

# DATA SCIENCE AFRICA (DSA)

DSA 2026, Makerere University, Uganda

THEME

*"Foundational and Practical Data Science in the Age of Generative AI"*

**Topic:** Multimodal AI, LLMs & Generative AI

Pre-Summer School Materials

By Jeremiah Ayock Ishaya and Regis N'Guessan

This pre-summer school guide is designed to help participants arrive with the right mental foundation before the Summer School. The summer school itself will be practical and interactive, so the goal here is not to master everything in advance. The goal is simpler: understand the basic language of modern AI, become comfortable with the core ideas, and come ready to build, question, and experiment.

The materials are grouped around the six lecture blocks:

1. Foundations of Generative AI
2. Multimodal AI
3. Generative Models
4. Prompt Engineering and Interaction
5. Building Applications
6. Evaluation and Ethics

Participants do not need to complete every resource. A recommended pathway is provided at the end.

## How to Use This Guide

For beginners, focus first on the “Essential” resources. These are enough to follow the workshop comfortably.

For MSc, PhD, and industry participants, the “Go Deeper” resources will provide stronger technical grounding and prepare you for advanced discussion.

### 1. Foundations of Generative AI

#### 1.1 Google Machine Learning Crash Course

**Link:** <https://developers.google.com/machine-learning/crash-course>

**Level:** Beginner

**Recommended time:** 1–2 hours before the workshop

**Priority:** Essential

The course is a practical introduction to the core ideas behind machine learning. It explains how models learn from data, how predictions are made, and why data quality matters. It includes short lessons, visual explanations, and exercises.

Before discussing LLMs and multimodal AI, participants need a basic understanding of what a model is, what training means, and why machine learning depends on examples. This resource gives that foundation without being too mathematical.

What participants should focus on

- What machine learning is
- Training vs inference
- Features and labels
- Generalization
- Model evaluation basics

## 1.2 Microsoft AI for Beginners

**Link:** <https://microsoft.github.io/AI-For-Beginners/>

**Level:** Beginner

**Recommended time:** 1 hour selected reading

**Priority:** Optional but helpful

The course is a beginner-friendly curriculum that introduces artificial intelligence, machine learning, neural networks, computer vision, natural language processing, and ethical issues. It includes lessons, quizzes, and practical examples.

It gives participants a broader map of AI before narrowing into LLMs and multimodal systems. It is especially useful for BSc students or industry participants who are new to AI vocabulary.

What participants should focus on

- What AI is
- Machine learning vs deep learning
- Neural networks
- Introduction to NLP
- Introduction to computer vision

## 1.3 Microsoft Generative AI for Beginners

**Link:** <https://github.com/microsoft/generative-ai-for-beginners>

**Level:** Beginner to intermediate

**Recommended time:** 1–2 selected lessons

**Priority:** Essential for those interested in applications

This course introduces the fundamentals of generative AI and explains how to build applications using modern AI models. It includes lessons on LLMs, prompt engineering, chat applications, retrieval-augmented generation, and responsible AI.

The workshop will move from theory to application. This resource helps participants understand how generative AI can be used in real software systems, not just as a chatbot.

What participants should focus on

- Introduction to generative AI and LLMs

- Prompt engineering basics
- Building chat applications
- Retrieval-augmented generation
- Responsible AI

## 2. Transformers, LLMs, Tokenisation, Embeddings, and Attention

### 2.1 The Illustrated Transformer by Jay Alammar

**Link:** <https://jalammar.github.io/illustrated-transformer/>

**Level:** Beginner to intermediate

**Recommended time:** 45–60 minutes

**Priority:** Essential

This video is one of the clearest visual explanations of the Transformer architecture. It explains attention, encoders, decoders, and how Transformers process sequences.

Transformers are the foundation of most modern LLMs and many multimodal AI systems. Participants do not need to understand every mathematical detail, but they should understand the intuition: attention allows the model to decide which parts of the input are important.

#### What participants should focus on

- Why Transformers replaced many older sequence models
- What attention means
- How information moves through Transformer layers
- The difference between encoder and decoder ideas

### 2.2 3Blue1Brown: Attention in Transformers, Step-by-Step

**Link:** <https://www.3blue1brown.com/lessons/attention/>

**Level:** Beginner to intermediate

**Recommended time:** 30–45 minutes

**Priority:** Essential for visual learners

This visual lesson explains how attention works inside Transformers. It focuses on intuition, geometric meaning, and how a model uses context to refine the meaning of words.

Many participants hear '**attention**' but do not really understand what it means. This resource makes the concept less abstract by showing attention visually.

#### What participants should focus on

- How words become vectors
- Why context changes meaning
- How attention helps resolve ambiguity
- Why attention is central to LLMs

## 2.3 Hugging Face LLM Course

**Link:** <https://huggingface.co/learn/llm-course/en/chapter1/1>

**Level:** Beginner to intermediate

**Recommended time:** 1–2 hours selected reading

**Priority:** Essential for technical participants

This course introduces large language models and natural language processing using the Hugging Face ecosystem. It explains Transformers, tokenisers, datasets, model loading, and practical NLP workflows.

Hugging Face is one of the most important platforms for open-source AI models. Participants who want to move beyond using commercial chatbots should understand how models, datasets, and pipelines are organised.

What participants should focus on

- What NLP is
- What Transformers are used for
- How tokenisers convert text into numerical form
- How pre-trained models are loaded and used
- Basic inference with existing models

## 2.4 Hugging Face Tokenisers Lesson

**Link:** <https://huggingface.co/learn/llm-course/chapter2/4>

**Level:** Beginner to intermediate

**Recommended time:** 20–30 minutes

**Priority:** Essential

This lesson explains tokenisation, one of the most important but often overlooked parts of LLM systems. It shows how text is split into tokens and converted into numbers that models can process.

Participants often assume LLMs read text the way humans do. They do not. They read tokens. Understanding tokenisation helps participants understand context windows, model cost, multilingual limitations, and why some African languages may be handled poorly by mainstream models.

What participants should focus on

- What tokens are
- Why models need numerical input
- Why tokenisation affects cost and performance
- Why tokenisation matters for low-resource languages

### 3. Multimodal AI

#### 3.1 Hugging Face Computer Vision Course: Multimodal Tasks and Models

**Link:**

<https://huggingface.co/learn/computer-vision-course/en/unit4/multimodal-models/tasks-models-part1>

**Level:** Beginner to intermediate

**Recommended time:** 30–45 minutes

**Priority:** Essential

This section introduces multimodal tasks such as visual question answering and visual reasoning. It explains how AI systems combine image understanding with language understanding.

The workshop will discuss text, image, and audio integration. This resource helps participants see what multimodal AI looks like in practice: an image is not just classified; it can be questioned, described, interpreted, and connected to language.

**What participants should focus on**

- Visual question answering
- Visual reasoning
- Image-text understanding
- Why multimodal AI is more than image classification

#### 3.2 Hugging Face Visual Question Answering Guide

**Link:** [https://huggingface.co/docs/transformers/en/tasks/visual\\_question\\_answering](https://huggingface.co/docs/transformers/en/tasks/visual_question_answering)

**Level:** Intermediate

**Recommended time:** 20–30 minutes

**Priority:** Optional for beginners, useful for technical participants

This guide explains how visual question answering systems work. A model receives an image and a question, processes both, and returns an answer based on visual and language understanding.

It gives participants a concrete example of multimodal AI. Instead of only saying, “AI can process images and text,” this shows the actual task design.

**What participants should focus on**

- Image + question as input
- Model processing with a vision-language architecture
- Answer generation
- Practical use cases

### 3.3 Hugging Face Image Captioning Guide

**Link:** [https://huggingface.co/docs/transformers/en/tasks/image\\_captioning](https://huggingface.co/docs/transformers/en/tasks/image_captioning)

**Level:** Beginner to intermediate

**Recommended time:** 20–30 minutes

**Priority:** Essential for the hands-on session

This guide introduces image captioning: generating a textual description of an image. It explains the task and shows how such systems can be used, including accessibility applications.

Image captioning is one of the simplest and most intuitive multimodal tasks. It helps participants understand how visual inputs can be converted into language outputs.

What participants should focus on

- What image captioning is
- Image input -> text output
- Accessibility use cases
- Practical model inference

### 3.4 Deep Learning. AI: Open Source Models with Hugging Face

**Link:** <https://www.deeplearning.ai/courses/open-source-models-hugging-face>

**Level:** Beginner to intermediate

**Recommended time:** 1 hour selected lessons

**Priority:** Optional but highly recommended

This course introduces how to use open-source models from Hugging Face for NLP, audio, image, and multimodal tasks. It also introduces Gradio and Hugging Face Spaces for building simple applications.

Participants will not only learn concepts; they will see how to turn models into usable applications. This is especially useful for industry professionals and students who want portfolio projects.

What participants should focus on

- Choosing open-source models
- Using models for text, image, and audio tasks
- Building simple apps with Gradio
- Sharing demos through Hugging Face Spaces

## 4. Generative Models: LLMs, Diffusion Models, and GANs

### 4.1 Hugging Face Diffusion Models Course

**Link:** <https://huggingface.co/learn/diffusion-course/unit0/1>

**Level:** Beginner to intermediate

**Recommended time:** 45–60 minutes selected reading

**Priority:** Essential for image generation concepts

This course introduces diffusion models and the Hugging Face Diffusers library. It explains how diffusion models generate images by learning to reverse a noise process.

Participants will hear about diffusion models in the lecture. They do not need to become image-generation experts, but they should understand the broad idea: diffusion models learn to denoise and gradually create images.

What participants should focus on

- What diffusion models are
- Why noise is involved
- How image generation differs from text generation
- The role of the Diffusers library

### 4.2 Hugging Face Unit 1: Introduction to Diffusion Models

**Link:** <https://huggingface.co/learn/diffusion-course/en/unit1/1>

**Level:** Beginner to intermediate

**Recommended time:** 30 minutes

**Priority:** Essential if participants want to understand image generation

This unit introduces the basic mechanics of diffusion models and shows how simple diffusion systems can be created using the Diffusers library.

The lecture will explain diffusion conceptually. This resource gives participants a more practical view of what happens behind image generation tools.

What participants should focus on

- Noise addition and denoising
- Model training intuition
- Image generation pipeline
- Difference between understanding an image and generating one

### 4.3 Google Developers: Introduction to Machine Learning and Generative AI Concepts

**Link:** <https://developers.google.com/machine-learning>

**Level:** Beginner

**Recommended time:** 30–45 minutes selected reading

**Priority:** Optional

This page collects Google’s foundational machine learning materials, including introductory ML resources and updated learning pathways.

It is helpful for participants who need a broader foundation before going into LLMs, diffusion models, and multimodal systems.

What participants should focus on

- ML foundations
- Neural network basics
- Model training
- Responsible AI themes

## 5. Prompt Engineering and Human-AI Interaction

### 5.1 OpenAI Prompt Engineering Guide

**Link:** <https://developers.openai.com/api/docs/guides/prompt-engineering>

**Level:** Beginner to intermediate

**Recommended time:** 30–45 minutes

**Priority:** Essential

This guide explains how to write effective instructions for AI models. It discusses why prompting is both art and science and how clear instructions improve model output.

Prompt engineering is a major practical skill. Participants will learn how to move from vague prompts to structured prompts that specify role, task, constraints, context, and output format.

What participants should focus on

- Giving clear instructions
- Specifying output format
- Providing context
- Iterating prompts
- Evaluating model responses

### 5.2 OpenAI Text Generation Guide

**Link:** <https://developers.openai.com/api/docs/guides/text>

**Level:** Beginner to intermediate

**Recommended time:** 30 minutes

**Priority:** Essential for application builders

This guide explains how to use the OpenAI API for text generation. It introduces messages, prompts, structured formats, and model-generated responses.

Participants will build or reason about simple LLM-powered applications. Understanding the basic input-output structure of an API request makes the practical session easier.

#### What participants should focus on

- Model input and output
- Messages and prompts
- Generating text from instructions
- Structured outputs

### 5.3 OpenAI Developer Quickstart

**Link:** <https://developers.openai.com/api/docs/quickstart>

**Level:** Beginner to intermediate

**Recommended time:** 30–45 minutes

**Priority:** Essential for hands-on participants

This quickstart explains how to create an API key, set up a development environment, and make a first API call. It introduces the OpenAI API as a simple interface for text, vision, and other AI tasks.

The hands-on session will include simple application workflows. Participants who complete this before the workshop will spend less time struggling with setup and more time understanding the AI pipeline.

#### What participants should focus on

- API key setup
- First request
- Basic response handling
- How APIs connect applications to models

### 5.4 Prompt Engineering Guide by DAIR.AI

**Link:** <https://www.promptingguide.ai/>

**Level:** Beginner to advanced

**Recommended time:** 45 minutes selected reading

**Priority:** Optional but valuable

This guide provides a broad introduction to prompt engineering, including zero-shot prompting, few-shot prompting, chain-of-thought prompting, prompt design, and advanced prompting methods.

It is useful for participants who want a richer prompt engineering reference beyond one vendor's documentation. It also helps researchers think about prompting as a method, not just a trick.

What participants should focus on

- Zero-shot prompting
- Few-shot prompting
- Chain-of-thought prompting
- Prompt reliability
- Prompt patterns for practical tasks

## 6. Building Applications: APIs, RAG, Chatbots, and Multimodal Assistants

### 6.1 OpenAI Developer Quickstart

**Link:** <https://developers.openai.com/api/docs/quickstart>

**Level:** Beginner to intermediate

**Recommended time:** 30–45 minutes

**Priority:** Essential

The quickstart provides the first steps for using the OpenAI API in an application. It explains how developers can call models from their code and receive generated outputs.

The workshop's application section will explain how AI systems are built around a simple structure: input -> model -> output. The quickstart makes this structure concrete.

What participants should focus on

- Setting up a simple AI request
- Understanding API calls
- Reading model responses
- Connecting a model to a basic application

### 6.2 Microsoft Generative AI for Beginners: Application Lessons

**Link:** <https://github.com/microsoft/generative-ai-for-beginners>

**Level:** Beginner to intermediate

**Recommended time:** 1 hour selected lessons

**Priority:** Essential

This course includes practical lessons on building AI applications, including chat experiences, retrieval-augmented generation, and responsible development.

It helps participants understand that LLM applications are not only prompts. They often involve data pipelines, retrieval, tools, interfaces, and evaluation.

What participants should focus on

- Chat application design

- Prompting inside applications
- RAG concepts
- Responsible AI checks

### 6.3 Hugging Face Learn Hub

**Link:** <https://huggingface.co/learn>

**Level:** Beginner to advanced

**Recommended time:** Browse for 20 minutes

**Priority:** Optional but useful

The Hugging Face Learn Hub provides courses on LLMs, agents, computer vision, audio, diffusion, reinforcement learning, and open-source AI workflows.

Participants can use this as a continuing learning hub after the workshop. It is especially helpful for those who want to build portfolio projects or research prototypes.

What participants should focus on

- LLM Course
- Computer Vision Course
- Audio Course
- Diffusion Course
- Open-Source AI Cookbook

### 6.4 Hugging Face Image Captioning Guide

**Link:** [https://huggingface.co/docs/transformers/en/tasks/image\\_captioning](https://huggingface.co/docs/transformers/en/tasks/image_captioning)

**Level:** Beginner to intermediate

**Recommended time:** 20–30 minutes

**Priority:** Essential for the multimodal application demo

This guide shows image captioning as a practical task: converting image content into text.

Participants will understand one of the easiest ways to build a multimodal application: accept an image, process it with a vision-language model, and return a useful text output.

What participants should focus on

- Image input
- Caption generation
- Accessibility use cases
- Practical inference workflow

## 7. Evaluation, Reliability, Bias, and Ethics

### 7.1 Stanford AI Index Report 2025

**Link:** <https://hai.stanford.edu/ai-index/2025-ai-index-report>

**Level:** Beginner to advanced

**Recommended time:** 30–45 minutes selected reading

**Priority:** Essential for MSc, PhD, and industry participants

The Stanford AI Index tracks major AI trends, including model performance, investment, responsible AI, safety, policy, and public attitudes. The 2025 report highlights the uneven development of responsible AI evaluation and the rise of AI-related incidents.

Participants need to understand AI beyond demos. This report helps frame generative AI as a social, economic, and governance issue, not only a technical achievement.

What participants should focus on

- Responsible AI trends
- AI incidents
- Evaluation gaps
- Public trust
- Policy and governance issues

### 7.2 Stanford HAI: Responsible AI Section

**Link:** <https://hai.stanford.edu/ai-index/2025-ai-index-report/responsible-ai>

**Level:** Beginner to advanced

**Recommended time:** 20–30 minutes

**Priority:** Essential

This section focuses specifically on responsible AI, including evaluation, safety, incidents, and the need for stronger assessment practices.

It helps participants see that AI evaluation is not only about accuracy. Modern systems must also be evaluated for safety, factuality, fairness, robustness, and real-world harm.

What participants should focus on

- Why responsible AI evaluation matters
- Why standardised evaluation is still difficult
- How AI risks are measured
- Why organisations need governance processes

### 7.3 UNESCO Recommendation on the Ethics of Artificial Intelligence

**Link:** <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>

**Level:** Beginner to intermediate

**Recommended time:** 30 minutes selected reading

**Priority:** Essential

UNESCO's Recommendation is a global framework for ethical AI. It emphasises human rights, dignity, transparency, fairness, privacy, accountability, and human oversight.

The workshop includes African and LMIC contexts. Ethical AI is not only about avoiding technical errors. It is about asking who benefits, who is excluded, who is represented, and who carries the risk when AI systems fail.

What participants should focus on

- Human rights and dignity
- Transparency and fairness
- Human oversight
- Data governance
- Inclusion and sustainability

## 7.4 OpenAI Prompt Engineering and Text Generation Guides

**Links:**

<https://developers.openai.com/api/docs/guides/prompt-engineering>

<https://developers.openai.com/api/docs/guides/text>

**Level:** Beginner to intermediate

**Recommended time:** 30–45 minutes

**Priority:** Essential

These guides explain how models respond to instructions and how generated outputs can vary. They also help participants understand why prompt design, constraints, and structured outputs are important.

Reliability begins at the interaction level. Poor prompts create unclear outputs. But even good prompts cannot guarantee truth. Participants should understand both the value and limits of prompting.

What participants should focus on

- Clear task instructions
- Output constraints
- Structured responses
- Why generative outputs are non-deterministic
- Why validation is necessary

## 8. Recommended Pre-Workshop Pathways

### Pathway A: Absolute Beginner

Complete these before the workshop:

1. Google Machine Learning Crash Course  
<https://developers.google.com/machine-learning/crash-course>
2. The Illustrated Transformer  
<https://jalammar.github.io/illustrated-transformer/>
3. OpenAI Prompt Engineering Guide  
<https://developers.openai.com/api/docs/guides/prompt-engineering>
4. Hugging Face Image Captioning Guide  
[https://huggingface.co/docs/transformers/en/tasks/image\\_captioning](https://huggingface.co/docs/transformers/en/tasks/image_captioning)
5. UNESCO AI Ethics Recommendation  
<https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>

### Pathway B: BSc Student with Some Programming Background

Complete these before the workshop:

1. Google Machine Learning Crash Course  
<https://developers.google.com/machine-learning/crash-course>
2. Hugging Face LLM Course, Chapter 1  
<https://huggingface.co/learn/llm-course/en/chapter1/1>
3. Hugging Face Tokenisers Lesson  
<https://huggingface.co/learn/llm-course/chapter2/4>
4. OpenAI Developer Quickstart  
<https://developers.openai.com/api/docs/quickstart>
5. Hugging Face Computer Vision Course: Multimodal Tasks and Models  
<https://huggingface.co/learn/computer-vision-course/en/unit4/multimodal-models/tasks-models-part1>

### Pathway C: MSc or PhD Participant

Complete these before the workshop:

1. The Illustrated Transformer  
<https://jalammar.github.io/illustrated-transformer/>

2. 3Blue1Brown Attention Lesson  
<https://www.3blue1brown.com/lessons/attention/>
3. Hugging Face LLM Course  
<https://huggingface.co/learn/llm-course/en/chapter1/1>
4. Hugging Face Diffusion Course  
<https://huggingface.co/learn/diffusion-course/unit0/1>
5. Stanford AI Index 2025  
<https://hai.stanford.edu/ai-index/2025-ai-index-report>
6. UNESCO AI Ethics Recommendation  
<https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>

#### Pathway D: Industry Participant

Complete these before the workshop:

1. Microsoft Generative AI for Beginners  
<https://github.com/microsoft/generative-ai-for-beginners>
2. OpenAI Prompt Engineering Guide  
<https://developers.openai.com/api/docs/guides/prompt-engineering>
3. OpenAI Developer Quickstart  
<https://developers.openai.com/api/docs/quickstart>
4. Microsoft Generative AI Application Lessons  
<https://github.com/microsoft/generative-ai-for-beginners>
5. Stanford HAI Responsible AI Section  
<https://hai.stanford.edu/ai-index/2025-ai-index-report/responsible-ai>

#### 9. Pre-Workshop Practical Preparation

Participants who want to take part in the hands-on session should prepare the following:

##### Required

- A laptop
- Internet access
- Python 3.10 or above
- A code editor such as VS Code, Cursor, or Jupyter Notebook
- Basic familiarity with running Python scripts

Recommended Python packages

**Install these before the workshop if possible:**

*pip install openai transformers gradio pillow pandas python-dotenv*

*For participants interested in retrieval-augmented generation:*

*pip install chromadb sentence-transformers langchain*

**For participants interested in image generation and diffusion models:**

*pip install diffusers accelerate torch torchvision*

## 10. Final Message to Participants

This workshop is not only about learning how to use AI tools. It is about understanding how modern AI systems work, where they fail, and how we can design them responsibly.

Come curious. Come critically. Come ready to build.

The most important mindset is not “How do I prompt ChatGPT?”

The better question is:

**How do I design an AI system that is useful, reliable, ethical, and meaningful in my own context?**