Model Context Protocol (MCP) – Documentation

Introduction

The Model Context Protocol (MCP) is a fundamental concept in working with Large Language Models (LLMs). At its core, MCP is about how we provide input (context) to a model so that it can generate high-quality and accurate outputs. This documentation explores the core idea of 'context', its limitations, and the best practices for structuring it.

What is Model Context?

A Large Language Model (LLM) is essentially a text predictor—it predicts the next word based on the words it has seen so far. The 'context' is the total amount of text that the model receives as input in a single request. Think of it as having a conversation with someone who has a short-term memory. You must tell them everything they need to know upfront in order for them to understand and respond accurately.