AMAZON NOVA CAPABILITIES (which go beyond simple chatbots)

Language models have become ubiquitous, and it is easy to think that their capabilities are limited to the dialogue interfaces we use every day.

However, the most transformative features of generative artificial intelligence often lie beneath the surface, far beyond simple chatbots. These advanced capabilities are the building blocks for a new class of autonomous, multimodal AI applications.

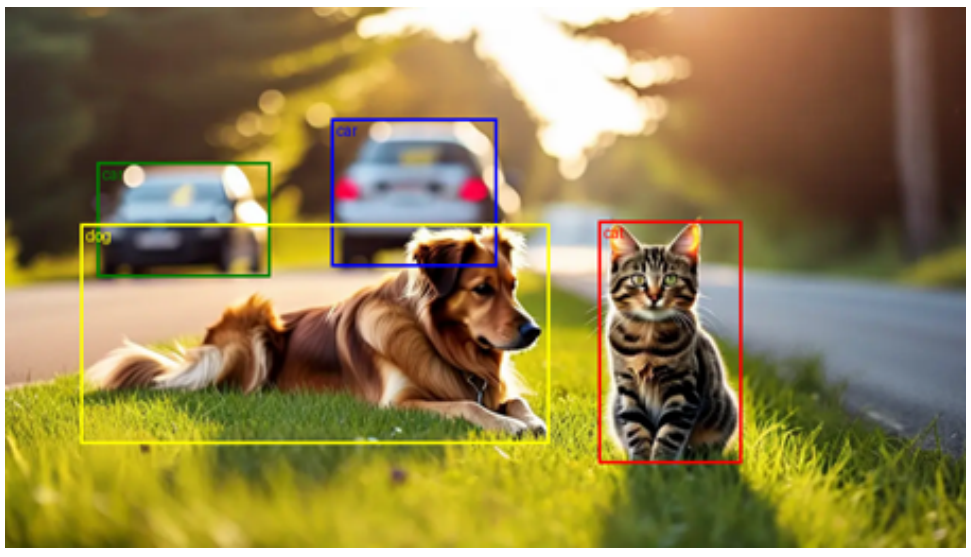
The Amazon Nova family of models is not just another large language model; it is a powerful ecosystem designed for complex, scalable applications that require more than just conversation. It is a suite of tools that push the boundaries of what AI can achieve.

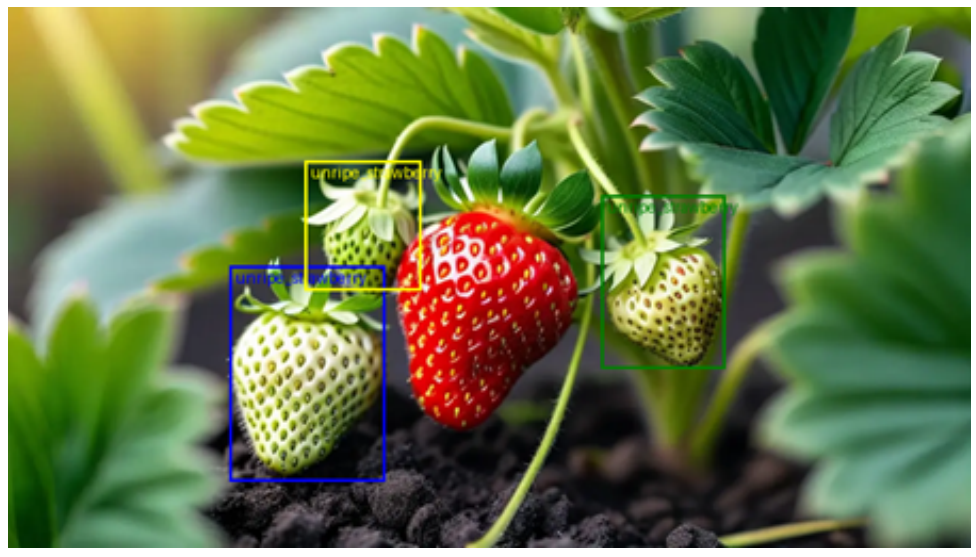
This article reveals six of the most impactful and sometimes counterintuitive capabilities of the Nova ecosystem. We will present them in list form so you can discover how these features can unlock new possibilities for your projects.

1. It can define any category of objects and draw bounding boxes on demand.

Nova models have powerful image anchoring and object detection capabilities across a wide range of domains.

Users can customise detection by specifying which specific categories of objects to look for via the prompt. Examples of categories include common objects such as "traffic lights", "bags" and "faces", or specific items such as "unripe_strawberry".





The model returns the coordinates of the bounding boxes. These coordinates are produced on a normalised scale of [0, 1000) and must be resized to the size of the original image for visualisation.

This feature is essential for use cases such as quality control, retail inventory management, or surveillance, including the ability to use detection to blur specific categories (such as faces) in the processed image.

2. It understands videos over time, not just still images.

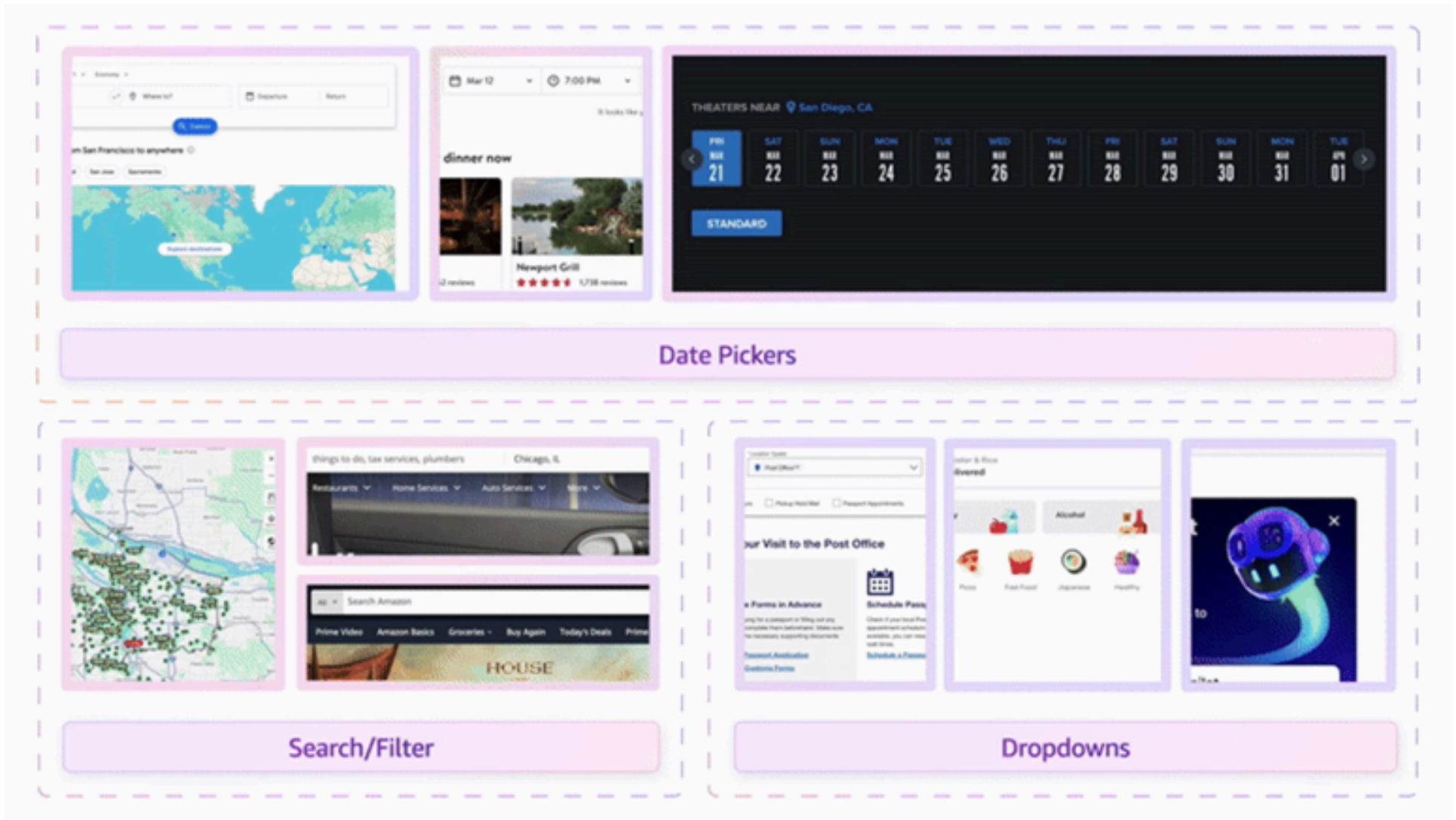
The real strength of Amazon Nova Premier in video analysis does not lie in its ability to process individual images extracted from a video. Its most surprising capability is its understanding of temporal relationships, the sequence of events, and the narrative flow. It does not just see what is in the video, it understands when things happen.
Examples from demonstrations show the model's accuracy with impressive granularity:

- It can identify the exact moment when rain starts to fall in a clip, specifying the timestamp 0:31.
- It can detect the first appearance of a character at a specific moment, such as at 0:38 in the reflection of a rear-view mirror.

To take this further, the model can generate a complete timeline of actions in a structured JSON format. This capability is extremely valuable for programmatic applications, as it allows unstructured video to be transformed into a series of events that can be exploited by automated systems.

Objectware

3. It can act on the web for you (Really!)



Here's a game-changing capability: Amazon Nova Act. It's not just a language model, but an agent specifically designed to free AI from API constraints and enable it to interact with the web as it was designed for humans.

Demonstrations show capabilities that go far beyond simple information extraction:

- Visiting a website, finding relevant information, and saving it directly to a database.
- Placing a complete order on an e-commerce site by interacting with forms, entering information, and using tool calls after an internal ‘thinking’ phase.

4. Agent architecture has evolved: Think ‘microservices for AI’

The Nova ecosystem encourages a major architectural evolution, moving from a ‘monolithic model’ to an ‘agent architecture’. This transition is comparable to that of monolithic software architectures to microservices, which revolutionised traditional application development. The idea is to break down complex tasks into smaller, modular and reusable services.
At the heart of this approach is the MCP (Model Context Provider), a standardised protocol for communication between agents and external tools.

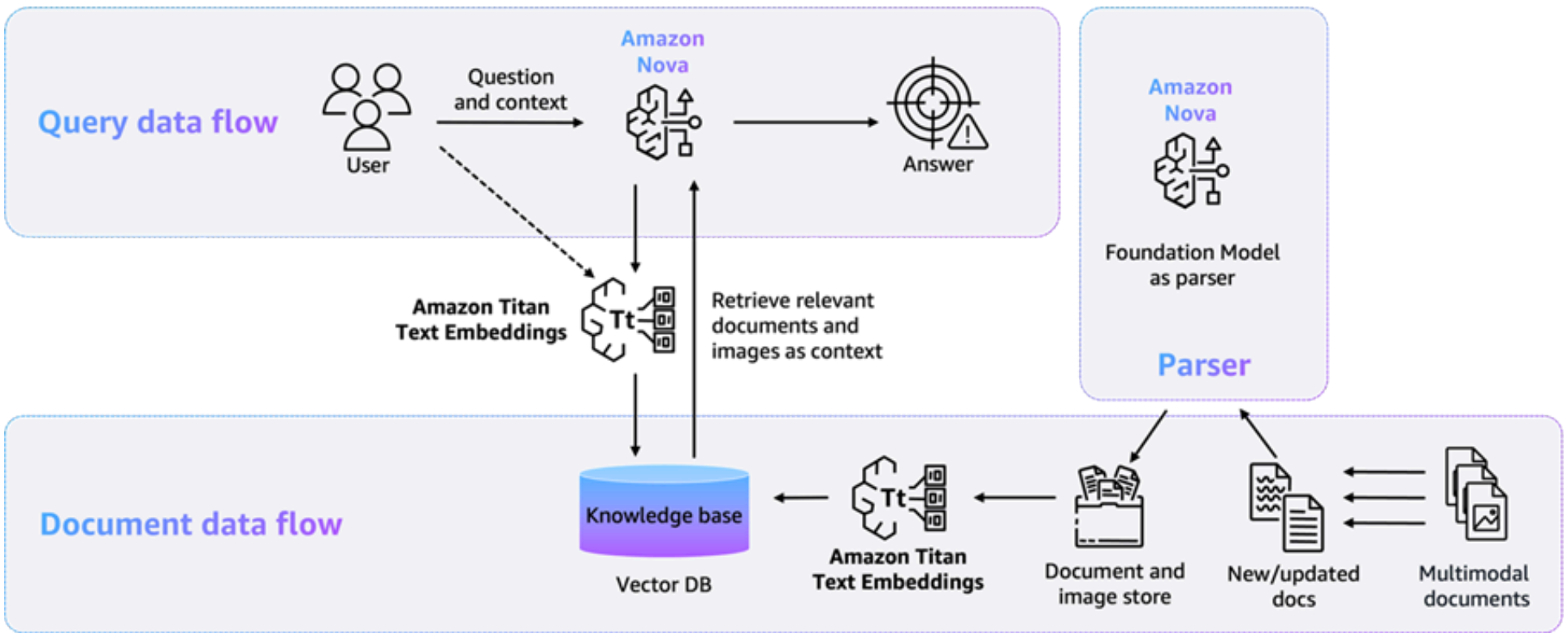
This separation between tool implementation and agent orchestration offers significant advantages:

- Decoupled development (modularity): Teams can develop, test and deploy tools independently, accelerating innovation cycles.
- Economies of scale (reusability): A single tool, such as an ‘inventory checker’, can be exposed via MCP and consumed by dozens of different agents for various applications.
- Scalability: Tool collections can grow and evolve without requiring changes to the orchestrator agent's code.

This concept is perfectly illustrated by multi-agent systems, where a planning agent (such as Nova Premier) can orchestrate specialised agents: Nova Act for browser automation and Nova Lite for analysing collected data.

5. RAG is no longer limited to text and is becoming multimodal.

The RAG (Retrieval-Augmented Generation) technique, which allows models to draw on external data to generate responses, links forward to Amazon Nova. Amazon Bedrock knowledge bases can now ingest and understand documents containing both text and complex visual elements such as graphs, images, or plans.



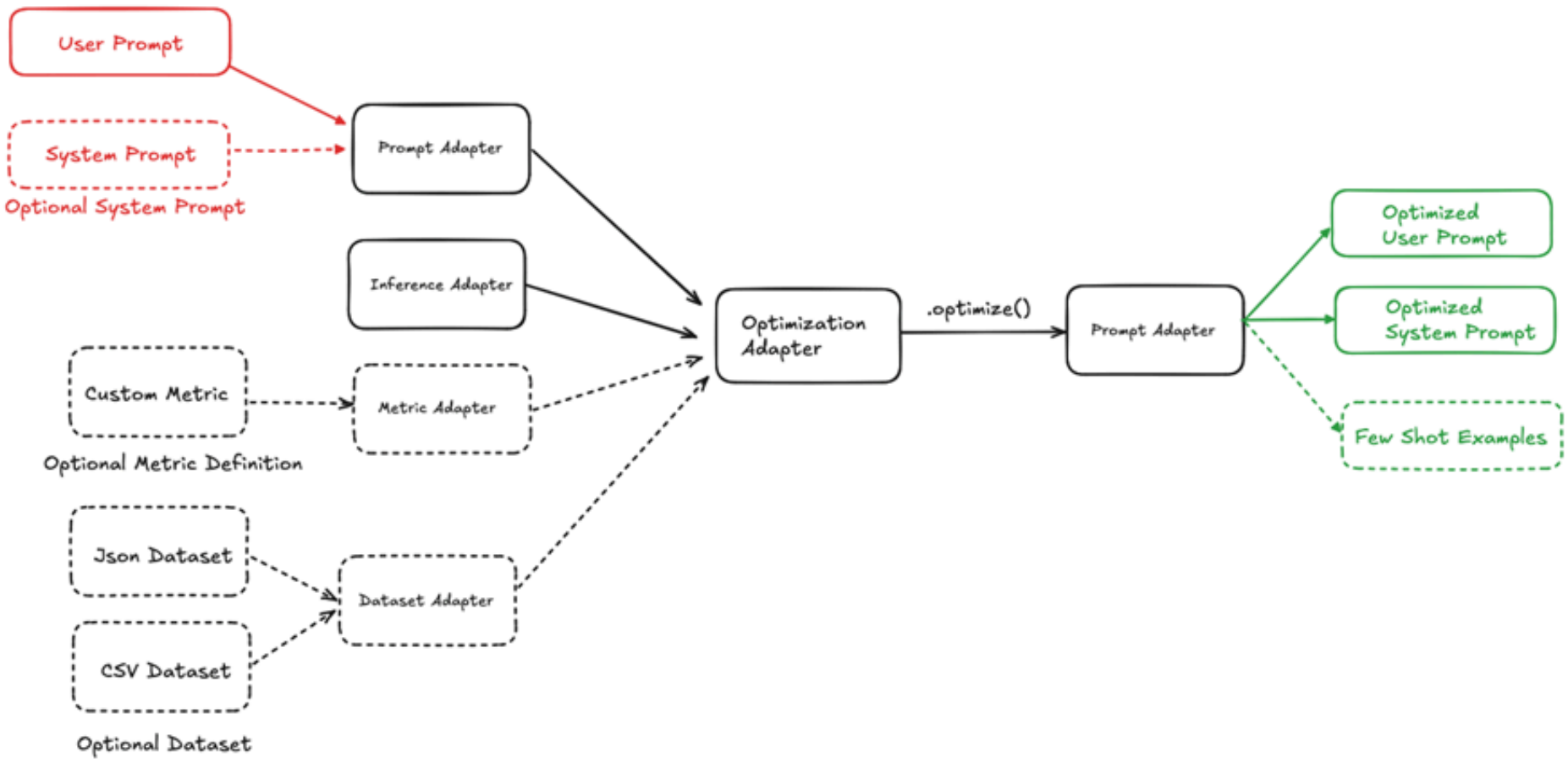
The process is particularly ingenious and takes place in two stages. First, a model such as 'Amazon Nova' Lite is used as an analyser to interpret and summarise the content of the images in a document. Then, a second model, such as 'Amazon Nova Pro', uses this enriched context (original text + image summaries) to generate a complete and accurate final response.

A concrete use case is that of a hotel booking assistant. By ingesting a financial report in PDF format, the assistant can answer specific questions based on the analysis of a financial graph contained in the document, for example: 'What was the revenue in 2022 compared to 2023?'.

This two-step process enables the system to transform unstructured and visually dense documents, such as annual reports in PDF format, into an interactive and searchable source of knowledge.

6. Prompt engineering can be automated (with measurable results)

Prompt engineering, often perceived as a manual art, becomes an exact science with the Nova Prompt Optimizer. This tool transforms what was once a trial-and-error process into efficient, data-driven optimisation.



Its operation is simple yet powerful: it automatically improves prompts by drawing on the client's own datasets.
The Nova Prompt Optimizer refines system instructions and adds relevant ‘a few times’ examples to guide the model towards the most accurate responses possible. For a developer, this means that the tool automates two of the most time-consuming and complex parts of prompt engineering: writing the perfect master instruction and selecting the most effective examples from user data.
The effectiveness of this approach is proven by quantifiable results.
In one case study, the optimiser increased the overall evaluation score of a prompt from 69.4% to 81.3%, an improvement of 17.9%.
For development teams, this translates into considerable time savings, more reliable performance, and the ability to optimise AI applications in a systematic and reproducible manner, ensuring the best possible results for each specific use case.

Objectware

Conclusion

From a model that understands the temporal flow of a video to an agent capable of acting on the web, through micro-service architectures for AI, multimodal RAG, and prompt engineering automation, it is clear that Amazon Nova is much more than just a chatbot. It is an ecosystem of powerful tools designed to build the next generation of sophisticated AI applications.
Amazon Nova provides the tools to automate tasks previously confined to human perception and interaction. The question is no longer whether you can automate your most complex workflows, but which one you will revolutionise first.

Article written by Bingqian SHU

Objectware