

Outline

- Transformers, Language models (Fantastic Beasts and Where to Find Them)
- Prompts for tech writing (AI as a tool for content creation)
- Lessons learned while creating prompts
- How does it affect me? (A glimpse into the future)

The transformer architecture

- A **transformer** is a deep learning model (neural network) that
 - transforms input sequences into output sequences
 - adopts the mechanism of self-attention
 - capturing the relationships between words of the same sentence
 - “The **man** walked to the **river bank**, and **he** ate a sandwich”
- introduced in **2017** by a team at Google Brain (“*Attention is all you need*”)
- training parallelization allows training on larger datasets

Large language models

- use transformer neural networks
- trained on massive amounts of text
- understand and generate text in a human-like fashion
- examples:
 - Google **BERT** (*Bidirectional Encoder Representations from Transformers*)
 - **LLaMA** from Meta
 - OpenAI **GPT-3**, **GPT-4** (*Generative Pre-trained Transformer*)
 - OpenAI **ChatGPT** (fine-tuned from GPT-3.5, optimized for dialogue)

What is a prompt

- A *prompt* represents instructions and context given to a language model to perform a task
 - “*You are a translation assistant and you will translate the following text: I want to eat an apple.*”
- *Prompt engineering* is the development and *refining* of prompts to *best* perform a task

The context window

- the range of tokens or words the AI model can access and consider when generating responses to prompts
 - instructions + context + response

- 1 token \approx 4 chars in English
- 1 token \approx $\frac{3}{4}$ words
- 100 tokens \approx 75 words

Or

- 1-2 sentence \approx 30 tokens
- 1 paragraph \approx 100 tokens
- 1,500 words \approx 2048 tokens

Model	Maximum text length
gpt-3.5-turbo	4,096 tokens (~5 pages)
gpt-4	8,192 tokens (~10 pages)
gpt-4-32k	32,768 tokens (~40 pages)

GPT-4	Model	Prompt	Completion
	8K context	\$0.03 / 1K tokens	\$0.06 / 1K tokens
	32K context	\$0.06 / 1K tokens	\$0.12 / 1K tokens

Chat	Model	Usage
	gpt-3.5-turbo	\$0.002 / 1K tokens

Fine tune

- Ability to train on more examples than can fit in a prompt
- Token savings due to shorter prompts
- Lower latency requests
- *ChatGPT doesn't support fine tuning yet*

<https://platform.openai.com/docs/guides/fine-tuning/prepare-training-data>

https://github.com/openai/openai-cookbook/blob/main/examples/Question_answering_using_embeddings.ipynb

<https://openai.com/pricing>

Limitations and perils

- Slow due to network latency and servers load
- Content sent to OpenAI Servers
 - REST API requests are not used in training – careful with <https://chat.openai.com/>

Will OpenAI use my content to improve models and services?


Data submitted through the OpenAI API is not used to train OpenAI models or improve OpenAI's service offering. Data submitted through non-API consumer services ChatGPT or DALL·E may be used to improve our models.

Limitations and perils

- Hallucination -

- No source of truth, finds the most plausible word
- Confidence, no clarifications
- Mitigation through prompt engineering


AL name the 3 founders of syncro soft

 The three founders of Syncro Soft, a software company specializing in XML content authoring and editing tools, are:

1. Adrian Buza
2. Alex Jitianu
3. Radu Coravu

These individuals played key roles in establishing Syncro Soft and developing their flagship product, Oxygen XML Editor.

AL tell me how to make a zip with a password on windows

 To create a password-protected ZIP file on Windows, you can follow these steps:

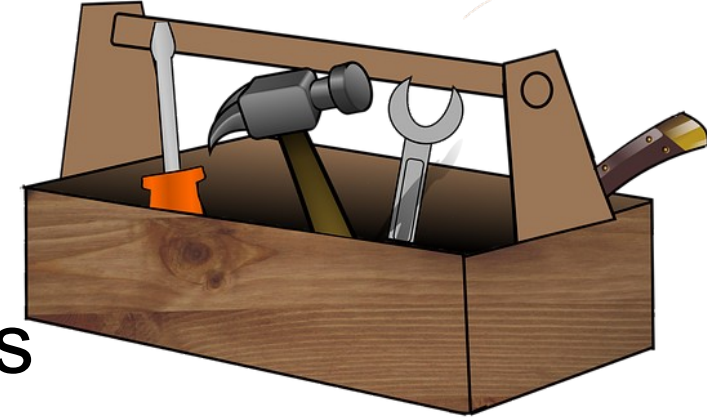
1. Select the files or folders you want to include in the ZIP file. You can hold down the **Ctrl** key and click on multiple items to select them.
2. Right-click on the selected items, hover over the "Send to" option, and choose "Compressed (zipped) folder" from the submenu. This will create a new ZIP file with the same name as the first selected item.
3. Double-click on the newly created ZIP file to open it.
4. In the ZIP file window, click on the "File" menu at the top-left corner and choose "Add Password" (or a similar option depending on your Windows version).

AI Tools

- For software developers (GitHub copilot)
 - <https://github.com/features/copilot>
- For documentation consumers
 - content written by writer is consumed by AI to give answers
 - OpenAI ChatGPT plugins :
<https://platform.openai.com/docs/plugins/introduction>
 - <https://markprompt.com/docs>
- For content creators

Prompts for Tech Docs

- the toolbox -



- Generate a first draft from raw SME input
- Create procedure from a rough set of notes
- Impose a style, tone-of-voice, improve readability
- AI could become the primary documentation user interface
 - *How/Does this <topic> answer this <question>?*
 - *What questions will bring people to this topic?*
 - *Where in my documentation is this question answered?*

Prompts for Tech Docs (II)

- the toolbox -

- AI works with **human language** instructions
- Schematron has **human-readable** error messages
 - *“Short descriptions should not exceed 50 words.”*

Care and Preparation

Short Description: When caring for your flower garden you want to feed your plants properly, control pests and weeds. Good soil is a must to successful gardening, large amounts of water are needed to keep plants healthy. You have to balance the soil structure with nutrients and regulate the pH to cover your plants. Many flower gardens fail because they just don't get enough of your attention.

Validation:
Short descriptions should not exceed 50 characters
Press F2 for focus

When caring for your flower garden, large amounts of water are needed to keep plants healthy. You have to balance the soil structure with nutrients and regulate the pH to cover your plants. Many flower gardens fail because they just don't get enough of your attention.

AI Positron Chat

Me
AI Fix
Fix the problem using AI.

AI Positron
<shortdesc>When caring for your flower garden, feed plants, control pests and weeds, balance soil nutrients & pH. Remember, gardens fail without attention.</shortdesc>

[Replace](#) [Preview](#) [Copy](#)

control pests and weeds. Good soil is a must to successful gardening, large amounts of water are needed to keep plants healthy. You have to balance the soil structure with nutrients and regulate the pH to cover your plants. Many flower gardens fail because they just don't get enough of your attention.

Prompt engineering steps

1) *Hint*: Clearly defining the goal or objective

“You get what you ask for, not what you want!”

2) Create an initial prompt





→ 3) Test the prompt

4) Analyze the response

└─ 5) Refine the prompt

6) Optional: *“Introspection”*

AL are you sure?

 Apologies for the confusion. It seems I provided incorrect instructions earlier. Windows does not have built-in functionality to password-protect ZIP files. However, you can use third-party software like 7-Zip or WinRAR to achieve this. Here's how you can do it using 7-Zip:   

Prompt engineering - iteration -

- *“You are John, a technical writer assistant that is tasked with correcting errors inside plain text or XML content.....
The problem you need to fix is”*
 - *“Here's the corrected content with the "format" attribute added: ...”*
- *“You are John, a technical writer assistant that is tasked with correcting errors inside plain text or XML content.....
John answers to the point without other explanations. The problem you need to fix is”*

Prompt engineering

- iteration -

(1) You are a translation system; you are translating XML structured documents...

Playground

Lo

SYSTEM You are a translation system; you are translating XML structured documents to German while maintaining the markup. Rules: <ul style="list-style-type: none">- the translation should include all the initial XML markup, tags, attributes and values;- do not translate text inside XML elements: <code><codeblock></code>, <code><pre></code>, <code><codeph></code>, <code><filepath></code>, <code><lines></code>, <code><programlisting></code>;- do not alter text already in German;- preserve the XML markup applied to the current text	USER <u>Je veux manger une poir</u>
	ASSISTANT <code><root></code> <code><sentence></code> Je veux manger une <code><term type="fruit">poire</term></code> <code></sentence></code> <code></root></code>
	+ Add message

*(2) You are a translation system; you are translating XML structured documents **or plain text**...*

Prompt engineering

- best practices -

- Give it an identity
 - *“I want you to act as a technical documentation writer.”*
 - *“You will act as Ava. Ava is a technical documentation writer. Ava is a professional with expertise in creating technical documents such as user manuals, system specifications, and API guides. Ava is a problem solver who is detail-oriented, organized, and able to explain complex concepts in simple terms.”*

Prompt engineering

- best practices -

- Give it an identity
- Start with zero-shot, then few-shot (examples)

AL Add DITA markup to this menu cascade: File > Export

Sure! Here's the menu cascade with DITA markup added:

```
xml Copy code
<menucascade>
  <menuseq>
    <menuchoice>
      <guimenu>
        <guimenuitem>File</guimenuitem>
      </guimenu>
    </menuchoice>
    <menuchoice>
      <guisubmenu>
        <guimenu>
          <guimenuitem>Export</guimenuitem>
        </guimenu>
      </guisubmenu>
    </menuchoice>
  </menuseq>
</menucascade>
```

AL Add DITA markup to a menu cascade

Text: File > Export
Markup: <menucascade>
 <uicontrol>File</uicontrol>
 <uicontrol>Export</uicontrol>
 </menucascade>

Text: Edit > Copy
Markup:

```
<menucascade>
  <uicontrol>Edit</uicontrol>
  <uicontrol>Copy</uicontrol>
</menucascade>
```

Prompt engineering

- best practices -

- Give it an identity
 - Start with zero-shot, then few-shot (examples)
 - Reduce “fluffy” and imprecise descriptions
 - *“The description for this product should be fairly short, a few sentences only, and not too much more.”*
 - *“Use a 3 to 5 sentence paragraph to describe this product.”*
- 1 token \approx 4 chars in English
 - 1 token \approx $\frac{3}{4}$ words
 - 100 tokens \approx 75 words

Prompt engineering

- best practices -

- Give it an identity
- Start with zero-shot, then few-shot (example), if neither of them works, then fine-tune
- Reduce “fluffy” and imprecise descriptions
- Instead of just saying what **not to do**, say what **to do** instead
- Temperature (random, creative)
- Save the prompts
 - **Experimenting with prompts is a time consuming process**
 - **New model releases might affect existing prompts**

Prompt engineering - development environment -

- <https://chat.openai.com/>
- <https://platform.openai.com/playground>

How does this affect my job (I)

- Will we still need content? Yes! AI empowers documentation.
 - AI could become the primarily documentation user interface
 - OpenAI ChatGPT plugins :
<https://platform.openai.com/docs/plugins/introduction>
 - AI could become a support specialist that searches for answers in the documentation.
 - Write for AI consumption
 - The quality of an AI answers depends on the quality of the ingested documentation.

How does this affect my job (II)

- Will we still need to publish?
 - a classic documentation site could become a secondary user interface
 - PDF still needed for print
- Will we still need to translate?
 - just ask the AI powered search in your native language

How does this affect my job (III)

- **Review will be important.**
 - Either human or through AI introspection
 - <https://learn.microsoft.com/en-us/azure/analysis-services/analysis-services-create-terraform>
 - The models are probabilistic, no guarantees (it might look good, can be subtly incorrect)
- **Long term metamorphosis**
 - Tech writers might transform into information reviewers or fact reviewers
 - Software developers might transform into prompt engineers

How does this affect my job (IV)

- **Humans are still in the decision loop**
 - strategy, collaboration with others, can follow a long term specific goal
- **brings a disadvantage for those that worked to have great English skills and will not be compensated accordingly**

How does this affect my job(V)

- The year 2023 is a turning point. AI-generated content is released.
- *LLM is a new species. Not known in depth.*
 - *AI alignment* - steer AI systems towards humans' intended goals, preferences, or ethical principles
 - https://en.wikipedia.org/wiki/AI_alignment
 - The “do-not-translate-codeblock” case

THANK YOU!



Alex Jitianu

alex_jitianu@oxygenxml.com

[@AlexJitianu](#)