Machine Learning –

Multimodal Machine Learning

In such tasks as computer vision or natural language processing that involve interaction between the model and the real world, the model often has to rely only on one type of data, be it images or text. But in real life, we perceive the world around us through many senses: smell, hearing, feeling textures, and flavors.

Multimodal machine learning suggests using the fact that the world around us can be experienced in multiple ways (called modalities) to build better models. The term “multimodal” in AI describes how to build ML models that can perceive the event in multiple modalities at a time, just like humans do.

Building an MML can be achieved through combining different types of information and using them in training. For example, matching images with audio and text labels to make them easier to recognize. Multimodal machine learning is so far a young field that is yet to be developed and advanced in 2023, but many believe that it can be key to achieving general AI.