

# Generative AI for journalists: Discovering what data can do

---

**Course Dates: November 20 to December 17, 2023**

**Instructor: Sil Hamilton, researcher-in-residence at Hacks/Hackers**

## About the course

---

Welcome to the Knight Center's new online course, **Generative AI for Journalists**, organized by the Knight Center for Journalism in the Americas in partnership with **Hacks/Hackers**. During this four-week online course held from **November 20 to December 17, 2023**, students will develop and implement their own generative AI workflows suitable for work. This course will equip you with the skills and knowledge necessary for navigating the next generation of computing.

## Goal

---

Through hands-on tutorials over the next four weeks, we want to help you get in-the-know by training you on practical applications and concepts integral to generative technologies. This course will introduce you to the machine learning domain by building up specific skills week over week, leaving you prepared for the next generation of generative AI applications and workflows only now entering newsrooms.

## Who can enroll? (Who is this Course for?)

---

This course is for journalists who may have heard of generative AI before and would like to begin engaging with these technologies on a more practical basis, whether to be better prepared for the future of computing or to improve their data journalism practice with new capabilities made possible by machine learning.

No coding experience is necessary, nor will technical skills be assumed. Course videos will introduce all prerequisite skills at your pace, and we'll make sure your computing environment is properly set up to download and run your own language models.

We'll also provide you with plenty of exercises to practice the skills using your own data at your own pace. These exercises will be available asynchronously, and the instructor will be around to answer any questions you may have.

## Objectives

---

We will build up your AI expertise such that you will be able to participate in AI policy formulation and implementation in your organization. Finishing this course will allow you to:

- Understand what generative AI is and is not
- Be able to clearly articulate when and where to deploy generative AI technologies
- Convert your data to formats suitable for language models
- Learn the basics of prompt engineering
- Embed your documents in a vector database to search through them with natural language
- Quickly develop prototype workflows to assess potency

## Tools/Applications Required

---

Students will need a computer with an internet connection. The computer should be a laptop or desktop running an operating system like MacOS, Windows, or Linux. Mobile devices like phones and tablets are not recommended, as the tools we will be using do not support mobile platforms.

We will be using the below resources:

[JupyterLab Desktop](#), an all-in-one application for running language models in a Python environment. You can download the program for Windows and MacOS, and it is recommended to do this before the course starts. For those whose computers are not modern, we recommend using the completely online [Google Colaboratory](#). This will require a Google account.

[Hugging Face](#), a website for accessing language and image models. We recommend making a free account to access certain features and models to be demonstrated in this course.

## How the course works

---

Our BOCs (Big Online Courses) offer a flexible learning experience. You have the freedom to log in to the course at your convenience and complete activities throughout the week at your own pace.

Although our BOCs are asynchronous, we value interactive learning, and to facilitate this, we host live office hours with our dedicated instructor(s). Attending our live events is optional but highly encouraged. All live sessions and office hours are recorded to ensure that those who are unable to attend can access them at a later date.

The material is organized into **four** weekly modules. Each module will be taught by **Sil Hamilton, researcher-in-residence at Hacks/Hackers**, and will cover a different topic through videos, presentations, readings and discussion forums. There will be a quiz each week to test the knowledge you've gained through the course materials. The weekly quizzes, and weekly participation in the discussion forums, are the basic requirements for earning a certificate of participation at the end of the course.

This course is very flexible, and if you are behind with the materials, you have the entire length of the course to complete them. We do recommend you complete each of the following before the end of each week so you don't fall behind:

- Video lectures
- Readings and handouts/exercises
- Participation in the discussion forums
- Quizzes covering concepts from video lectures and/or readings

## The course is divided into four weekly modules (Course outline)

---

### Introduction Module - Generative AI For Journalists

Welcome to the course! We'll begin by diving into the recent history of generative AI through a study of successful AI projects instructor Sil Hamilton has observed while working with newsrooms and organizations across the industry. Next, we'll get you set up with the required tools we'll be using to discover AI during the course. We'll also set aside time to go through what you'll be learning — the exercises and discussions throughout the course will encourage you to try these techniques on your own datasets.

*This module will cover:*

- Defining generative AI and understanding what makes a successful implementation
- An overview of the course structure
- Getting set up with our required tools and applications
- Tips on how to make the best of this course

## **Module 1 - But What Are Models? (November 20 - 26, 2023)**

What is called generative AI today is built on the success of machine learning models capable of understanding the world around us through text and images. We'll develop an intuitive understanding of what is, and is not, possible with generative AI models today by looking at what makes these models tick.

**Office Hours:** Wednesday at 2 PM CST.

*This module will cover:*

- Prediction tasks: how generative models are trained
- Natural language processing fundamentals
- How ChatGPT works — and why
- Why understanding modeling matters

## **Module 2 - Discover The Data In Your Documents (November 27 - December 3, 2023)**

Generative models talk to each other through text. Learn how to see your data in new ways by making your data — and your newsroom — “AI ready” by converting your unstructured documents into structured formats via optical character recognition (OCR) and embeddings, the fundamental unit of meaning for generative AI models. Embed your articles, documents, sources, and more.

**Office Hours:** Wednesday at 2 PM CST.

**In Conversation:** John Keefe, weather data editor at the New York Times.

*This module will cover:*

- What sorts of data machine learning models expect
- Converting your non-textual data to structured formats suitable for language models
- Ways to “embed” your data with the help of embedding models and vector stores

## **Module 3 - Run And Use AI Models (December 4 - 10, 2023)**

With your data cleaned and structured, it is now time to use generative models to transform your data in interesting and useful ways. Learn how to run a variety of multimodal models both in the cloud and on your local computer with LangChain, a framework for learning language models into conversational “agents” capable of many things: trawling your archives, summarizing documents, and rearranging your sources in new ways.

**Office Hours:** Wednesday at 2 PM CST.

*This module will cover:*

- Creating an agent with LangChain, a framework for developing applications with AI
- Plugging your new agent into your vector store to create your very own research assistant
- Giving your agent a custom personality
- Extending your agent with new capabilities via tools and external APIs

## **Module 4 - Putting It All Together (December 11 - 17, 2023)**

Now that you've created your very own agent using LangChain, learn how to share it with the wider world by packaging and deploying it with the help of Hugging Face Spaces — an easy-to-use hosting platform for machine learning applications suitable for use in your newsroom.

**In Conversation:** Freddy Boulton, software developer at Hugging Face.

**Office Hours:** Wednesday at 2 PM CST.

*This module will cover:*

- Giving your LangChain application a stylish interface with the help of Gradio
- Customizing and styling front-end
- Hosting your application online on your very own Hugging Face space