```
!pip install -q condacolab
import condacolab
condacolab.install()

⬇ Downloading
https://github.com/jaimergp/miniforge/releases/download/24.11.2-
1_colab/Miniforge3-colab-24.11.2-1_colab-Linux-x86_64.sh...
📦 Installing...
📌 Adjusting configuration...
🩹 Patching environment...
⏲ Done in 0:00:16
🔁 Restarting kernel...

!wget https://cs.stanford.edu/~minalee/zip/chi2022-coauthor-v1.0.zip
!unzip -q chi2022-coauthor-v1.0.zip
!rm chi2022-coauthor-v1.0.zip
!unzip -q creativity_index-main.zip

--2025-02-26 18:17:12--  https://cs.stanford.edu/~minalee/zip/chi2022-
coauthor-v1.0.zip
Resolving cs.stanford.edu (cs.stanford.edu)... 171.64.64.64
Connecting to cs.stanford.edu (cs.stanford.edu)|171.64.64.64|:443...
connected.
HTTP request sent, awaiting response... 200 OK
Length: 49956179 (48M) [application/zip]
Saving to: 'chi2022-coauthor-v1.0.zip'

chi2022-coauthor-v1 100%[===================>]  47.64M  8.81MB/s    in
5.7s

2025-02-26 18:17:18 (8.34 MB/s) - 'chi2022-coauthor-v1.0.zip' saved
[49956179/49956179]

unzip:  cannot find or open creativity_index-main.zip,
creativity_index-main.zip.zip or creativity_index-main.zip.ZIP.
```

Coauthor cleanup

```
import os

dataset_dir = './coauthor-v1.0'
paths = [
    os.path.join(dataset_dir, path)
    for path in os.listdir(dataset_dir)
    if path.endswith('jsonl')
]
```

```python
print(f'Successfully downloaded {len(paths)} writing sessions in
CoAuthor!')
```

Successfully downloaded 1447 writing sessions in CoAuthor!

```python
import json

def read_writing_session(path):
    events = []
    with open(path, 'r') as f:
        for event in f:
            events.append(json.loads(event))
    print(f'Successfully read {len(events)} events in a writing
session from {path}')
    return events

events = read_writing_session(paths[0])
```

Successfully read 2405 events in a writing session from ./coauthor-
v1.0/608e5b73341a4f3ca937316f99dae32d.jsonl

```python
from collections import defaultdict

def reconstruct_user_text(events):
    user_texts = []
    current_text = ""

    # Sort events by eventNum
    sorted_events = sorted(events, key=lambda e: e["eventNum"])

    for event in sorted_events:
        event_type = event["eventName"]
        text_delta = event["textDelta"]

        if event_type == "text-insert" and text_delta:
            extracted_text = ""
            for op in text_delta['ops']:
                if 'insert' in op:
                    extracted_text += op['insert']
            current_text += extracted_text

        elif event_type == "text-delete" and text_delta:
            delete_len = 0
            for op in text_delta['ops']:
                if 'delete' in op:
                    delete_len += op['delete']
                elif 'retain' in op:
                    pass
            current_text = current_text[:-delete_len]
```

```python
        # Store finalized text if user completes a thought
        if event_type == "suggestion-get" or event_type ==
"suggestion-select":
            if current_text.strip():
                user_texts.append(current_text.strip())
                current_text = ""

    return user_texts

# Process all valid events
user_texts = reconstruct_user_text(events)

print(f"Total reconstructed user-written sentences:
{len(user_texts)}")
print("Example user-written sentence:", user_texts)  # Show first 3
examples
```

Total reconstructed user-written sentences: 5
Example user-written sentence: ['it is the responsibility of a good
citizen to keep up with the news.', "I think it is important to know
what is going on in the world because it can affect us. For example,
there are many pieces of legislation that are passed that can have bad
current and future affects on the population. Due to people not
keeping up with the news, the public isn't aware of the effects of the
legislation until it is too late. There are many instance where the
public iecomes aware of a bad bill and they are able to apply pressure
to their government representatives to make changes.", 'before it is
too late\n\nKeeping up with the news also allows the public to stay
informed about issues such as climate change.', 'We are far behind
when it comes to dealing with climate change. I believe we would be
mounting a better challenge to climate change if more people were
knowledgeable about the effects. This knowledge could be garnered
from the news. When you have areas of the world that are experiencing
historical high water levels due to climate change, there are often
news reports being done on these subjects. If the public is not
paying attention, youhe public will probably not put pressure on their
government representatives to act.', 'I believe that people who go
through life oblivious to the problems that surround them are not good
people. The news often focuses on problems that may not be a problem
for certain individuals--yet.']

```python
def extract_selected_ai_suggestions(events):
    ai_suggestions = []  # Stores accepted AI-generated suggestions
    last_suggestion_open = None  # Store the most recent `suggestion-
open` event

    for event in events:
        if event["eventName"] == "suggestion-open" and
event.get("currentSuggestions"):
            last_suggestion_open = event  # Save latest `suggestion-
```

```
open` event

        elif event["eventName"] == "suggestion-select":
            if last_suggestion_open and
last_suggestion_open.get("currentSuggestions"):
                selected_index = event.get("currentSuggestionIndex", -
1)

                # Ensure selected index is valid
                if 0 <= selected_index <
len(last_suggestion_open["currentSuggestions"]):
                    selected_suggestion =
last_suggestion_open["currentSuggestions"][selected_index]["trimmed"]
                    ai_suggestions.append(selected_suggestion)

    return ai_suggestions

# Process all valid events
ai_suggestions = extract_selected_ai_suggestions(events)

print(f"Total AI Suggestions Accepted: {len(ai_suggestions)}")
print("Example AI Suggestion:", ai_suggestions[:3])  # Show first 3
examples
```

```
Total AI Suggestions Accepted: 2
Example AI Suggestion: ['I think it is important to know what is going
on in the world because it can affect us.', 'For example, the Stop
Online Piracy Act (SOPA) was a bill that would have allowed the
government to shut down websites that were suspected']
```

```
def extract_acceptance_status(events):
    """Track if a suggestion was accepted or rejected."""
    acceptance_status = []  # Stores 'accepted' or 'rejected'
    last_suggestion_open = None  # Store the most recent `suggestion-
open`

    for event in events:
        if event["eventName"] == "suggestion-open" and
event.get("currentSuggestions"):
            last_suggestion_open = event  # Save the latest
`suggestion-open`

        elif event["eventName"] == "suggestion-select" and
last_suggestion_open:
            # Suggestion selected
            acceptance_status.append('accepted')
            last_suggestion_open = None  # Reset, since we handled the
acceptance

        elif event["eventName"] == "suggestion-close" and
```

```python
last_suggestion_open:
            # Suggestion closed without selection (rejected)
            acceptance_status.append('rejected')
            last_suggestion_open = None  # Reset, since we handled the
rejection

    return acceptance_status

import pandas as pd

jsd_results = []
small = paths[100:110]

def extract_acceptance_status(events):
    """Track if a suggestion was accepted or rejected."""
    acceptance_status = []  # Stores 'accepted' or 'rejected'
    last_suggestion_open = None  # Store the most recent `suggestion-
open`

    for event in events:
        if event["eventName"] == "suggestion-open" and
event.get("currentSuggestions"):
            last_suggestion_open = event  # Save the latest
`suggestion-open`

        elif event["eventName"] == "suggestion-select" and
last_suggestion_open:
            # Suggestion selected
            acceptance_status.append('accepted')
            last_suggestion_open = None  # Reset, since we handled the
acceptance

        elif event["eventName"] == "suggestion-close" and
last_suggestion_open:
            # Suggestion closed without selection (rejected)
            acceptance_status.append('rejected')
            last_suggestion_open = None  # Reset, since we handled the
rejection

    return acceptance_status

!conda env create -f environment_infini.yml
!conda env create -f environment_vllm.yml

Channels:
 - conda-forge
 - defaults
Platform: linux-64
Collecting package metadata (repodata.json): - \ | / - \ | / - \ | / -
\ | / - \ | / - \ | / - \ | / - \ | / - \ | / - \ | / - \ | /
done
```

```
Solving environment: \ | / failed

PackagesNotFoundError: The following packages are not available from
current channels:

  - xz==5.2.6=h57fd34a_0
  - wasabi==1.1.2=py38h10201cd_0
  - tk==8.6.13=h5083fa2_1
  - thinc==8.2.2=py38h24577cc_0
  - srsly==2.4.8=py38he333c0f_1
  - sqlite==3.45.1=hf2abe2d_0
  - spacy==3.7.2=py38h2105574_0
  - sentencepiece-spm==0.1.96=h09ea954_4
  - sentencepiece-python==0.1.96=py38h36d6a8b_4
  - sentencepiece==0.1.96=h150bfb4_4
  - readline==8.2=h92ec313_1
  - python==3.8.12=hd949e87_3_cpython
  - pydantic-core==2.16.2=py38h51b07bf_1
  - preshed==3.0.9=py38he333c0f_1
  - openssl==3.2.1=h0d3ecfb_0
  - numpy==1.24.4=py38ha84db1f_0
  - ncurses==6.4=h463b476_2
  - murmurhash==1.0.10=py38he333c0f_1
  - markupsafe==2.1.5=py38h336bac9_0
  - llvm-openmp==17.0.6=hcd81f8e_0
  - libzlib==1.2.13=h53f4e23_5
  - libsqlite==3.45.1=h091b4b1_0
  - libsentencepiece==0.1.96=h09ea954_4
  - libprotobuf==3.21.12=ha614eb4_2
  - libopenblas==0.3.26=openmp_h6c19121_0
  - liblapack==3.9.0=21_osxarm64_openblas
  - libgfortran5==13.2.0=hf226fd6_3
  - libgfortran==5.0.0=13_2_0_hd922786_3
  - libffi==3.4.2=h3422bc3_5
  - libcxx==16.0.6=h4653b0c_0
  - libcblas==3.9.0=21_osxarm64_openblas
  - libblas==3.9.0=21_osxarm64_openblas
  - libabseil==20220623.0=cxx17_h28b99d4_6
  - cython-blis==0.7.10=py38h5912bca_2
  - cymem==2.0.8=py38he333c0f_1
  - confection==0.1.4=py38h472928c_0
  - catalogue==2.0.10=py38h10201cd_0
  - ca-certificates==2024.2.2=hf0a4a13_0
  - brotli-python==1.1.0=py38he333c0f_1

Current channels:

  - https://conda.anaconda.org/conda-forge
  - https://repo.anaconda.com/pkgs/main
  - https://repo.anaconda.com/pkgs/r
```

To search for alternate channels that may provide the conda package you're
looking for, navigate to

    https://anaconda.org

and use the search bar at the top of the page.

Channels:
 - defaults
 - conda-forge
Platform: linux-64
Collecting package metadata (repodata.json): - \ | / - \ done
Solving environment: / - done

==> WARNING: A newer version of conda exists. <==
    current version: 24.11.2
    latest version: 25.1.1

Please update conda by running

    $ conda update -n base -c conda-forge conda

Downloading and Extracting Packages:
python-3.10.0 | 23.5 MB | :   0% 0/1 [00:00<?, ?it/s]
libgcc-ng-11.2.0 | 5.3 MB | :   0% 0/1 [00:00<?,
?it/s]pl_linux-64-2.4 | 710 KB | :   0% 0/1 [00:00<?, ?it/s]p-
11.2.0 | 474 KB | :   0% 0/1 [00:00<?, ?it/s]pl_linux-64-2.4
| 710 KB | :   2% 0.022543831405604585/1 [00:00<00:19, 19.50s/it]p-
11.2.0 | 474 KB | :   3% 0.03377134671077719/1 [00:00<00:13,
13.77s/it]pl_linux-64-2.4 | 710 KB | : 100% 1.0/1 [00:00<00:00,
19.50s/it]   p-11.2.0 | 474 KB | : 100% 1.0/1
[00:00<00:00, 13.77s/it]   ore hidden) ...pl_linux-64-2.4
| 710 KB | : 100% 1.0/1 [00:02<00:00, 2.01s/it]pl_linux-64-2.4 |
710 KB | : 100% 1.0/1 [00:02<00:00, 2.01s/it]p-11.2.0 | 474
KB | : 100% 1.0/1 [00:02<00:00, 2.08s/it]p-11.2.0 | 474 KB
| : 100% 1.0/1 [00:02<00:00, 2.08s/it]ore hidden) ...ore
hidden) ...ents:
['/usr/local/envs/vllm/bin/python', '-m', 'pip', 'install', '-U', '-
r', '/content/condaenv.v2x4sx6t.requirements.txt', '--exists-
action=b']
Pip subprocess output:
Collecting accelerate==1.0.1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 1))
  Downloading accelerate-1.0.1-py3-none-any.whl.metadata (19 kB)

```
Collecting aiohappyeyeballs==2.4.3 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 2))
  Downloading aiohappyeyeballs-2.4.3-py3-none-any.whl.metadata (6.1
kB)
Collecting aiohttp==3.10.10 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 3))
  Downloading aiohttp-3.10.10-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (7.6 kB)
Collecting aiosignal==1.3.1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 4))
  Downloading aiosignal-1.3.1-py3-none-any.whl.metadata (4.0 kB)
Collecting alembic==1.14.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 5))
  Downloading alembic-1.14.0-py3-none-any.whl.metadata (7.4 kB)
Collecting annotated-types==0.7.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 6))
  Downloading annotated_types-0.7.0-py3-none-any.whl.metadata (15 kB)
Collecting anyio==4.6.2.post1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 7))
  Downloading anyio-4.6.2.post1-py3-none-any.whl.metadata (4.7 kB)
Collecting asttokens==2.4.1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 8))
  Downloading asttokens-2.4.1-py2.py3-none-any.whl.metadata (5.2 kB)
Collecting async-timeout==4.0.3 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 9))
  Downloading async_timeout-4.0.3-py3-none-any.whl.metadata (4.2 kB)
Collecting attrs==24.2.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 10))
  Downloading attrs-24.2.0-py3-none-any.whl.metadata (11 kB)
Collecting autofaiss==2.17.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 11))
  Downloading autofaiss-2.17.0-py3-none-any.whl.metadata (15 kB)
Collecting beautifulsoup4==4.12.3 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 12))
  Downloading beautifulsoup4-4.12.3-py3-none-any.whl.metadata (3.8 kB)
Collecting cbor==1.0.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 13))
  Downloading cbor-1.0.0.tar.gz (20 kB)
  Preparing metadata (setup.py): started
  Preparing metadata (setup.py): finished with status 'done'
Collecting cbor2==5.6.5 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 14))
  Downloading cbor2-5.6.5-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (6.0 kB)
Collecting certifi==2024.8.30 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 15))
  Downloading certifi-2024.8.30-py3-none-any.whl.metadata (2.2 kB)
Collecting charset-normalizer==3.4.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 16))
```

```
  Downloading charset_normalizer-3.4.0-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (34 kB)
Collecting click==8.1.7 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 17))
  Downloading click-8.1.7-py3-none-any.whl.metadata (3.0 kB)
Collecting cloudpickle==3.1.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 18))
  Downloading cloudpickle-3.1.0-py3-none-any.whl.metadata (7.0 kB)
Collecting colorlog==6.9.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 19))
  Downloading colorlog-6.9.0-py3-none-any.whl.metadata (10 kB)
Collecting compressed-tensors==0.6.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 20))
  Downloading compressed_tensors-0.6.0-py3-none-any.whl.metadata (6.8
kB)
Collecting contourpy==1.3.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 21))
  Downloading contourpy-1.3.0-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (5.4 kB)
Collecting cramjam==2.9.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 22))
  Downloading cramjam-2.9.0-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (4.9 kB)
Collecting cycler==0.12.1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 23))
  Downloading cycler-0.12.1-py3-none-any.whl.metadata (3.8 kB)
Collecting datasets==3.0.2 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 24))
  Downloading datasets-3.0.2-py3-none-any.whl.metadata (20 kB)
Collecting decorator==5.1.1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 25))
  Downloading decorator-5.1.1-py3-none-any.whl.metadata (4.0 kB)
Collecting dill==0.3.8 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 26))
  Downloading dill-0.3.8-py3-none-any.whl.metadata (10 kB)
Collecting diskcache==5.6.3 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 27))
  Downloading diskcache-5.6.3-py3-none-any.whl.metadata (20 kB)
Collecting distro==1.9.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 28))
  Downloading distro-1.9.0-py3-none-any.whl.metadata (6.8 kB)
Collecting einops==0.8.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 29))
  Downloading einops-0.8.0-py3-none-any.whl.metadata (12 kB)
Collecting embedding-reader==1.7.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 30))
  Downloading embedding_reader-1.7.0-py3-none-any.whl.metadata (8.5
kB)
Collecting exceptiongroup==1.2.2 (from -r
```

```
/content/condaenv.v2x4sx6t.requirements.txt (line 31))
  Downloading exceptiongroup-1.2.2-py3-none-any.whl.metadata (6.6 kB)
Collecting executing==2.1.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 32))
  Downloading executing-2.1.0-py2.py3-none-any.whl.metadata (8.9 kB)
Collecting faiss-cpu==1.9.0.post1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 33))
  Downloading faiss_cpu-1.9.0.post1-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (4.4 kB)
Collecting fastapi==0.115.3 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 34))
  Downloading fastapi-0.115.3-py3-none-any.whl.metadata (27 kB)
Collecting fastparquet==2024.11.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 35))
  Downloading fastparquet-2024.11.0-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (4.2 kB)
Collecting filelock==3.16.1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 36))
  Downloading filelock-3.16.1-py3-none-any.whl.metadata (2.9 kB)
Collecting fire==0.5.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 37))
  Downloading fire-0.5.0.tar.gz (88 kB)
  Preparing metadata (setup.py): started
  Preparing metadata (setup.py): finished with status 'done'
Collecting fonttools==4.54.1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 38))
  Downloading fonttools-4.54.1-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (163 kB)
Collecting frozenlist==1.4.1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 39))
  Downloading frozenlist-1.4.1-cp310-cp310-
manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_17_x86_64.manylinux
2014_x86_64.whl.metadata (12 kB)
Collecting fsspec==2024.9.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 40))
  Downloading fsspec-2024.9.0-py3-none-any.whl.metadata (11 kB)
Collecting gguf==0.10.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 41))
  Downloading gguf-0.10.0-py3-none-any.whl.metadata (3.5 kB)
Collecting greenlet==3.1.1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 42))
  Downloading greenlet-3.1.1-cp310-cp310-
manylinux_2_24_x86_64.manylinux_2_28_x86_64.whl.metadata (3.8 kB)
Collecting h11==0.14.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 43))
  Downloading h11-0.14.0-py3-none-any.whl.metadata (8.2 kB)
Collecting httpcore==1.0.6 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 44))
  Downloading httpcore-1.0.6-py3-none-any.whl.metadata (21 kB)
```

```
Collecting httptools==0.6.4 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 45))
  Downloading httptools-0.6.4-cp310-cp310-
manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_17_x86_64.manylinux
2014_x86_64.whl.metadata (3.6 kB)
Collecting httpx==0.27.2 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 46))
  Downloading httpx-0.27.2-py3-none-any.whl.metadata (7.1 kB)
Collecting huggingface-hub==0.26.1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 47))
  Downloading huggingface_hub-0.26.1-py3-none-any.whl.metadata (13 kB)
Collecting idna==3.10 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 48))
  Downloading idna-3.10-py3-none-any.whl.metadata (10 kB)
Collecting ijson==3.3.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 49))
  Downloading ijson-3.3.0-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (21 kB)
Collecting importlib-metadata==8.5.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 50))
  Downloading importlib_metadata-8.5.0-py3-none-any.whl.metadata (4.8
kB)
Collecting indxr==0.1.5 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 51))
  Downloading indxr-0.1.5-py3-none-any.whl.metadata (8.6 kB)
Collecting inscriptis==2.5.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 52))
  Downloading inscriptis-2.5.0-py3-none-any.whl.metadata (25 kB)
Collecting interegular==0.3.3 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 53))
  Downloading interegular-0.3.3-py37-none-any.whl.metadata (3.0 kB)
Collecting ipython==8.28.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 54))
  Downloading ipython-8.28.0-py3-none-any.whl.metadata (5.0 kB)
Collecting ir-datasets==0.5.9 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 55))
  Downloading ir_datasets-0.5.9-py3-none-any.whl.metadata (12 kB)
Collecting jedi==0.19.1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 56))
  Downloading jedi-0.19.1-py2.py3-none-any.whl.metadata (22 kB)
Collecting jinja2==3.1.4 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 57))
  Downloading jinja2-3.1.4-py3-none-any.whl.metadata (2.6 kB)
Collecting jiter==0.6.1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 58))
  Downloading jiter-0.6.1-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (5.2 kB)
Collecting joblib==1.4.2 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 59))
```

```
  Downloading joblib-1.4.2-py3-none-any.whl.metadata (5.4 kB)
Collecting jsonschema==4.23.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 60))
  Downloading jsonschema-4.23.0-py3-none-any.whl.metadata (7.9 kB)
Collecting jsonschema-specifications==2024.10.1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 61))
  Downloading jsonschema_specifications-2024.10.1-py3-none-
any.whl.metadata (3.0 kB)
Collecting kiwisolver==1.4.7 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 62))
  Downloading kiwisolver-1.4.7-cp310-cp310-
manylinux_2_12_x86_64.manylinux2010_x86_64.whl.metadata (6.3 kB)
Collecting krovetzstemmer==0.8 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 63))
  Downloading KrovetzStemmer-0.8.tar.gz (112 kB)
  Preparing metadata (setup.py): started
  Preparing metadata (setup.py): finished with status 'done'
Collecting lark==1.2.2 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 64))
  Downloading lark-1.2.2-py3-none-any.whl.metadata (1.8 kB)
Collecting llvmlite==0.43.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 65))
  Downloading llvmlite-0.43.0-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (4.8 kB)
Collecting lm-format-enforcer==0.10.6 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 66))
  Downloading lm_format_enforcer-0.10.6-py3-none-any.whl.metadata (16
kB)
Collecting lxml==5.3.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 67))
  Downloading lxml-5.3.0-cp310-cp310-
manylinux_2_28_x86_64.whl.metadata (3.8 kB)
Collecting lz4==4.3.3 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 68))
  Downloading lz4-4.3.3-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (3.7 kB)
Collecting mako==1.3.6 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 69))
  Downloading Mako-1.3.6-py3-none-any.whl.metadata (2.9 kB)
Collecting markdown-it-py==3.0.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 70))
  Downloading markdown_it_py-3.0.0-py3-none-any.whl.metadata (6.9 kB)
Collecting markupsafe==3.0.2 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 71))
  Downloading MarkupSafe-3.0.2-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (4.0 kB)
Collecting matplotlib==3.9.2 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 72))
  Downloading matplotlib-3.9.2-cp310-cp310-
```

```
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (11 kB)
Collecting matplotlib-inline==0.1.7 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 73))
  Downloading matplotlib_inline-0.1.7-py3-none-any.whl.metadata (3.9
kB)
Collecting mdurl==0.1.2 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 74))
  Downloading mdurl-0.1.2-py3-none-any.whl.metadata (1.6 kB)
Collecting mistral-common==1.4.4 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 75))
  Downloading mistral_common-1.4.4-py3-none-any.whl.metadata (4.6 kB)
Collecting mpmath==1.3.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 76))
  Downloading mpmath-1.3.0-py3-none-any.whl.metadata (8.6 kB)
Collecting msgpack==1.1.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 77))
  Downloading msgpack-1.1.0-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (8.4 kB)
Collecting msgspec==0.18.6 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 78))
  Downloading msgspec-0.18.6-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (6.9 kB)
Collecting multidict==6.1.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 79))
  Downloading multidict-6.1.0-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (5.0 kB)
Collecting multipipe==0.1.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 80))
  Downloading multipipe-0.1.0-py3-none-any.whl.metadata (4.0 kB)
Collecting multiprocess==0.70.16 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 81))
  Downloading multiprocess-0.70.16-py310-none-any.whl.metadata (7.2
kB)
Collecting nest-asyncio==1.6.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 82))
  Downloading nest_asyncio-1.6.0-py3-none-any.whl.metadata (2.8 kB)
Collecting networkx==3.4.2 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 83))
  Downloading networkx-3.4.2-py3-none-any.whl.metadata (6.3 kB)
Collecting nltk==3.9.1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 84))
  Downloading nltk-3.9.1-py3-none-any.whl.metadata (2.9 kB)
Collecting numba==0.60.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 85))
  Downloading numba-0.60.0-cp310-cp310-
manylinux2014_x86_64.manylinux_2_17_x86_64.whl.metadata (2.7 kB)
Collecting numpy==1.26.4 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 86))
  Downloading numpy-1.26.4-cp310-cp310-
```

```
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (61 kB)
Collecting nvidia-cublas-cu12==12.1.3.1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 87))
  Downloading nvidia_cublas_cu12-12.1.3.1-py3-none-
manylinux1_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cuda-cupti-cu12==12.1.105 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 88))
  Downloading nvidia_cuda_cupti_cu12-12.1.105-py3-none-
manylinux1_x86_64.whl.metadata (1.6 kB)
Collecting nvidia-cuda-nvrtc-cu12==12.1.105 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 89))
  Downloading nvidia_cuda_nvrtc_cu12-12.1.105-py3-none-
manylinux1_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cuda-runtime-cu12==12.1.105 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 90))
  Downloading nvidia_cuda_runtime_cu12-12.1.105-py3-none-
manylinux1_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cudnn-cu12==9.1.0.70 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 91))
  Downloading nvidia_cudnn_cu12-9.1.0.70-py3-none-
manylinux2014_x86_64.whl.metadata (1.6 kB)
Collecting nvidia-cufft-cu12==11.0.2.54 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 92))
  Downloading nvidia_cufft_cu12-11.0.2.54-py3-none-
manylinux1_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-curand-cu12==10.3.2.106 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 93))
  Downloading nvidia_curand_cu12-10.3.2.106-py3-none-
manylinux1_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cusolver-cu12==11.4.5.107 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 94))
  Downloading nvidia_cusolver_cu12-11.4.5.107-py3-none-
manylinux1_x86_64.whl.metadata (1.6 kB)
Collecting nvidia-cusparse-cu12==12.1.0.106 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 95))
  Downloading nvidia_cusparse_cu12-12.1.0.106-py3-none-
manylinux1_x86_64.whl.metadata (1.6 kB)
Collecting nvidia-ml-py==12.560.30 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 96))
  Downloading nvidia_ml_py-12.560.30-py3-none-any.whl.metadata (8.6
kB)
Collecting nvidia-nccl-cu12==2.20.5 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 97))
  Downloading nvidia_nccl_cu12-2.20.5-py3-none-
manylinux2014_x86_64.whl.metadata (1.8 kB)
Collecting nvidia-nvjitlink-cu12==12.4.127 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 98))
  Downloading nvidia_nvjitlink_cu12-12.4.127-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
```

```
Collecting nvidia-nvtx-cu12==12.1.105 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 99))
  Downloading nvidia_nvtx_cu12-12.1.105-py3-none-
manylinux1_x86_64.whl.metadata (1.7 kB)
Collecting oneliner-utils==0.1.2 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 100))
  Downloading oneliner_utils-0.1.2-py3-none-any.whl.metadata (656
bytes)
Collecting openai==1.55.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 101))
  Downloading openai-1.55.0-py3-none-any.whl.metadata (24 kB)
Collecting opencv-python-headless==4.10.0.84 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 102))
  Downloading opencv_python_headless-4.10.0.84-cp37-abi3-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (20 kB)
Collecting optuna==4.1.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 103))
  Downloading optuna-4.1.0-py3-none-any.whl.metadata (16 kB)
Collecting orjson==3.10.11 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 104))
  Downloading orjson-3.10.11-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (50 kB)
Collecting outlines==0.0.46 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 105))
  Downloading outlines-0.0.46-py3-none-any.whl.metadata (15 kB)
Collecting packaging==24.1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 106))
  Downloading packaging-24.1-py3-none-any.whl.metadata (3.2 kB)
Collecting pandas==2.2.3 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 107))
  Downloading pandas-2.2.3-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (89 kB)
Collecting parso==0.8.4 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 108))
  Downloading parso-0.8.4-py2.py3-none-any.whl.metadata (7.7 kB)
Collecting partial-json-parser==0.2.1.1.post4 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 109))
  Downloading partial_json_parser-0.2.1.1.post4-py3-none-
any.whl.metadata (6.2 kB)
Collecting pexpect==4.9.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 110))
  Downloading pexpect-4.9.0-py2.py3-none-any.whl.metadata (2.5 kB)
Collecting pillow==10.4.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 111))
  Downloading pillow-10.4.0-cp310-cp310-
manylinux_2_28_x86_64.whl.metadata (9.2 kB)
Collecting prometheus-client==0.21.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 112))
  Downloading prometheus_client-0.21.0-py3-none-any.whl.metadata (1.8
kB)
```

```
Collecting prometheus-fastapi-instrumentator==7.0.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 113))
  Downloading prometheus_fastapi_instrumentator-7.0.0-py3-none-
any.whl.metadata (13 kB)
Collecting prompt-toolkit==3.0.48 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 114))
  Downloading prompt_toolkit-3.0.48-py3-none-any.whl.metadata (6.4 kB)
Collecting propcache==0.2.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 115))
  Downloading propcache-0.2.0-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (7.7 kB)
Collecting protobuf==5.28.2 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 116))
  Downloading protobuf-5.28.2-cp38-abi3-
manylinux2014_x86_64.whl.metadata (592 bytes)
Collecting psutil==6.1.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 117))
  Downloading psutil-6.1.0-cp36-abi3-
manylinux_2_12_x86_64.manylinux2010_x86_64.manylinux_2_17_x86_64.manyl
inux2014_x86_64.whl.metadata (22 kB)
Collecting ptyprocess==0.7.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 118))
  Downloading ptyprocess-0.7.0-py2.py3-none-any.whl.metadata (1.3 kB)
Collecting pure-eval==0.2.3 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 119))
  Downloading pure_eval-0.2.3-py3-none-any.whl.metadata (6.3 kB)
Collecting py-cpuinfo==9.0.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 120))
  Downloading py_cpuinfo-9.0.0-py3-none-any.whl.metadata (794 bytes)
Collecting pyairports==2.1.1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 121))
  Downloading pyairports-2.1.1-py3-none-any.whl.metadata (1.7 kB)
Collecting pyarrow==15.0.2 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 122))
  Downloading pyarrow-15.0.2-cp310-cp310-
manylinux_2_28_x86_64.whl.metadata (3.0 kB)
Collecting pycountry==24.6.1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 123))
  Downloading pycountry-24.6.1-py3-none-any.whl.metadata (12 kB)
Collecting pydantic==2.9.2 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 124))
  Downloading pydantic-2.9.2-py3-none-any.whl.metadata (149 kB)
Collecting pydantic-core==2.23.4 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 125))
  Downloading pydantic_core-2.23.4-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (6.6 kB)
Collecting pygments==2.18.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 126))
  Downloading pygments-2.18.0-py3-none-any.whl.metadata (2.5 kB)
```

```
Collecting pyparsing==3.2.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 127))
  Downloading pyparsing-3.2.0-py3-none-any.whl.metadata (5.0 kB)
Collecting pystemmer==2.2.0.3 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 128))
  Downloading PyStemmer-2.2.0.3-cp310-cp310-
manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_17_x86_64.manylinux
2014_x86_64.whl.metadata (3.2 kB)
Collecting python-dateutil==2.9.0.post0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 129))
  Downloading python_dateutil-2.9.0.post0-py2.py3-none-
any.whl.metadata (8.4 kB)
Collecting python-dotenv==1.0.1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 130))
  Downloading python_dotenv-1.0.1-py3-none-any.whl.metadata (23 kB)
Collecting pytz==2024.2 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 131))
  Downloading pytz-2024.2-py2.py3-none-any.whl.metadata (22 kB)
Collecting pyyaml==6.0.2 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 132))
  Downloading PyYAML-6.0.2-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (2.1 kB)
Collecting pyzmq==26.2.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 133))
  Downloading pyzmq-26.2.0-cp310-cp310-
manylinux_2_28_x86_64.whl.metadata (6.2 kB)
Collecting ranx==0.3.20 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 134))
  Downloading ranx-0.3.20-py3-none-any.whl.metadata (17 kB)
Collecting ray==2.37.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 135))
  Downloading ray-2.37.0-cp310-cp310-manylinux2014_x86_64.whl.metadata
(16 kB)
Collecting referencing==0.35.1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 136))
  Downloading referencing-0.35.1-py3-none-any.whl.metadata (2.8 kB)
Collecting regex==2024.9.11 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 137))
  Downloading regex-2024.9.11-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (40 kB)
Collecting requests==2.32.3 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 138))
  Downloading requests-2.32.3-py3-none-any.whl.metadata (4.6 kB)
Collecting retriv==0.2.3 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 139))
  Downloading retriv-0.2.3-py3-none-any.whl.metadata (10 kB)
Collecting rich==13.9.4 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 140))
  Downloading rich-13.9.4-py3-none-any.whl.metadata (18 kB)
```

```
Collecting rpds-py==0.20.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 141))
  Downloading rpds_py-0.20.0-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (4.2 kB)
Collecting safetensors==0.4.5 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 142))
  Downloading safetensors-0.4.5-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (3.8 kB)
Collecting scikit-learn==1.5.2 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 143))
  Downloading scikit_learn-1.5.2-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (13 kB)
Collecting scipy==1.14.1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 144))
  Downloading scipy-1.14.1-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (60 kB)
Collecting seaborn==0.13.2 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 145))
  Downloading seaborn-0.13.2-py3-none-any.whl.metadata (5.4 kB)
Collecting sentencepiece==0.2.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 146))
  Downloading sentencepiece-0.2.0-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (7.7 kB)
Collecting six==1.16.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 147))
  Downloading six-1.16.0-py2.py3-none-any.whl.metadata (1.8 kB)
Collecting sniffio==1.3.1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 148))
  Downloading sniffio-1.3.1-py3-none-any.whl.metadata (3.9 kB)
Collecting soupsieve==2.6 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 149))
  Downloading soupsieve-2.6-py3-none-any.whl.metadata (4.6 kB)
Collecting sqlalchemy==2.0.36 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 150))
  Downloading SQLAlchemy-2.0.36-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (9.7 kB)
Collecting stack-data==0.6.3 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 151))
  Downloading stack_data-0.6.3-py3-none-any.whl.metadata (18 kB)
Collecting starlette==0.41.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 152))
  Downloading starlette-0.41.0-py3-none-any.whl.metadata (6.0 kB)
Collecting sympy==1.13.1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 153))
  Downloading sympy-1.13.1-py3-none-any.whl.metadata (12 kB)
Collecting tabulate==0.9.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 154))
  Downloading tabulate-0.9.0-py3-none-any.whl.metadata (34 kB)
Collecting termcolor==2.5.0 (from -r
```

```
/content/condaenv.v2x4sx6t.requirements.txt (line 155))
  Downloading termcolor-2.5.0-py3-none-any.whl.metadata (6.1 kB)
Collecting threadpoolctl==3.5.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 156))
  Downloading threadpoolctl-3.5.0-py3-none-any.whl.metadata (13 kB)
Collecting tiktoken==0.7.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 157))
  Downloading tiktoken-0.7.0-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (6.6 kB)
Collecting tokenizers==0.20.1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 158))
  Downloading tokenizers-0.20.1-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (6.7 kB)
Collecting torch==2.4.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 159))
  Downloading torch-2.4.0-cp310-cp310-manylinux1_x86_64.whl.metadata
(26 kB)
Collecting torchaudio==2.5.1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 160))
  Downloading torchaudio-2.5.1-cp310-cp310-
manylinux1_x86_64.whl.metadata (6.4 kB)
Collecting torchvision==0.19.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 161))
  Downloading torchvision-0.19.0-cp310-cp310-
manylinux1_x86_64.whl.metadata (6.0 kB)
Collecting tqdm==4.66.5 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 162))
  Downloading tqdm-4.66.5-py3-none-any.whl.metadata (57 kB)
Collecting traitlets==5.14.3 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 163))
  Downloading traitlets-5.14.3-py3-none-any.whl.metadata (10 kB)
Collecting transformers==4.45.2 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 164))
  Downloading transformers-4.45.2-py3-none-any.whl.metadata (44 kB)
Collecting trec-car-tools==2.6 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 165))
  Downloading trec_car_tools-2.6-py3-none-any.whl.metadata (640 bytes)
Collecting triton==3.0.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 166))
  Downloading triton-3.0.0-1-cp310-cp310-
manylinux2014_x86_64.manylinux_2_17_x86_64.whl.metadata (1.3 kB)
Collecting typing-extensions==4.12.2 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 167))
  Downloading typing_extensions-4.12.2-py3-none-any.whl.metadata (3.0
kB)
Collecting tzdata==2024.2 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 168))
  Downloading tzdata-2024.2-py2.py3-none-any.whl.metadata (1.4 kB)
Collecting unidecode==1.3.8 (from -r
```

```
/content/condaenv.v2x4sx6t.requirements.txt (line 169))
  Downloading Unidecode-1.3.8-py3-none-any.whl.metadata (13 kB)
Collecting unlzw3==0.2.2 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 170))
  Downloading unlzw3-0.2.2-py3-none-any.whl.metadata (2.4 kB)
Collecting urllib3==2.2.3 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 171))
  Downloading urllib3-2.2.3-py3-none-any.whl.metadata (6.5 kB)
Collecting uvicorn==0.32.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 172))
  Downloading uvicorn-0.32.0-py3-none-any.whl.metadata (6.6 kB)
Collecting uvloop==0.21.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 173))
  Downloading uvloop-0.21.0-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (4.9 kB)
Collecting vllm==0.6.3.post1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 174))
  Downloading vllm-0.6.3.post1-cp38-abi3-
manylinux1_x86_64.whl.metadata (10 kB)
Collecting warc3-wet==0.2.5 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 175))
  Downloading warc3_wet-0.2.5-py3-none-any.whl.metadata (2.2 kB)
Collecting warc3-wet-clueweb09==0.2.5 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 176))
  Downloading warc3-wet-clueweb09-0.2.5.tar.gz (17 kB)
  Preparing metadata (setup.py): started
  Preparing metadata (setup.py): finished with status 'done'
Collecting watchfiles==0.24.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 177))
  Downloading watchfiles-0.24.0-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (4.9 kB)
Collecting wcwidth==0.2.13 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 178))
  Downloading wcwidth-0.2.13-py2.py3-none-any.whl.metadata (14 kB)
Collecting websockets==13.1 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 179))
  Downloading websockets-13.1-cp310-cp310-
manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_17_x86_64.manylinux
2014_x86_64.whl.metadata (6.8 kB)
Collecting xformers==0.0.27.post2 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 180))
  Downloading xformers-0.0.27.post2-cp310-cp310-
manylinux2014_x86_64.whl.metadata (1.0 kB)
Collecting xxhash==3.5.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 181))
  Downloading xxhash-3.5.0-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (12 kB)
Collecting yarl==1.16.0 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 182))
```

```
  Downloading yarl-1.16.0-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (63 kB)
Collecting zipp==3.20.2 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 183))
  Downloading zipp-3.20.2-py3-none-any.whl.metadata (3.7 kB)
Collecting zlib-state==0.1.9 (from -r
/content/condaenv.v2x4sx6t.requirements.txt (line 184))
  Downloading zlib_state-0.1.9-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (3.3 kB)
INFO: pip is looking at multiple versions of retriv to determine which
version is compatible with other requirements. This could take a
while.

The conflict is caused by:
    The user requested pystemmer==2.2.0.3
    retriv 0.2.3 depends on pystemmer==2.0.1

To fix this you could try to:
1. loosen the range of package versions you've specified
2. remove package versions to allow pip to attempt to solve the
dependency conflict


Pip subprocess error:
ERROR: Cannot install -r /content/condaenv.v2x4sx6t.requirements.txt
(line 139) and pystemmer==2.2.0.3 because these package versions have
conflicting dependencies.
ERROR: ResolutionImpossible: for help visit
https://pip.pypa.io/en/latest/topics/dependency-resolution/#dealing-
with-dependency-conflicts

| failed

CondaEnvException: Pip failed
```

Anil R. Doshi, Oliver P. Hauser ,Generative AI enhances individual creativity but reduces the collective diversity of novel content.Sci. Adv.10,eadn5290(2024).DOI:10.1126/sciadv.adn5290

Doshi and Hauser didn't present a single "magic-formula" but rather defined creativity as a combination of two dimensions—novelty and usefulness—each measured by averaging three constituent ratings. In their study:

- The **novelty index** is computed as the mean of ratings on:
  – **Novelty**
  – **Originality**
  – **Rarity**
- The **usefulness index** is computed as the mean of ratings on:
  – **Appropriateness** (for the target audience)
  – **Feasibility** (of further development)

- **Publishability** (likelihood a publisher would take it on)

Usefulness is fully captured in the other features of helpfulness.

Inputs and Data: AI-Suggested Text (Candidate)

x: The newly generated text from the AI model.

Context Text

c: The preceding user text, conversation history, or story background. Used to measure appropriateness and novelty relative to context.

Reference Corpus

A large dataset representing "generic" writing or domain-specific text. Used to measure how common or rare certain phrases, syntactic patterns, or ideas are.

```
!pip install transformers
!pip install sacremoses
!pip install numpy
!pip install nltk

Requirement already satisfied: transformers in
/usr/local/lib/python3.11/dist-packages (4.48.3)
Requirement already satisfied: filelock in
/usr/local/lib/python3.11/dist-packages (from transformers) (3.17.0)
Requirement already satisfied: huggingface-hub<1.0,>=0.24.0 in
/usr/local/lib/python3.11/dist-packages (from transformers) (0.28.1)
Requirement already satisfied: numpy>=1.17 in
/usr/local/lib/python3.11/dist-packages (from transformers) (1.26.4)
Requirement already satisfied: packaging>=20.0 in
/usr/local/lib/python3.11/dist-packages (from transformers) (24.2)
Requirement already satisfied: pyyaml>=5.1 in
/usr/local/lib/python3.11/dist-packages (from transformers) (6.0.2)
Requirement already satisfied: regex!=2019.12.17 in
/usr/local/lib/python3.11/dist-packages (from transformers)
(2024.11.6)
Requirement already satisfied: requests in
/usr/local/lib/python3.11/dist-packages (from transformers) (2.32.3)
Requirement already satisfied: tokenizers<0.22,>=0.21 in
/usr/local/lib/python3.11/dist-packages (from transformers) (0.21.0)
Requirement already satisfied: safetensors>=0.4.1 in
/usr/local/lib/python3.11/dist-packages (from transformers) (0.5.2)
Requirement already satisfied: tqdm>=4.27 in
/usr/local/lib/python3.11/dist-packages (from transformers) (4.67.1)
Requirement already satisfied: fsspec>=2023.5.0 in
/usr/local/lib/python3.11/dist-packages (from huggingface-
hub<1.0,>=0.24.0->transformers) (2024.10.0)
Requirement already satisfied: typing-extensions>=3.7.4.3 in
/usr/local/lib/python3.11/dist-packages (from huggingface-
hub<1.0,>=0.24.0->transformers) (4.12.2)
```

```
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.11/dist-packages (from requests->transformers)
(3.4.1)
Requirement already satisfied: idna<4,>=2.5 in
/usr/local/lib/python3.11/dist-packages (from requests->transformers)
(3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.11/dist-packages (from requests->transformers)
(2.3.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.11/dist-packages (from requests->transformers)
(2025.1.31)
Collecting sacremoses
  Downloading sacremoses-0.1.1-py3-none-any.whl.metadata (8.3 kB)
Requirement already satisfied: regex in
/usr/local/lib/python3.11/dist-packages (from sacremoses) (2024.11.6)
Requirement already satisfied: click in
/usr/local/lib/python3.11/dist-packages (from sacremoses) (8.1.8)
Requirement already satisfied: joblib in
/usr/local/lib/python3.11/dist-packages (from sacremoses) (1.4.2)
Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-
packages (from sacremoses) (4.67.1)
Downloading sacremoses-0.1.1-py3-none-any.whl (897 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 897.5/897.5 kB 8.5 MB/s eta
0:00:00
oses
Successfully installed sacremoses-0.1.1
Requirement already satisfied: numpy in
/usr/local/lib/python3.11/dist-packages (1.26.4)
Requirement already satisfied: nltk in /usr/local/lib/python3.11/dist-
packages (3.9.1)
Requirement already satisfied: click in
/usr/local/lib/python3.11/dist-packages (from nltk) (8.1.8)
Requirement already satisfied: joblib in
/usr/local/lib/python3.11/dist-packages (from nltk) (1.4.2)
Requirement already satisfied: regex>=2021.8.3 in
/usr/local/lib/python3.11/dist-packages (from nltk) (2024.11.6)
Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-
packages (from nltk) (4.67.1)
```

To measure creativity, we'll break this down into two components.

1. Through `DJ_search_exact` by Lu et al. 2024 to find exact matches against a large corpus. We can get how much of the generated suggestions are directly "non-creative." The output is a score from 0-1, 0 if no spans are found, 1 if all tokens up to the last relevant index are matched.

2. We introduce our self-made `Stylescore` to measure how stylistiscally "creative" the text really is. To achieve this, we will measure

• Syntactic complexity.

- Figurative or Descriptive Language use.

```
!pip install nltk
!conda init
!conda activate infini-gram
!pip install unidecode

Requirement already satisfied: nltk in /usr/local/lib/python3.11/dist-
packages (3.9.1)
Requirement already satisfied: click in
/usr/local/lib/python3.11/dist-packages (from nltk) (8.1.8)
Requirement already satisfied: joblib in
/usr/local/lib/python3.11/dist-packages (from nltk) (1.4.2)
Requirement already satisfied: regex>=2021.8.3 in
/usr/local/lib/python3.11/dist-packages (from nltk) (2024.11.6)
Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-
packages (from nltk) (4.67.1)
/bin/bash: line 1: conda: command not found
/bin/bash: line 1: conda: command not found
Collecting unidecode
  Downloading Unidecode-1.3.8-py3-none-any.whl.metadata (13 kB)
Downloading Unidecode-1.3.8-py3-none-any.whl (235 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 235.5/235.5 kB 3.5 MB/s eta
0:00:00
```

The following is adapted from Lu et al. 2024.

```
!pip install --upgrade numpy

Requirement already satisfied: numpy in
/usr/local/lib/python3.11/site-packages (2.0.2)
Collecting numpy
  Using cached numpy-2.2.3-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (62 kB)
Using cached numpy-2.2.3-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl (16.4 MB)
Installing collected packages: numpy
  Attempting uninstall: numpy
    Found existing installation: numpy 2.0.2
    Uninstalling numpy-2.0.2:
      Successfully uninstalled numpy-2.0.2
ERROR: pip's dependency resolver does not currently take into account
all the packages that are installed. This behaviour is the source of
the following dependency conflicts.
tensorflow 2.18.0 requires numpy<2.1.0,>=1.26.0, but you have numpy
2.2.3 which is incompatible.
Successfully installed numpy-2.2.3

import os
import nltk
```

```python
import json
import time
import requests
import argparse
import numpy as np
from tqdm import tqdm
from typing import List, Callable
from dataclasses import dataclass
from unidecode import unidecode
from sacremoses import MosesDetokenizer
from transformers import AutoTokenizer

md = MosesDetokenizer(lang='en')
API_URL = 'https://api.infini-gram.io/'
HF_TOKEN = "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"
tokenizer = AutoTokenizer.from_pretrained("meta-llama/Llama-2-7b-hf",
token=HF_TOKEN,
                                          add_bos_token=False,
add_eos_token=False)
@dataclass
class Document:
    doc_id: str
    tokens: List[str]

@dataclass
class Span:
    start_index: int
    end_index: int
    span_text: str
    occurrence: int

class Hypothesis:
    def __init__(self, target_doc: Document, min_ngram: int) -> None:
        self.target_doc = target_doc
        self.min_ngram = min_ngram
        self.spans = []
        self.finished = False

    def add_span(self, new_span: Span) -> None:
        self.spans.append(new_span)
        if new_span.end_index >= len(self.target_doc.tokens):
            self.finished = True

    def replace_span(self, new_span: Span) -> None:
        self.spans = self.spans[:-1] + [new_span]
        if new_span.end_index >= len(self.target_doc.tokens):
            self.finished = True

    def get_score(self) -> float:
        if not self.spans:
```

```python
            return 0.0
        progress_len = self.spans[-1].end_index if not self.finished
else len(self.target_doc.tokens)
        flags = [False] * progress_len
        for span in self.spans:
            span_length = span.end_index - span.start_index
            flags[span.start_index:span.end_index] = [True] *
span_length
        coverage = sum(flags) / len(flags)
        return coverage

    def format_span(self) -> str:
        return ' | '.join([s.span_text for s in self.spans])

    def __hash__(self) -> int:
        return hash(self.format_span())

    def __eq__(self, other) -> bool:
        if isinstance(other, Hypothesis):
            return self.format_span() == other.format_span()
        return NotImplemented

    def get_avg_span_len(self) -> float:
        if not self.spans:
            return 0.0
        span_lengths = [s.end_index - s.start_index for s in
self.spans]
        return sum(span_lengths) / len(span_lengths)

    def export_json(self) -> dict:
        matched_spans = [{
            'start_index': s.start_index,
            'end_index': s.end_index,
            'span_text': s.span_text,
            'occurrence': s.occurrence
        } for s in self.spans]
        return {
            'matched_spans': matched_spans,
            'coverage': self.get_score(),
            'avg_span_len': self.get_avg_span_len(),
        }

def find_exact_match(detokenize: Callable, doc: Document, min_ngram:
int) -> dict:
    hypothesis = Hypothesis(doc, min_ngram)
    first_pointer, second_pointer = 0, min_ngram
    while second_pointer <= len(doc.tokens):
        span_text =
detokenize(doc.tokens[first_pointer:second_pointer])
        request_data = {
```

```python
            'corpus': 'v4_rpj_llama_s4',
            'engine': 'c++',
            'query_type': 'count',
            'query': span_text,
        }
        time.sleep(0.01)
        search_result = requests.post(API_URL,
json=request_data).json()
        occurrence = search_result.get('count', 0)

        if occurrence:
            matched_span = Span(
                start_index=first_pointer,
                end_index=second_pointer,
                span_text=span_text,
                occurrence=occurrence
            )
            if not hypothesis.spans:
                hypothesis.add_span(matched_span)
            else:
                last_span = hypothesis.spans[-1]
                if matched_span.start_index <= last_span.start_index
and last_span.end_index <= matched_span.end_index:
                    hypothesis.replace_span(matched_span)
                else:
                    hypothesis.add_span(matched_span)
            second_pointer += 1

            #
print("****************************************************************
***********************************")
            # print(hypothesis.format_span())
            # print(f'score: {hypothesis.get_score():.4f}
avg_span_length: {hypothesis.get_avg_span_len()}')
            #
print("****************************************************************
***********************************")
        else:
            if second_pointer - first_pointer > min_ngram:
                first_pointer += 1
            elif second_pointer - first_pointer == min_ngram:
                first_pointer += 1
                second_pointer += 1
            else:
                raise ValueError("Invalid state in span detection.")

    hypothesis.finished = True
    return hypothesis.export_json()

def process_text(text: str, min_ngram: int = 6, lm_tokenizer: bool =
```

```python
False) -> dict:
    # Choose the appropriate tokenizer/detokenizer.
    if not lm_tokenizer:
        tokenize_func = lambda x:
nltk.tokenize.casual.casual_tokenize(x)
        detokenize = lambda tokens: md.detokenize(tokens)
    else:
        tokenize_func = lambda x: tokenizer.tokenize(x)
        detokenize = lambda tokens:
tokenizer.decode(tokenizer.convert_tokens_to_ids(tokens))

    # Preprocess and tokenize the plain text.
    processed_text = unidecode(text)
    tokens = tokenize_func(processed_text)

    if len(tokens) <= min_ngram:
        raise ValueError("Input text is too short for the specified
min_ngram.")

    # Build a Document and perform the exact match search.
    doc = Document(doc_id="input_text", tokens=tokens)
    result = find_exact_match(detokenize, doc, min_ngram)
    return result

print (process_text("I'm the best in the world, laughing out loud.")
['coverage'])
```

1.0

```python
!pip install spacy
!python -m spacy download en_core_web_sm
```

Collecting spacy
  Downloading spacy-3.8.4-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (27 kB)
Collecting spacy-legacy<3.1.0,>=3.0.11 (from spacy)
  Downloading spacy_legacy-3.0.12-py2.py3-none-any.whl.metadata (2.8
kB)
Collecting spacy-loggers<2.0.0,>=1.0.0 (from spacy)
  Downloading spacy_loggers-1.0.5-py3-none-any.whl.metadata (23 kB)
Collecting murmurhash<1.1.0,>=0.28.0 (from spacy)
  Downloading murmurhash-1.0.12-cp311-cp311-
manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_17_x86_64.manylinux
2014_x86_64.whl.metadata (2.1 kB)
Collecting cymem<2.1.0,>=2.0.2 (from spacy)
  Downloading cymem-2.0.11-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (8.5 kB)
Collecting preshed<3.1.0,>=3.0.2 (from spacy)
  Downloading preshed-3.0.9-cp311-cp311-
manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_17_x86_64.manylinux

```
2014_x86_64.whl.metadata (2.2 kB)
Collecting thinc<8.4.0,>=8.3.4 (from spacy)
  Downloading thinc-8.3.4-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (15 kB)
Collecting wasabi<1.2.0,>=0.9.1 (from spacy)
  Downloading wasabi-1.1.3-py3-none-any.whl.metadata (28 kB)
Collecting srsly<3.0.0,>=2.4.3 (from spacy)
  Downloading srsly-2.5.1-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (19 kB)
Collecting catalogue<2.1.0,>=2.0.6 (from spacy)
  Downloading catalogue-2.0.10-py3-none-any.whl.metadata (14 kB)
Collecting weasel<0.5.0,>=0.1.0 (from spacy)
  Downloading weasel-0.4.1-py3-none-any.whl.metadata (4.6 kB)
Collecting typer<1.0.0,>=0.3.0 (from spacy)
  Downloading typer-0.15.1-py3-none-any.whl.metadata (15 kB)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in
/usr/local/lib/python3.11/site-packages (from spacy) (4.67.1)
Requirement already satisfied: numpy>=1.19.0 in
/usr/local/lib/python3.11/site-packages (from spacy) (2.0.2)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in
/usr/local/lib/python3.11/site-packages (from spacy) (2.32.3)
Collecting pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4 (from spacy)
  Downloading pydantic-2.10.6-py3-none-any.whl.metadata (30 kB)
Collecting jinja2 (from spacy)
  Downloading jinja2-3.1.5-py3-none-any.whl.metadata (2.6 kB)
Requirement already satisfied: setuptools in
/usr/local/lib/python3.11/site-packages (from spacy) (65.6.3)
Requirement already satisfied: packaging>=20.0 in
/usr/local/lib/python3.11/site-packages (from spacy) (24.2)
Collecting langcodes<4.0.0,>=3.2.0 (from spacy)
  Downloading langcodes-3.5.0-py3-none-any.whl.metadata (29 kB)
Collecting language-data>=1.2 (from langcodes<4.0.0,>=3.2.0->spacy)
  Downloading language_data-1.3.0-py3-none-any.whl.metadata (4.3 kB)
Collecting annotated-types>=0.6.0 (from pydantic!=1.8,!
=1.8.1,<3.0.0,>=1.7.4->spacy)
  Using cached annotated_types-0.7.0-py3-none-any.whl.metadata (15 kB)
Collecting pydantic-core==2.27.2 (from pydantic!=1.8,!
=1.8.1,<3.0.0,>=1.7.4->spacy)
  Downloading pydantic_core-2.27.2-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (6.6 kB)
Requirement already satisfied: typing-extensions>=4.12.2 in
/usr/local/lib/python3.11/site-packages (from pydantic!=1.8,!
=1.8.1,<3.0.0,>=1.7.4->spacy) (4.12.2)
Requirement already satisfied: charset_normalizer<4,>=2 in
/usr/local/lib/python3.11/site-packages (from requests<3.0.0,>=2.13.0-
>spacy) (3.4.1)
Requirement already satisfied: idna<4,>=2.5 in
/usr/local/lib/python3.11/site-packages (from requests<3.0.0,>=2.13.0-
>spacy) (3.10)
```

```
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.11/site-packages (from requests<3.0.0,>=2.13.0-
>spacy) (2.3.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.11/site-packages (from requests<3.0.0,>=2.13.0-
>spacy) (2024.12.14)
Collecting blis<1.3.0,>=1.2.0 (from thinc<8.4.0,>=8.3.4->spacy)
  Downloading blis-1.2.0-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (7.7 kB)
Collecting confection<1.0.0,>=0.0.1 (from thinc<8.4.0,>=8.3.4->spacy)
  Downloading confection-0.1.5-py3-none-any.whl.metadata (19 kB)
Requirement already satisfied: click>=8.0.0 in
/usr/local/lib/python3.11/site-packages (from typer<1.0.0,>=0.3.0-
>spacy) (8.1.8)
Collecting shellingham>=1.3.0 (from typer<1.0.0,>=0.3.0->spacy)
  Downloading shellingham-1.5.4-py2.py3-none-any.whl.metadata (3.5 kB)
Requirement already satisfied: rich>=10.11.0 in
/usr/local/lib/python3.11/site-packages (from typer<1.0.0,>=0.3.0-
>spacy) (13.9.4)
Collecting cloudpathlib<1.0.0,>=0.7.0 (from weasel<0.5.0,>=0.1.0-
>spacy)
  Downloading cloudpathlib-0.20.0-py3-none-any.whl.metadata (14 kB)
Collecting smart-open<8.0.0,>=5.2.1 (from weasel<0.5.0,>=0.1.0->spacy)
  Downloading smart_open-7.1.0-py3-none-any.whl.metadata (24 kB)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.11/site-packages (from jinja2->spacy) (3.0.2)
Collecting marisa-trie>=1.1.0 (from language-data>=1.2-
>langcodes<4.0.0,>=3.2.0->spacy)
  Downloading marisa_trie-1.2.1-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (9.0 kB)
Requirement already satisfied: markdown-it-py>=2.2.0 in
/usr/local/lib/python3.11/site-packages (from rich>=10.11.0-
>typer<1.0.0,>=0.3.0->spacy) (3.0.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in
/usr/local/lib/python3.11/site-packages (from rich>=10.11.0-
>typer<1.0.0,>=0.3.0->spacy) (2.19.1)
Requirement already satisfied: wrapt in
/usr/local/lib/python3.11/site-packages (from smart-
open<8.0.0,>=5.2.1->weasel<0.5.0,>=0.1.0->spacy) (1.17.2)
Requirement already satisfied: mdurl~=0.1 in
/usr/local/lib/python3.11/site-packages (from markdown-it-py>=2.2.0-
>rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy) (0.1.2)
Downloading spacy-3.8.4-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl (30.6 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 30.6/30.6 MB 43.9 MB/s eta
0:00:00
em-2.0.11-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl
(218 kB)
Downloading langcodes-3.5.0-py3-none-any.whl (182 kB)
```

```
Downloading murmurhash-1.0.12-cp311-cp311-
manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_17_x86_64.manylinux
2014_x86_64.whl (134 kB)
Downloading preshed-3.0.9-cp311-cp311-
manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_17_x86_64.manylinux
2014_x86_64.whl (157 kB)
Downloading pydantic-2.10.6-py3-none-any.whl (431 kB)
Downloading pydantic_core-2.27.2-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl (2.0 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2.0/2.0 MB 62.5 MB/s eta
0:00:00
anylinux_2_17_x86_64.manylinux2014_x86_64.whl (1.1 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.1/1.1 MB 46.8 MB/s eta
0:00:00
anylinux_2_17_x86_64.manylinux2014_x86_64.whl (3.9 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 3.9/3.9 MB 85.8 MB/s eta
0:00:00
anylinux_2_17_x86_64.manylinux2014_x86_64.whl (11.7 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 11.7/11.7 MB 72.7 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 5.4/5.4 MB 56.0 MB/s eta
0:00:00
-1.5.4-py2.py3-none-any.whl (9.8 kB)
Downloading smart_open-7.1.0-py3-none-any.whl (61 kB)
Downloading marisa_trie-1.2.1-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl (1.4 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.4/1.4 MB 42.3 MB/s eta
0:00:00
em, wasabi, spacy-loggers, spacy-legacy, smart-open, shellingham,
pydantic-core, murmurhash, marisa-trie, jinja2, cloudpathlib,
catalogue, blis, annotated-types, srsly, pydantic, preshed, language-
data, typer, langcodes, confection, weasel, thinc, spacy
Successfully installed annotated-types-0.7.0 blis-1.2.0 catalogue-
2.0.10 cloudpathlib-0.20.0 confection-0.1.5 cymem-2.0.11 jinja2-3.1.5
langcodes-3.5.0 language-data-1.3.0 marisa-trie-1.2.1 murmurhash-
1.0.12 preshed-3.0.9 pydantic-2.10.6 pydantic-core-2.27.2 shellingham-
1.5.4 smart-open-7.1.0 spacy-3.8.4 spacy-legacy-3.0.12 spacy-loggers-
1.0.5 srsly-2.5.1 thinc-8.3.4 typer-0.15.1 wasabi-1.1.3 weasel-0.4.1
Collecting en-core-web-sm==3.8.0
  Downloading
https://github.com/explosion/spacy-models/releases/download/en_core_we
b_sm-3.8.0/en_core_web_sm-3.8.0-py3-none-any.whl (12.8 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 12.8/12.8 MB 95.6 MB/s eta
0:00:00

Successfully installed en-core-web-sm-3.8.0
✔ Download and installation successful
You can now load the package via spacy.load('en_core_web_sm')
```

```python
import spacy
from collections import deque

# Load spaCy model
nlp = spacy.load("en_core_web_sm")

def get_parse_tree_depth(sent):
    """
    Calculate parse tree depth for a sentence using a BFS
    starting from the root token.
    """
    roots = [token for token in sent if token.head == token]
    if not roots:
        return 0
    root = roots[0]
    max_depth = 0
    depths = {root: 0}
    queue = deque([root])

    while queue:
        node = queue.popleft()
        for child in node.children:
            depths[child] = depths[node] + 1
            max_depth = max(max_depth, depths[child])
            queue.append(child)
    return max_depth

def count_subordinate_clauses(doc):
    """
    Count subordinate clauses by looking for tokens with dependency
labels
    typically marking clause boundaries:
      - 'mark' (subordinating conjunctions),
      - 'advcl' (adverbial clauses), and
      - 'ccomp' (clausal complements).
    """
    count = 0
    for token in doc:
        if token.dep_ in {"mark", "advcl", "ccomp"}:
            count += 1
    return count

def count_passive_sentences(doc):
    """
    Identify and count passive sentences. A sentence is considered
passive
    if it contains any token with the dependency 'nsubjpass'.
    """
    passive_count = 0
    for sent in doc.sents:
```

```python
        if any(token.dep_ == "nsubjpass" for token in sent):
            passive_count += 1
    return passive_count

def syntactic_complexity(text, d_max=10,
max_subordinate_per_sentence=2):
    """
    Calculate a normalized syntactic complexity score in the range
[0,1].
    It is very knowledge-based.
    This function computes three components:
      1. Depth Component: Average parse tree depth normalized by
d_max.
         The average is capped at d_max to avoid outlier effects.
      2. Subordinate Clause Component: The average number of
subordinate clauses
         per sentence normalized by an assumed maximum (default=2).
      3. Passive Voice Component: The fraction of sentences exhibiting
passive voice.

    The final score is the average of these three components.
    """
    doc = nlp(text)
    sentences = list(doc.sents)
    num_sentences = len(sentences)

    if num_sentences == 0:
        return 0.0

    # Depth Component
    total_depth = sum(get_parse_tree_depth(sent) for sent in
sentences)
    avg_depth = total_depth / num_sentences
    depth_component = min(avg_depth, d_max) / d_max  # normalize to
[0,1]

    # Subordinate Clause Component
    num_subordinate = count_subordinate_clauses(doc)
    subordinate_component = (num_subordinate / num_sentences) /
max_subordinate_per_sentence
    subordinate_component = min(subordinate_component, 1.0)

    # Passive Voice Component
    passive_sentences = count_passive_sentences(doc)
    passive_component = passive_sentences / num_sentences  # already
in [0,1]

    # Combine components equally to yield a final score between 0 and
1
    normalized_score = (depth_component + subordinate_component +
```

```
passive_component) / 3.0
    return normalized_score

text = (
    "The old mansion, which had been abandoned for years, stood
silent. Its walls, marked by time and neglect, told stories as if they
were whispering. The garden was overgrown, and nature had reclaimed
it, making it look like a scene from a forgotten fairy tale."
)

score = syntactic_complexity(text)
print("Normalized Syntactic Complexity Score:", score)

Normalized Syntactic Complexity Score: 0.5777777777777777
```

Successfully read 2215 events in a writing session from ./coauthor-v1.0/bac237f54fe449e18334a685c6381657.jsonl
User-written sentences: ['dating apps and technology in general have paved the way to bridge closer connections together than', 'ever before. Because of dating apps such as Tinder, Facebook', ', Instagram, Snapchat, and other social media platforms, people are able to connect with each other in ways that were not possible before. Imagine before where two strangers could only meet by chance on a particular instance, nowadays we are', "able to connect with people from all over the world. It's almost as if it's cheating to find the one and what most people call,", 'your soulmate.\n\nOf course being too distracted and neglecting to spend time with your loved ones if you use these dating apps too often can be a negative. But so is', 'anything else that you do in excess. So the usage of technology must come with a great', 'balance and moderation. This way,', 'you can still enjoy the benefits of technology while still being able to maintain a healthy relationship with your significant other.\n\nW remember being a kid and having no internet in the house, in order for me to contact a loved one that is far away is to', 'write a letter and send it through snail mail. It was a slow, expensive, and', 'a time consuming process. We almost take it for granted today that', 'we can send a message to someone in another country in a matter of seconds that is also free of cost, considering you already have the equipment to do so. I feel like in the future, it would just be improved and we would all feel very connected more than less.\n\nBeing in a video call or a phone call, something may seem missing.', "It's not the same as being in person. Being intimate and feeling the touch of someone is still something that the technology can give. That's why", 'I am optimistic that in the future, this can be', 'improved and we can have a more intimate connection with each other. Having a device that lets you', "feel the touch of someone is something that I would like to see in the future. That way, there's no need for", 'you to be in the same room as someone in order to feel their touch. This eliminates the

issue of', 'having to travel long distances just to see someone.\n\nSo just to clarify my stance with technology and dating apps, I feel like they are a bridge instead of', 'a barrier. Using it in moderation, I feel like the tools the dating apps and social media provide is very', 'useful and can help people connect with each other in ways that were not possible before. It is also actively trying to fulfill the further needs of humanity and making me optimistic that in the future,', "we will be able to have a more intimate connection with each other. Have you contacted a loved one recently? If not, I do urge you to go ahead and make sure they're"]

AI-generated suggestions: ['ever before.', ', Instagram, Snapchat, and other social media platforms, people are able to connect with each other in ways that were not possible before.', 'able to connect with people from all over the world.', 'balance and moderation.', 'you can still enjoy the benefits of technology while still being able to maintain a healthy relationship with your significant other.', 'write a letter and send it through snail mail.', 'a time consuming process.', 'we can send a message to someone in another country in a matter of seconds.', "It's not the same as being in person.", 'improved and we can have a more intimate connection with each other.', 'feel the touch of someone is something that I would like to see in the future.', 'you to be in the same room as someone in order to feel their touch.', 'having to travel long distances just to see someone.', 'a barrier.', 'useful and can help people connect with each other in ways that were not possible before.', 'we will be able to have a more intimate connection with each other.', 'doing well.']

Successfully read 2355 events in a writing session from ./coauthor-v1.0/8393a8ed335f438c9ad5ca3d25a947b6.jsonl

User-written sentences: ['And the pig was like "That\'s right, dude. I saw the shaped charges you used on the 91st pig\'s steel reinforced concrete house, and that\'s not gonna cut it with mine."', 'The wolf was like, "Whatever, dude. I\'m tired of this game anyway, you build the craziest most reinforced house you like, because I\'m not huffing, or puffing, or shooting. or dynamiting, or anything else to try to get in there." Then,', 'the wolf went off to fhe forest to live a normal wolf life, he joined a pack,', 'and became a normal wolf.', 'He met a lady wolf, they had cubs, and', 'ate deer, and rabbits, and it was', 'a very happy, normal wolf life.\n\nMeanwhile, the pig', 'sat around his depleted uranium house, and worried. He', 'was like, "What if the wolf hadn\'t given up?" He was constantly waiting for the attack which never came.', 'He lived in fear for seven years. Over those seven years, he', 'ate his food without savoring it, he never laughed, he had no friends or family. He searched his house for', "hidden ways in, and when he couldn't find any, he searched for hidden ways out. That was a bust. He checked the yard for signs", 'of the wolf, he installed motion sensing cameras around his house, he bought a wolf-call for the phone. He did constant Google searches related to home security, and incidentally found out that depleted uranium', "was a little dangerous, and known to raise the incidence of cancer in human tank

operators.\n\nHe told himself that that didn't apply to pigs, and in any event, it was worth it to be wolf-free. So, he", 'sat, and he waited, and he ate. He got fat as a pig, which in this case, him already being fat as a pig, meant he was as fat as', 'a pig that had gotten fatter than a normal pig. In other words, as fat as two', 'norma After seven years of this, he realized he was', 'wasting his life, and he decided he would leave home for the first time in seven years to see what was going on outside. slowly killing himself, and \n\nHe left his depleted uranium home, and ventured out to the forest, where', 'he saw the normal wolf pack which had been living there, completely oblivious to his existence. He spoke to the wolf. "Hey, you look pretty happy, I guess not eating me was a pretty good deal after all?" The wolf was silent, because now he was a normal wolf who couldn\'t talk.\n\nThe pig continued "So, would it be okay if I moved out here, and lived a normal life you like you have?" The wolf moved his head a bit, which the pig took as an assent, and the pig began rooting among the shrubs for stuff to eat.\n\nThe wolf pack paused for a moment, surprised by thehesitatedis strange creature that talked, and seemed to have no fear of them. After that moment, of course, they']

AI-generated suggestions: ['And the wolf was like "dude."', 'a few hours later when he went to go huff the house, he was like, "What is this stuff?', 'and got some other wolves to do his dirty work for him.', 'all good.', 'was like, "What if the wolf hadn\'t given up?"', 'Until one day, the wolf came back.', 'ate his food without savoring it, he never laughed, he had no friends or family.', 'radiation leaks constantly.', "of tunneling, and when he couldn't find any there either, he checked the moon for signs of wolves.", "is worse than he had thought, it's not just a little bad for you, but a lot bad for you.", 'sat, and he waited, and he ate.', 'a pig could be.', 'normal pigs.', 'just as fat and miserable as he would have been if the wolf had come and eaten him, so one day he got up, ate his breakfast without', 'he saw the normal wolf pack which had been living there, completely oblivious to his existence.', 'tore him to pieces.']

Successfully read 2150 events in a writing session from ./coauthor-v1.0/994c929ac0db414cbb96423bb245eb7c.jsonl

User-written sentences: ['Scientists realize that this drug actually causes', 'people to become particularly voracious about their sexuality.', 'The drug is recalled, and the government is forced to pay for the cost of the drugabout satiating basic human needs and desires like food, shelter, clothing, and sex.\n\n But already an entire human population has been impacted by the drug. The only people saved are those who are your age or older since you were the only ones that did not receive the drug.', "You are one of the few people who are not affected by the drug.\n But you have seen how this drug has totally decimated others. For example, a good friend of yours who is younger than you is obese because she cannot stop stuffing food into her face. Another one of your friends has gone bankrupt buying herself 100 houses since she feels extreme anxiety about not having enough

housing to live in. And another one of your friends has also become a hoarder since she has so many clothes bursting at the seams of her house. Finally, one of your friends is in a constant state of postpartum depression since she's now been through forty-three pregnancies. Her sexual appetite is just unending.", 'You have decided to take action.', 'You want to save humanity from itself.', 'You want to stop this drug from ever being used again.', 'You have decided to create a new drug that will counteract the effects of the immortality drug.', 'You call it the "Anti-Immortality Drug."', 'You spend many nights and days working in your lab trying to perfect the drug. When it has finally been perfected you', "decide to test it out on yourself. You don't see any short-term side effects or reactions so you feel good about your creation. You then administer the drug to your friends. The friend of yours who has the high libido instantly feels that her sexual urges are quelled. The friend who is bankrupt from all her houses starts selling them off like crazy, thus vastly improving her net worth. The friend who is a hoarder gives all her clothes to the homeless. And the friend who is obese learns moderation when it comes to her diet.\n\nEveryone in society who takes your drug", 'is now living a normal life. They all cheer for you as their personal hero.']

AI-generated suggestions: ['The drug is recalled, and the government is forced to pay for the cost of the drug and the side effects.', 'You are one of the few people who are not affected by the drug.', 'You have decided to take action.', 'You want to save humanity from itself.', 'You want to stop this drug from ever being used again.', 'You have decided to create a new drug that will counteract the effects of the immortality drug.', 'You call it the "Anti-Immortality Drug."', 'decide to test it out on yourself.', 'is now living a normal life.', 'You are hailed as the savior of humanity.']

Successfully read 3792 events in a writing session from ./coauthor-v1.0/b6acb37208be4f3e96965f989fd6f9fb.jsonl

User-written sentences: ['I was 28 when English scientists discovered a miracle drug that stopped humanity from aging. Humanity had finally discovered the key to living forever and avoiding the one debt that every human being owes upon birth, death. Scientists had warned that the drug was only meant for citizens under 26, which left billions of people, including myself, without any hope of taking the drug. People around the world were furious and saddened to hear the news that they would not live forever. Meanwhile, the younger generations were divided on whether or not they wanted to live forever. Taking the drug meant living forever or outliving loved ones and watching them pass on.', 'The rest of the world was not as hesitant to take the drug, and many saw immortality as a chance for humanity to expand and achieve incredible things. Taking the drug would allow people to study whatever they wanted and help to improve society for the better. No longer did people have to settle for boring desk jobs, they could pursue any field of study they wanted.', 'Doctors no longer had to worry about sicknesses and disease, they could now devote their time

towards helping people. among a large portion of society who could not take the drugThe "Lost Generation" continued to age, while the young took the drug and became immortal. Families were torn apart, knowing that they would leave loved ones behind. They feared that their loved ones would forget about them as the centuries carried on.', 'While the young celebrated their immortality, the old feared that they would be forgotten. As time passed, I grew older, finally reaching the ripe old age of 85,. Society had definitely improved through the efforts of an immortal, optimistic generation of people with all the time in the world on their shoulders.,', 'They set their sights on new frontiers that would be impossible for a normal human lifetime.had Technology had improved by leaps and bounds, providing comfort beyond imagination for people from all walks of life. Medicinal advances provided the "Lost Generation" with cures for ailments that only affected an aging body. I began to wonder if taking the drug was worth it. At a certain point, the "Lost Generation" would be no more, ceasing to exist in time, leaving the immortals behind.', 'Was living forever worth it? Were all of the human advances worth it in the end?', 'The immortals were only prolonging the inevitable.', 'I was concerned with how the immortals would feel after living for hundreds of years, or even thousands. Finally, in a shocking turn of events, scientists discovered that the drug resulted in tragic side effects. The immortals began to lose their sight, hearing, taste, sense of smell, aspects of human life. all No longer could they enjoy what made life so pleasurable, the biological features that made life worth living.', 'The drug had robbed them of all senses, leaving them to a life where they were no longer huma They could not rely on "The Lost Generatthe", for they too needed help. One generation was dying naturally, the other were becoming walking husks, devoid of humanity.', 'The immortals were now a burden, and the "Lost Generation" was stuck in between.', 'The immortals were now at the mercy of the "Lost Generation", for they needed help. The "Lost Generation" could not help, they could only provide wisdom to ease the pain of living without their senses.', 'The immortals were not the picture of humanity, they were a nightmare. Zombies walking amongst the living and the aging.', 'The immortals were not living, they were merely existing in the world. Their hope and dreams of living forevers dashed against the stones by the grim reaper. The "Lost Generation" died out one by one, I was one of the remaining few a bygone era. of']
AI-generated suggestions: ['The rest of the world was not as hesitant to take the drug, and many saw immortality as a chance for humanity to expand and achieve incredible things.', 'Doctors no longer had to worry about sicknesses and disease, they could now devote their time towards helping people.', 'While the young celebrated their immortality, the old feared that they would be forgotten.', 'They set their sights on new frontiers that would be impossible for a normal human lifetime.', 'Was living forever worth it?', 'The immortals were only prolonging the inevitable.', 'I was concerned with how the immortals would feel after living for hundreds of years, or even

thousands.', 'The drug had robbed them of all senses, leaving them to a life where they were no longer human.', 'The immortals were now a burden, and the "Lost Generation" was stuck in between.', 'The immortals were now at the mercy of the "Lost Generation", for they needed help.', 'The immortals were not the picture of humanity, they were a nightmare.', 'The immortals were not living, they were merely existing in the world.', 'I had witnessed the downfall of humanity, and luckily for me, I was spared from the pain of loss.']
Successfully read 1748 events in a writing session from ./coauthor-v1.0/12c0273868a84564a5f001c3c086f69f.jsonl
User-written sentences: ["I can't blow that down. That's not fair. Where did you even get the uranium from?", "I have no idea how to blow that down.\n\nThe 92nd little pig was ecstatic because he had solved the issue that the wolf couldn't fix. All of his brethren before had their houses blown down. He outsmarted the wolf."]
AI-generated suggestions: ['So the wolf went and he built a house out of depleted uranium.', 'He had a house of super strength.']
Successfully read 1402 events in a writing session from ./coauthor-v1.0/63ab89c6ca594ea0a80519442ac4c27b.jsonl
User-written sentences: ["Now, it's too late. It is rumored that a new drug has been created.", "But the government isn't giving it away for free. It is too expensive for most citizens so people are going great lengths to obtain it illegally.", 'To get the drug, you and your friends decide to find a way into the headquarters of the government facility where it is supposedly being held.\n The risks are high, but the group is going for it anyway.', 'They planned carefully and the big night had finally arrived.', 'They began their journey and successfully make it into the facility. They follow the map to the vault where the drugs are being stored.', "But it turns out there is an alarm. They hadn't planned for this. They had to either turn around and leave now, or find a way to bypass the alarm.", 'They chose to continue and discover an elevator that leads to a secret floor deep within the building. The elevator descended to the floor below, and then kept going and going.', "It was not to the lower floor, but to another section of the facility. It went down so far they couldn't fathom how far underground they had gone.", 'Then, the elevator stopped and the doors opened.', 'They were in a strange laboratory. "This must be where they make the drug."', 'you thought. The team searched on. "We shouldn\'t be down here. We should just find the product and go. We shouldn\'t mess with this stuff."', 'your best friend said. Suddenly, the elevator doors closed and the elevator lifted up away, leaving them in the deep basement.', "They were stuck. They went to press the elevator buttons, but there weren't any.", 'A few moments later, the doors opened and a mysterious figure walked in.', 'He was the leader of the facility. "I understand you want access to our product. I have a better offer."', 'he said.', '"I will give you access, but you must agree to my terms." They agreed.', '"You will bring me an item every week for the rest of your lives, and I will give you your drug.']

AI-generated suggestions: ["Now, it's too late.", "But the government isn't giving it away for free.", "When you're 28, you have a normal life.", "The night is still young, and with any luck, they'll all be strong enough to live through the night.", 'But it turns out there is an alarm.', 'They chose to continue and discover an elevator that leads to a secret floor deep within the building.', 'As the elevator finally stopped, they discovered the secret floor was covered in discarded prototypes.', 'Then, the elevator stopped and the doors opened.', 'They stepped out to find themselves in a totally different part of the facility.', 'They thought.', '"Yeah, but it could be our only chance to live forever."', 'They were stuck.', 'A few moments later, the doors opened and a mysterious figure walked in.', 'He was the leader of the facility.', 'he said.', 'He was an older man, but he appeared stronger and more powerful than you.', '"You will bring me an item every week for the rest of your lives, and I will give you your drug.', 'If you fail, I will use your friends as my new test subjects."']
Successfully read 1445 events in a writing session from ./coauthor-v1.0/3b6e7a4c9d65411f9342254120abcde7.jsonl
User-written sentences: ["Choosing depleted uranium wasn't 92's first choice of building material,", 'nor his second, or even his third. But, after realizing that this was a material that none of his other sibling would use, he accepted this as a challenge. \n\nFor generations, Wolf has been the enemy to the pigs and their mother. But after meeting 60 or 70 pigs, Wolf realized he could learn from their determination. Maybe he could even help them from the past mistakes of their elders.', 'He could use his knowledge to meet their needs, and make a mutually beneficial deal.', "The wolf started to teach the pigs how to be shrewd business people. and one of those lessons came in the form of a town who had shut down its nuclear power plant decades ago, but left the uranium on site. \n\nTo be fair, the house itself isn't built out of the uranium itself, but the steel containers that hold the product.", "The 92nd pig found out that steel containers were either thrown away, bur. These were stackable barrels that exuded a faint glow. Pigs being pigs, housekeeping wasn't always the best, yet bugs never entered 92's property."]
AI-generated suggestions: ['nor his second, or even his third.', 'He could use his knowledge to meet their needs, and make a mutually beneficial deal.', 'The wolf started to teach the pigs how to be shrewd business people.', 'The 92nd pig found out that steel containers were either thrown away, buried, or sold for scrap.', 'Choosing this material also helped 92 make a profit.']
Successfully read 1080 events in a writing session from ./coauthor-v1.0/3778eebfc9df42429aa1f23db1934a9d.jsonl
User-written sentences: ['the most important things students should learn in school is social skills academic knowledge as well as .', 'Academic knowledge is important because it helps students to learn how to solve problems and think critically.', 'This will help students to cope with their future jobs and daily problems they encounter.',

'Academic knowledge learnt at school is fundamental to everything. The knowledge will also help students to distinguibetween the truth and what is true and what is untrue.', ', which is very important when the students make decisions on their owncertain . Besides, getting along with their classmates at school also helps the students develop social skills, which are necessary for their survival in the society.']
AI-generated suggestions: ['Academic knowledge is important because it helps students to learn how to solve problems and think critically.', 'The most important thing I have learned in school is the importance of being honest.']
Successfully read 1929 events in a writing session from ./coauthor-v1.0/8b3585b594e24e0dbfd53eae933d2f4a.jsonl
User-written sentences: ['technology has greatly improved dating.', "I've found that dating apps in particular increase the odds of finding a relationship, at least for me. Just having the ability to use Tinder or Bumble takes away the nerves from introducing yourself to someone in person. Breaking that ice is what I find the most difficult. Being able to talk to someone via a messaging app gives me a good indication on whether or not we are compatible.\n\nObviously there is still stress when it comes to dating even with the advancement of technology. Ultimately, no matter how well you know someone via the internet or via messaging apps, that first meeting is person is very nerve wracking.", "I think technology does a really good job in helping alleviate anxiety, but it never fully prepares you for a first impression.\n I may be extremely comfortable speaking to someone via text or messenger but may be too nervous in person and those nerves could show.\n\nPersonally, I have had a ton of experience as I am now 30 years old. I felt that over time I got less nervous. It's kind of like anything. The more you do something the more confident you get. I remember my first date I actually spoke to a girl online a ton. Meeting her I was so nervous because I really liked her and felt she liked me as well. We finally met and things went pretty smoothly, much better than I was expecting considering how nervous I was. I look back at that first date and laugh because I was putting so much pressure on myself for no reason.\n\nTechnology lets us break the ice and then ultimately when you're on the date you just have to be yourself. The other person is probably just as nervous as you are. After all, nobody wants to embarrass themselves or come across negatively on a first date. Of course there are first dates that just don't go well. That's a part of life."]
AI-generated suggestions: ['technology enables us to connect in ways that we otherwise wouldn't be able to.', 'Its availability in cities, and ease of use has increased the pool of potential romantic partners, makes finding people from different backgrounds and interests much easier.', 'I think technology does a really good job in helping alleviate anxiety, but it never fully prepares you for a first impression.', 'And believe me, most of the time I wish I had just remained friends with the person we went out with rather than actually going on a date and']

Successfully read 2474 events in a writing session from ./coauthor-v1.0/bf03f184f45441258d7e0658a1742332.jsonl
User-written sentences: ['This, after all, is not the first time that Matt Damon has gone missing.', 'The first time Matt Damon went missing, humanity had practically stopped in its tracks to try to find him.', 'The alien is not aware that Matt Damon is a very important person to the human race.\n This is because Matt Damon has incredible talents. He is funny, personable and can even cast magical spells.', 'He is also a master of disguise and can be anyone he wants to be. For example, Matt Damon can morph into a clown, or a doctor, or a lawyer, or a schoolteacher, or anyone else depending on his mood.', 'With all these talents, skills and abilities, Matt Damon is arguably the most powerful human on planet Earth.\n\nSo when the alien kidnaps Matt Damon, he really has no idea what he is up against.', "He is unaware that the human race will stop at nothing to retrieve Matt Damon. In fact, a group of humans based on Switzerland is immediately notified of Matt Damon's disappearance.", 'They are called the "Matt Damon Retrieval Team" and they are the most elite group of humans on the planet. This Swiss team has a lot of powerful weapons at their disposal to try to get Matt Damon back on planet Earth.', 'The alien is not aware of this, but he is about to get a rude awakening.\n The Swiss team sets out to find the alien in a spaceship. They set their sights on the planet where they know the alien is: Planet Borg. Planet Borg is zillions of lightyears away from the Swiss, but they are totally undeterred. They truly stop at nothing for Matt Damon. And so, as a result, the alien is not prepared for the blasts of rain and fire that he is hit with all of a sudden on Planet Borg. The Swiss have arrived!', 'The Swiss team is not going to let the alien have his way with Matt Damon.', 'They are going to get Matt Damon back, no matter what.', 'They are going to get Matt Damon back, even if they have to destroy Planet Borg in the process. Matt Damon is relieved to see the Swiss. After all, the alien had been confining him to a small isolation room, where he was only allowed to eat hot dogs and watch the same Netflix show over and over again, on repeat.', "The alien had been very cruel to Matt Damon. So when the Swiss start blasting Planet Borg, Matt Damon knows he's in good hands. He is promptly freed and gives the alien a very sarcastic goodbye.", 'The alien is not happy.']
AI-generated suggestions: ['The alien is not aware that Matt Damon is a very important person to the human race.', 'He is also a master of disguise and can be anyone he wants to be.', 'He is unaware that the human race will stop at nothing to retrieve Matt Damon.', 'They are called the "Matt Damon Retrieval Team" and they are the most elite group of humans on the planet.', 'The alien is not aware of this, but he is about to get a rude awakening.', 'The Swiss team is not going to let the alien have his way with Matt Damon.', 'They are going to get Matt Damon back, no matter what.', 'They are going to get Matt Damon back, even if they have to destroy Planet Borg in the process.', 'The alien had been very cruel to Matt Damon.', 'The alien is not happy.',

```
'He is not happy that he has been defeated by the Swiss.']
Processed 1447 paths. Results saved to emotion_similarity_results.csv.

def extract_acceptance_status(events):
    """Track if a suggestion was accepted or rejected."""
    acceptance_status = []  # Stores 'accepted' or 'rejected'
    last_suggestion_open = None  # Store the most recent `suggestion-
open`

    for event in events:
        if event["eventName"] == "suggestion-open" and
event.get("currentSuggestions"):
            last_suggestion_open = event  # Save the latest
`suggestion-open`

        elif event["eventName"] == "suggestion-select" and
last_suggestion_open:
            # Suggestion selected
            acceptance_status.append('accepted')
            last_suggestion_open = None  # Reset, since we handled the
acceptance

        elif event["eventName"] == "suggestion-close" and
last_suggestion_open:
            # Suggestion closed without selection (rejected)
            acceptance_status.append('rejected')
            last_suggestion_open = None  # Reset, since we handled the
rejection

    return acceptance_status

# Loop through each path (limited to paths 100)
jsd_results = []
for path in paths[:100]:
    events = read_writing_session(path)

    if len(events) < 3000:

        # Extract user-written text and AI-generated suggestions
        user_texts = reconstruct_user_text(events)
        ai_suggestions = extract_selected_ai_suggestions(events)
        acceptance_status = extract_acceptance_status(events)

        # Ensure we only compare matching pairs
        for user_text, ai_text, status in zip(user_texts,
ai_suggestions, acceptance_status):
            if len(ai_text.split()) < 6:
                continue
            user_score = syntactic_complexity(user_text)
            ai_score = syntactic_complexity(ai_text)
```

```python
        coverage = process_text(ai_text)['coverage']
        # Store results, including acceptance status
        jsd_results.append({
            "path": path,
            "user_text": user_text,
            "ai_suggestion": ai_text,
            "acceptance_status": status,
            "user_score": user_score,
            "ai_score": ai_score,
            "coverage": coverage
        })
```

Successfully read 2405 events in a writing session from ./coauthor-v1.0/608e5b73341a4f3ca937316f99dae32d.jsonl
Successfully read 2830 events in a writing session from ./coauthor-v1.0/2d9358657a364ccab47e9a7538cb6650.jsonl
Successfully read 1075 events in a writing session from ./coauthor-v1.0/8c11358444974bf0b5224183acd8149d.jsonl
Successfully read 1377 events in a writing session from ./coauthor-v1.0/ea10c484cd6245f6a5840eea4b4b143b.jsonl
Successfully read 2381 events in a writing session from ./coauthor-v1.0/87396211fee244ccbda4b0f4bcefc8e9.jsonl
Successfully read 1312 events in a writing session from ./coauthor-v1.0/f4d98feb9e2f478abc3817bdfef5cca0.jsonl
Successfully read 1968 events in a writing session from ./coauthor-v1.0/6938b4a468aa482ba459d2fbe140cb30.jsonl
Successfully read 2910 events in a writing session from ./coauthor-v1.0/b609372842a94db2b278d4ad7f7e366b.jsonl
Successfully read 2329 events in a writing session from ./coauthor-v1.0/7b7b1b40027c4974ade1850596e6726f.jsonl
Successfully read 2250 events in a writing session from ./coauthor-v1.0/ee15d49584a54e8ab79b140e400276ea.jsonl
Successfully read 2262 events in a writing session from ./coauthor-v1.0/ba6d32a6c29a4bfeacf1e25a87a1f804.jsonl
Successfully read 1615 events in a writing session from ./coauthor-v1.0/0ff12af0651e4b0eaf88e22e82058137.jsonl
Successfully read 863 events in a writing session from ./coauthor-v1.0/9c4942cdf0d94b6d9ce49896e0093da6.jsonl
Successfully read 1803 events in a writing session from ./coauthor-v1.0/09770cd9061f456fa6d82db00b43a0bd.jsonl
Successfully read 2178 events in a writing session from ./coauthor-v1.0/9ec10cba9d8c48428d656fceff57e834.jsonl
Successfully read 2235 events in a writing session from ./coauthor-v1.0/e7107e7319eb4ab18a7e3fc4445b2d91.jsonl
Successfully read 2947 events in a writing session from ./coauthor-v1.0/eec20f24484d418eb4ba564fbfbc2ab6.jsonl
Successfully read 2805 events in a writing session from ./coauthor-v1.0/6d6a8d63190243f685256031a59166dd.jsonl
Successfully read 1357 events in a writing session from ./coauthor-v1.0/a87286598bc54313b39f6ce6013541bf.jsonl

Successfully read 1626 events in a writing session from ./coauthor-v1.0/a215d4b3242a4e999ee6fafdd793cac6.jsonl
Successfully read 2433 events in a writing session from ./coauthor-v1.0/79a6126076714a1ba4fc9abc52c9201d.jsonl
Successfully read 2397 events in a writing session from ./coauthor-v1.0/f0b4803428a846c6b22e2d420e980a33.jsonl
Successfully read 1516 events in a writing session from ./coauthor-v1.0/1ec7d03923814ad590d270c260b37fad.jsonl
Successfully read 1194 events in a writing session from ./coauthor-v1.0/e29599231e66474b8a524c23db775e47.jsonl
Successfully read 3366 events in a writing session from ./coauthor-v1.0/0a83297c1ce94d15b358d79e533396c7.jsonl
Successfully read 1078 events in a writing session from ./coauthor-v1.0/6fb96c88e7b942e6b225baa611dd9738.jsonl
Successfully read 1877 events in a writing session from ./coauthor-v1.0/0bcb6a7180d64d7c9af732f3ad5aff51.jsonl
Successfully read 2686 events in a writing session from ./coauthor-v1.0/3be583425f4c4656a906a036b13e963e.jsonl
Successfully read 977 events in a writing session from ./coauthor-v1.0/d4b47a8a66b84e8ebdf986e6d4edf139.jsonl
Successfully read 1809 events in a writing session from ./coauthor-v1.0/3959b06dd9344c6aa484b3d6b14b4a27.jsonl
Successfully read 1966 events in a writing session from ./coauthor-v1.0/c1714313080c448d857756a2e0ad6071.jsonl
Successfully read 1537 events in a writing session from ./coauthor-v1.0/ace79657baa04aaca50a076720ee1fe8.jsonl
Successfully read 1021 events in a writing session from ./coauthor-v1.0/36bc101319da4b3590b96bce76f7c02c.jsonl
Successfully read 2355 events in a writing session from ./coauthor-v1.0/8393a8ed335f438c9ad5ca3d25a947b6.jsonl
Successfully read 2155 events in a writing session from ./coauthor-v1.0/b768132809f240b2a3efa1719be82ae4.jsonl
Successfully read 1033 events in a writing session from ./coauthor-v1.0/e2cdf3ffc1fb4396a67368c8331d4f58.jsonl
Successfully read 1830 events in a writing session from ./coauthor-v1.0/423daa7fb2214d1093f73d4966240646.jsonl
Successfully read 2828 events in a writing session from ./coauthor-v1.0/d714bc54e0dc4819bafa77a93927817b.jsonl
Successfully read 2573 events in a writing session from ./coauthor-v1.0/105bf88bb4bc42688e06a54644e2989b.jsonl
Successfully read 2257 events in a writing session from ./coauthor-v1.0/09644506674341f5b14e0d62c88ee70a.jsonl
Successfully read 1902 events in a writing session from ./coauthor-v1.0/b30082b36e764dbd859c020b10fe78b5.jsonl
Successfully read 1361 events in a writing session from ./coauthor-v1.0/8d6e9dac68324fae884571881f5b21f1.jsonl
Successfully read 2672 events in a writing session from ./coauthor-v1.0/25742b852b764843b15266008afdc15d.jsonl
Successfully read 2426 events in a writing session from ./coauthor-

v1.0/b65e8358d12a498fa442f3d1072c4bb2.jsonl
Successfully read 1266 events in a writing session from ./coauthor-
v1.0/de90bdd73b72426daacc29cd1e1a2e97.jsonl
Successfully read 1399 events in a writing session from ./coauthor-
v1.0/b75ef56faed74be0b4b59e15138e18f4.jsonl
Successfully read 1603 events in a writing session from ./coauthor-
v1.0/adb2b2bbd3fb48fa8433ee84a26949a5.jsonl
Successfully read 1621 events in a writing session from ./coauthor-
v1.0/3cfab79e841a41a7b6edf77e7b28886b.jsonl
Successfully read 1851 events in a writing session from ./coauthor-
v1.0/e8cdd0418ddb4263ae2799ded6316693.jsonl
Successfully read 2560 events in a writing session from ./coauthor-
v1.0/5773114053f147a69bd6790843ea4fd8.jsonl
Successfully read 1590 events in a writing session from ./coauthor-
v1.0/d374cad6be364ca08df8064807b4696e.jsonl
Successfully read 3397 events in a writing session from ./coauthor-
v1.0/c326ab6f04ca475a93d6322122ee6eb0.jsonl
Successfully read 2085 events in a writing session from ./coauthor-
v1.0/2194c0e020934c74b8ab6c28d758feb2.jsonl
Successfully read 1968 events in a writing session from ./coauthor-
v1.0/3969517100834dbc813b628a27fed790.jsonl
Successfully read 1498 events in a writing session from ./coauthor-
v1.0/749e032914384e4da8f2392a19066b89.jsonl
Successfully read 2062 events in a writing session from ./coauthor-
v1.0/876a292223a74b1a8ad5525d93b26d69.jsonl
Successfully read 708 events in a writing session from ./coauthor-
v1.0/8d5bc658213f46b881b2a862bfd5a105.jsonl
Successfully read 1787 events in a writing session from ./coauthor-
v1.0/f84ff332dc06486c9d61698be8666941.jsonl
Successfully read 1733 events in a writing session from ./coauthor-
v1.0/c0bda7b135404137b66b5ade3bf1c8bc.jsonl
Successfully read 805 events in a writing session from ./coauthor-
v1.0/8637ce05b848420f920d5ca02f88a89a.jsonl
Successfully read 1839 events in a writing session from ./coauthor-
v1.0/29056a31430b414582f488535aa0fe91.jsonl
Successfully read 1476 events in a writing session from ./coauthor-
v1.0/409ea6cd2f8347dbb41344d1243c7e17.jsonl
Successfully read 2005 events in a writing session from ./coauthor-
v1.0/d4a257392aa84174be6166c9c664caae.jsonl
Successfully read 1227 events in a writing session from ./coauthor-
v1.0/9e408240b6c34495b3d20597eaa1dd80.jsonl
Successfully read 2045 events in a writing session from ./coauthor-
v1.0/09c68920f9e5434bb35a22e1909f324a.jsonl
Successfully read 1502 events in a writing session from ./coauthor-
v1.0/4efb9d8419ee42a3b2ec99f01f94e0a6.jsonl
Successfully read 1247 events in a writing session from ./coauthor-
v1.0/08b1a2012f4d4115b56cc8231449ff48.jsonl
Successfully read 1432 events in a writing session from ./coauthor-
v1.0/66b3ac872822426e81bab531fb15d8a0.jsonl

Successfully read 2619 events in a writing session from ./coauthor-v1.0/5c1905e6a05c4203b624c4db10ede78a.jsonl
Successfully read 2564 events in a writing session from ./coauthor-v1.0/dcb51651d76c424ebd58a8802629f05c.jsonl
Successfully read 1925 events in a writing session from ./coauthor-v1.0/7e400fae43e543b0a6719cc2974029de.jsonl
Successfully read 955 events in a writing session from ./coauthor-v1.0/1059f00ce0f042808d27139c2a2ca9c5.jsonl
Successfully read 1757 events in a writing session from ./coauthor-v1.0/f6c2fa4230df43b9b9857e71b9d68a89.jsonl
Successfully read 1377 events in a writing session from ./coauthor-v1.0/931c030acd5e41b1bfac856da1aeb8e3.jsonl
Successfully read 2762 events in a writing session from ./coauthor-v1.0/13b2ae3d425e4b76b887182418ba4f58.jsonl
Successfully read 1275 events in a writing session from ./coauthor-v1.0/5f020c0bdea94db5aad0478ba02f2924.jsonl
Successfully read 1494 events in a writing session from ./coauthor-v1.0/f151863fbea2417c9bbec7082abaefd6.jsonl
Successfully read 1776 events in a writing session from ./coauthor-v1.0/40d132e1bc0146c0920d7b38c5b9f7c7.jsonl
Successfully read 1893 events in a writing session from ./coauthor-v1.0/fe677c513394449e987eb3ccce62d03c.jsonl
Successfully read 1355 events in a writing session from ./coauthor-v1.0/fbd087c5c3ab48f1b6540c80976e3285.jsonl
Successfully read 2196 events in a writing session from ./coauthor-v1.0/0a08f7fd748246c8ad4e353edb7e2394.jsonl
Successfully read 3177 events in a writing session from ./coauthor-v1.0/4d06b9aa68764d749ef6fffb4fcb54f4.jsonl
Successfully read 1805 events in a writing session from ./coauthor-v1.0/65936cbe394149c6b24802cad630e2d7.jsonl
Successfully read 975 events in a writing session from ./coauthor-v1.0/c790c4b915e84dbf8e283b598105ea90.jsonl
Successfully read 1725 events in a writing session from ./coauthor-v1.0/4181c7d35eff4eceb4aad545dd2a9273.jsonl
Successfully read 1190 events in a writing session from ./coauthor-v1.0/787e032a5802459e9ac5731cdeb7b077.jsonl
Successfully read 1243 events in a writing session from ./coauthor-v1.0/a23600c207bd470c929852b33b9a2946.jsonl
Successfully read 1259 events in a writing session from ./coauthor-v1.0/864b77f0d5f34f2b9946e518cdeee98a.jsonl
Successfully read 3376 events in a writing session from ./coauthor-v1.0/f3ecf0ce872940238b0bd429a3e36322.jsonl
Successfully read 460 events in a writing session from ./coauthor-v1.0/1524055efcec4b43b192c6b706e44b1e.jsonl
Successfully read 935 events in a writing session from ./coauthor-v1.0/b15bdf5fdb764c3b97f9ec4ab3d348e3.jsonl
Successfully read 1986 events in a writing session from ./coauthor-v1.0/36eead66360847afba6dc9ebf6d99593.jsonl
Successfully read 2125 events in a writing session from ./coauthor-

```
v1.0/3c2459fc295b4836b027ff004c6b88c1.jsonl
Successfully read 3457 events in a writing session from ./coauthor-
v1.0/5de04cda76194f70ba4a49ba5c34ad14.jsonl
Successfully read 1729 events in a writing session from ./coauthor-
v1.0/a9aa499623e3453dbcd52de1c862f71b.jsonl
Successfully read 2966 events in a writing session from ./coauthor-
v1.0/64e545e776c1466687f0fe0dc8bb0d88.jsonl
Successfully read 1309 events in a writing session from ./coauthor-
v1.0/947065accb8b402c98fb2aa62a497842.jsonl
Successfully read 3460 events in a writing session from ./coauthor-
v1.0/4bdb5a30d17d40d7b160d942552c190b.jsonl
Successfully read 1689 events in a writing session from ./coauthor-
v1.0/c95024d4d29d4d90a8aa665aef6397aa.jsonl
Successfully read 1540 events in a writing session from ./coauthor-
v1.0/ff21eabc9606411ba8cd583fdcdfe72c.jsonl

df_results_with_acceptances = pd.DataFrame(jsd_results)
df_results_with_acceptances.to_csv("creativity1.csv", index=False)

print(f"Processed. Results saved to creativity1.csv.")

Processed. Results saved to creativity1.csv.
```

show 5 highest syntactical complexity.

```python
# highest and lowest 5 similarity scores and text
#highest five after similarity of 1
highest_5 =
df_results_with_acceptances[df_results_with_acceptances['ai_score'] <
1].nlargest(5, 'ai_score')
lowest_5 = df_results_with_acceptances.nsmallest(5, 'ai_score')
print("Highest 5 Similarity Scores:")
for index, row in highest_5.iterrows():
    print(f"Score: {row['ai_score']}")
    print(f"User Text: {row['user_text']}")
    print(f"AI Suggestion: {row['ai_suggestion']}")


print("\nLowest 5 Similarity Scores:")
for index, row in lowest_5.iterrows():
    print(f"Score: {row['ai_score']}")
    print(f"User Text: {row['user_text']}")
    print(f"AI Suggestion: {row['ai_suggestion']}")
highest_5cover =
df_results_with_acceptances[df_results_with_acceptances['coverage'] <
1].nlargest(5, 'coverage')
lowest_5cover= df_results_with_acceptances.nsmallest(5, 'coverage')
print("Highest 5 coverages:")
for index, row in highest_5.iterrows():
    print(f"Score: {row['coverage']}")
```

```
    print(f"User Text: {row['user_text']}")
    print(f"AI Suggestion: {row['ai_suggestion']}")

print("lowest 5 coverages:")
for index, row in lowest_5cover.iterrows():
    print(f"Score: {row['coverage']}")
    print(f"User Text: {row['user_text']}")
    print(f"AI Suggestion: {row['ai_suggestion']}")
```

Highest 5 Similarity Scores:
Score: 0.9666666666666667
User Text: Obama gasps and cannot believe what he sees.
AI Suggestion: JFK goes on to tell Obama that the society is governed by a group of twelve men (all former presidents), he is one of them
Score: 0.9
User Text: that they're just too much of it, with the news being reported daily, is that the news has become more biased than ever. A person that is not keeping up with the news is more responsible and more
AI Suggestion: inclined to be a well-rounded citizen, but what's happening now is that uninformed citizens are being taken advantage of by the media and the government
Score: 0.9
User Text: I do not believe that schools should provide free pads and tampons for students. The school should not have to worry about feminine products for the students. This should be the responsiblity of the guardiansiparental student's.
AI Suggestion: I believe that pads and tampons should be treated like toilet paper, soap, band-aids and other products that are already provided
Score: 0.9
User Text: The shapeshifter himself has always been very talented, and is a doctor by trade. He has decided to shape shift into a more handsome doctor, kind of like a George Clooney from ER type, because he knows she loves the show. The shapeshifter, Jim, has done his research and found out that the woman, Karen, goes to her primary care physician at the Coastal Medical institute.
AI Suggestion: He also knows that she's been going to Dr. John, who is a specialist, for her asthma.
Score: 0.9
User Text: Two men with black suits come up in a black car and tell Obama to get in.
AI Suggestion: He tells him that he was not the only president who was apart of the society, but since his retirement, Obama is now being asked to join.

Lowest 5 Similarity Scores:
Score: 0.03333333333333333
User Text: "Mom, I've been trying to get a girl to like me, but nothing seems to wor"
```

AI Suggestion: Why don't you just be yourself?"
Score: 0.04444444444444446
User Text: "Oh, thank goodness you're back!" says Matt Damon. "How did
you do that? And where did you go?"
AI Suggestion: "Oh, thank goodness you're back!" says Matt Damon.
Score: 0.06666666666666667
User Text: From there it was a mad dash back to the bunker, followed
by days of hushed talk and whispered rumors.
 The kid could do things now, but he had almost no control over them.
And peopel started talking about the old books, how all the little
tricks people could do - light a candle, make a gift invisible, etc. -
used to be real poeople So other peopel started touching the stone,.
AI Suggestion: And they started feeling things too.
Score: 0.06666666666666667
User Text: and they will check it to see if they have missed
anything.sneak away    The pandemic has made thi worse,
AI Suggestion: , are not a good thing.
Score: 0.06666666666666667
User Text: DAD #1 DAD #2 DAD.
AI Suggestion: DAD #1 DAD #2 DAD #3
Highest 5 coverages:
Score: 0.7142857142857143
User Text: Obama gasps and cannot believe what he sees.
AI Suggestion: JFK goes on to tell Obama that the society is governed
by a group of twelve men (all former presidents), he is one of them
Score: 0.9230769230769231
User Text: that they're just too much of it, with the news being
reported daily, is that the news has become more biased than ever. A
person that is not keeping up with the news is more responsible and
more
AI Suggestion: inclined to be a well-rounded citizen, but what's
happening now is that uninformed citizens are being taken advantage of
by the media and the government
Score: 0.8260869565217391
User Text: I do not believe that schools should provide free pads and
tampons for students.  The school should not have to worry about
feminine products for the students.  This should be the responsiblity
of the guardiansiparental  student's.
AI Suggestion: I believe that pads and tampons should be treated like
toilet paper, soap, band-aids and other products that are already
provided
Score: 0.5714285714285714
User Text: The shapeshifter himself has always been very talented, and
is a doctor by trade. He has decided to shape shift into a more
handsome doctor, kind of like a George Clooney from ER type, because
he knows she loves the show. The shapeshifter, Jim, has done his
research and found out that the woman, Karen, goes to her primary care
physician at the Coastal Medical institute.
AI Suggestion: He also knows that she's been going to Dr. John, who is

a specialist, for her asthma.
Score: 0.833333333333334
User Text: Two men with black suits come up in a black car and tell
Obama to get in.
AI Suggestion: He tells him that he was not the only president who was
apart of the society, but since his retirement, Obama is now being
asked to join.
lowest 5 coverages:
Score: 0.0
User Text: The steady hum of the power generators, the quiet
conversation of her fellow survivors, and a familiar loping thump as
Herman ran by outside. It seemed like just any other day. But she knew
it wasn't. It was time.
AI Suggestion: With a heavy sigh, Morgan pushed herself up.
Score: 0.0
User Text: From there it was a mad dash back to the bunker, followed
by days of hushed talk and whispered rumors.
 The kid could do things now, but he had almost no control over them.
And peopel started talking about the old books, how all the little
tricks people could do - light a candle, make a gift invisible, etc. -
used to be real poeople So other peopel started touching the stone,.
AI Suggestion: And they started feeling things too.
Score: 0.0
User Text: But it what way? Before...she was always a bit strongr with
her trickse, almost powerful by t standards.
AI Suggestion: Was she about to become...something else?
Score: 0.0
User Text: Whatever the reason, Morgan was grateful not to have to
worry about any other distractions. SHe ignored the one observer in
the room, and walked straight to the stone.
AI Suggestion: as her hand connected, and a pulse of light flowed out.
Score: 0.0
User Text: it seems like people portray characters of Asian descent
offensively far too often.
AI Suggestion: Asian women are portrayed as either a submissive exotic
beauty or as an overbearing femme fatale.

```python
import matplotlib.pyplot as plt
# Statistical analysis of ai_score
ai_score_stats = df_results_with_acceptances['ai_score'].describe()
print("AI Score Statistics:")
print(ai_score_stats)

# Group ai_score by acceptance status
ai_score_by_acceptance =
df_results_with_acceptances.groupby('acceptance_status')
['ai_score'].describe()
print("\nAI Score by Acceptance Status:")
print(ai_score_by_acceptance)
```

```python
# Plotting ai_score distribution by acceptance status (enhanced)
plt.figure(figsize=(10, 6))
plt.hist(df_results_with_acceptances[df_results_with_acceptances['acce
ptance_status'] == 'accepted']['ai_score'],
         bins=35, color='green', alpha=0.7, label='Accepted')
plt.hist(df_results_with_acceptances[df_results_with_acceptances['acce
ptance_status'] == 'rejected']['ai_score'],
         bins=35, color='red', alpha=0.7, label='Rejected')

# Add descriptive statistics to the plot
accepted_ai_mean =
df_results_with_acceptances[df_results_with_acceptances['acceptance_st
atus'] == 'accepted']['ai_score'].mean()
rejected_ai_mean =
df_results_with_acceptances[df_results_with_acceptances['acceptance_st
atus'] == 'rejected']['ai_score'].mean()
plt.axvline(accepted_ai_mean, color='green', linestyle='dashed',
linewidth=1,
            label=f'Accepted Mean: {accepted_ai_mean:.2f}')
plt.axvline(rejected_ai_mean, color='red', linestyle='dashed',
linewidth=1,
            label=f'Rejected Mean: {rejected_ai_mean:.2f}')

plt.title('Distribution of AI Score by Acceptance Status')
plt.xlabel('AI Score')
plt.ylabel('Frequency')
plt.legend()
plt.grid(True)  # add gridlines for better readability
plt.show()
```

```
AI Score Statistics:
count    668.000000
mean       0.319428
std        0.205482
min        0.033333
25%        0.133333
50%        0.266667
75%        0.466667
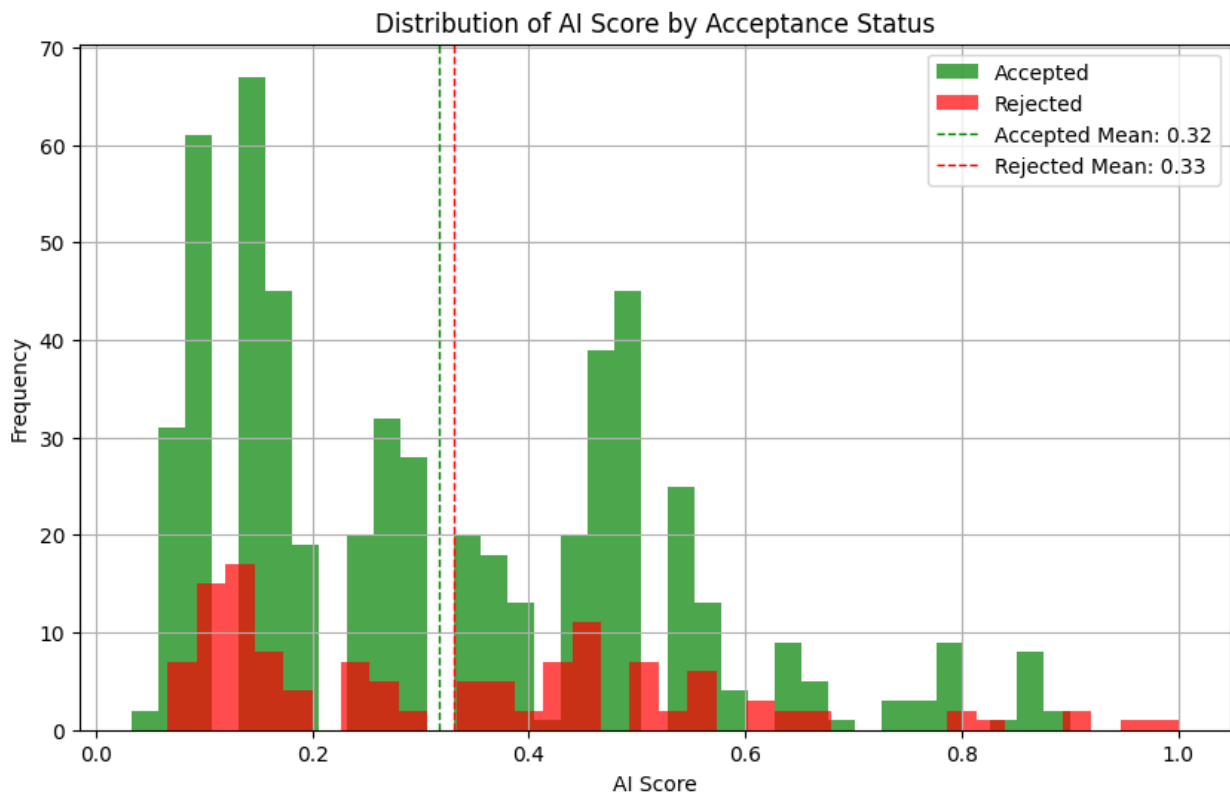max        1.000000
Name: ai_score, dtype: float64

AI Score by Acceptance Status:
                   count       mean        std        min        25%
50%  \
acceptance_status

accepted           544.0   0.316932   0.201711   0.033333   0.133333
0.266667
rejected           124.0   0.330376   0.221802   0.066667   0.133333
0.266667
```

```
                            75%   max
acceptance_status
accepted             0.466667   0.9
rejected             0.466667   1.0
```


Distribution of AI Score by Acceptance Status

```
# Statistical analysis of coverage
coverage_stats = df_results_with_acceptances['coverage'].describe()
print("Coverage Statistics:")
print(coverage_stats)

# Group coverage by acceptance status
coverage_by_acceptance =
df_results_with_acceptances.groupby('acceptance_status')
['coverage'].describe()
print("\nCoverage by Acceptance Status:")
print(coverage_by_acceptance)


# Plotting coverage distribution by acceptance status (enhanced)
plt.figure(figsize=(10, 6))
plt.hist(df_results_with_acceptances[df_results_with_acceptances['acce
ptance_status'] == 'accepted']['coverage'], bins=20, color='green',
alpha=0.7, label='Accepted')
```

```python
plt.hist(df_results_with_acceptances[df_results_with_acceptances['acce
ptance_status'] == 'rejected']['coverage'], bins=20, color='red',
alpha=0.7, label='Rejected')

# Add descriptive statistics to the plot
accepted_mean =
df_results_with_acceptances[df_results_with_acceptances['acceptance_st
atus'] == 'accepted']['coverage'].mean()
rejected_mean =
df_results_with_acceptances[df_results_with_acceptances['acceptance_st
atus'] == 'rejected']['coverage'].mean()
plt.axvline(accepted_mean, color='green', linestyle='dashed',
linewidth=1, label=f'Accepted Mean: {accepted_mean:.2f}')
plt.axvline(rejected_mean, color='red', linestyle='dashed',
linewidth=1, label=f'Rejected Mean: {rejected_mean:.2f}')

plt.title('Distribution of Coverage by Acceptance Status')
plt.xlabel('AI Coverage')
plt.ylabel('Frequency')
plt.legend()
plt.grid(True) # add gridlines for better readability
plt.show()


# Correlation analysis (example)
correlation =
df_results_with_acceptances['coverage'].corr(df_results_with_acceptanc
es['ai_score'])
print(f"\nCorrelation between Coverage and AI Score: {correlation}")

Coverage Statistics:
count    668.000000
mean       0.717964
std        0.303438
min        0.000000
25%        0.587010
50%        0.800000
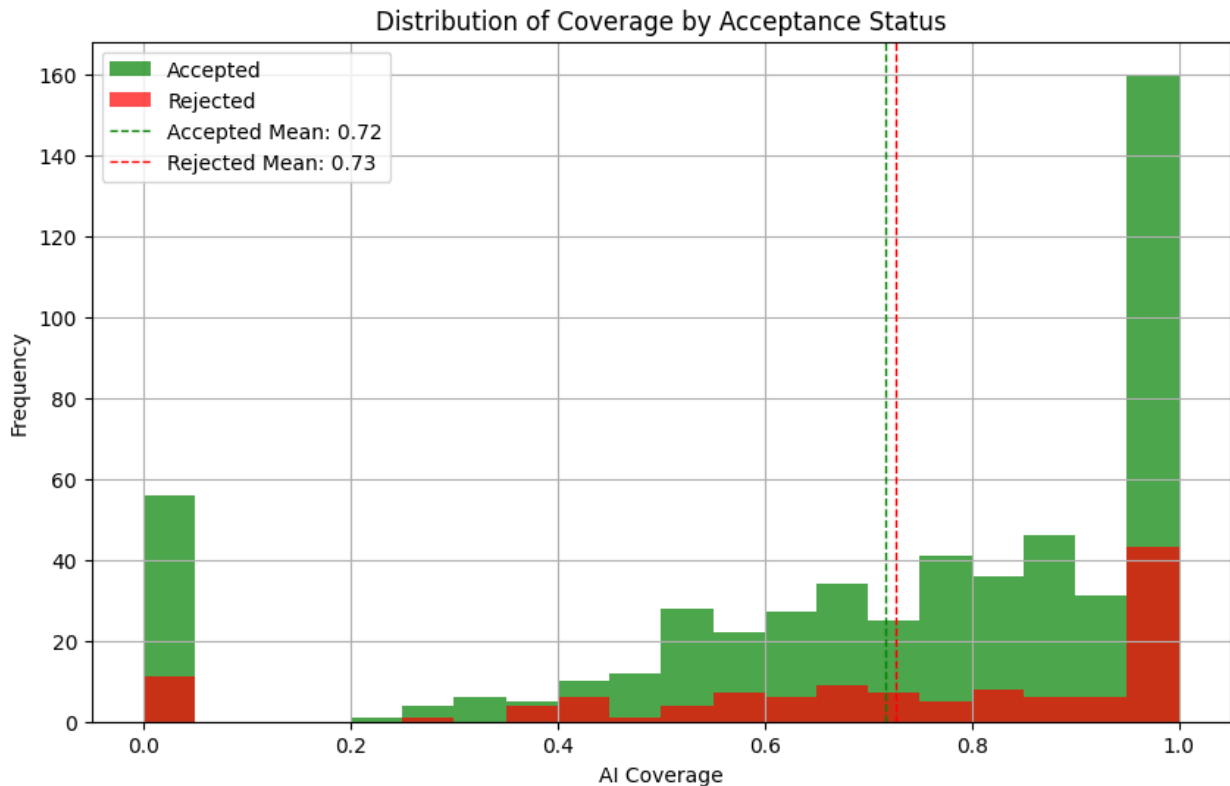75%        1.000000
max        1.000000
Name: coverage, dtype: float64

Coverage by Acceptance Status:
                  count       mean       std  min       25%  50%  75%
max
acceptance_status

accepted          544.0  0.716087  0.304044  0.0  0.587010  0.8  1.0
1.0
rejected          124.0  0.726199  0.301852  0.0  0.595833  0.8  1.0
1.0
```

Distribution of Coverage by Acceptance Status

Correlation between Coverage and AI Score: 0.11951377105708061

too high coverage might be not good for analysis

```
# Filter out rows where coverage > 0.9
df_filtered =
df_results_with_acceptances[df_results_with_acceptances['coverage'] <=
0.9]

# Statistical analysis of coverage
coverage_stats = df_filtered['coverage'].describe()
print("Coverage Statistics:")
print(coverage_stats)

# Group coverage by acceptance status
coverage_by_acceptance = df_filtered.groupby('acceptance_status')
['coverage'].describe()
print("\nCoverage by Acceptance Status:")
print(coverage_by_acceptance)

# Plotting coverage distribution by acceptance status (enhanced)
plt.figure(figsize=(10, 6))
plt.hist(df_filtered[df_filtered['acceptance_status'] == 'accepted']
['coverage'], bins=20, color='green', alpha=0.7, label='Accepted')
```

```python
plt.hist(df_filtered[df_filtered['acceptance_status'] == 'rejected']
['coverage'], bins=20, color='red', alpha=0.7, label='Rejected')

# Add descriptive statistics to the plot
accepted_mean = df_filtered[df_filtered['acceptance_status'] ==
'accepted']['coverage'].mean()
rejected_mean = df_filtered[df_filtered['acceptance_status'] ==
'rejected']['coverage'].mean()
plt.axvline(accepted_mean, color='green', linestyle='dashed',
linewidth=1, label=f'Accepted Mean: {accepted_mean:.2f}')
plt.axvline(rejected_mean, color='red', linestyle='dashed',
linewidth=1, label=f'Rejected Mean: {rejected_mean:.2f}')

plt.title('Distribution of Coverage by Acceptance Status')
plt.xlabel('AI Coverage')
plt.ylabel('Frequency')
plt.legend()
plt.grid(True)
plt.show()

# Correlation analysis (example)
correlation = df_filtered['coverage'].corr(df_filtered['ai_score'])
print(f"\nCorrelation between Coverage and AI Score: {correlation}")
```

```
Coverage Statistics:
count    431.000000
mean       0.569908
std        0.283726
min        0.000000
25%        0.461538
50%        0.666667
75%        0.777778
max        0.900000
Name: coverage, dtype: float64

Coverage by Acceptance Status:
                  count      mean       std  min       25%       50%
\
acceptance_status

accepted          356.0  0.573186  0.285929  0.0  0.465385  0.666667

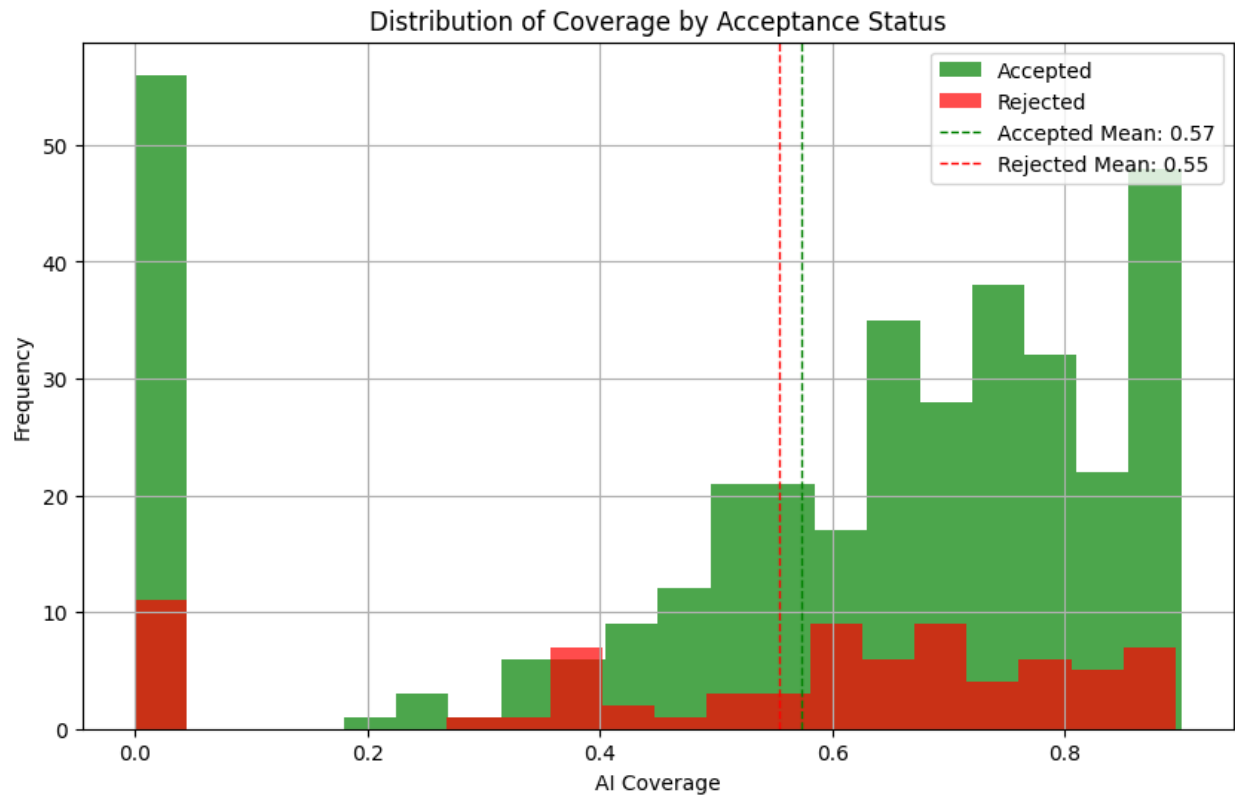rejected           75.0  0.554352  0.274353  0.0  0.400000  0.625000


                      75%       max
acceptance_status
accepted          0.783385  0.900000
rejected          0.743421  0.894737
```

Distribution of Coverage by Acceptance Status

Correlation between Coverage and AI Score: 0.21174872318229554