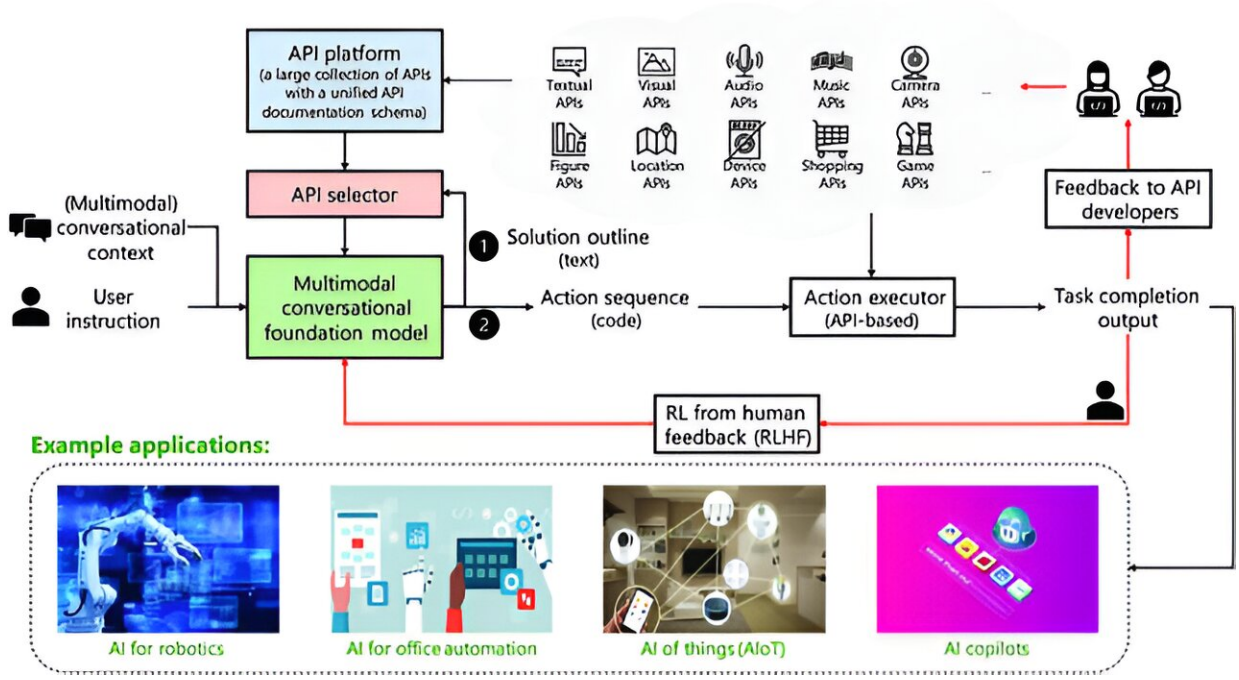
TaskMatrix.AI: Making big models do small jobs with application programming interfaces

March 11 2024



Overview of TaskMatrix.AI. Credit: *Intelligent Computing* (2023). DOI: 10.34133/icomputing.0063

A research team at Microsoft has designed an efficiency tool called TaskMatrix.AI that can be used to accomplish a wide variety of specific AI tasks. TaskMatrix.AI connects general-purpose foundation models like GPT-4, the model behind ChatGPT, with specialized models suitable for certain tasksâ€"much like a human project manager. This research was published in *Intelligent Computing*.

Foundation models and specialized models usually have different mechanisms and, thus, are not easily compatible. Rather than modifying and integrating existing models, TaskMatrix.AI bridges the gaps between them through application programming interfaces, or APIs, which enable software components to communicate.

The research team envisioned an AI ecosystem applicable to office automation, robotics, the Internet of Things, and other domains. Accordingly, their TaskMatrix.AI can perform various digital and physical tasks, give interpretable responses, and learn continuously.

TaskMatrix.AI has four key components: a conversational foundation model that understands user inputs across various modalities (such as text and images) and generates executable action code as input for APIs; an API platform that holds a vast repository of APIs and their documentation; an API selector that chooses the most suitable APIs for the foundation model and an action executor that executes the code given by the model.

As the ecosystem evolves, API developers can improve the documentation based on user feedback.

The team demonstrated the use of TaskMatrix.AI for processing images and automatically making PowerPoint slides.

During the image processing task, a human interacted with TaskMatrix.AI by typing natural language instructions for complex visual tasks such as image generation, editing, and description. TaskMatrix.AI demonstrated its ability to understand human intentions through text-based inputs and provided satisfactory output.

For example, with a tiny input image of a pink flower with a green background and a single instruction to "extend it to 2048 Ã- 4096," TaskMatrix.AI generated a convincing image of vibrant, colorful flowers against lush green leaves through question-answering, captioning, and object replacement APIs.

The PowerPoint automation task required TaskMatrix.AI to create a set of slides, each introducing a different tech company. ChatGPT served as the foundation model for understanding complex user instructions, such as inserting text, resizing and relocating images, and changing the theme for the PowerPoint slides. For example, TaskMatrix.AI successfully inserted and resized five company logos, which it obtained from the Internet, by calling several relevant APIs.

Despite the preliminary validation of TaskMatrix.AI, the team pointed out some challenges ahead, such as finding and adjusting a powerful foundation model, building and maintaining an ideal API platform and addressing user-level concerns like data security, privacy, and customization needs.

More information: Yaobo Liang et al, TaskMatrix.AI: Completing Tasks by Connecting Foundation Models with Millions of APIs, *Intelligent Computing* (2023). DOI: 10.34133/icomputing.0063

TechXplore

Provided by Intelligent Computing

Citation: TaskMatrix.AI: Making big models do small jobs with application programming interfaces (2024, March 11) retrieved 9 May 2024 from https://techxplore.com/news/2024-03-taskmatrixai-big-small-jobs-application.html

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.