



**ResearchInChina**  
[www.researchinchina.com](http://www.researchinchina.com)

**Automotive AI Agent Product  
Development and  
Commercialization Research  
Report, 2024**

Aug. 2024

# Automotive AI Agent product development: How to enable “cockpit endorser” via foundation models?

According to OPEN AI's taxonomy of AI (a total of 5 levels), AI Agent is at L3 in the AI development path:

**OPEN AI's Taxonomy of AI**

AI Level	Definition	Feature
L1	Chatbots	AI with the ability to converse
L2	Reasoners	AI that performs reasoning like humans and can solve a variety of complex problems
L3	Agents	AI that automatically takes actions and performs tasks on behalf of users
L4	Innovators	AI that helps humans complete new inventions
L5	Organizations	AI that operates organizations

Source: OPEN AI

Limited by interaction modes and tool usage capabilities, popular foundation models in 2023 can only reach L2 (Reasoners) at most. In contrast, developing the automotive AI system by building automotive agents is a more appropriate goal: Agent improves weak links of application of foundation models in scenarios by way of calling active intelligent features and multiple tools/foundation models, further improving cockpit intelligence level.

# Agent is the endorser of emotional cockpits

"Emotional cockpit" has been nothing new for multiple years, but actually realizing it still starts with the introduction of foundation models in vehicles. Under specific triggering conditions, voice assistant chats with the user through preset emotional corpus, but it cannot adapt to human dialogue logic in real chat scenarios. After being applied to vehicles, Agent integrated with multiple foundation model bases can recognize the environment more accurately, and more tool library interfaces further enhances its generalization capability to cope with chat and Q&A in diversified scenarios, truly realizing the warm companionship of the "cockpit endorser".

## Emotional Care Cases Provided by Some Agents in Different Scenarios

Scenario	Agent's Emotional Caring Moves
Mobility scenarios	<ul style="list-style-type: none"><li>• According to users' mobility habits and weather forecast, remind users whether they need to bring an umbrella or extra warm clothing.</li><li>• When waiting at a red light or in traffic jams, Agent will recommend music, podcasts or audiobooks that users like to relieve their anxiety of waiting.</li><li>• If Agent detects the driver's fatigue, it will provide games or interactions, play music or release fragrance to keep the driver awake or happy.</li></ul>
Chat scenarios	<ul style="list-style-type: none"><li>• If Agent recognizes that the user is nervous through voice analysis, it will play relaxing music or tell jokes to help the user relax.</li><li>• On the user's birthday or special anniversary, Agent will send a greeting and may suggest some celebratory activities.</li></ul>
Q&A scenarios	<ul style="list-style-type: none"><li>• When the user mentions a relevant place, Agent will perceive his/her status and make inquiries, for example, if "hospital" is mentioned, it will monitor vital signs and provide caring through emotional corpus.</li><li>• Perceive user needs at the intervals between questions and answers, and proactively provide solutions or channels.</li></ul>

Source: ResearchInChina

# Agent Emotion Technology/Products of Some OEMs and Tier1s

The design of mainstream emotional interaction scenarios focuses on emotion recognition, user memory, and behavior arrangement. Some OEMs and Tier1s have also launched technologies or products to enhance the emotional value of Agents:

Agent Emotion Technology/Products of Some OEMs and Tier1s

OEM	Agent Emotion Technology/Products
NIO	The emotion engine, involving situational intelligence, exclusive intelligence and emotional expression, can screen, infer, calculate and precipitate interactions between users and NOMI, transform them into personnel exclusive memories, and use a streaming prediction model similar to GPT for search and association.
Xiaomi	<ul style="list-style-type: none"><li>• Install the "emotional dialogue system", and use the combination of "emotion category + emotion reason" to cover a variety of fine-grained emotions and improve Xiaoai Tongxue's emotion recognition accuracy.</li><li>• Flexibly integrate multiple dredging strategies when replying to make the conversation more coherent, and increase users' willingness to interact, more in line with real-life conversations.</li></ul>
Chery	Launch an emotional interaction model solution to analyze and judge current emotional state of the driver or passengers, and enter a specific intelligent emotional interaction scenario to achieve emotional interaction through audio and video interaction information.
SenseTime	Launch a large anthropomorphic model which makes quick response and supports long conversation memory, allowing Agent to accurately memorize the content of more than dozens of historical conversation rounds. It also supports setting of different levels of conversation intimacy, providing users with different conversation experiences at different stages.
iFLYTEK	Launch multi-emotion super-anthropomorphic synthesis technology, combined with the sound replication function, to enhance the perceptibility of emotional expression.

Source: ResearchInChina

# Steps to Build Xiaoai Tongxue's "Emotional Dialogue System"

For example, Xiaoai Tongxue's "emotional dialogue system" is built in three steps:

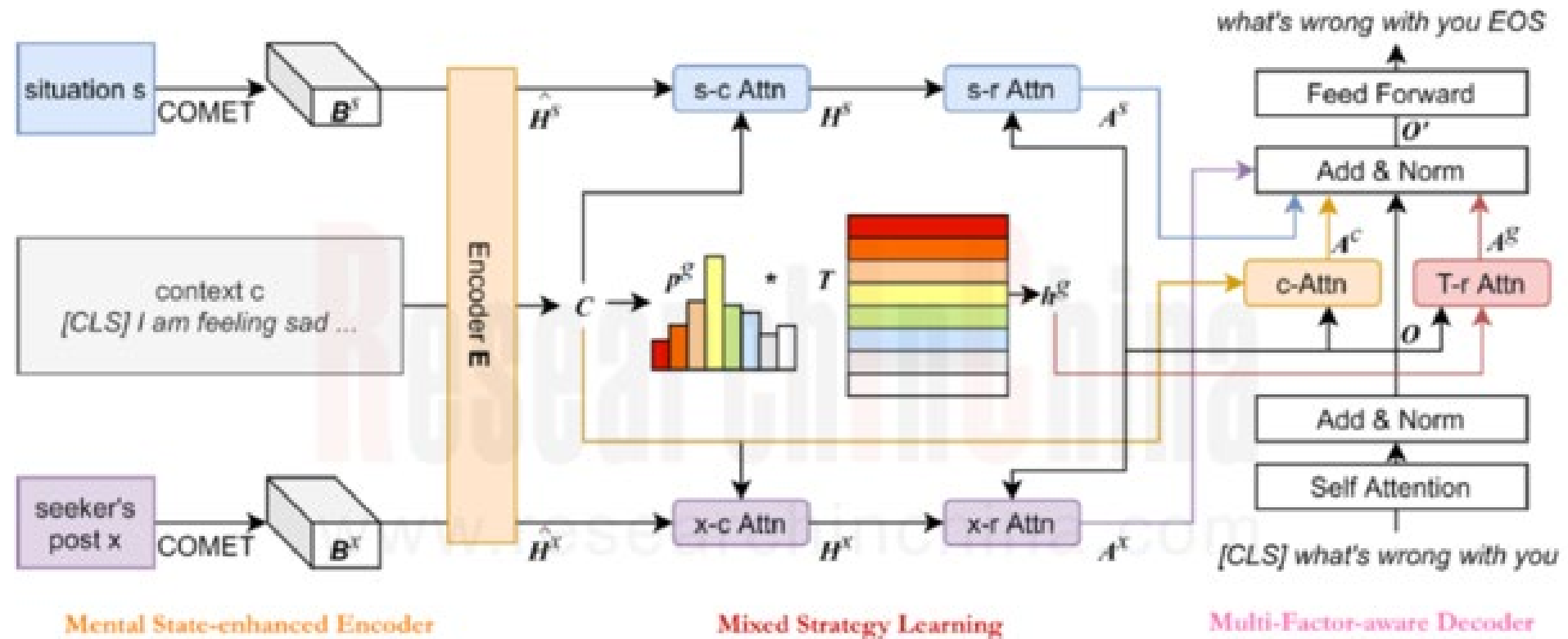
**Steps to Build Xiaoai Tongxue's "Emotional Dialogue System"**

No.	Implementation Step	Features
1	Emotion recognition and understanding	<ul style="list-style-type: none"><li>• "Emotion category + emotion reason" taxonomy, covering a variety of fine-grained emotions</li><li>• Use constrained optimization objectives to improve algorithms to solve the problem that general taxonomy models cannot distinguish similar emotional expressions.</li></ul>
2	Timely response and empathetic response	Use empathetic response algorithms based on common sense graph.
3	Long-term interaction and emotional dredging	Propose a mixed dredging strategy model which integrates multiple dredging strategies when replying, enables more coherent conversations, and makes users more willing to interact, more in line with real-life conversations.

Source: ResearchInChina

# The mixed strategy dredging model of Xiaomi

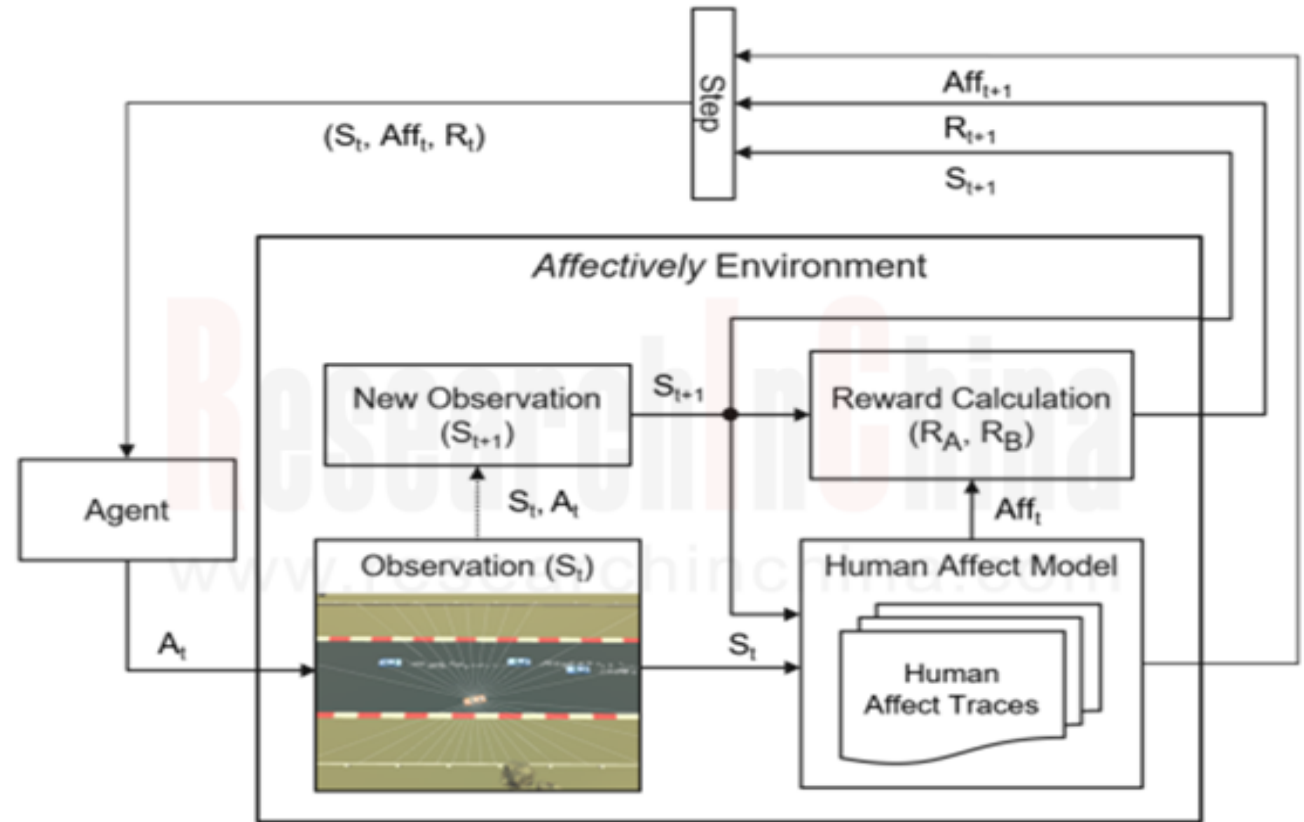
The mixed strategy dredging model is composed of three important components: mental state-enhanced encoder, mixed strategy learning module, and multi-factor-aware decoder.



Source: Xiaomi

The Institute of Digital Games at University of Malta proposes the Affectively Framework, establishes an emotional model, and adopts behavior reward and affective reward mechanisms in the training process to help Agents better understand human emotions and interact with humans more naturally.

Affective Reward-based Affectively Framework



Source: Institute of Digital Games, University of Malta

# Sore points that need to be solved to improve user experience

Imagine that an intelligent cockpit can not only understand and execute instructions given by the car owner, but also predict the owner's needs, just like a thoughtful personal assistant. Will this make car owners more excited? Compared to buying a traditional car and having to explore each function on one's own, everyone wants a cockpit "endorser" which can help manage all cockpit functions as they just say a few words. Agent is a time-saving and trouble-free solution.

Currently most Agents introduced in vehicles still serve as an assistant and a companion listing functions for specific scenarios. Yet compared with foundation models, Agents feature greater potential, motivated autonomy, and outstanding tool-using capabilities, more fit with the label of "active intelligence", and can even make up for the limitations of foundation models in practical applications.

## Some Differences between Foundation Models and Agents

Product	Difference at Interaction Level	Difference at Tool Using Level	Difference at Capability Level
Foundation Model	Passive interaction, relying on the outside world to complete tasks/goals, lack of autonomy	Lack of real-time performance	Single skills in professional fields, for example, LLM only deals with language tasks
Agent	Actively perceive the interaction via sensors, realize self-prompt and actively complete tasks	Connect to multiple tool libraries updated in real time	Call multiple and multiple types of foundation models, with rich and diverse capabilities and technical approaches

Source: ResearchInChina



## Sore points that need to be solved to improve user experience

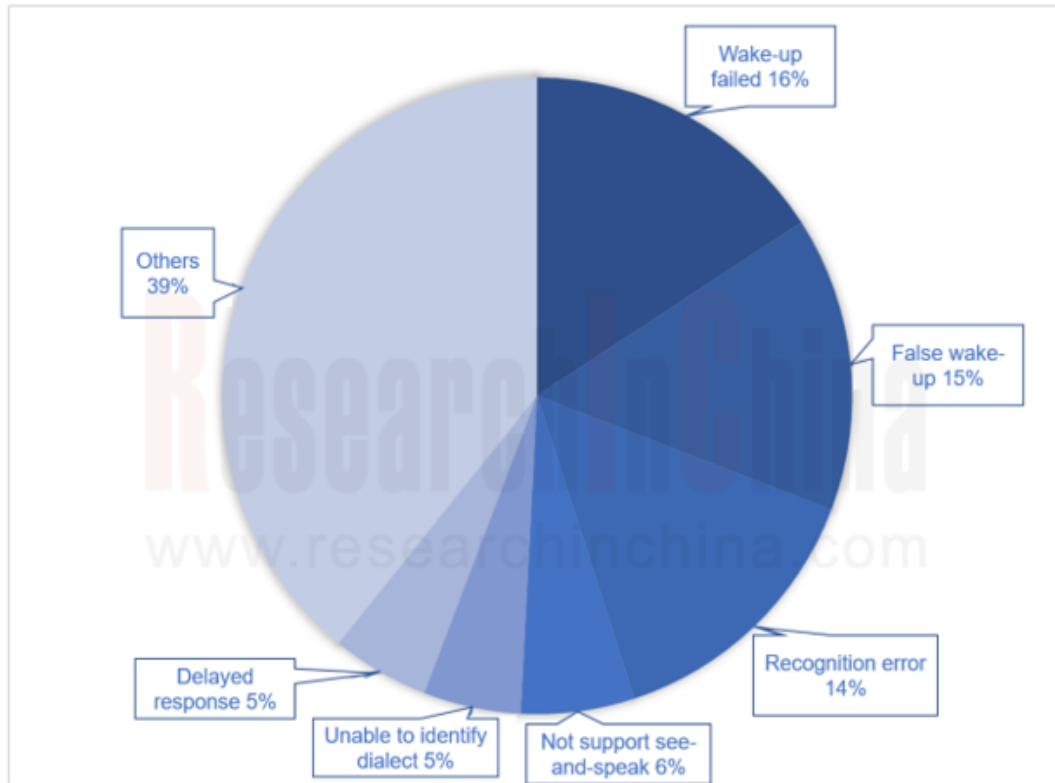
There is however still a long way to go in technology development to make automotive agents truly "active and intelligent" and meet users' experience value. Agent needs to be more precise in active perception, data processing, state recognition, etc., accurately understand the environment, judge real needs of people in the car, and then adopt corresponding strategies. Wherein, one of challenges lies in Agent's accurate judgment of user needs. Compared with passive interaction in normal circumstances, active intention recognition lacks voice commands. In the process of environment/personnel/vehicle state recognition, it may not be possible to obtain a description that is extremely close to the current scenario through vector feature matching, and the preset solution may not satisfy the real intentions of people in the car.

At present, most recommended functions are just to execute preset instructions. This limits "active and intelligent" capabilities of Agent and leads to frequent sore points in the reasoning process. For example, if Agent fails to accurately understand the current scenario, it may not make recommendations as expected, for instance, recommending music or navigation at a wrong time. The final result is to affect user experience and make the Agent become a "guessing machine" to users.

In addition, Agent also has shortcomings in perception when receiving voice commands. According to ResearchInChina's incomplete statistics on sore points in automotive agent use cases of some car owners, the most frequent sore points are wake-up failure, recognition error, and false wake-up.

# Sore Points in Using Automotive Agents based on Incomplete Statistics

Sore Points in Using Automotive Agents based on Incomplete Statistics



Source: ResearchInChina

Among the 120 cases, wake-up failure, recognition error, and false wake-up are mentioned 19, 18, and 17 times respectively, namely, accounting for 16%, 15%, and 14%. Other sore points also include unavailability of see-and-speak, semantic clarification and continuous commands, inability to recognize dialects, and delayed response, totaling 89 sore points in voice link, or 74.2% of the total in this statistical survey.

Furthermore, a range of problems caused by unreasonable Agent architecture/scenario design also include irrational scenario triggering conditions, secondary wake-up of foundation models, failure of long/short-term memory, and recommended actions made autonomously according to owners' habits but failing to meet expectations, which respectively reflect limitations of Agent in scenario setting, architecture deployment, memory module, and reflection module.

In summary, sore points of users are concentrated in the perception and reasoning links:

- Perception: wake-up failure, false wake-up, recognition error, unavailability of see-and-speak, delayed response, etc.
- Reasoning: object recognition error, autonomous recommendation failing to meet user expectations, etc.

# Quick-response multi-agent framework

To enable all the functions of the "endorser" in cockpit, it is very critical to design the service framework of Agent in diversified scenarios. Agent framework is relatively flexible in construction. The simplest "receiver + executer" architecture can be used, or a more complex multi-agent architecture can be built. Its design principle is very simple: as long as it can solve user problems in a specific scenario, it is a good framework design. As a qualified "cockpit endorser", automotive Agent not only needs to act as an independent thinker, make decisions and solve problems on its own, but also quickly and freely adopts human behavior patterns, acting as a human.

A typical example is NIO Nomi. It uses a multi-agent architecture, calling different tools in different scenarios, and using multiple agents with different functions to perform specific duties and jointly complete the process of understanding needs, making decisions, executing tasks, and reflecting on iterations. The multi-agent architecture allows Nomi to not only make quick response, but also react more naturally like a human. Its seamless integration with other vehicle functions brings smoother experiences.

Compared with single-agent systems, multi-agent systems are more suitable for executing complex instructions. They are like a small community in which each "agent" has its own tasks, but can cooperate to complete more complex tasks. For example, one agent is responsible for understanding your instructions, another is responsible for making a decision, and there are special agents to perform tasks. This design makes automotive AI Agent systems more flexible and allows them to handle more diverse tasks. For example, the Commonwealth Scientific and Industrial Research Organization (CSIRO) of Australia proposed a multi-agent system that uses both collaboration agents and execution agents:

The entire Agent framework is divided into 6 modules, namely, Understanding & Interaction, Reasoning, Tool Use, Multi-Agent Collaboration, Reflection, and Alignment. It embraces mainstream Agent design patterns, and covers the entire process from active perception, reasoning and decision, tool calling to generation and execution, reflection and iteration, and alignment with human values. This framework features a multi-agent system where different Agents can play different roles (distribution/decision/actuation) in the entire process, making best use of each Agent to improve task execution efficiency.

In addition, in diversified scenarios, Agent deployment methods and tool calling capabilities also affect whether or not user needs can be quickly and accurately executed. Take NIO Nomi as an example:

Nomi Agents are deployed at the end and cloud sides. End-side model and NomiGPT are deployed at the end and cloud sides, respectively. Deeply integrated with SkyOS, the end-side model can call atomic capabilities in time and schedule resources (data, vehicle control hardware/software, etc.) across domains to speed up response. NomiGPT on the cloud connects more cloud tool resource interfaces to further enhance Nomi Agents' capability of calling tools. Nomi Agents' architecture is arranged in SkyOS middleware layer. Combining with SkyOS, it makes the process of calling atomic APIs, hardware/software and data more natural, coordinated, and faster.

# Table of Content (1)

## 1 Overview of Automotive AI Agent

- 1.1 Definition of Agent
- 1.2 Development History of Agent
- 1.3 Foundation Models Regain Vitality Using the Agent Concept
- 1.4 Differences between Foundation Models, Agents, and AIGC
- 1.5 Automotive AI Agent Product Definition
- 1.6 Automotive AI Agent based on Multi-agent System: Module Design
- 1.6 Automotive AI Agent based on Multi-agent System: Component Functions
- 1.6 Automotive AI Agent based on Multi-agent System: Component Characteristics (1)
- 1.6 Automotive AI Agent based on Multi-agent System: Component Characteristics (2)
- 1.6 Automotive AI Agent based on Multi-agent System: Component Characteristics (3)
- 1.6 Automotive AI Agent based on Multi-agent System: Component Characteristics (4)
- 1.6 Automotive AI Agent based on Multi-agent System: Component Characteristics (5)
- 1.6 Automotive AI Agent based on Multi-agent System: Component Characteristics (6)
- 1.6 Automotive AI Agent based on Multi-agent System: Component Characteristics (7)
- 1.6 Automotive AI Agent based on Multi-agent System: Component Characteristics (8)
- 1.7 Automotive AI Agent Reference Architecture (by Functional Module and Component)
- 1.7 Automotive AI Agent Reference Architecture (by Deployment Level)
- 1.8 Agent Architecture Case (1): Original Diagram of NIO (Nomi) Architecture
- 1.8 Agent Architecture Case (1): Original Diagram of NIO (Nomi) Deployment

- 1.8 Agent Architecture Case (1): NIO (Nomi) Module Design
- 1.8 Agent Architecture Case (1): NIO (Nomi) Module Design - Multimodal Perception
- 1.8 Agent Architecture Case (1): NIO (Nomi) Module Design - Command Distribution
- 1.8 Agent Architecture Case (1): NIO (Nomi) Module Design - Scenario Customization and Creation Process
- 1.8 Agent Architecture Case (1): Highlights of NIO (Nomi) (1)
- 1.8 Agent Architecture Case (1): Highlights of NIO (Nomi) (2)
- 1.8 Agent Architecture Case (1): Highlights of NIO (Nomi) (3)
- 1.8 Agent Architecture Case (2): Original Diagram of Li Auto (Lixiang Tongxue) Architecture
- 1.8 Agent Architecture Case (2): Li Auto (Lixiang Tongxue) Module Design
- 1.8 Agent Architecture Case (2): Li Auto (Lixiang Tongxue) Supporting Facilities - Data/Training Platform
- 1.8 Agent Architecture Case (2): Li Auto (Lixiang Tongxue) Supporting Facilities - Reasoning Engine
- 1.8 Agent Architecture Case (3): Original Diagram of Xiaomi (Xiaoai Tongxue) Architecture
- 1.8 Agent Architecture Case (3): Xiaomi (Xiaoai Tongxue) Module Design
- 1.8 Agent Architecture Case (4): Zeekr Agent Module Design
- 1.8 Agent Architecture Case (5): Original Diagram of Neta Agent Architecture Deployment
- 1.8 Agent Architecture Case (5): Neta Agent Module Design
- 1.8 Agent Architecture Case (6): Original Diagram of BAIC Agent Architecture Deployment
- 1.8 Agent Architecture Case (6): BAIC Agent Module Design
- 1.8 Agent Architecture Case (7): Huawei (Pangu Agent) Module Design
- 1.8 Agent Architecture Case (8): Original Diagram of AISpeech Agent Architecture Deployment
- 1.8 Agent Architecture Case (8): AISpeech Agent Module Design

# Table of Content (2)

- 1.8 Agent Architecture Case (9): Original Diagram of Lenovo Agent Architecture Deployment
- 1.8 Agent Architecture Case (10): Original Diagram of Zhipu Agent Architecture Deployment
- 1.8 Agent Architecture Case (10): Zhipu Agent Module Design
- 1.8 Agent Architecture Case (11): Original Diagram of Tinnove Agent Architecture Deployment
- 1.8 Agent Architecture Case (11): Tinnove Agent Module Design
- 1.9 Agent Architecture Design Process: Framework Selection
- 1.9 Agent Architecture Design Process: Tool Calling Method
- 1.10 Comparison of Automotive AI Agent Architecture

## **2 Key Issues in Development of Automotive AI Agent Products - User Sore Points and Technical Difficulties**

- 2.1 Classification of Automotive AI Agent Scenario: Typical Commands in Different Scenarios
- 2.1 Classification of Automotive AI Agent Scenario: Case (1) NIO
- 2.1 Classification of Automotive AI Agent Scenario: Case (2) Li Auto
- 2.1 Classification of Automotive AI Agent Scenario: Case (3) Xiaomi
- 2.2 Automotive AI Agent Scenario Design Case (1) Q&A Scenario
- 2.2 Automotive AI Agent Scenario Design Case (2) Q&A Scenario
- 2.2 Automotive AI Agent Scenario Design Case (3) Mobility Scenario
- 2.2 Automotive AI Agent Scenario Design Case (4) Chat Scenario
- 2.2 Automotive AI Agent Scenario Design Case (5) Chat Scenario
- 2.2 Automotive AI Agent Scenario Design Case (6) Chat Scenario
- 2.2 Automotive AI Agent Scenario Design Case (7) Q&A/Office Scenario
- 2.3 User Sore Points in Different Agent Usage Scenarios: Summary
- 2.4 User Sore Points (1): Vehicle Control Scenario
- 2.4 User Sore Points (2): Mobility Scenario

- 2.4 User Sore Points (3): Q&A Scenario
- 2.4 User Sore Points (4): Entertainment Scenario
- 2.5 Agent Technical Difficulties
- 2.6 Agent Technology Case: Intent Recognition (Case 1)
- 2.6 Agent Technology Case: Intent Recognition (Case 2)
- 2.6 Agent Technology Case: Intent Recognition (Case 3)
- 2.6 Agent Technology Case: Intent Recognition (Case 4)
- 2.6 Agent Technology Case: Reasoning Acceleration (Case 1)
- 2.6 Agent Technology Case: Reasoning Acceleration (Case 2)
- 2.6 Agent Technology Case: Reasoning Acceleration (Case 3)
- 2.6 Agent Technology Case: Streaming Voice (Case 1)
- 2.6 Agent Technology Case: Streaming Voice (Case 2)
- 2.6 Agent Technology Case: Streaming Voice (Case 3)
- 2.6 Agent Technology Case: Emotional Interaction (Case 1)
- 2.6 Agent Technology Case: Emotional Interaction (Case 2)
- 2.6 Agent Technology Case: Emotional Interaction (Case 3)
- 2.7 Agent Technology Trends (1): Two Keys to Achieving Active Intelligence
- 2.7 Agent Technology Trends (2):
- 2.7 Agent Technology Trends (3): Two Mainstream Design Methods for Emotional Anthropomorphism

## **3 OEMs' AI Agent Investment, Development, and Operation**

- 3.1 Comparison of Automotive AI Agent Development Support
- 3.2 OEMs' Planning for Automotive AI Agents
- 3.3 Comparison between Three Automotive AI Agent Development Modes: Advantages/Disadvantages
- 3.3 Comparison between Three Automotive AI Agent Development Modes: Cost
- 3.4 Position Setting of OEMs' AI Agent Team

# Table of Content (3)

- 3.4 Case of OEMs' AI Agent Team Position Setting (1): Positions Recruited by Chery AI Agent Team
- 3.4 Case of OEMs' AI Agent Team Position Setting (2): Positions Recruited by Geely AI Agent Team
- 3.4 Case of OEMs' AI Agent Team Position Setting (3): Positions Recruited by Li Auto AI Agent Team
- 3.4 Case of OEMs' AI Agent Team Position Setting (4): Positions Recruited by NIO AI Agent Team
- 3.4 Case of OEMs' AI Agent Team Position Setting (5): Positions Recruited by Xiaomi AI Agent Team
- 3.5 AI Agent Development Cycle and Operation Mode
- 3.6 AI Agent Business: OEMs' Profit Model
- 3.6 AI Agent Business: Suppliers' Profit Model
- 3.6 AI Agent Business: Suppliers' Charging Standards
- 3.7 Commercial Development Trends of Automotive AI Agents (1)
- 3.7 Commercial Development Trends of Automotive AI Agents (2)

## **4 Automotive AI Agent Suppliers and Their Supply Relationships**

- 4.1 Cockpit Base Foundation Model: Model Configurations
- 4.1 Cockpit Base Foundation Model: Selection Reference Factors
- 4.2 Cockpit Base Foundation Model Suppliers (1)
- 4.2 Cockpit Base Foundation Model Suppliers (2)
- 4.2 Cockpit Base Foundation Model Suppliers (3)
- 4.2 Cockpit Base Foundation Model Suppliers (4)
- 4.2 Cockpit Base Foundation Model Suppliers (5)
- 4.2 Cockpit Base Foundation Model Suppliers (6)
- 4.2 Cockpit Base Foundation Model Suppliers (7)
- 4.2 Cockpit Base Foundation Model Suppliers (8)
- 4.2 Cockpit Base Foundation Model Suppliers (9)

- 4.2 Cockpit Base Foundation Model Suppliers (10)
- 4.3 Industry Chain of Vector Database Suppliers
- 4.4 Comparison between Vector Database Products: Chinese Vector Databases
- 4.4 Comparison between Vector Database Products: Foreign Vector Databases
- 4.5 Vector Database Supplier Cases (1)
- 4.5 Vector Database Supplier Cases (2)
- 4.5 Vector Database Supplier Cases (3)
- 4.5 Vector Database Supplier Cases (4)
- 4.5 Vector Database Supplier Cases (5)
- 4.5 Vector Database Supplier Cases (6)
- 4.5 Vector Database Supplier Cases (7)
- 4.5 Vector Database Supplier Cases (8)
- 4.6 Comparison between Voice ASR Module Suppliers
- 4.7 ASR Module Supplier Cases (1)
- 4.7 ASR Module Supplier Cases (2)
- 4.7 ASR Module Supplier Cases (3)
- 4.7 ASR Module Supplier Cases (4)
- 4.7 ASR Module Supplier Cases (5)
- 4.7 ASR Module Supplier Cases (6)
- 4.7 ASR Module Supplier Cases (7)
- 4.7 ASR Module Supplier Cases (8)
- 4.7 ASR Module Supplier Cases (9)
- 4.8 Cockpit Data Collection Sensors: Mainstream Configurations/Data Collection Regulations
- 4.9 Sensor Data Processing Cases (1)
- 4.9 Sensor Data Processing Cases (2)
- 4.9 Sensor Data Processing Cases (3)
- 4.9 Sensor Data Processing Cases (4)



## Beijing Headquarters

TEL: 13718845418

Email: [report@researchinchina.com](mailto:report@researchinchina.com)

Website: [ResearchInChina](http://ResearchInChina.com)

WeChat: Zuosiqiche



## Chengdu Branch

TEL: 028-68738514

FAX: 028-86930659

