

The Paradigm Shift in Multimodal Reasoning

Think about Images



Think with Text

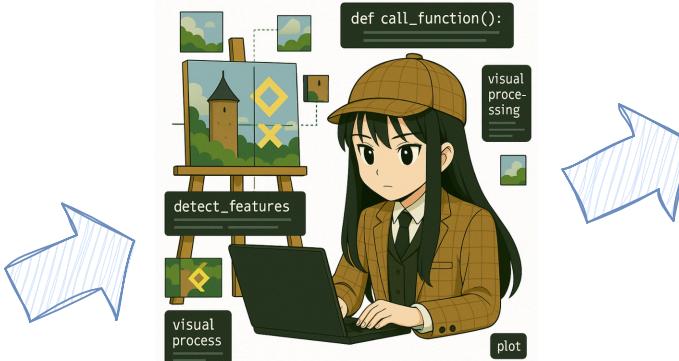
Textual reasoning over a static visual context.



Stage 1: Tool-Driven Visual Exploration
Commanding a fixed toolkit for visual analysis.

Think with Images

Intelligence Increasing



Stage 2: Programmatic Visual Manipulation

Creating bespoke visual operations via code generation.



Stage 3: Intrinsic Visual Imagination

Internally generating visual thoughts within a cognitive loop.