Utilizing GenAI to Enhance Early-Stage Technical Content Creation

Sorin Carbunaru
Software Developer, Syncro Soft
sorin_carbunaru@oxygenxml.com

ETC Sofia, 2024

syncro soft

<oXygen/> XML Editor

Agenda

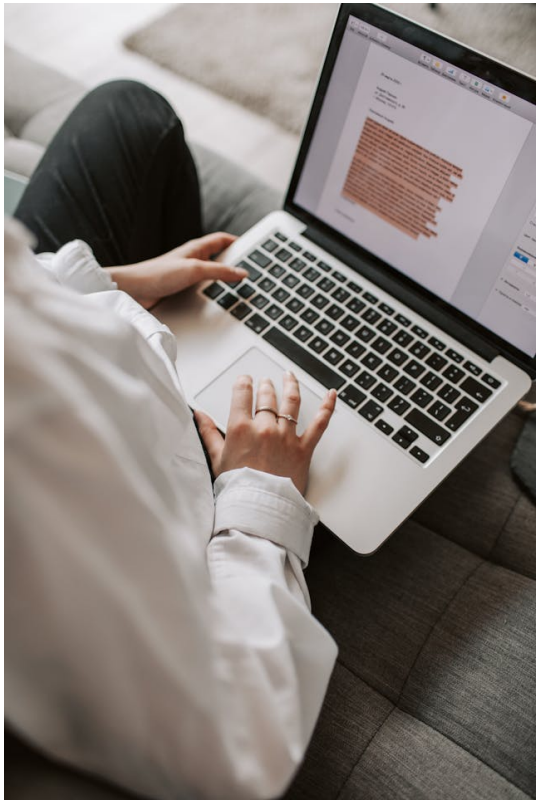
- Introduction (GenAI brief overview & my goals)
- GenAI in Technical Content Creation
- GenAI in Early-Stage Technical Content Creation
- OpenAI GPT-4 Turbo & GPT-4o
- GenAI in Drafting Documentation using Image-to-Text
- Tailoring the Drafting Process
- Conclusions and Possible Future Directions

Introduction

- Generative Artificial Intelligence (GenAI) is a category of AI focused on **creating content** (text, images, videos, music, etc.), as opposed to traditional AI, which aims primarily at understanding or classifying data.
- Therefore, the significance of GenAI in technical writing cannot be overstated. Creating content (technical documentation included) is one of the **main purposes** for which GenAI has been created and exists.
- My goals with this presentation are for you to:
  1. Acquire a sense of how GenAI can enhance your productivity in the early-stage technical content creation.
  2. Develop an understanding of the power of image-to-text GenAI in documentation drafting.

GenAI in Technical Content Creation

- Technical content creation can be split in 4 stages / 4 **R**'s:
  1. **R**esearching: gather information, understand the subject matter, identify the right keywords for SEO, etc.
  2. W**r**iting: put down the ideas resulted from the research and convert them into actual content.
  3. **R**eviewing: check for grammar and spelling mistakes, logic problems, wrong markup, lack of clarity, bad readability, confusing structure, etc.
  4. **R**efining: polish and perfect the content by incorporating the feedback from the review stage.



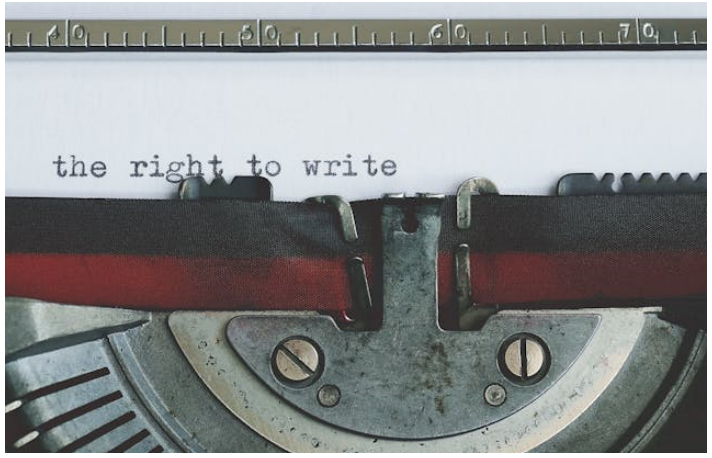
GenAI in Technical Content Creation

- How can GenAI help:
  1. Researching:

  - Quickly gather information
  - Get summaries, explanations, clarifications for the existing content
  - Brainstorming

GenAI in Technical Content Creation

- How can GenAI help:
  2. Writing:



  - Generate a high quality initial draft
  - Generate topic parts: short descriptions, index terms, image captions, etc.
  - Convert content from one text format to another (e.g. plain text to DITA)

GenAI in Technical Content Creation

- How can GenAI help:
  3. Reviewing:

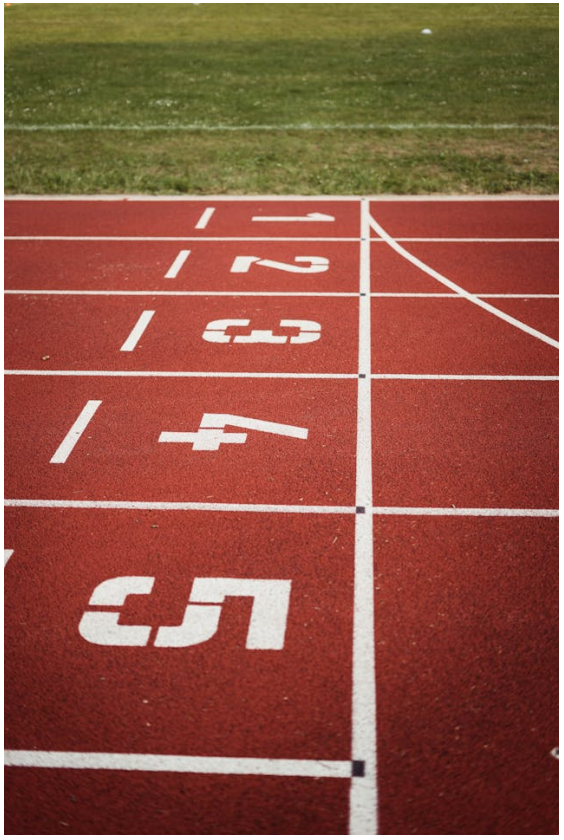
  - Almost replaces the need to ask someone else to scrutinize your content.
  - From spelling and grammar mistakes to markup and readability issues, GenAI can detect pretty much everything.

GenAI in Technical Content Creation

- How can GenAI help:
  4. Refining:

  - If GenAI is able to find the issues in your content, be sure it's also able to fix them.
  - GenAI can actually merge the reviewing and refining stages into a single stage, where it both detects issues and proposes ways to correct them.

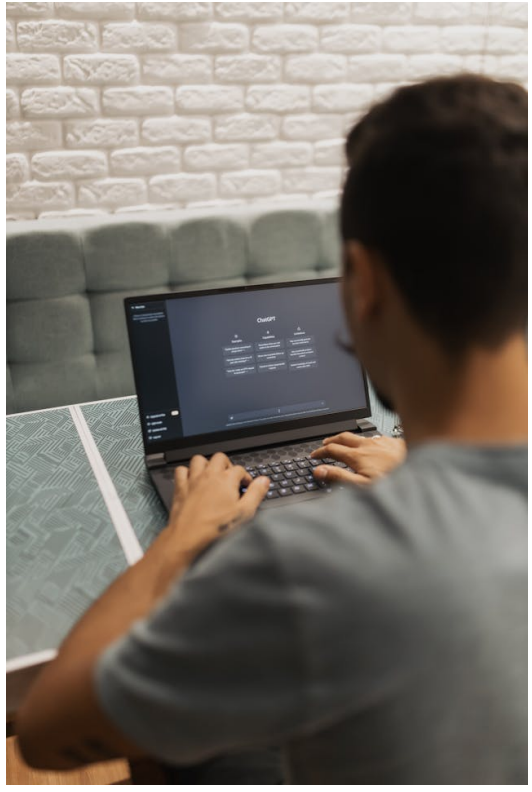
GenAI in Early-Stage Technical Content Creation



- Early-stage = right before starting to write. You know... That moment when you procrastinate as much as possible...
- At this point, information should have been gathered, screenshots should have been taken, code snippets should have been collected. In other words, everything should be ready for the initial draft.
- In this stage, GenAI can help with at least 2 major use cases:
  1. Generating a documentation draft from scratch.
  2. Updating the existing text content based on recent changes in a product.
- In both cases, the image analysis capabilities of the latest GenAI models, such as GPT-4 Turbo or GPT-4o, can be of utmost importance.

OpenAI GPT-4 Turbo & GPT-4o

- GPT-4 Turbo and GPT-4o both have vision capabilities, and they allow the AI to take in images and perform tasks (with text output) related to them.
- Use-cases:
  - Image description
  - Object detection
  - Text transcription
  - Data interpretation
  - Video narration
  - Assistance for the visually impaired
- **Note:** there are alternatives to GPT-4 Turbo and GPT-4o, but I experimented with just one, and only a few times.



GenAI in Drafting Documentation using Image-to-Text

- It is said that an image is worth 1000 words. Then... Why not generate 1000 words from an image?

The previous image was actually worth 1096 words, 1948 syllables, 7132 characters :)

The image depicts a bustling city street drenched in the vibrant glow of countless neon lights, painting an atmospheric picture of urban nightlife in a Japanese metropolis. The scene is quintessential in its portrayal of a lively district, likely representative of areas such as Shibuya, Shinjuku, or even Osaka's Dotonbori, renowned for their ceaseless energy and kaleidoscopic displays.

Dominating the scene are the myriad neon signs that span both sides of the street, climbing the buildings like vertical ladders of light. These signs are an eclectic mix, featuring advertisements for various businesses, from karaoke bars and restaurants to pachinko parlors and retail shops. The characters displayed on the signs are primarily in Japanese, contributing to the authentic cultural ambiance. The kanji and katakana characters light up in a dazzling array of colors—vibrant reds, bold whites, luminous purples, and bright yellows. Some of the signs incorporate both text and images, ensuring that they capture and hold the attention of passersby. These signs serve as a vibrant canvas showcasing the diverse commercial life thriving in this urban setting.

Along the street, there are multiple types of businesses open and inviting with their lit-up entrances and posted menus. To the left side, a building marked "B1" seems to advertise an internet café or potentially a gaming center, given the illustrated signs featuring a computerized image. Nearby, there's a taller building indicating a karaoke bar with its prominent bilingual signage. This establishment likely caters to both local patrons and tourists, providing a common form of entertainment where groups can rent private rooms to sing popular songs. Karaoke, after all, is a beloved pastime in Japan, often involving friends, family, or coworkers in boisterous and lively gatherings.

The street level offers an array of eateries, visible through their open-door policies and brightly lit interiors. One can catch sight of restaurants specializing in Japanese cuisine, including sushi bars, ramen shops, and likely izakayas—informal pubs where patrons might gather for drinks and small dishes. The signage here is meticulously detailed, often showing enticing images of dishes, ideal for drawing in hungry customers. These eateries form the culinary heart of the district, offering everything from quick bites to elaborate multi-course meals.

In the center-ground of the image, the street is moderately crowded with a mix of locals and tourists, each contributing to the dynamic atmosphere. The people are variously dressed, with some perhaps wearing business attire suggesting they may be unwinding after a day's work, while others are more casually dressed, likely out for an evening of leisure. Amid the throngs, a few individuals are seen holding umbrellas, hinting perhaps at a recent or impending rain. The density of the crowd increases the feeling of an active, lived-in environment, where every corner holds the potential for new experiences. The ebb and flow of foot traffic add a kinetic energy to the scene, as people move in and out of establishments, converse in small groups, or simply take in the sights.

The street itself is narrow, flanked by modern buildings that stretch upwards to the darkened sky. The architectural style blends traditional aspects with contemporary structures, commonplace in metropolitan Japan where history and modern advancements go hand in hand. Street lamps provide additional illumination, their glow intermingling with the neon lights to bathe the area in a warm, inviting light. Suspended above the street, an assortment of hanging lanterns adds an extra layer of charm and authenticity, painting a vivid picture of city life in Japan. These lanterns often evoke traditional Japanese aesthetics, a nod to the culture's rich heritage amidst the urban hustle.

The darkening sky indicates the photograph was taken during evening or night when the city transforms into a haven of lights. The skyline isn't predominantly visible due to the density of the buildings, but one can infer the urban sprawl extends far beyond what is immediately visible. This bustling snapshot encapsulates the relentless pulse of city life where the night holds as much appeal as the day. The absence of stars in the sky further highlights the overpowering brilliance of the city lights, a common sight in densely populated urban areas.

Further to the right side of the image, an establishment spread across multiple floors might catch the eye, identified by large vertical signs hosting kanji characters. This kind of vertical expansion is commonplace in Japanese cities where space is at a premium, making upward growth a practical necessity. Each floor appears to be home to different businesses, possibly including restaurants, bars, and entertainment facilities, each vying for the attention of potential customers through their vividly illuminated advertisements. The verticality of these buildings adds to the multifaceted character of the street, where one structure can house an entire ecosystem of varied experiences.

In the background, the convergence of sights and lights evokes a sense of endless opportunity and exploration. Akin to many urban experiences, this Japanese street is a microcosm of the broader cultural and social landscape—vibrant, diverse, and continually evolving. The intermingling of traditional and modern elements is striking—from the architectural details to the variety of commercial offerings, illustrating the dynamic cultural fabric of the city. The traditional charm interspersed with modern conveniences creates a unique atmosphere that is both nostalgic and forward-looking.

The pedestrian-friendly nature of the street highlights the importance placed on walkability in urban Japanese design, with narrow side roads and alleyways likely branching off from the main street, each harboring its own secrets, be it a cozy café, a hidden bar, or a unique shop. These alleyways often serve as quiet refuges from the main thoroughfare's hustle and bustle, offering intimate spaces for discovery and relaxation.

The image subtly emphasizes the value of communal spaces in urban settings. The street, bustling with life and energy, represents a social hub where people connect, relax, and share experiences, underscoring the role of public spaces in fostering community spirit in a city as vast and varied as this. It's a testament to the human need for interaction and shared experiences amidst the anonymity of city life.

In essence, this image serves as a window into the exhilarating yet routine nightlife of a Japanese city—a celebration of lights, culture, and human connection. Whether one is a local or a visitor, such streets offer an immersive experience that is as invigorating as it is intriguing, capturing the essence of urban life in one of the world's most fascinating places. The vibrancy and diversity of the scene highlight the unique charm of Japanese urban environments, where every glance reveals a story, every step invites exploration, and every night promises a new adventure.

The previous image was actually worth 1096 words, 1948 syllables, 7132 characters :)

The image depicts a bustling city street drenched in the vibrant g scene is quintessential in its portrayal of a lively district, likely representative of areas such as Shibuya, Shinjuku, or even Osaka'

Dominating the scene are the myriad neon signs that span both s advertisements for various businesses, from karaoke bars and restaurants to pachinko parlors and retail shops. The characters katakana characters light up in a dazzling array of colors—vibrant reds, bold whites, luminous purples, and bright yellows. S ersby. These signs serve as a vibrant canvas showcasing the diverse commercial life thriving in this urban setting.

Along the street, there are multiple types of businesses open and tise an internet café or potentially a gaming center, given the illustrated signs featuring a computerized image. Nearby, there's o both local patrons and tourists, providing a common form of entertainment where groups can rent private rooms to sing popu terous and lively gatherings.

The street level offers an array of eateries, visible through their o ncluding sushi bars, ramen shops, and likely izakayas—informal pubs where patrons might gather for drinks and small di hungry customers. These eateries form the culinary heart of the district, offering everything from quick bites to elaborate mul

In the center-ground of the image, the street is moderately crow lressed, with some perhaps wearing business attire suggesting they may be unwinding after a day's work, while others are more orellas, hinting perhaps at a recent or impending rain. The density of the crowd increases the feeling of an active, lived-in er a kinetic energy to the scene, as people move in and out of establishments, converse in small groups, or simply take in the si

The street itself is narrow, flanked by modern buildings that stret es, commonplace in metropolitan Japan where history and modern advancements go hand in hand. Street lamps provide ad Suspended above the street, an assortment of hanging lanterns adds an extra layer of charm and authenticity, painting a vivid pic heritage amidst the urban hustle.

The darkening sky indicates the photograph was taken during eve e density of the buildings, but one can infer the urban sprawl extends far beyond what is immediately visible. This bustling sna ce of stars in the sky further highlights the overpowering brilliance of the city lights, a common sight in densely populated

Further to the right side of the image, an establishment spread a vertical expansion is commonplace in Japanese cities where space is at a premium, making upward growth a practical necessi ent facilities, each vying for the attention of potential customers through their vividly illuminated advertisements. The n entire ecosystem of varied experiences.

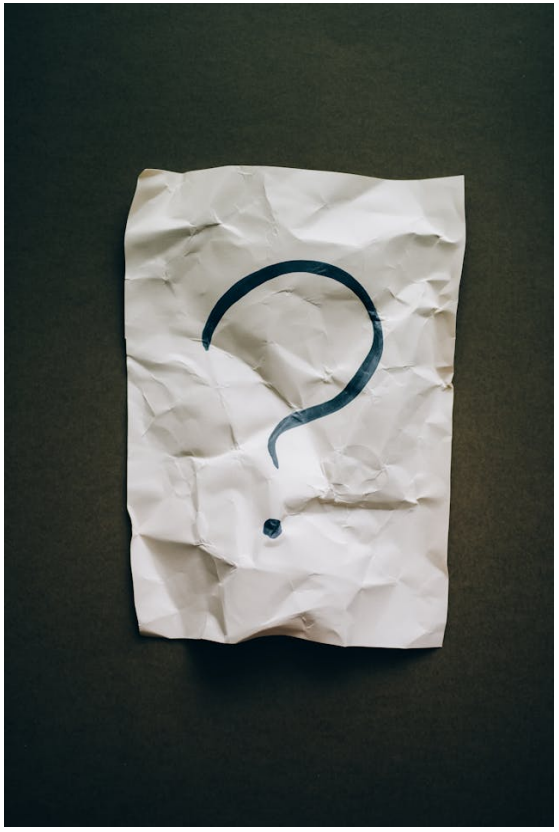
In the background, the convergence of sights and lights evokes a cosm of the broader cultural and social landscape—vibrant, diverse, and continually evolving. The intermingling of traditional ating the dynamic cultural fabric of the city. The traditional charm interspersed with modern conveniences creates a unique

The pedestrian-friendly nature of the street highlights the import f from the main street, each harboring its own secrets, be it a cozy café, a hidden bar, or a unique shop. These alleyways often nd relaxation.

The image subtly emphasizes the value of communal spaces in u d share experiences, underscoring the role of public spaces in fostering community spirit in a city as vast and varied as this. It's

In essence, this image serves as a window into the exhilarating ye a local or a visitor, such streets offer an immersive experience that is as invigorating as it is intriguing, capturing the essence of u nique charm of Japanese urban environments, where every glance reveals a story, every step invites exploration, and every night promises a new adventure.

GenAI in Drafting Documentation using Image-to-Text



- GenAI with image-to-text capabilities can be used in drafting documentation for:
  1. Transcribing hand-written notes or scanned documents.
  2. Converting charts and diagrams to text summaries.
  3. Describing a process based on a sketch or on multiple screenshots with the specific steps.
  4. Documenting UI elements from screenshots, components of a complex mechanism, etc.
  5. Generating titles, captions, alternate texts for images.
  6. Detecting discrepancies between text and images (diagrams, screenshots, etc.) and updating the text accordingly.
- **Enough with the talk! Let's see some magic!**

GenAI in Drafting Documentation using Image-to-Text

```
<?xml version="1.0" encoding="UTF-8"?>
<doc-draft>
  <title>XML Refactoring Doc Draft</title>
  <instructions>
    <draft-summary>The XML Refactoring tool in Oxygen XML Editor (available in the Tools menu) helps
      manage the structure of XML documents by allowing users to perform various structure
      management tasks such as inserting, deleting, or renaming attributes in all instances of a
      particular element. The tool is presented as a wizard with multiple pages, including options
      to configure operation parameters and select the scope and filters for the operation. Users
      can preview the changes before applying them, but there is no automatic way to revert the
      operations once they are finished. Additionally, the tool cannot preserve CDATA sections in
      the XML documents.</draft-summary>
    <instruction>Add each image in the documentation and present the flow of the operation as
      suggested by these images. Add them in the correct order of the steps. You should document
      each UI element shown in the images.</instruction>
    <image href="img/xmlrefactoring_step1.png"/>
    <image href="img/xmlrefactoring_step2.png"/>
    <image href="img/xmlrefactoring_step3.png"/>
    <instruction>For the previous image, make sure to document all the available options using DITA definition
lists.</instruction>
    <instruction>Also add an important note about the fact that the operation results can be previewed,
    and add the following image as an example, with the caption "Isn't it cool?".</instruction>
    <image href="img/xmlrefactoring_preview.png"/>
  </instructions>
  <relationship-context>
    <similar-topic href="similar-converting-between-schema-languages.dita"/>
  </relationship-context>
</doc-draft>
```

GenAI in Drafting Documentation using Image-to-Text

Using the XML Refactoring Tool

Metadata

Keywords:

(XML Refactoring) (Oxygen XML Editor)

(Structure Management)

The **XML Refactoring** tool within the Oxygen XML Editor offers a powerful solution for managing the structure of XML documents. This tool provides a variety of operations, such as inserting, deleting, or renaming attributes across all instances of a particular element.

To access the **XML Refactoring** tool, navigate to the **Tools** menu in Oxygen XML Editor.

Refactoring Operation Steps

Follow these steps to perform a refactoring operation:

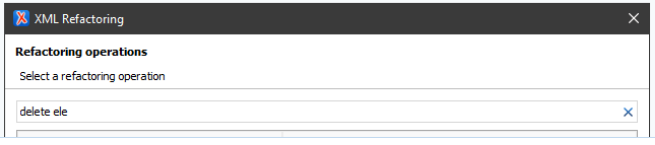
1. Select the desired refactoring operation from the list presented in the wizard.
2. Specify any necessary operation parameters, such as element names or XPath expressions.
3. Choose the scope and set any filters for the refactoring operation to determine which documents will be affected.
4. Preview the changes before applying them to ensure accuracy.
5. Once you have confirmed the changes, apply the refactoring operation.

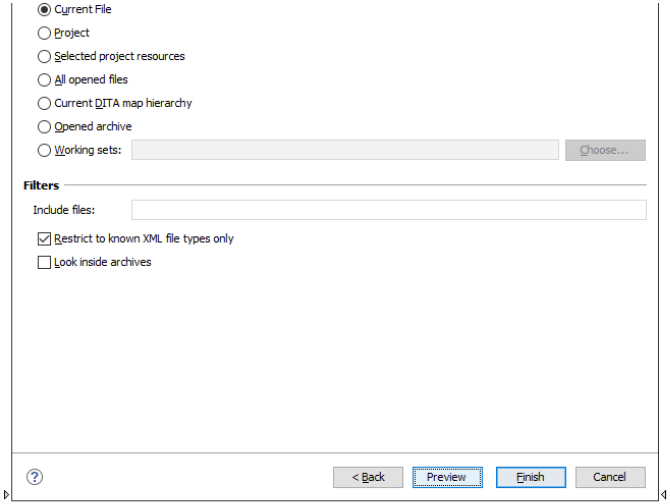
Note: There is no automatic way to revert operations once they are applied, and CDATA sections cannot be preserved during the refactoring process.

XML Refactoring Wizard

The XML Refactoring tool is presented as a multi-page wizard, guiding you through each step of the refactoring process. Below are the images depicting each step in the wizard and the components within:

Refactoring Operations Selection





Scope

Choose where the refactoring operation will apply:

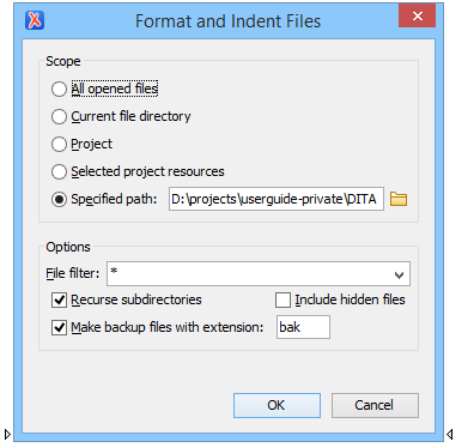
- **Current File** - Applies changes to the file currently open.
- **Project** - Applies changes to all files within the current project.
- **Selected project resources** - Applies changes to project files selected by the user.
- **All opened files** - Applies changes to all files open in the editor.
- **Current DITA map hierarchy** - Applies changes to files structured in the current DITA map.
- **Opened archive** - Applies changes to files within an open archive file.
- **Working sets** - Applies changes to a set of user-defined files.

Filters

Refine which files to include in the operation:

- **Restrict to known XML file types only** - Limits the operation to recognized XML formats.
- **Look inside archives** - Includes files contained within archive files (e.g., ZIP).

GenAI in Drafting Documentation using Image-to-Text

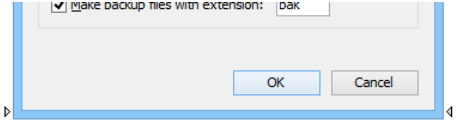


The **Scope** section allows you to choose from the following scopes:

- **All opened files** - The *pretty-print* is performed in all opened files.
- **Current file directory** - All the files in the folder of the currently edited file.

The **Options** section includes the following options:

- **File filter** - Allows you to filter the files from the selected scope.
- **Recurse subdirectories** - When selected, the *pretty-print* is performed recursively for the specified scope. The one exception is that this option is ignored if the scope is set to **All opened files**.



The **Scope** section allows you to choose from the following scopes:

- **All opened files** - The *pretty-print* is performed in all opened files.
- **Current file directory** - All the files in the folder of the currently edited file.
- **Project** - Pretty-print is performed on all files within the current project.
- **Selected project resources** - Allows you to select specific resources within the project for pretty-printing.
- **Specified path** - Enables pretty-printing for files in a specified directory path.

The **Options** section includes the following options:

- **File filter** - Allows you to filter the files from the selected scope.
- **Recurse subdirectories** - When selected, the *pretty-print* is performed recursively for the specified scope. The one exception is that this option is ignored if the scope is set to **All opened files**.
- **Include hidden files** - If checked, hidden files will also be included in the pretty-print process.
- **Make backup files with extension** - Allows you to specify an extension for creating backup copies of the files before formatting.

Tailoring the Drafting Process

- **Refine the prompt.** This is an art. Too little or too much context, and the prompt becomes sub-optimal.
- **Provide a similar document.** It makes a big difference in the quality of the result if you provide a document similar to what you expect the AI service to generate.

  Even if you don't have a similar topic, provide a couple of topic examples showing how you structure a document, what markup you use, what tone you adopt, etc.
- In more complex prompts, instead of adding all the instructions in the system message and the images as user messages, **group instructions and images** as content pieces in a user message.
- **Try tweaking the temperature** (the randomness level).

BONUS: Prompt Engineering Best Practices

- The quality of the results depends on the quality of the prompt. So here are a few extra prompt engineering best practices:
  1. Be clear, specific, detailed.
  2. Provide examples.
  3. Split complex tasks into simpler ones.
  4. Give positive instructions (dos, not don'ts).
  5. Don't forget to be clear, specific, detailed.
  6. Include context.
  7. Ask the model to adopt a persona.
  8. Clarify the output format.
  9. Start simple, test, adapt.

Conclusions

- Image-to-text GenAI empowers you with a ***vision-****ary* way of drafting: transforming pixels into phrases.
- This visual-to-verbal magic can be summoned both when starting fresh and when needing to adapt existing content.
- And, in the end, it's all about enhancing efficiency and productivity.

Possible Future Directions

- Theoretically, this is the worst GenAI will ever be from here on out.
- One interesting direction is using the new OpenAI Assistants API with Vision, which is more powerful, allowing for a code interpreter, online search, knowledge retrieval from uploaded files, etc.
- Another fascinating idea is trying to generate drafts using audio or video clips.

Resources

- OpenAI Vision
  https://platform.openai.com/docs/guides/vision
- Processing and narrating a video
  https://cookbook.openai.com/examples/gpt_with_vision_for_video_understanding
- Prompt engineering
  https://platform.openai.com/docs/guides/prompt-engineering

THANK YOU!

Looking forward to your feedback :)

Sorin Carbunaru

sorin_carbunaru@oxygenxml.com