How AI Learns from Data

Transcript

Let’s look at how AI learns from data.

Think about the last time you used an internet search engine, or when your phone suggested the next word in a text message, or when your music streaming service suggested a new playlist. All of these use Artificial Intelligence. But how are different AI models actually trained to perform tasks?

Understanding this matters because the advice, suggestions, and decisions offered by the AI you use will be directly influenced by how it learned - and from what data.

Simply put: the inputs affect the outputs.

So, how does AI learn from data? How do different types of AI models acquire their knowledge? And what might this mean for the work that you do?

AI systems learn using something called Machine Learning which, broadly speaking, is when computers learn from examples instead of being explicitly programmed with rules.

Think of it like this: instead of writing detailed instructions for every possible situation, AI tools and systems are shown thousands of examples so that it can work out the patterns for itself.

We all encounter AI trained through machine learning every day and this can include in things we use for work such as email spam filters, voice assistants, and at home smart speakers and robotic vacuum cleaners.

But not all AI models learn in the same way or at the same scale.

Traditional Machine Learning uses simpler algorithms like decision trees or basic statistical methods. These might power tools like email suggested responses, simple risk assessment calculators, or systems that help determine what action to take based on clear criteria that has been set.

Deep Learning uses more complex systems called neural networks - computer systems loosely inspired by how the human brain processes information, where multiple layers can recognise increasingly complex patterns. These power more sophisticated tools like image recognition systems, voice assistants, or advanced document management software that can analyse multiple factors simultaneously.

Large Language Models are a specific type of deep learning system trained on massive amounts of text from the internet, published works, and other sources. These are AI systems that can understand and generate human-like text in response to ordinary speech or written questions. They are often adapted for professional applications like report writing, data analysis, creating designs and images, and can answer complex questions. The most well-known example at the moment is OpenAI's ChatGPT.

The information used to train all of these systems is called Training Data and this is absolutely critical, regardless of which type of model is being used. If any AI system is trained using data that has intentional or unintentional bias, or excludes certain groups or situations, it could undermine the objectives that people are trying to achieve and, in some contexts, lead to very damaging consequences. And it's not just about having diverse data - it's also about having enough useful data. An AI trained on too few examples might not generalise well to real-world situations.

For example, if a public services tool is trained primarily on data from urban areas, it might assume users will have accessible public transport, completely failing people living in rural areas who may not have the same provision or access.

Or if a healthcare AI is trained mainly on data from one demographic group, it might miss important health factors affecting patients in other demographic groups.

The data shapes everything the AI can do - and cannot do.

Let's look at each of these approaches with practical examples that show both their strengths and their limitations.

All AI systems - whether simple algorithms or deep learning models, learn using three main approaches. The learning method doesn't depend on how complex the AI is, but rather on what kind of data is available and what the user wants the system to do.

Supervised Learning is like having an experienced supervisor teach a new colleague.

The system is shown examples with known outcomes - like emails already labelled as 'spam' or 'not spam,' actions marked ‘urgent’ or ‘routine’, or cases marked as 'high risk' or 'low risk.' The AI learns to spot the patterns that led to each outcome.

For instance, a medical AI might be shown thousands of patient cases along with their diagnoses and treatment outcomes. It learns that 'when a patient has these particular symptoms and test results, this diagnosis is likely, and this is treatment typically followed.'

Unsupervised Learning is when AI finds patterns in data without being told what to look for. It's like asking someone to organise a messy drawer - they'll create sections or categories based on groups of things or patterns they notice, which might be different from how you would have organised it. We’ve all seen how differently people organise their books.

This might help identify previously unknown factors that influence outcomes or discover unexpected connections in data.

For example, in public services, an AI system might analyse family services data and discover that people who engage with community gardening projects also tend to engage more consistently with other family support services. Nobody told the AI to look for this connection: it found the pattern itself by looking at which people were supported and what they had in common.

So, unsupervised learning can reveal valuable insights that humans might miss, but the patterns AI finds might not always be meaningful - sometimes it's just coincidence. The AI might also create categories that don't match how humans would organise things, or prioritise factors that you wouldn't consider important.

Reinforcement Learning works through trial, error, and feedback. The AI learns what works by receiving 'rewards' for good decisions and 'penalties' for poor ones, gradually improving over time.

For example, a scheduling system might learn that offering morning appointments to certain groups leads to better attendance, while evening appointments work better for others. Each time an appointment is kept, it receives a 'reward'; when appointments are missed, it receives a 'penalty and this drives the system to become more efficient.

While these are distinct approaches, modern AI systems often combine multiple learning methods. For example, ChatGPT initially trains using unsupervised learning on massive text datasets, then uses supervised learning to fine-tune its responses and then reinforcement learning to refine its responses and make them more helpful.

Understanding how AI systems learn helps us ask the right questions about any AI tool we encounter, regardless of its complexity. Here are three key things to consider:

1. Know the Type and Its Limitations

Different models have different weaknesses. A supervised learning model might be quite rigid, trying to squeeze new situations into familiar categories it learned during training.

An unsupervised learning model might group things in unexpected ways that don't match how humans would categorise them.

A reinforcement learning model might prioritise 'winning' its rewards over other important factors like fairness or efficiency.

2. Question the Data

It is really important to know what kind of data the AI you are using was trained on, and if that data is representative of the people and situations you're working with. For example, if a tool used in a cardiology department to predict heart conditions was trained primarily on data from male heart disease patients, how might it perform when it’s analysing symptoms in female patients?

On the other hand, if the AI is a generalist Large Language Model, like ChatGPT, trained broadly on internet text rather than specialised medical literature, does it have the deep specialist knowledge required to make important decisions about a person's health?

Who is represented in the data – and the quality of the data are fundamental questions. It’s important that there is transparency about what data is being used and how a system learned.

3. Understand When and How It Might Fail

It’s also important to understand and question how AI handles information that isn’t like the information it was trained on. What happens when it encounters something new? Understanding the learning method reveals the AI's limitations, biases, and appropriate uses. Different learning approaches have different failure modes - knowing which type you're dealing with tells you what to watch out for.

Understanding these learning approaches doesn't just help you spot problems - it helps you use AI more effectively. When you know how a system learned, you can better judge when to trust its recommendations and when to apply your own expertise and judgment.