

# GenAI & Wiki-Workflows: A practical workshop

**With:**

Sam Klein

Shani Evenstein Sigalov

Ilario Valdelli

Andrew Lih

**WIKIMANIA**  
**KATOWICE**



# Why this workshop?

Harnessing GenAI tools to  
improve wiki-workflows →

Saving volunteers' time and  
making our work more efficient

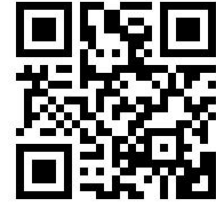
**WIKIMANIA**  
**KATOWICE**



# Existing channels for engagement

Telegram group:

<https://t.me/joinchat/ELlyMRaZrlcFyCvZOKhfLw>



Telegram  
group

Facebook group:

<https://www.facebook.com/groups/571301011510344>



Facebook  
group




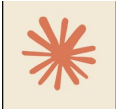




**1: Wikipedia**

**SJ & Shani**

**WIKIMANIA  
KATOWICE**

# Major AI tools to know

- ❖ Many large companies (and some open communities) have their own chat-based models. Better models tend to cost more to run.

- ❖ ChatGPT (OpenAI) 
- ❖ Claude (Anthropic) 
- ❖ Gemini (Google) 
- ❖ Copilot (Microsoft) 
- ❖ Llama (Meta/Facebook) 
- ❖ Huggingface – a commons of AI tools + elements 

➤ Ex: <https://huggingface.co/datasets/wikimedia/wikipedia>



# Tools to explore

- ❖ **Modalities:** Summarization, Proofreading, Translation, Style Drafting edits, Drafting articles
  - [General models](#) are good for all of this.
- ❖ **Edits: WP Add A Fact:** a Future Audiences production
  - <https://bit.ly/wp-add-af>
- ❖ **Articles: STORM** - aiming for well-referenced articles
  - <https://storm.genie.stanford.edu/>
  - <https://huggingface.co/spaces/multimodalart/storm>



# Shani's focus: tools for academic lit review and references

- ❖ [Research Rabbit](#) - [tutorial here](#)
  - ❖ [Perplexity](#) - [tutorial here](#)
  - ❖ [SciSpace](#) - [tutorial here](#)
  
  - ❖ [Scite](#) - [tutorial here](#)
  - ❖ [Consensus](#) - [tutorial here](#)
  - ❖ [Elicit](#) - [tutorial here](#)
  
  - ❖ [ChatPDF](#) - [tutorial here](#)
  - ❖ [Jenni AI](#) - [tutorial here](#)
- and here, specifically [connecting it to zotero](#)
- (use MAREK20 for an additional 20% off)



# Shani's insights

- ❖ These tools are great as a **great starting point**, but you still need to use your Wiki-instincts and critical thinking in the process!
- ❖ **Never trust one tool alone!** I always work with a few to get to a place I feel I have a good handle of the material
- ❖ These are just **a few options out of many**. There are new tools and capabilities popping all the time. It can be overwhelming and hard to follow. But worth the effort, as it saves TONS of time.
- ❖ Like with every technology – **there's a learning curve**. Embrace it. Experiment. Share. And enjoy the process.

**GenAI is here to stay.**







## 2: Wiki Commons

Ilario

# Ilario's tools to explore

- ❖ **Craiyon - quite free but slow and not qualitative high**
- ❖ **Nightcafe - Limited free version**
- ❖ **Adobe Firefly - Limited free version**
- ❖ **ChatGPT4o - Good quality but paid version**



# Ilario's tools to explore

- ❖ How to consider the images generated by AI?
- ❖ The copyright protects the human generated images, but AI is trained using images that can be copyrighted.
- ❖ [https://commons.wikimedia.org/wiki/Commons:AI-generated\\_media](https://commons.wikimedia.org/wiki/Commons:AI-generated_media)
- ❖ {{PD-algorithm}}



Craiyon (<https://www.craiyon.com/>)

**Input:** *Create an image of  
wikipedians meeting together.  
Relevance of wikipedia logo.*

There is to select the style:

- Digital art
- Photo
- Drawing



# Nightcafe (<https://creator.nightcafe.studio/>)

**Input:** *Create an image of wikipedians meeting together.  
Relevance of wikipedia logo.  
Digital art.*

Additional functionalities to refine the result.

Limited free version, a lot of credits to earn.

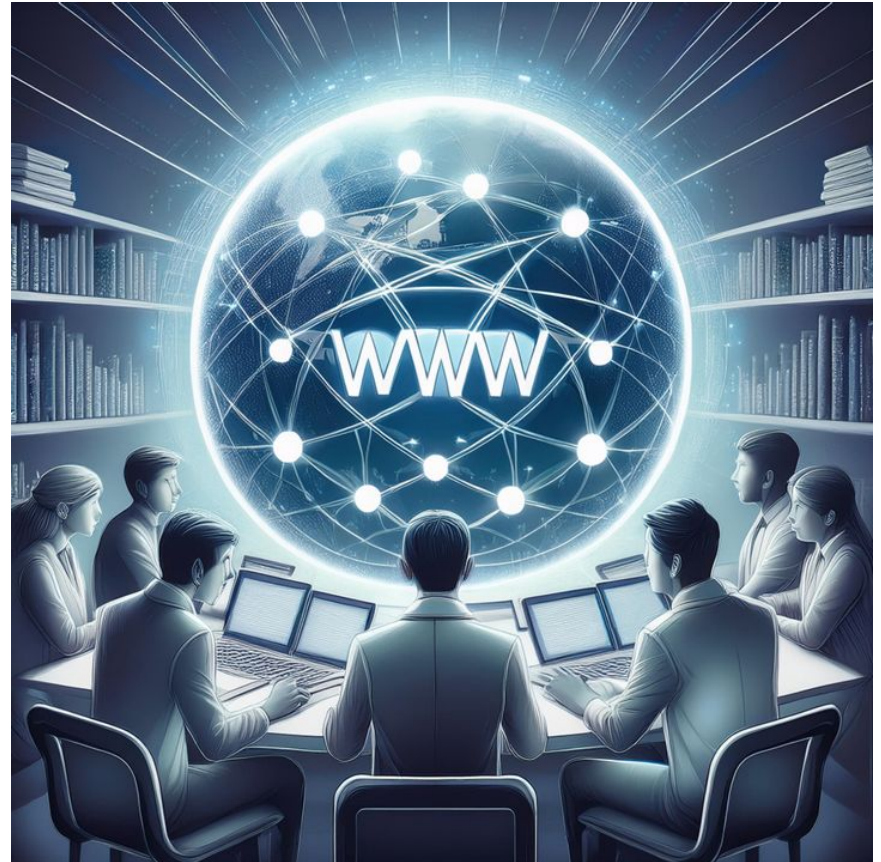


# Adobe Firefly (<https://firefly.adobe.com/>)

**Input:** *reate an image of wikipedians meeting together. Relevance of wikipedia logo. Add programmers, editors and librarians.*

Several styles. I used:

- Art+digital art



# CHATGPT 4o

create an image of wikipedians meeting together



To create an image of Wikipedians meeting together, I'll depict a diverse group of people sitting around a table in a modern conference room. They will have laptops, notebooks, and possibly some posters or whiteboards with Wikipedia logos or articles on them. The scene will convey a collaborative and friendly atmosphere, with individuals engaged in discussion and working together. Let's generate the image.



# CHATGPT 4o

Can we change the setting theme?

Of course! What kind of setting theme would you like for the meeting? We can change it to something more casual, a different type of environment like a café or an outdoor area, or even something more unique like a historical or futuristic setting. Let me know your preference!

In a café

Let's create an image of Wikipedians meeting together in a cozy café setting. The scene will include comfortable seating, a warm atmosphere, and a mix of people working on laptops, enjoying coffee, and engaging in discussion. I'll depict the café with a relaxed vibe, featuring some typical elements like coffee cups, wooden tables, and maybe some bookshelves or art on the walls. Let's generate this new image.

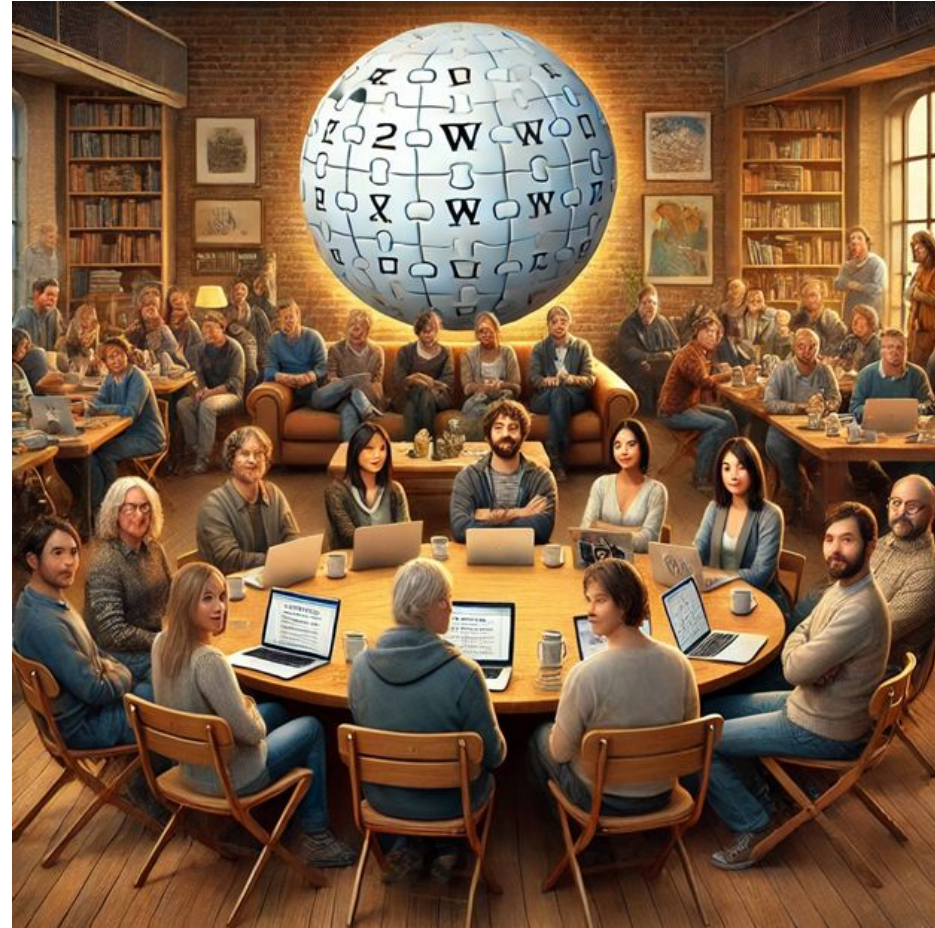




# CHATGPT 4o

Stress more the wikipedia logo

Let's enhance the presence of the Wikipedia logo in the café setting. I'll ensure that the logo is more prominently displayed, perhaps on laptops, notebooks, and even as a decoration within the café, such as on the wall or as a small sign. This will help emphasize the connection to Wikipedia. I'll create the updated image with these adjustments.



# CHATGPT 4o

Create a cartoon with the same style of:

- \* two wikipedians meeting in the street
- \* they greeting each other
- \* they go together
- \* they join a big meeting with other wikipedians



Let's create a four-panel cartoon with the following scenes:

- Two Wikipedians Meeting in the Street:** The first panel will show two characters, representing Wikipedians, encountering each other in a city street. They will have friendly expressions and be recognizable as part of the Wikipedia community, perhaps with logos or badges.
- Greeting Each Other:** The second panel will depict the two Wikipedians greeting each other warmly, perhaps with a handshake or a wave.
- Going Together:** In the third panel, the two Wikipedians will walk together, heading toward a common destination, with a sense of camaraderie.
- Joining a Big Meeting with Other Wikipedians:** The final panel will show the two joining a large group of diverse Wikipedians in a meeting or event setting, highlighting the collaborative spirit of the community.

I'll create this cartoon with a cohesive style and a friendly, approachable tone. Let's go ahead and generate the cartoon.



# CHATGPT 4o

Can we see the characters closer up?

Let's create a close-up image of the two characters from the cartoon, highlighting their features and expressions. This will allow for a better view of their design and details, maintaining the same cartoonish and friendly style. I'll focus on a moment when they are interacting, perhaps during the greeting or while walking together. Let's generate this closer view.



# CHATGPT 4o

Create a new cartoon concept.

Let's create a new cartoon concept! Here's an idea:

## Title: "The Wiki Adventures"

### Panel 1: "The Discovery"

A character stumbles upon an old, dusty book in a library titled "The Book of Knowledge." As they open it, they're surprised to see a glowing Wikipedia logo on the first page. The book's glow indicates something magical and intriguing. ↓



A decorative floral wreath with a green leafy border. Inside the wreath, there are two white birds with blue and orange wings and tails, facing each other. The wreath is adorned with various colorful flowers, including pink, blue, yellow, and red blooms, some with intricate patterns.

# 3: Wikidata

**Andrew**

**WIKIMANIA  
KATOWICE**

# Case: Unstructured to Structured

- ❖ Art history is messy
- ❖ Vague, soft descriptions of dates
  - 1743
  - circa 1820
  - 100 B.C.–200 A.D.
  - early 19th century
  - probably late 17th century
  - 19th Dynasty (Egypt)



**Period**

**305 BCE-395 CE**

**National Museum of Asian Art (US)**  
[https://asia.si.edu/explore-art-culture/collections/search/edanmdm:fsg\\_F1905.237/](https://asia.si.edu/explore-art-culture/collections/search/edanmdm:fsg_F1905.237/)

```
"objectDate": "ca. 1194–1188 B.C.",  
"objectBeginDate": -1194,
```

**Metropolitan Museum of Art API**  
<https://collectionapi.metmuseum.org/public/collection/v1/objects/561265>



# Case: Unstructured to Structured

- ❖ Challenge: convert "loose and lexical" to "structured statements"
- ❖ Typical Wikidata statements
  - P571 - inception
  - P1326 - latest date
  - P1480 - sourcing circumstances
  - Date granularity
    - 18th century - +1701-00-00T00:00:00Z/7
    - 1939 - +1939-00-00T00:00:00Z/9



# Training session

- ❖ Train ChatGPT-4o as a transformer ([link](#))

I want you to act as a helper to transform dates in an English language description to a more precise date notation. The input will be a text string that describes a date, and the output will be a notation for Wikidata Quickstatements. Do you know how to work with Wikidata Quickstatements, and how would I get you to do that by showing you examples of the input descriptions and the output values?



# It seems to know QS

## ❖ Response ([link](#))

Yes, I can help with transforming date descriptions into precise date notations suitable for Wikidata Quickstatements.

Wikidata Quickstatements use a specific format to add or edit data. For dates, the format typically involves specifying the year, month, and day, with optional precision.

### Here's how you can guide me to help:

1. **Provide Example Inputs:** You can show me text descriptions of dates in natural language, such as "early 2023", "mid-July 1990", or "around 1850".
2. **Provide Corresponding Outputs:** Show the precise date notation for Wikidata Quickstatements corresponding to those descriptions, such as ``+2023-01-01T00:00:00Z/11``, ``+1990-07-15T00:00:00Z/10``, or ``+1850-00-00T00:00:00Z/9``.
3. **Ask for Transformations:** Once I have a few examples, I can begin transforming new date descriptions into the precise Quickstatement notation.

# How to proceed

- ❖ Response ([link](#))

## Example:

Input: "mid-July 1990"

Output: ``+1990-07-15T00:00:00Z/10``

Input: "early 2023"

Output: ``+2023-01-01T00:00:00Z/11``

You can continue to provide examples, and I can help generate the correct Wikidata Quickstatements format for different types of date descriptions.

# Let's train!

- ❖ Training data in CSV ([link](#))

Find it in the Etherpad:

[https://etherpad.wikimedia.org/p/WM2024\\_Day4\\_Ochrid\\_-\\_Room\\_9](https://etherpad.wikimedia.org/p/WM2024_Day4_Ochrid_-_Room_9)

also Wikimania wiki:  
2024:Program#Session\_notes


Here is a CSV test showing input and output values:

```
1967,P571|+1967-00-00T00:00:00Z/9
1942,P571|+1942-00-00T00:00:00Z/9
1939,P571|+1939-00-00T00:00:00Z/9
18th century,P571|+1701-00-00T00:00:00Z/7
19th century,P571|+1801-00-00T00:00:00Z/7
20th century,P571|+1901-00-00T00:00:00Z/7
1950s,P571|+1950-00-00T00:00:00Z/8
1920s,P571|+1920-00-00T00:00:00Z/8
1960s,P571|+1960-00-00T00:00:00Z/8
fourth quarter 19th century,P571|+1875-00-
00T00:00:00Z/9|P1326|+1900-00-00T00:00:00Z/9
ca. 1900,P571|+1900-00-
00T00:00:00Z/9|P1480|Q5727902
early 20th century,P571|+1901-00-
00T00:00:00Z/7|P1326|+1949-00-00T00:00:00Z/9
```

# Impressive results

## ❖ Impressive results from ChatGPT 4o

Based on that CSV file, predict the output given the input "probably late 18th century"

 Given the input "probably late 18th century," the predicted output for Wikidata Quickstatements would be:

```
makefile Copy code  
P571 | +1750-00-00T00:00:00Z/9 | P1326 | +1800-00-00T00:00:00Z/9 | P1480 | Q5
```

This follows the pattern where "probably" is indicated by adding `P1480 | Q56644435`, and "late 18th century" is represented with a start date of 1750 and an end date of 1800, both with precision 9.

<https://quickstatements.toolforge.org/>

1	init	<a href="#">Wikidata Sandbox</a> [Q4115189]	<a href="#">ADD</a>	Statement	<a href="#">inception</a> [P571] : 1750	
2	init	<a href="#">Wikidata Sandbox</a> [Q4115189]	<a href="#">ADD</a>	Qualifier to	<a href="#">inception</a> [P571] : 1750	<a href="#">latest date</a> [P1326] : 1800
3	init	<a href="#">Wikidata Sandbox</a> [Q4115189]	<a href="#">ADD</a>	Qualifier to	<a href="#">inception</a> [P571] : 1750	<a href="#">sourcing circumstances</a> [P1480] : <a href="#">probably</a> [Q56644435]

# Challenge

- ❖ How far can you go with other date conversion scenarios?
  - Handle "A.D." as well as "C.E."
- ❖ Variations of these:
  - probably
  - approximately
  - presumably
- ❖ Other date formats? 2024-0808
- ❖ Different languages for month names?
- ❖ Ask GPT to generate code to do the same thing?
  - Python and PAWS?





**Thanks for  
engaging!**

**WIKIMANIA  
KATOWICE**



# Extras

WIKIMANIA  
KATOWICE



# 1: Wikipedia Extras SJ & Shani

WIKIMANIA  
KATOWICE



# SJ's Extra tools to play with

- 



# Shani's Extra tools to play with





# 2: Wiki Commons Extras Ilario

# Ilario's Extra tools to play with



A decorative wreath made of green leaves and various colorful flowers (pink, blue, yellow, red) surrounds the central text. Two stylized white birds with blue and orange accents are perched on the wreath, one on the left and one on the right.

# 3: Wikidata Extras Andrew

WIKIMANIA  
KATOWICE

# Andrew's Extra tools to play with

