

Generative AI

Katrina Kirby

LISUM43

4/27/2025

Generative AI has been used and created to take raw data and learn to generate statistically probable outcomes when prompted. It is a deep-learning model that has been trained to generate these outputs. It can create a variety of content such as text, images, audio, and even video and it learns how to do all of this from existing data. According to Zewe (2023), “these powerful machine-learning models draw on research and computational advances that go back more than 50 years.”

Generative AI has been created to help people work faster and more efficiently while also being creative. We have so many tools right now such as ChatGPT, OpenAI, HuggingFace, and many more. Users can use this software to do just about anything. For example, users can go into ChatGPT to create prompts and ask questions on something they want to learn about or gather more information on. You can ask AI to create a poem, song, paper, and much more. With the help of analyzing data, coding, and math, AI can do so much.

There are a variety of GenAI tools that developers can use such as Python, PyTorch, Midjourney, DeepAI, and others. A lot of researchers have credited these GenAI tools to be “critical to the future of science communication and data analysis” while others have advised caution in the use of GenAI and the tools it offers. When developers use and produce their work with GenAI in journals, they require specific disclosure of how GenAI tools were used. There has been concern on the use of GenAI to support peer-reviewing papers because this can be seen

as a breach of confidentiality. GenAI tools are typically trained through reinforcement learning which allows for continuous ingesting of data to refine the model.

Large Language Models otherwise known as LLMs are the brains behind GenAI and help them work with languages that help develop the outcomes like writing essays, answering questions, or chatting to the AI like a human. These are powerful AI systems that have been trained on large amounts of text such as books, articles, and websites. Software engineers train these models using programming tools such as Python and machine learning libraries like PyTorch.

Diffusion Models are used to help create realistic images from scratch with the help of GenAI. These models essentially start from scratch and with a process of many steps, it can form a creative image based on what the model learned in training. GANs also known as Generative Adversarial Networks is like two artists competing, one tries to create fakes and the other tries to catch fakes. This is another way Generative AI learns to generate images, videos, and music. If we were looking at GANs in the context of creating an image, one part is trying to make the perfect picture while the other is trying to spot any mistakes. These have been used to generate photorealistic human faces, deepfake videos, and even piece of art.

Python is a primary programming language that has been used for AI and machine learning projects. It has a massive library, strong community, and it is connected to all parts of GenAI from training models to handling data and deploying applications. When developers train models, it is always important to prepare, clean, and organize the data that they are using. Pandas is another powerful library that helps developers handle large datasets so that they can clean and transform them to make sure the information is ready for the AI to be trained on and learn. Sklearn is very important for smaller tasks in AI projects such as data preprocessing, feature

selection, and building simpler models. Sklearn also known as scikit-learn offers a fast and reliable tool to help prepare and experiment before going to LLMs. All together, these tools are an essential part to Generative AI.

Here we can see a brief snippet of my own example of code involving some of the tools we have gone over:

```
[4]: !pip install scikit-learn

Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: scikit-learn in c:\users\katki\appdata\local\packages\pythonsoftwarefoundation.python.3.13_qbz5n2kfra8p0\localcache\local-packages\python313\site-packages (1.6.1)
Requirement already satisfied: numpy>=1.19.5 in c:\users\katki\appdata\local\packages\pythonsoftwarefoundation.python.3.13_qbz5n2kfra8p0\localcache\local-packages\python313\site-packages (from scikit-learn) (2.2.3)
Requirement already satisfied: scipy>=1.6.0 in c:\users\katki\appdata\local\packages\pythonsoftwarefoundation.python.3.13_qbz5n2kfra8p0\localcache\local-packages\python313\site-packages (from scikit-learn) (1.15.2)
Requirement already satisfied: joblib>=1.2.0 in c:\users\katki\appdata\local\packages\pythonsoftwarefoundation.python.3.13_qbz5n2kfra8p0\localcache\local-packages\python313\site-packages (from scikit-learn) (1.4.2)
Requirement already satisfied: threadpoolctl>=3.1.0 in c:\users\katki\appdata\local\packages\pythonsoftwarefoundation.python.3.13_qbz5n2kfra8p0\localcache\local-packages\python313\site-packages (from scikit-learn) (3.5.0)

[notice] A new release of pip is available: 25.0.1 -> 25.1
[notice] To update, run: C:\Users\katki\AppData\Local\Microsoft\WindowsApps\PythonSoftwareFoundation.Python.3.13_qbz5n2kfra8p0\python.exe -m pip install --upgrade pip
```

```
[5]: import pandas as pd
from sklearn.feature_extraction.text import CountVectorizer
```

```
[6]: # Step 1: Load simple text data using Pandas
data = {
    'Text': [
        "Generative AI creates amazing images.",
        "Large Language Models can write essays.",
        "GANs generate realistic pictures.",
        "Diffusion Models turn noise into art."
    ]
}
```

```
[5]: import pandas as pd
from sklearn.feature_extraction.text import CountVectorizer
```

```
[6]: # Step 1: Load simple text data using Pandas
data = {
    'Text': [
        "Generative AI creates amazing images.",
        "Large Language Models can write essays.",
        "GANs generate realistic pictures.",
        "Diffusion Models turn noise into art."
    ]
}

df = pd.DataFrame(data)
```

```
[7]: # Step 2: Convert the text into numerical features using scikit-Learn
vectorizer = CountVectorizer()
X = vectorizer.fit_transform(df['Text'])
```

```
[8]: # Step 3: Show the result
print("Feature Names:", vectorizer.get_feature_names_out())
print("Numerical Representation:\n", X.toarray())

Feature Names: ['ai' 'amazing' 'art' 'can' 'creates' 'diffusion' 'essays' 'gans'
'generate' 'generative' 'images' 'into' 'language' 'large' 'models'
'noise' 'pictures' 'realistic' 'turn' 'write']
Numerical Representation:
[[1 1 0 0 1 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0]
 [0 0 0 1 0 0 1 0 0 0 0 0 1 1 1 0 0 0 0 1]
 [0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 1 1 0 0]
 [0 0 1 0 0 1 0 0 0 0 0 1 0 0 1 1 0 0 1 0]]
```

From these images, you can see I started off by making sure the scikit-learn was installed into the environment. Pandas typically does not need to be installed since it is one of the most common libraries which was already installed. The next step was to import those libraries and then load simple text data by using Pandas. Step 2 involved converting the text into numerical features using scikit-learn. Afterwards, we printed the results to show the output.

References

Adam Zewe. (2023, November 9). *Explained: Generative AI*. MIT News | Massachusetts Institute of Technology. <https://news.mit.edu/2023/explained-generative-ai-1109>

Author links open overlay panelRobin R. White 1, 1, & AbstractThe launch of generative artificial intelligence (GenAI) tools has catalyzed considerable discussion about the potential impacts of these systems within the scientific article preparation process. This symposium paper seeks to summarize current re. (2025, January 6). *Generative Artificial Intelligence Tools in journal article preparation: A preliminary catalog of ethical considerations, opportunities, and Pitfalls*. JDS Communications. <https://www.sciencedirect.com/science/article/pii/S2666910224002011>

Davenport, T. H., & Mittal, N. (2022, November 14). *How generative AI is changing creative work*. Harvard Business Review. <https://hbr.org/2022/11/how-generative-ai-is-changing-creative-work>

Hugging face – the AI community building the future. Hugging Face –. (n.d.). <https://huggingface.co/>

Introducing Chatgpt | openai. (n.d.). <https://openai.com/index/chatgpt/>

Martineau, K. (2024, September 1). *What is Generative Ai?*. IBM Research. <https://research.ibm.com/blog/what-is-generative-AI>

Park, S. H. (2023, August). *Use of generative artificial intelligence, including large language models such as CHATGPT, in scientific publications: Policies of kjr and prominent authorities*. Korean journal of radiology. <https://pmc.ncbi.nlm.nih.gov/articles/PMC10400373/>