What Artificial Intelligence Is (And Is Not)

AiCE Africanism & Pan-African Studies — Demystifying AI

Demystifying AI: A Non-Technical Explanation

Artificial intelligence is a broad term for computer systems that can perform tasks requiring what we would consider intelligence if done by humans. Within AI, the most important subfield today is machine learning — a technique in which computers learn to perform tasks by analysing large amounts of data, rather than by following explicit instructions written by programmers.

To understand machine learning, consider a spam email filter. A traditional approach would require a programmer to write rules: 'if the email contains the words FREE MONEY or CLICK HERE, mark it as spam.' Machine learning takes a different approach: feed the system thousands of examples of spam emails and non-spam emails, and let it learn from these examples what features distinguish spam from legitimate email. The system discovers patterns in the data that no programmer explicitly specified. This learned ability then applies to new emails the system has never seen before.

The same principle — learning patterns from large amounts of data — underlies every major AI application. A medical AI that diagnoses cancer from radiology images has learned to recognise cancer patterns by studying thousands of labelled scans. A language AI that generates human-like text has learned the patterns of human language by processing billions of words. A credit scoring AI has learned to predict creditworthiness by studying thousands of past borrowers and their repayment behaviour. The power of machine learning is that it can discover patterns in data too complex for human experts to specify explicitly — and apply those patterns to new cases faster and more consistently than any human.

Key AI Concepts for Non-Technical Readers
Machine Learning (ML): the technique by which AI systems learn from data rather than explicit programming rules — the foundation of virtually all contemporary AI applications.

Key AI Concepts for Non-Technical Readers
Deep Learning: a powerful subtype of machine learning using artificial neural networks with many layers — enables breakthroughs in image recognition, speech recognition, and natural language processing.
Natural Language Processing (NLP): AI capability for understanding and generating human language — powers chatbots, translation, sentiment analysis, and large language models.
Large Language Model (LLM): an AI system trained on vast amounts of text that can understand and generate human language with remarkable sophistication — GPT-4, Claude, and Gemini are examples.
Computer Vision: AI capability for interpreting visual information — enables medical image diagnosis, crop disease detection, and facial recognition.
Generative AI: AI systems that generate new content — text, images, audio, video, code — rather than merely classifying or predicting; the current frontier of AI capability.
Training Data: the large dataset used to train an AI model — its quality, diversity, and representativeness determines the quality and fairness of the AI system.
Algorithmic Bias: when an AI system produces systematically unfair outcomes for certain groups — typically resulting from biased training data or biased design choices.

What AI Can and Cannot Do

One of the most important things to understand about AI is the gap between what it actually does and what popular discourse suggests it does. AI systems are extraordinary at specific, well-defined tasks for which large amounts of training data are available. They are very limited at tasks requiring common sense, causal reasoning, genuine understanding of context, or the ability to operate reliably in situations significantly different from their training data.

Current AI systems do not understand the world in any meaningful sense. A medical AI that correctly diagnoses a rare cancer from a scan is not applying medical knowledge the way a doctor does. It is recognising patterns that correlate with cancer in its training data. If the patterns in a new patient's scan differ sufficiently from its training data —

perhaps because the patient belongs to a demographic group that was underrepresented in training — the AI may fail, even on a case that a skilled doctor would correctly diagnose. This brittleness — the tendency of AI systems to fail unpredictably when they encounter situations outside their training distribution — is the most important limitation to understand when evaluating AI applications in African contexts.

Key Takeaways

Machine learning — the technique of learning patterns from data rather than following explicit rules — underlies virtually every contemporary AI application, from spam filters to medical diagnosis to language models. AI's central limitation is brittleness: it fails unpredictably outside its training distribution, which is precisely why the representativeness of training data matters so much for African populations. The next lesson examines where these capabilities can help Africa most.