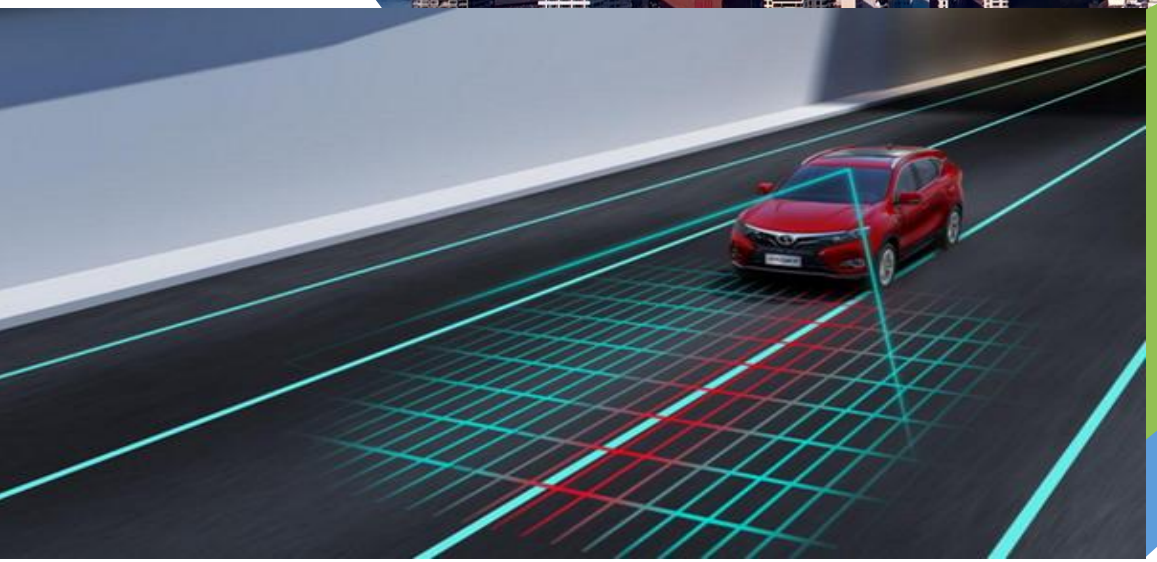


ResearchInChina
www.researchinchina.com

Cockpit AI Agent Research Report, 2024

July 2024

Cockpit AI Agent: Autonomous scenario creation becomes the first step to personalize cockpits

In AI Foundation Models' Impacts on Vehicle Intelligent Design and Development Research Report, 2024, ResearchInChina mentioned that the core of an AI Agent uses a large language model (LLM) as its core computing engine (LLM OS). In the AI service framework, the LLM acts as AI core and the Agent acts as AI APP. With the help of reasoning and generation capabilities of AI foundation model, the Agent can create more cockpit scenarios, while further improving current multimodal interaction, voice processing and other technologies in the cockpit.

Service Framework of AI Agent Based on LLM

Source: Rutgers University

Correspondence between AI OS and Traditional OS

Compared with the traditional OS, AI OS is programmed in natural language, which is more suitable for human operation and convenient for customization and personalized human-computer interaction. The Agent responds to user prompts or instructions in natural language, and can perform a lot of tasks according to its interaction with physical or digital environment.

Correspondence between AI OS and Traditional OS

AI OS Parts	Traditional OS Parts
LLM	OS kernel
Agent	Application
Context window	RAM
External storage	File system
Hardware	Peripheral equipment
Software	Programming library

Source: ResearchInChina

Foundation models spawn Agent customized scenarios

Since AI foundation models were installed in cars, they have been widely used in multimodal interaction scenarios. OEMs have started from voice assistants to further improve intelligent processing of user commands:
Before foundation models got on cars, the wake-up rate and effective instructions of voice assistants in cockpits are at a low level;
After foundation models were available in cars, not only has voice recognition become more intelligent, but the application of expression monitoring and gesture control have also been improved, with effective commands and daily active rates increasing significantly.

AI foundation models further enhance the user experience of smart cockpits through strong context processing and reasoning capabilities. Please refer to the following dialogue:

(Mobility scenario)
--**Tongxue! (the name of a voice assistant)
-I am here!
-There seems to be a prairie ahead. Please take a picture for me!
-Wait a moment, and I will capture the beautiful scenariory of the prairie for you. I'll let you know when the photo is taken. Is there anything else I can help you with?
-Send it to Moments immediately. I want my friends to have a look at this beautiful scenariory.
-Ok, according to your authorized account, please choose the material and filter style.
-Just this one, with a caption.
-Ok, three captions have been generated for you. Which one do you like?
-Just this one.

(You enter the office and receive a phone call)
-There is a phone call from a customer, and I have transferred it to the IVI for you. Would you like to answer it?
-Yes, and turn on the office mode.
-Ok, I will generate an avatar of the customer for you and start the conversation. Do you have any questions or things to discuss? I will record the conversation and provide you with the minutes of the meeting after it is over.
(The avatar is generated)
-Now you can start the conversation.

The above scenarios will not be materialized in the cockpit until 2024 when foundation models are installed on vehicles by some OEMs.

IM L6 has built Carlog and City Drive scenarios

For example, IM L6 has built Carlog and City Drive scenarios to enable the AI foundation models to proactively recommend food and attractions and allow users to post them on social media:

Carlog: Actively perceive the scenario during driving through AI vision foundation model, mobilize four cameras to take photos, automatically save and edit them, and support one-click share in Moments.

City Drive: Cooperate with Volcengine to model nearby food, scenic spots and landmarks in real time in the digital screen, and push them according to users' habits and preferences.

IM Carlog

Source: IM

Applicability of foundation models in various scenarios

The applicability of foundation models in various scenarios has stimulated users' demand for intelligent agents that can uniformly manage cockpit functions. In 2024, OEMs such as NIO, Li Auto, and Hozon successively launched Agent frameworks, using voice assistants as the starting point to manage functions and applications in cockpits.

Agent service frameworks can not only manage cockpit functions in a unified way, but also provide more abundant scenario modes according to customers' needs and preferences, especially supporting customized scenarios for users, which accelerates the advent of the cockpit personalization era.

Some Application Scenarios of AI Agents

No.	Scenarios	Agent functions
1	Mobility	Multimodal technology can provide navigation guidance, play music and adjust temperature inside the car, which makes driving experience more convenient and personalized.
2	Office	As intelligent assistants, AI Agents can help drivers and passengers to deal with work tasks, such as scheduling, mail receiving and dispatching, data analysis, etc.
3	Education	24-hour online education resources help drivers and passengers learn new knowledge on the road, and digital people's lectures are both entertaining and educational.
4	Parenting	Educational content and entertainment activities suitable for children are provided, such as storytelling, playing nursery rhymes, etc. Some cockpits support voiceprint replication, so that when the child's emotions are sensed to be abnormal, the parent's voice can be used to comfort the child.
5	Movie watching	Personalized movies are recommended. Voice commands are recognized to control the interior ambient lighting and seat angles. AI noise reduction, AI sound field and other technologies create the best viewing environment.
6	Music cockpit	Personalized music is recommended and wallpaper is generated in real time based on the lyrics. Music can be continued and improvisation is supported.

Source: ResearchInChina

NIO’s NOMI GPT

For example, NIO’s NOMI GPT allows users to set an AI scenario with just one sentence:

NIO’s "AI Scenario" Generation Case

Source: NIO

Core competence of cockpit Agents

AI Agents in the era of foundation models are based on LLMs, whose powerful reasoning expands the applicable scenarios of AI Agents that can improve the thinking capability of foundation models through feedback obtained during operation. In the cockpit, the Agent capability paradigm can be roughly divided into "Understanding" + "Planning" + "Tool Use" + "Reflection".

Cockpit Agent Capability Paradigm

Cockpit Agent capability module	Description
Understanding	• For users: get instructions and analyze intentions (passive perception) • For environment: Understand the environmental changes and the state of the object/car owner (**active perception**).
Planning	• Task decomposition: Break down instructions into sub-tasks. • Solution selection: Construct the task implementation path and select the optimal path
Tool Use	Use external API tools
Reflection	• Learning feedback: Reflect through user feedback, improve details, and enhance planning capabilities for similar events • Enhanced planning: Enhance **generalization** on the basis of learning feedback, and take the optimal path when faced with tasks with the same logic or characteristics.

Source: ResearchInChina

Xiaomi's voice interaction process

When Agents first get on cars, cognitive and planning abilities are more important. The understanding of task goals and the choice of implementation paths directly determine the accuracy of performance results, which in turn affect the scenario utilization rate of Agents.

For example, in Xiaomi's voice interaction process, semantic understanding is the difficulty of the entire automotive voice processing process. XiaoAi handles semantic parsing through a semantic parsing model.

XiaoAi's Voice Command Understanding and Planning

Step	Specific process
Understanding	• Acquire user voice • Send it to the semantic parsing system for recognition. • The semantic parsing system carries out semantic analysis through a parsing model. • Set the confidence of weight evaluation • Output instruction intention
Planning	• Determine the task category • Disassemble the task and plan the path • Decide whether to call an external tool library.

Source: Xiaomi; ResearchInChina

Lixiang Tongxue offered by Li Auto supports the creation of one-sentence scenarios

After the mass production of Agents, the personalized cockpits that support users to customize scenario modes become the highlight, and Reflection becomes the most important core competence at this stage, so it is necessary to build an Agentic Workflow that is constantly learning and optimizing.

For example, Lixiang Tongxue offered by Li Auto supports the creation of one-sentence scenarios. It is backed by Mind GPT's built-in memory network and online reinforcement learning capabilities. Mind GPT can remember personalized preferences and habits based on historical conversations. When similar scenarios recur, it can automatically set scenario parameters through historical data to fit the user's original intentions.

The Agent that can reflect and optimize is designed by Agentic Workflow.

Source: Angus Wu

Z-One accesses the LLM kernel (LLM OS) at the kernel layer

At the AI OS architecture setting level, we take SAIC Z-One as an example:

Z-One accesses the LLM kernel (LLM OS) at the kernel layer, which controls the interfaces of AI OS SDK and ASF with the original microkernel respectively, in which AI OS SDK receives the scheduling of the LLM to promote the Agent service framework of the application layer. The Z-One AI OS architecture highly integrates AI with CPU. Through SOA atomic services, AI is then connected to the vehicle's sensors, actuators and controllers. This architecture, based on a terminal-cloud foundation model, can enhance the computing power of the terminal-side foundation model and reduce operational latency.

Application Difficulty of Cockpit AI Agents

Agents connect to users and execute commands. In the application process, in addition to the technical difficulties of putting foundation models on cars, they also face scenario difficulties. In the process of command reception-semantic analysis-intention reasoning-task execution, the accuracy of the performance results and the delay in human-computer interaction directly affect the user's riding experience.

Some Difficulties of Using Agents in Cockpits

No.	Problems	Status quo
1	AI OS architecture design	Agents are in the early stages of being installed on vehicles. Combining LLMs with automotive software platforms requires changing the architecture design.
2	Foundation model performance	• The personalized services for users put forward new requirements for the reasoning and feedback learning capabilities of foundation models. • RLHF, CoT and other methods can not completely eliminate the illusion of foundation models, and the accuracy of task execution results determines the user experience and the reputation of Agents.
3	Generalization capability	The mode of "creating a task in one sentence" and the "custom" mode require foundation models to strengthen their capabilities to deal with unexpected scenarios and feature universality.
4	Data privacy	User data security is involved
5	Cockpit platform selection and integration mode	Both platform effect and cost should be taken into account
6	Response time	The response time of AI Agents is the key to user experience.
7	Update and maintenance	In order to save cost, it is necessary to strengthen the updating and iteration of foundation models.
8	Humanization of interaction	Users need Agents to provide services with more human feelings and "consideration".

Source: ResearchInChina

Humanization of interaction

For example, in the "emotional consultant" scenario, Agents should resonate emotionally with car owners and perform anthropomorphism. Generally, there are three forms of anthropomorphism of AI Agents: physical anthropomorphism, personality anthropomorphism, and emotional anthropomorphism.

Three Anthropomorphism Methods of AI Agents

Anthropomorphism method	Features	Typical path
Physical anthropomorphism	Feature a humanoid shell	-
Personality anthropomorphism	Add social elements and give social status and position.	Database and corpus
Emotional anthropomorphism	Simulate human emotional characteristics, such as tone of voice and facial expressions.	Memory

Source: ResearchInChina

NIO's NOMI GPT uses "personality anthropomorphism" and "emotional anthropomorphism"

NIO's NOMI GPT uses "personality anthropomorphism" and "emotional anthropomorphism":

Anthropomorphism method	Application	Implementation path
Physical anthropomorphism	×	-
Personality anthropomorphism	√	• Through the emotional engine, a unique personality is established, including character, outlook on life, world view, values, dreams, etc. Every sentence and expression of NOMI should conform to NOMI's basic persona. • The state of task execution is displayed on the interactive interface, mainly the thinking state, such as "thinking" and "generating", to simulate people's thinking state.
Emotional anthropomorphism	√	• Through the cognitive center, the capability to recognize and understand voice context is enhanced, and then task-based interaction is called. For example, it supports users to say the destination after an interruption and then start navigation directly without the user having to repeat the complete command again. • Equipped with an emotion engine, it uses long-term and short-term memory to remember the user's recent topics, people and things the user mentioned before, the user's family and friends, the user's preferences, etc., and continuously iterates through real-time feedback, post-event reflection, manual training, etc., and builds NomiGPT's emotion output through the optimization of foundation model reasoning and the update of the database.

Source: NIO; ResearchInChina

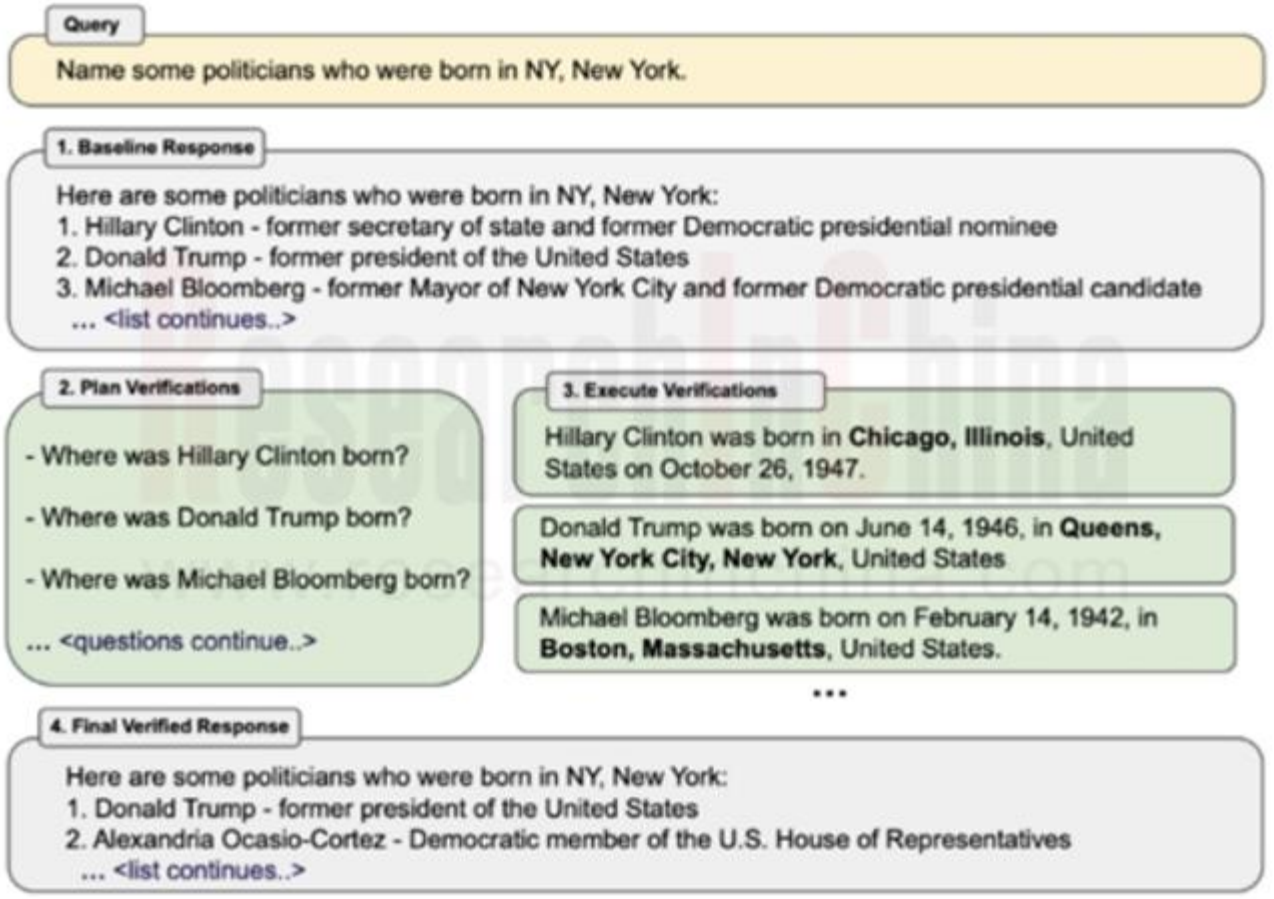
Foundation model performance

In the "encyclopedia question and answer" scenario, Agents may be unable to answer the user's questions, especially open questions, accurately because of LLM illusion after semantic analysis, database search, answer generation and the like.

Current solutions include advanced prompting, RAG+knowledge graph, ReAct, CoT/ToT, etc., which cannot completely eliminate “LLM illusion”. In the cockpit, external databases, RAG, self-consistency and other methods are more often used to reduce the frequency of “LLM illusion”.

Some foundation model manufacturers have improved the above solutions. For example, Meta has proposed to reduce “LLM illusion” through Chain-of-Verification (CoVe). This method breaks down fact-checking into more detailed sub-questions to improve response accuracy and is consistent with the human-driven fact-checking process. It can effectively improve the FACTSCORE indicator in long-form generation tasks.

CoVe includes four steps: query, plan verification, execute verification and final verified response.



Source: Meta

Table of Content (1)

Table of Content (2)

Table of Content (3)

Table of Content (4)

Contact

Beijing Headquarters

TEL: 13718845418
Email: report@researchinchina.com

Website: ResearchInChina

WeChat: Zuosiqiche

Chengdu Branch

TEL: 028-68738514
FAX: 028-86930659

zuosiqiche