibt mir nur übrig zu erklären daß sie sich wider mein Wissen und Wollen vollzogen hat<sup>1</sup>. Ich bin immer stolz gewesen auf die Ehrungen die mir von wissenschaftlichen Instituten Italiens erwiesen wurden. Das

```
<body>
  <div subtype="original" type="letter" xml:id="L.XXXX" xml:lang="de">
    <pb n="1"/>
    <dateline> <name type="place">Graz</name> <date when="1920-06-06">6. 6. '20</date> </dateline>
    <opener> <salute>Lieber Freund,</salute> </opener>
    <p> Verzeihe daß ich erst jetzt Deinen lieben Brief vom 2. April beantworte. Ich bin
      alt, krank, matt, und über die Augen habe auch ich zu klagen; und besonders
      leide ich alljährlich nach Beginn der Gewitterperiode an dauerndem Kopfdruk. </p>
    <p> Die Angelegenheit, wegen deren ich Dir zuletzt schrieb, war schon damals eine
      vollendete Tatsache, und es bleibt mir nur übrig zu erklären daß sie sich wider
      mein Wissen und Wollen vollzogen hat <note n="1" place="foot" type="editorial"
      >Schuchardt si riferiva alla cancellazione, nell'annuario dell'Università di
      Graz, dei suoi titoli di socio di accademie scientifiche appartenenti a Stati
      nemici dell'Austria-Ungheria durante la guerra da poco conclusa. Nella lettera
      precedente, D'Ovidio aveva nettamente smentito l'espulsione di Schuchardt o di
      altri soci stranieri dall'istituzione che presiedeva.</note>. </p>
    <pb n="2"/>
    <p> Mein Herz ist inbezug auf Italien stets treu geblieben. Aber Treue um Treue! Ich
      müßte die Verachtung der Italiener befürchten, wenn ich über die Wegnahme
      deutschen Gebietes nicht ebenso denken und fühlen würde wie sie selbst über dem
```

[2]

Mein Herz ist  
die Verachtung  
Gebietes nicht

Das TEI XML vom “nächste” Brief wurde von GPT4 alleine generiert.

GPT4 liefert mit dem richtigen Prompt ein solides, valides TEI XML und kann Textstrukturen (p, pb, opener, salute etc.), aber auch Named Entities annotieren.

- <date> wurde erkannt und korrekt normalisiert.
- Graz als Ort erkannt
- Fußnote erkannt.



# Naja! Ist cool, aber sooo berauschend ist das nicht...

- Das ist das Ergebnis von zwei Nachmittagen!
- Die Technologie steckt noch in den Kinderschuhen und wird ständig weiterentwickelt! (Über GPT-5 wird schon geredet, GPT-4.2 sicher schon in Arbeit)
- Was hindert uns daran, das fertige TEI-XML von GPT erneut verarbeiten zu lassen und dann die Ausgabe von GPT weiterverarbeiten und die Ergebnisse von GPT überprüfen zu lassen...?
- Wir können...
- ... Software als Tools in GPT laufen lassen!
- ... eigene Vektordatenbanken anlegen, um projektspezifisches oder genrespezifisches Kontextwissen in GPT zu nutzen!
- ... am Input-Text Dinge vorab algorithmisch bearbeiten und dann GPT arbeiten lassen!
- ... Arbeitsschritte automatisiert vorgenerieren, die dann von den Editor\*innen bearbeitet werden und uns viel mühsame Arbeitszeit ersparen!
- ... auch ein Tool entwickeln, das KI-gestützten Text in TEI transformiert und ständig Feedback von den Editor\*innen bekommt und seine TEI immer weiter anpasst!
- ... noch viel am Prompt Engineering drehen!
- ...

# Prompt Engineering

## 1.2 Importance of prompt engineering in maximizing the effectiveness of ChatGPT

Prompt engineering is the art of crafting effective prompts that guide ChatGPT to generate desired responses. It plays a crucial role in maximizing the effectiveness of the language model by bridging the gap between user intent and model understanding. A well-engineered prompt can significantly improve the quality and relevance of ChatGPT's output, while a poorly crafted one may lead to unsatisfactory or erroneous responses.

For example, consider the following prompts and their corresponding ChatGPT responses:

**Prompt 1:** "Tell me about AI."

**Response 1:** "Artificial intelligence (AI) is a branch of computer science that deals with the creation and development of machines that can perform tasks requiring human-like intelligence. This includes learning, reasoning, problem-solving, and understanding natural language."

**Prompt 2:** "Provide a concise summary of artificial intelligence, its subfields, and current applications."

**Response 2:** "Artificial intelligence (AI) is the study of creating intelligent agents capable of performing tasks autonomously. Key subfields include machine learning, natural language processing, and robotics. Current applications range from virtual assistants and recommendation systems to autonomous vehicles and medical diagnosis."

The second prompt yields a more informative and focused response due to its clarity and specificity.

Ekin, Sabit (2023): Prompt Engineering For ChatGPT: A Quick Guide To Techniques, Tips, And Best Practices. TechRxiv. Preprint.  
<https://doi.org/10.36227/techrxiv.22683919.v1>

ChatGPT Prompt Engineering for Developers, <https://learn.deeplearning.ai/chatgpt-prompt-eng>

# Prompt Engineering Principles

- Seien Sie spezifisch und klar in Prompts.
- Geben Sie dem Model Zeit zum “Nachdenken”.
- Geben Sie, wenn nötig, Kontext und Beispiele an. Verwenden Sie beispielbasiertes Lernen, indem Sie GPT einige Beispiele für Eingabe/Ausgabe-Paare zur Verfügung stellen.
- Erwägen Sie einen iterativen Ansatz zur Verfeinerung der Prompts - iterative Entwicklung von Prompts.
- Fähigkeiten von GPT: Zusammenfassen, ableiten, umwandeln, erweitern.
- Explizite Einschränkungen verwenden.
- Kontrolle des Detailgrads der Ausgabe.
- Das gewünschte Format und die Struktur angeben.
- Zu viele Aufgaben und zu komplexe Aufgaben auf einmal sind ein Problem.

Ekin, Sabit (2023): Prompt Engineering For ChatGPT: A Quick Guide To Techniques, Tips, And Best Practices. TechRxiv. Preprint.  
<https://doi.org/10.36227/techrxiv.22683919.v1>

ChatGPT Prompt Engineering for Developers, <https://learn.deeplearning.ai/chatgpt-prompt-eng>

# Vorgehensweise

- Analysieren der Struktur des Briefes
- Identifizieren von relevanten Elementen und Attributen
- Verstehen des Mappings zwischen unstrukturiertem Text und TEI-XML
- Beispiel von unstrukturierten Text und entsprechendem TEI-XML ("Ground Truth")
- Ausprobieren, scheitern, erneut ausprobieren, was lernen, Prompt anpassen, ...

## Analysieren der Struktur des Briefes und Identifizieren von relevanten Elementen und Attributen

I want to convert unstructured text into TEI XML, without programming language and only with GPT. It is a correspondence from Hugo Schuchardt. I give you the unstructured text (id text.1) between the delimiters ``` and the corresponding TEI XML with id (tei.1) into which other texts of any structure are to be converted. The TEI XML (tei.1) is also located in the delimiters ``` . First identify the structure of the text (text.1) and make a list of it. Then compare the list with the corresponding TEI XML elements from tei.1. We will skip the teiHeader and focus only on the <text> element. Ask me then as many questions as you need to know to understand the domain.

text.1: ``` {text.1} ```

tei.1: ``` {tei.1} ```



# Klare Anweisung der Aufgabe und klare Formulierung des Outputs

I want to convert unstructured text into TEI XML, without programming language and only with GPT. It is a correspondence from Hugo Schuchardt. I give you the unstructured text id (text.1) between the delimiters ``` and the corresponding TEI XML with id (tei.1) into which other texts of any structure are to be converted. The TEI XML (tei.1) is also located in the delimiters ``` . First identify the structure of the text (text.1) and make a list of it. Then compare the list with the corresponding TEI XML elements from tei.1. We will skip the teiHeader and focus only on the <text> element. Ask me then as many questions as you need to know to understand the domain.

```
text.1: ``` {text.1} ```  
tei.1: ``` {tei.1} ```
```

# Klare Beschreibung des Kontexts und Identifikation von Bestandteilen des Prompts

I want to convert unstructured text into TEI XML, without programming language and only with GPT. It is a correspondence from Hugo Schuchardt. I give you the unstructured text (text.1) between the delimiters ``` and the corresponding TEI XML with id (tei.1) into which other texts of any structure are to be converted. The TEI XML (tei.1) is also located in the delimiters ``` . First identify the structure of the text (text.1) and make a list of it. Then compare the list with the corresponding TEI XML elements from tei.1. We will skip the `teiHeader` and focus only on the `<text>` element. Ask me then as many questions as you need to know to understand the domain.

text.1: ``` {text.1} ```

tei.1: ``` {tei.1} ```

# Klare Anleitung der Arbeitsschritte und klären, was jetzt noch unklar ist

I want to convert unstructured text into TEI XML, without programming language and only with GPT. It is a correspondence from Hugo Schuchardt. I give you the unstructured text id (text.1) between the delimiters ``` and the corresponding TEI XML with id (tei.1) into which other texts of any structure are to be converted. The TEI XML (tei.1) is also located in the delimiters ```. **First identify the structure of the text (text.1) and make a list of it. Then compare the list with the corresponding TEI XML elements from tei.1.** We will skip the teiHeader and focus only on the <text> element. Ask me then as many questions as you need to know to understand the domain.

```
text.1: ``` {text.1} ```  
tei.1: ``` {tei.1} ```
```



From the unstructured text (text.1), I can identify the following structure:

1. Location and Date: "Wien 14/2 79"
2. Address: "IV Victorg. 52"
3. Body of the letter: This is the main content of the letter, which includes the sender's response to the recipient's previous letter and other discussions.
4. Closing Salutation: "Mit hochachtungsvollem Gruss"
5. Signature: "Ihr ergebener Bdf"

<sup>1</sup>Wien 14/2 79

IV Victorg. 5<sup>2</sup>

Auf Ihre freundlichen Zeilen vom 8. d. M., die mir erst am 11. zukamen, verehrter Herr College, würde ich gern umgehend geantwortet haben, wenn sich nicht gerade in diesen Tagen sehr Viel gehäuft hätte. Leider bin ich ausser Stande das befremdliche Verhalten aufzuklären, das Ihnen einen so unwillkommenen Eindruck mit allem Recht macht. Ich kenne [E. Hübner](#)<sup>3</sup> fast gar nicht, bin überhaupt nur zwei Mal mit ihm persönlich zusammengekommen, sonst nur aber auch das selten in brieflichem Verkehr gewesen aus geschäftlichem Anlass.<sup>4</sup> Aber ich sollte meinen dass ein objectives offenes Wort den Zwischenfall von Ihrer Seite ohne Weiteres begleichen müsste.

Ein Missbehagen wie Sie es schildern [\[2\]](#) ist mir leider aus eigener Erfahrung nur zu bekannt. Möchte es sich Ihnen rasch beheben. Munterkeit brauchen wir zu unserm Beruf so gar mehr als viele andere.

Mit hochachtungsvollem Gruss

Ihr ergebener

[Bdf](#)

Comparing this with the TEI XML (tei.1), the corresponding elements are:

1. Location and Date: This is represented in the `**<dateline>**` element.
2. Address: This is also part of the `**<dateline>**` element, with additional information provided in a `**<note>**` element.
3. Body of the letter: This is represented in the `**<p>**` elements. The body of the letter can be split into multiple `**<p>**` elements, each representing a paragraph. The TEI XML also includes `**<persName>**` elements to denote names of people, and `**<note>**` elements to provide additional information.
4. Closing Salutation: This is represented in the `**<closer>**` element, specifically in the `**<salute>**` sub-elements.
5. Signature: This is represented in the `**<signed>**` element within the `**<closer>**` element.



Now, I have a few questions to better understand the domain:

1. How should we handle names in the text? In the TEI XML, names are represented with `**<persName>**` tags and include a reference link. Should we generate these reference links, or are they provided?
2. The TEI XML includes `**<note>**` elements that provide additional information. How should we handle these? Are they manually added by an editor, or should they be generated from the text?
3. The TEI XML includes a `**<pb>**` element to denote a page break. How should we identify page breaks in the unstructured text?
4. The TEI XML includes a `**subtype**` and `**type**` attribute in the `**<div>**` element. How should we determine these values?
5. The TEI XML includes `**xml:id**` and `**xml:lang**` attributes in the `**<div>**` element. How should we determine these values?
6. The TEI XML includes a `**fac**` attribute in the `**<pb>**` element. How should we determine this value?



1. stick to `<persName>` but ignore the reference link.
2. the editors will add the notes manually. so create an empty note with a `@n` for numbering



3. Page breaks are indicated by a vertical bar followed by a number, like "[2]". This should be converted into a `<pb>` element with the ``n`` attribute set to the page number.
4. For the ``subtype`` and ``type`` attributes in the `<div>` element, you can use "original" and "letter" respectively, as these seem to be standard for this kind of document.
5. For the ``xml:id`` attribute, you can use a unique identifier for each letter. This could be based on the date and sender of the letter, or it could be a simple incrementing number. The ``xml:lang`` attribute should be set to the language of the letter, which in this case is "de" for German.
6. The ``facts`` attribute in the `<pb>` element seems to be a reference to an image of the page. Since we don't have these images, you can leave this attribute out.

Versehentlich habe ich bei der Beantwortung auf Enter geklickt. Aber GPT-4 hat alle Fragen selbstständig richtig beantwortet.

GPT-4 ist in der Lage, Schlussfolgerungen zu ziehen. Es kann sogar eine "*Chain of Thought zur Begründung*" einer Schlussfolgerung angeben. Wir wissen aber nie, ob die Line Of Thought wirklich die Begründung für ein Ergebnis ist.

In diesem Fall fragt GPT-4 uns etwas, was es eigentlich eh schon weiß.

Aber: viel wichtiger ist, dass wir eigentlich all diese Informationen nun explizit in unserem Prompt einbeziehen können.

# Iteratives Prompt Engineering



Use the following text to create a prompt for GPT4 to transform unstructured text into TEI XML, including all of the following information:

I want to convert unstructured text into TEI XML, without programming language and only with GPT. It is a correspondence from Hugo Schuchardt. I give you the unstructured text (id

Den ganzen bis jetzt erzeugten Text erneut als Prompt übergeben, damit auch sicher nichts vergessen wird. GPT-4 verliert innerhalb einer Konversation die "Aufmerksamkeit".

# Wir generieren einen “perfekten” Prompt, der alle Mapping-Regeln beinhaltet.

Given the following unstructured text of a letter, convert it into TEI XML format. The text includes a location and date, an address, the body of the letter, a closing salutation, and a signature. In the TEI XML, these should be represented by the <dateline>, <p>, <closer>, and <signed> elements respectively. Names should be enclosed in <persName> tags without reference links. Page breaks are indicated by a vertical bar followed by a number, like "|2|", and should be converted into a <pb> element with the n attribute set to the page number. The subtype and type attributes in the <div> element should be set to "original" and "letter" respectively. The xml:id attribute should be a unique identifier for each letter, and the xml:lang attribute should be set to "de" for German. Ignore the facs attribute in the <pb> element. Also, include an empty <note> element with a n attribute for numbering, which will be filled in manually by editors later.

Unstructured text:

...

...

Please convert this into the corresponding TEI XML format.

# Der nächste Brief

<sup>1</sup>G. 16. Febr. 1879.

Verehrter Herr Kollege!

Im letzten Augenblick vor meiner Abreise<sup>2</sup> erhalte ich Ihre liebenswürdigen Zeilen und danke Ihnen bestens. Ich hatte mich getäuscht, als ich glaubte Sie seien ein spezieller Bekannter [H.](#)'s.<sup>3</sup> Unmittelbar nach Empfang des erwähnten Briefes schrieb ich einen sehr höflichen und ruhigen an H., indem ich nur bedauerte, dass er eine Missstimmung gegen mich hege und erklärte, dass ich mir durchaus nicht bewusst wäre, ihm irgend welchen Anlass dazu gegeben [\[2\]](#) zu haben. Ich erwartete darauf eine Antwort; und als ich dieselbe nach Verlauf von 8 Tagen nicht erhielt, schrieb ich einen zweiten – allerdings sehr maliciösen – Brief. [G. Meyer](#),<sup>4</sup> dem ich denselben vorlas, rieht mir ab, ihn fortzuschicken. Allein ich kann für meine Natur nicht; ich ertrüge tausend Mal lieber offene Feindschaft, als verhüllte Antipathieen. Und dieser Fall war mir – nachdem Ähnliches in diesem Winter mir mehrfach passirt war – doch zu stark. Hätte *ich* auch alle denkbare Ursache, Jemandem zu grollen, dem freundlich und [\[3\]](#) vertrauensvoll Bittenden würde *ich* nie einen solchen Ausdruck an den Kopf werfen, wie mir das Hübner gethan hat. Nochmals bitte ich Sie sehr um Entschuldigung, dass ich Sie mit dieser Sache behelligt habe. Habe ich irgend einen Strauss auszufechten, so rufe ich immer nach Zeugen; ich habe einen allzustarken Drang nach Öffentlichkeit, der mir in der That schon mehrfach verargt worden ist. Wenn Sie in oder aus Spanien irgend Etwas wünschen, ich stehe vollkommen zu Ihrer Verfügung.

In grösster Eile -

mit herzlichsten Grüssen

ihr ergebener

[H. Schuchardt](#)





Given the following unstructured text of a letter, convert it into TEI XML format. The text includes a location and date, an address, the body of the letter, a closing salutation, and a signature. In the TEI XML, these should be represented by the <dateline>, <p>, <closer>, and <signed> elements respectively. Names should be enclosed in <persName> tags without reference links. Page breaks are indicated by a vertical bar followed by a number, like "[2]", and should be converted into a <pb> element with the n attribute set to the page number. The subtype and type attributes in the <div> element should be set to "original" and "letter" respectively. The xmlid attribute should be a unique identifier for each letter, and the xml:lang attribute should be set to "de" for German. Ignore the facs attribute in the <pb> element. Also, include an empty <note> element with a n attribute for numbering, which will be filled in manually by editors later.

Unstructured text:

...

[1]

1G. 16. Febr. 1879.

Verehrter Herr Kollege!

Im letzten Augenblick vor meiner Abreise<sup>2</sup> erhalte ich Ihre liebenswürdigen Zeilen und danke Ihnen bestens. Ich hatte mich getäuscht, als ich glaubte Sie seien ein spezieller Bekannter H.'s.<sup>3</sup> Unmittelbar nach Empfang des erwähnten Briefes schrieb ich einen sehr höflichen und ruhigen an H., indem ich nur bedauerte, dass er eine Missstimmung gegen mich hege und erklärte, dass ich mir durchaus nicht bewusst wäre, ihm irgend welchen Anlass dazu gegeben [2] zu haben. Ich erwartete darauf eine Antwort; und als ich dieselbe nach Verlauf von 8 Tagen nicht erhielt, schrieb ich einen zweiten – allerdings sehr maliciösen – Brief. G. Meyer,<sup>4</sup> dem ich denselben vorlas, rieht mir ab, ihn fortzuschicken. Allein ich kann für meine Natur nicht; ich ertrüge tausend Mal lieber offene Feindschaft, als verhüllte Antipathien. Und dieser Fall war mir – nachdem Ähnliches in diesem Winter mir mehrfach passiert war – doch zu stark. Hätte ich auch alle denkbare Ursache, Jemandem zu grollen, dem freundlich und [3] vertrauensvoll Bittenden würde ich nie einen solchen Ausdruck an den Kopf werfen, wie mir das Hübner gethan hat. Nochmals bitte ich Sie sehr um Entschuldigung, dass ich Sie mit dieser Sache behelligt habe. Habe ich irgend einen Strauss auszufechten, so rufe ich immer nach Zeugen; ich habe einen allzustarken Drang nach Öffentlichkeit, der mir in der That schon mehrfach verargt worden ist. Wenn Sie in oder aus Spanien irgend Etwas wünschen, ich stehe vollkommen zu Ihrer Verfügung.

In grösster Eile -

mit herzlichsten Grüssen

ihr ergebener

H. Schuchardt

...

Please convert this into the corresponding TEI XML format.



```
<text>
  <body>
    <div subtype="original" type="letter" xml:id="L.7712" xml:lang="de">
      <pb n="1"/>
      <dateline> <note n="1" place="foot" type="editorial"/>G. 16. Febr. 1879. </dateline>
      <opener> <salute>Verehrter Herr Kollege!</salute> </opener>
      <p>Im letzten Augenblick vor meiner Abreise<note n="2" place="foot" type="editorial"
        /> erhalte ich Ihre liebenswürdigen Zeilen und danke Ihnen bestens. Ich hatte
        mich getäuscht, als ich glaubte Sie seien ein spezieller Bekannter
        <persName>H.</persName><note n="3" place="foot" type="editorial"/> Unmittelbar
        nach Empfang des erwähnten Briefes schrieb ich einen sehr höflichen und ruhigen
        an H., indem ich nur bedauerte, dass er eine Missstimmung gegen mich hege und
        erklärte, dass ich mir durchaus nicht bewusst wäre, ihm irgend welchen Anlass
        dazu gegeben</p>
      <pb n="2"/>
      <p>zu haben. Ich erwartete darauf eine Antwort; und als ich dieselbe nach Verlauf
        von 8 Tagen nicht erhielt, schrieb ich einen zweiten – allerdings sehr
        maliciösen – Brief. <persName>G. Meyer</persName><note n="4" place="foot"
        type="editorial"/>, dem ich denselben vorlas, rieht mir ab, ihn fortzuschicken.
        Allein ich kann für meine Natur nicht; ich ertrüge tausend Mal lieber offene
        Feindschaft, als verhüllte Antipathien. Und dieser Fall war mir – nachdem
        Ähnliches in diesem Winter mir mehrfach passiert war – doch zu stark. Hätte ich
        auch alle denkbare Ursache, Jemandem zu grollen, dem freundlich und</p>
      <pb n="3"/>
      <p>vertrauensvoll Bittenden würde ich nie einen solchen Ausdruck an den Kopf werfen,
        wie mir das Hübner gethan hat. Nochmals bitte ich Sie sehr um Entschuldigung,
        dass ich Sie mit dieser Sache behelligt habe. Habe ich irgend einen Strauss
        auszufechten, so rufe ich immer nach Zeugen; ich habe einen allzustarken Drang
        nach Öffentlichkeit, der mir in der That schon mehrfach verargt worden ist. Wenn
        Sie in oder aus Spanien irgend Etwas wünschen, ich stehe vollkommen zu Ihrer
        Verfügung.</p>
      <closer> <salute>In grösster Eile -</salute> <lb/> <salute>mit herzlichsten
        Grüssen</salute> <lb/> <signed> <persName>ihr ergebener</persName> <lb/>
        <persName>H. Schuchardt</persName> </signed> </closer>
    </div>
  </body>
</text>
```

```

<div subtype="original" type="letter" xml:id="L.7712" xml:lang="de">
  <pb n="1"/>
  <dateline> <note n="1" place="foot" type="editorial"> </note>G. 16. Febr.
    1879.</dateline>
  <opener>Verehrter Herr Kollege!</opener>
  <p>Im letzten Augenblick vor meiner Abreise<note n="2" place="foot" type="editorial"
    > </note> erhalte ich Ihre liebenswürdigen Zeilen und danke Ihnen bestens. Ich
    hatte mich getäuscht, als ich glaubte Sie seien ein spezieller Bekannter
    <persName ref="https://gams.uni-graz.at/o:hsa.persons#P.1805"
    >H.</persName>'s.<note n="3" place="foot" type="editorial"> </note> Unmittelbar
    nach Empfang des erwähnten Briefes schrieb ich eine
    an H., indem ich nur bedauerte, dass er eine Misss
    erklärte, dass ich mir durchaus nicht bewusst wäre,
    dazu gegeben <pb n="2"/> zu haben. Ich erwartete d
    ich dieselbe nach Verlauf von 8 Tagen nicht erhielt
    allerdings sehr maliciösen - Brief. <persName
    ref="https://gams.uni-graz.at/o:hsa.persons#P.2185'
    n="4" place="foot" type="editorial"> </note> dem ic
    ab, ihn fortzuschicken. Allein ich kann für meine l
    tausend Mal lieber offene Feindschaft, als verhüllt
    Fall war mir - nachdem Ähnliches in diesem Winter n
    doch zu stark. Hätte <hi rend="unknown" rendition='
    denkbare Ursache, Jemandem zu grollen, dem freundl:
    vertrauensvoll Bittenden würde <hi rend="unknown" >

```

## GPT TEI XML

```

<div subtype="original" type="letter" xml:id="L.7712" xml:lang="de">
  <pb n="1"/>
  <dateline> <note n="1" place="foot" type="editorial"/>G. 16. Febr. 1879. </dateline>
  <opener> <salute>Verehrter Herr Kollege!</salute> </opener>
  <p>Im letzten Augenblick vor meiner Abreise<note n="2" place="foot" type="editorial"
    /> erhalte ich Ihre liebenswürdigen Zeilen und danke Ihnen bestens. Ich hatte
    mich getäuscht, als ich glaubte Sie seien ein spezieller Bekannter
    <persName>H.</persName><note n="3" place="foot" type="editorial"/> Unmittelbar
    nach Empfang des erwähnten Briefes schrieb ich einen sehr höflichen und ruhigen
    an H., indem ich nur bedauerte, dass er eine Missstimmung gegen mich hege und
    erklärte, dass ich mir durchaus nicht bewusst wäre, ihm irgend welchen Anlass
    dazu gegeben</p>
  <pb n="2"/>
  <p>zu haben. Ich erwartete darauf eine Antwort; und als ich dieselbe nach Verlauf
    von 8 Tagen nicht erhielt, schrieb ich einen zweiten - allerdings sehr
    maliciösen - Brief. <persName>G. Meyer</persName><note n="4" place="foot"
    type="editorial"/>, dem ich denselben vorlas, rieth mir ab, ihn fortzuschicken.
    Allein ich kann für meine Natur nicht; ich ertrüge tausend Mal lieber offene
    Feindschaft, als verhüllte Antipathieen. Und dieser Fall war mir - nachdem
    Ähnliches in diesem Winter mir mehrfach passirt war - doch zu stark. Hätte ich
    auch alle denkbare Ursache, Jemandem zu grollen, dem freundlich und</p>
  <pb n="3"/>
  <p>vertrauensvoll Bittenden würde ich nie einen solchen Ausdruck an den Kopf werfen,
    wie mir das Hübner gethan hat. Nochmals bitte ich Sie sehr um Entschuldigung,
    dass ich Sie mit dieser Sache behelligt habe. Habe ich irgend einen Strauss

```

Original TEI XML  
(<note> habe ich für  
Übersichtlichkeit leer  
gemacht.)

# Analyse von Original und GPT-Erzeugnis

- Alle `<note>` richtig gesetzt.
- Das originale TEI-XML besteht aus einem `<p>`, in dem sich die 3 `<pb>` befinden. GPT-4 hat 3 `<p>` für jeden `<pb>` gemacht.
- `<pb>` sind richtig gesetzt.
- Zu `<hi rend="unknown" rendition="#none">` gibt es keine Info im Ausgangstext.
- `<closer>` ist genau gleich.
- GPT hat ein zusätzliches `<salute>` in den `<opener>` gegeben.
- “H.”, “G. Meyer” und “H. Schuchardt” richtig als Person identifiziert.  
Aber auch `<persName>ihr ergebener</persName> <lb/> <persName>H. Schuchardt</persName>`



# Feedback einarbeiten



Why did you put the whole text "Verehrter Herr Kollege!" in a `<salute>` in the `<opener>`?

Why did you annotate `<persName>her devoted</persName>` as `<persName>`?

Why did you create 3 `<p>`?



I apologize for the confusion. You're correct in pointing out these issues. Let's correct them:

1. The `<salute>` tag in the `<opener>` should only contain the salutation, not the entire opening line. In this case, "Verehrter Herr Kollege!" is indeed the salutation, so it was correct to include it within the `<salute>` tag.
2. The phrase "ihr ergebener" translates to "your devoted" and is part of the closing of the letter, not a person's name. It should be included within a `<salute>` tag in the `<closer>` section, not a `<persName>` tag.
3. The `<p>` tags are used to separate paragraphs in the text. In the original unstructured text, new paragraphs are indicated by a line break. However, in this case, it seems I may have misinterpreted some of the line breaks as new paragraphs. The text should be divided into paragraphs based on its content and context.

Here's the corrected TEI XML:

# Named Entities



Please extend the following TEI XML excerpt by performing the following tasks:

Identify all named entities within the text.

Conduct online research on these entities using resources like Wikipedia, GND or Wikidata.

For each identified entity, add the correct TEI XML element and include the URI (Uniform Resource Identifier) to identify the resource in an attribute ref.

Link any relevant resources you find (for instance, a Wikipedia article about the named entity) using the corresp attribute.

Identify all dates within the text and normalize them according to a consistent format.

The XML snippet you'll be working with is:

Prompts lässt man  
von GPT-4  
verbessern.

# Named Entities - GPT4 + Live Web via Bing



Finished browsing



The named entities in the original text are "H.", "G. Meyer", and "H. Schuchardt". Upon conducting research, I found the following:

- "H." could not be identified due to its generic nature.
- "G. Meyer" was associated with Günter Meyer, a German Geographer and Orientalist, known for his work on the Political Geography of the Arab World [\[19+source\]](#) . However, considering the time frame of the letter, this may not be the correct association.
- "H. Schuchardt" was associated with Hugo Ernst Mario Schuchardt, an eminent German linguist known for his work in the Romance languages, the Basque language, and in mixed languages, including pidgins, creoles, and the Lingua franca of the Mediterranean [\[13+source\]](#) .

Additional entity "Hübner" was identified and found to be a Germanic surname meaning an agricultural worker, a farmer, possibly and specifically one who worked a "hube", which was a piece of land roughly equivalent to the English measurement of a "hide", about 120 acres [\[25+source\]](#) .

The date within the text "16. Febr. 1879" was normalized to the ISO 8601 standard, resulting in "1879-02-16".

Here is the updated XML excerpt:

```

<text>
  <body>
    <div subtype="original" type="letter" xml:id="L.7712" xml:lang="de">
      <pb n="1"/>
      <dateline> <note n="1" place="foot" type="editorial"/><date when="1879-02-16">G. 16.
        Febr. 1879. </date></dateline>
      <opener> <salute>Verehrter Herr Kollege!</salute> </opener>
      <p>Im letzten Augenblick vor meiner Abreise<note n="2" place="foot" type="editorial"
        /> erhalte ich Ihre liebenswürdigen Zeilen und danke Ihnen bestens. Ich hatte
        mich getäuscht, als ich glaubte Sie seien ein spezieller Bekannter <persName
        ref="https://en.wikipedia.org/wiki/H%C3%BCbner"
        corresp="https://en.wikipedia.org/wiki/H%C3%BCbner">H.</persName><note n="3"
        place="foot" type="editorial"/> Unmittelbar nach Empfang des erwähnten Briefes
        irgend etwas wünschen, ich stehe vollkommen zu Ihrer Verfügung.</p>
      <closer> <salute>In grösster Eile -</salute> <lb/> <salute>mit herzlichsten
        Grüssen</salute> <lb/> <signed> <persName>ihr ergebener</persName> <lb/>
        <persName ref="https://en.wikipedia.org/wiki/Hugo_Schuchardt"
        corresp="https://en.wikipedia.org/wiki/Hugo_Schuchardt">H. Schuchardt</persName>
        </signed> </closer>
    </div>

```

- Datum richtig
- "H." falsch. → es war mein Fehler! Im Prompt muss ich sagen "Nur wenn du dir sicher bist, dann..."
- Schuchard mit dem richtigen Wikipedia-Artikel

# Die Ergebnisse werden instabiler, wenn man “zu viel” will oder es komplexer wird.

Given the following text includes a letter salutation, and a <dateline>, <p>, <p> <persName> tags with

Check the text for all named entities and annotate them. Search the web for URLs in Wikidata to normalise them. Normalise any data you find. Finally, analyse the TEI XML you have created and create a `teiHeader` using all the information you have.

followed by a number, like "|2|", and should be converted into a <pb> element with the `n` attribute set to the page number. The `subtype` and `type` attributes in the <div> element should be set to "original" and "letter" respectively. The `xml:id` attribute should be a unique identifier for each letter, and the `xml:lang` attribute should be set to "de" for German. Ignore the `fac` attribute in the <pb> element. Also, include an empty <note> element with a `n` attribute for numbering, which will be filled in manually by editors later. Check the text for any named entities and annotate them. Normalise any date you find. Finally, analyse the TEI XML you have created and create a `teiHeader` using all the information you have. Add the appropriate `correspDesc` to the `teiHeader`.

Unstructured text:

'''  
'''

Please convert this into the corresponding TEI XML format.



```

<text>
  <body>
    <div type="letter" subtype="original"> <opener> <dateline><placeName
      ref="https://www.wikidata.org/wiki/Q1741">Graz</placeName> 8. Febr. 79.</dateline>
    <salute>Verehrter Herr Kollege!</salute> </opener> <p>Verzeihen Sie, wenn ich mit
      einer Mittheilung beginne, von der Sie sagen werden, „das geht mich gar Nichts
      an.“</p> <p>Vor längerer Zeit schrieb ich an <persName
      ref="https://www.wikidata.org/wiki/Q68483">E. Hübner</persName> in <placeName
      ref="https://www.wikidata.org/wiki/Q64">Berlin</placeName> und bat ihn um
      Empfehlungen nach Spanien. Da ich keine Antwort erhielt, glaubte ich den Brief
      verloren und wiederholte die Bitte. Gestern beantwortete mir Hübner in möglichst
      knapper Weise die an ihn gerichteten Fragen, erklärte nicht in der Lage zu sein, mir
      Empfehlungen zu geben, da <pb n="2"/> die im Süden vorübergehend angeknüpften
      Beziehungen nicht ausreichten und schloss mit den Worten: „Im Uebrigen glaube ich
      der Belehrung darüber was von den politischen Gesinnungen der Nation zu halten sei,
      nicht zu bedürfen.“</p>.....<closer> <salute>Mit vorzüglicher Hochachtung Ihr
      ergebenster</salute> <signed><persName ref="https://www.wikidata.org/wiki/Q70878">H.
      Schuchardt</persName>.</signed> </closer> </div>
  </body>
</text>

```

- Q1741 ist nicht Graz, sondern Wien: <https://www.wikidata.org/wiki/Q1741>
- ... GPT beginnt, ab zu großen Prompts + Aufmerksamkeitsverlust + Komplexität der Aufgaben abzukürzen.

```
<body>
  <div type="letter" subtype="original" xml:id="schuchardt_1879">
    <opener> <dateline><placeName>Graz</placeName>, <date when="1879-02-08">8. Febr.
      79</date></dateline> <salute>Verehrter <persName>Herr
      Kollege</persName>!</salute> </opener>
    <p>Verzeihen Sie, wenn ich mit einer Mittheilung beginne, von der Sie sagen werden,
      „das geht mich gar Nichts an.“</p>
    <p>Vor längerer Zeit schrieb ich an <persName>E. Hübner</persName> in Berlin und bat
      ihn um Empfehlungen nach Spanien. Da ich keine Antwort erhielt, glaubte ich den
      Brief verloren und wiederholte die Bitte. Gestern beantwortete mir Hübner in
      möglichst knapper Weise die an ihn gerichteten Fragen, erklärte nicht in der
      Lage zu sein, mir Empfehlungen zu geben, da <pb n="2"/> die im Süden
      vorübergehend angeknüpften Beziehungen nicht ausreichten und schloss mit den
      Worten: „Im Uebrigen glaube ich der Belehrung darüber was von den politischen
      Gesinnungen der Nation zu halten sei, nicht zu bedürfen.“</p>
    <p>Sie wissen, dass diese Ausdrucksweise innerhalb der guten Gesellschaft immer und
      zweifellos als eine beleidigende gilt und welche Folgen ihr Gebrauch in der
      Konversation nach sich zu ziehen pflegt...</p>
    <note n="1"/>
    <p>...</p>
    <pb n="3"/>
    <p>...</p>
    <p>...</p>
    <pb n="4"/>
    <p>...</p>
    <p>...</p>
    <closer> <salute>Mit vorzüglicher Hochachtung Ihr ergebenster</salute>
      <signed><persName>H. Schuchardt</persName></signed> </closer>
  </div>
</body>
```

```

<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Letter from H. Schuchardt to unnamed recipient</title>
      </titleStmt>
      <publicationStmt>
        <publisher>Unknown</publisher>
        <date>1879</date>
      </publicationStmt>
      <sourceDesc>
        <p>Original letter written by H. Schuchardt</p>
      </sourceDesc>
    </fileDesc>
    <profileDesc>
      <correspDesc>
        <correspAction type="sent">
          <persName>H. Schuchardt</persName>
          <date when="1879-02-08">8th February 1879</date>
          <placeName>Graz</placeName>
        </correspAction>
        <correspAction type="received">
          <persName>Herr Kollege</persName>
        </correspAction>
      </correspDesc>
    </profileDesc>
  </teiHeader>

```

- correspDesc lässt sich einfach generieren

# Learnings

- Eine Task nach der anderen abarbeiten: z. B. Textstruktur, dann Named Entities, dann Metadaten. Also n Iterationen:  
Text → GPT → TEI → GPT → TEI → GPT → TEI  
Workflows und API (und nicht ChatGPT; GPT-4 gibt es noch (!) nicht als API).
- Einspielen von Feedback durch die Editor\*innen. Wie wird Feedback effektiv für GPT zurückgespielt? → Vektordatenbank?
- Mastering Prompt Engineering (von 81% auf 89%, <https://www.youtube.com/watch?v=wVzuvf9D9BU>)
  - “Step-by-Step Prompting” aka. “Chain of Thought”
  - “Finding its own Errors”
  - “Dialoguing with itself”
- Eigene Kontexte abbilden mit Vektordatenbanken?
- Wie wird die Datenqualität überprüft: im worst case lesen Editor\*innen die TEI im Oxygen Korrektur ... aber das wird eh sowieso gemacht.
- GPT-4-API ist noch nicht offen, es gibt absurde Tools wie Code Interpreter in GPT-4 und mit Orca kommen Open-Source-Technologien.

**... in 3 Jahren wird kein Mensch mehr TEI XML schreiben !?**

Anhang



# Prompt Engineering

GPT 4 is Smarter than You Think: Introducing SmartGPT.

<https://www.youtube.com/watch?v=wVzuvf9D9BU>

zhou-01-ins (instruction)	<i>Zhou-instruction</i>	"Let's work this out in a step by step way to be sure we have the right answer."
qa-10 (instruction)	<i>Plan</i>	"First think step by step - describe your plan for how to get to the right answer, written out in great detail. Then answer the question."
qa-12 (instruction)	<i>Articulate</i>	"Carefully read the question & work this out in a step by step way to be sure you have the right answer. Be certain to spell out your thoughts & reasoning so anyone can verify them. Spell out everything in painstaking detail & don't skip any steps!"
qa-13 (instruction)	<i>Rephrase</i>	"Instruction: First let's rephrase the question to be sure we understood it correctly. Second, let's work this out step by step by spelling out our thoughts & reasoning so anyone can verify them. Third, make sure we have the right answer."
qa-16 (instruction)	<i>Elaborate</i>	"Answer the following question through careful, concise step-by-step reasoning. First, complement the question with helpful knowledge and important additional facts. Second, generate sub-questions that are required to answer the original question, answer them until you can answer the original question."
qa-17 (instruction)	<i>Converse</i>	"Create a dialog between a professor and a student. The student asks sub-questions to the question. The professor works them out in a step by step way and makes sure that the student understood how they got to the right answer."
refl-01 (instruction)	<i>Self-critique</i>	"Answer the question, then critique the answer. Based on the critique, reconsider the other answer options and give a single final answer."

when the few-shot examples in the prompt contain the reasoning steps for reaching a conclusion (“chain-of-thought reasoning”) (Magister et al., 2022). Moreover, Kojima et al. (Kojima et al., 2022) showed that, even in the absence of exemplar demonstrations, instructing language models to think “step-by-step” enhances their reasoning performance. **There is not currently a theoretical understanding of why these prompting techniques are beneficial**, however some recent studies have explored the effects of compositional structure and local dependencies in training data on efficacy of these methods (Hahn and Goyal, 2023; Prystawski and Goodman, 2023).

**Prompt:**  
Read the scenario and answer the following question:

Scenario: “Anne made lasagna in the blue dish. After Anne left, Ian came home and ate the lasagna. Then he filled the blue dish with spaghetti and replaced it in the fridge.”

Q: Does Anne think the blue dish contains spaghetti?

A: When Anne left the blue dish contained lasagna. Ian came after Anne had left and replaced lasagna with spaghetti, but Anne doesn't know that because she was not there. So, the answer is: No, she doesn't think the blue dish contains spaghetti.

Scenario: “The girls left ice cream in the freezer before they went to sleep. Over night the power to the kitchen was cut and the ice cream melted.”

Q: When they get up, do the girls believe the ice cream is melted?

spaghetti and replaced it in the fridge.”

Q: Does Anne think the blue dish contains spaghetti?

A: Let's think step by step: When Anne left the blue dish contained lasagna. Ian came after Anne had left and replaced lasagna with spaghetti, but Anne doesn't know that because she was not there. So, the answer is: No, she doesn't think the blue dish contains spaghetti.

Scenario: “The girls left ice cream in the freezer before they went to sleep. Over night the power to the kitchen was cut and the ice cream melted.”

Q: When they get up, do the girls believe the ice cream is melted?

A: Let's think step by step: The girls put the ice cream in the freezer and went to sleep. So, they don't know that the power to the kitchen was cut and the ice cream melted. So, the answer is: No, the girls don't believe the ice cream is melted.

Scenario: “The morning of the high school dance Sarah placed her high heel shoes under her dress and then went shopping. That afternoon, her sister borrowed the shoes and later put them under Sarah's bed.”

Question: When Sarah gets ready, does she assume her shoes are under her dress?

A: Let's think step by step:

GPT 4 is Smarter than You Think: Introducing SmartGPT.

<https://www.youtube.com/watch?v=wVzuvf9D9BU>

**Figure 1. Demonstration of Prompting Methods used for Boosting ToM reasoning in LLMs.** Examples of 4 prompting types used to test the ToM performance of LLMs. Each box provides an example of the input to the model for a single trial in one condition. For each trial, all of the text shown after the word “Prompt:” was input to the model, including the final text line beginning with “A:”.

Im Kontext der Generierung von synthetischen Daten, wenn es wenig menschlich annotierte Daten gibt:

*“Researchers have recently shown that LLMs can outperform human crowd workers in data labeling (Gilardi et al., 2023; He et al., 2023), highlighting their potential for automating labor-intensive annotation tasks and improving label quality.”*

Møller, Anders Giovanni, Jacob Aarup Dalsgaard, Arianna Pera, and 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>.