AI BASICS // MULTI-MODAL AI

Multi-modal AI: how it's changing everything

From one type of data to many: why connecting the pieces is what makes AI personal.

> **Travis's short version:** early AI applications were trained on a single type of data, whether text, images or audio. Multi-modal AI draws on several data types at once and connects the pieces together, which improves accuracy and lets it produce insights and outputs personalised to your own wants and needs.

The evolution from single to multi-modal

Initially, AI applications were trained on, and generated outputs and insights from, a single type of data. Each of those applications is powerful in its own way, and we've all met them:

Single-modal AI	What it does	Where you'll have met it
Natural language processing	Works with text and language.	Powers the customer service chatbots we begrudgingly interact with.
Computer vision	Works with images.	Enables facial recognition, security and image generation capabilities.
Speech recognition	Works with audio.	The one that speaks for itself. Who doesn't love Siri?

What changes with multi-modal

With multi-modal AI, you can now draw from multiple data types to collaboratively generate outputs and insights that are personalised to your own individual wants and needs. You can feed in text, image, video and audio, and the model generates a layer of insights and outputs. It also improves accuracy by connecting the pieces together.

Single-modal	Multi-modal
One type of data in: text, images or audio. Each application is powerful in its own way, but it only sees its own slice of the picture.	Text, image, video and audio in together. The model connects the pieces, which improves accuracy and personalises the result.

Multi-modal AI lets you draw from many data types to produce outputs that are personalised to you.

A day-to-day example: dinner on a work night

Enough technical talk. To put the pieces together, here's how multi-modal AI could be used in your everyday life. Imagine you're cooking dinner and you ask an application to help you create an Asian-inspired dish, in under an hour.

1. **You ask for what you want.** Help me create an Asian-inspired dish in under an hour, because it's a work night.
2. **You snap a couple of pictures.** The fridge and the pantry: what you've actually got on hand.
3. **You let the app do its magic.** It combines your request with the photos, rather than treating them as separate inputs.
4. **You get a personalised answer.** A step-by-step Asian-inspired recipe you can cook in under an hour, using only the ingredients you already have.

That level of personalisation is what gives multi-modal AI its edge. The only problem now: you still have to cook it.

Why this matters for your business

The same idea applies beyond the kitchen. The goal of this series is to broaden your understanding of multi-modal AI, its applications and, most importantly, the value proposition to your industry. When a system can take in more than one kind of information and connect it together, the answers it gives are more accurate and fit your situation more closely.

The bottom line

Single-modal AI works with one type of data. Multi-modal AI works with text, image, video and audio together and connects the pieces. That's where the accuracy and the personalisation come from, and it's the starting point for understanding where AI can add value in your industry. **Next in the series:** the evolution of AI agents and how they're changing automation.

Want to talk this through for your business?

Book a free 15-minute call. We'll look at your situation and give you a straight answer, with no jargon and no sales pitch.

Let's Chat

Opens the request form on our website. Free, no obligation.