WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology

ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 8, Number 6, Year 2026
DOI: https://doi.org/10.62051/ijcsit.v8n6.05
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Research on Building a Vertical AI Foundation for Automotive Digital Marketing with Open-Source Large Models

Rui Shi *, Haonan Chen

CATARC Brand Technology (Tianjin) Co., Ltd., China
*Corresponding Author: shirui2022@catarc.ac.cn

ABSTRACT

As the automotive industry transitions from "incremental expansion" to "deepening existing markets," digital marketing is emerging as a critical battleground for cost reduction and efficiency enhancement. This study focuses on the "platform AI large model foundation" system, exploring how to integrate open-source large models with domain-specific automotive datasets to train vertical models tailored for digital marketing scenarios in the automotive sector, thereby supporting core capabilities such as semantic understanding and content generation. The research constructs a specialized automotive marketing corpus covering vehicle analysis, competitive benchmarking, user profiling, and marketing messaging. It proposes a dual-model training architecture combining LoRA-based parameter-efficient fine-tuning and RAG (Retrieval-Augmented Generation), and integrates knowledge graphs into the marketing decision-making pipeline. This work provides a practical technical foundation and systematic methodology for the intelligent transformation of digital marketing in the automotive industry.

KEYWORDS

Large language model; Automotive digital marketing; Vertical model; Supervised fine-tuning; Retrieval-augmented generation

1. INTRODUCTION

The global automotive industry is currently undergoing a profound paradigm shift. The ongoing waves of electrification, intelligence, and connectivity are continuously evolving, reshaping vehicle product forms, industrial chain divisions, and competitive landscapes. Meanwhile, the logic driving China's automotive market has transitioned from incremental expansion to a new phase focused on deepening existing demand. As growth benefits gradually fade, competition within the industry has intensified. Traditional, broad-based marketing models—centered on maximizing sales leads—are no longer sustainable. Automakers are systematically shifting their marketing focus toward lead conversion quality, operational efficiency, and deeper exploration of long-term user value. Amid this dual transformation driven by technological innovation and market evolution, artificial intelligence—particularly large language models—has achieved breakthrough progress, creating unprecedented opportunities for digital marketing transformation in the automotive sector. Leveraging the powerful natural language understanding, multi-turn dialogue, and data analytics capabilities of large models, automakers now have the opportunity to reconstruct key business processes such as lead nurturing, private domain operations, and end-to-end user lifecycle management, enabling a leapfrog upgrade toward smarter, more refined, and personalized marketing.

In industry practice, intelligent automotive marketing has begun exploring multiple pathways. DeepAgent, developed by Shenyan Intelligence in collaboration with DeepBlue Auto, employs a "large model + industry-specific small model" architecture to empower the entire lead management lifecycle. Kuka Technology leverages transaction data and knowledge graphs to enable intelligent generation of marketing content, while Qiyu's "New Chain" trains a vertical automotive large model using vast amounts of live-streaming private message data. These cases demonstrate that general-purpose large models must be deeply integrated with domain-specific automotive marketing knowledge to deliver tangible business value.

Although general large models excel in language understanding and generation, they still face four core challenges when directly applied to automotive digital marketing: first, insufficient industry-specific marketing knowledge—due to limitations in training data, model outputs lack automotive expertise and brand differentiation; second, inaccuracies in user intent recognition—automotive purchase decisions involve long cycles, making it difficult for general models to precisely determine a user's stage in the marketing funnel and match appropriate operational strategies; third, content production efficiency bottlenecks—high-frequency, multi-channel content demands place significant pressure on supply, while manual creation is costly; fourth, difficulties in building a closed-loop data system—the models struggle to deeply integrate with automakers' user platforms, failing to meet requirements for localized storage and privacy compliance. At its core, this reflects a misalignment between the capabilities of general large models and the specialized needs of automotive marketing scenarios. Therefore, building a vertical AI foundation infused with domain-specific automotive marketing knowledge has become an urgent technical challenge to overcome.

2. TECHNICAL FOUNDATION

2.1. Evolution of Large Language Model Technology

The development of large language models has evolved from statistical language models to neural network-based language models, and further to pre-trained language models. The introduction of the Transformer architecture marked a pivotal turning point in this evolution, as its core self-attention mechanism enables models to capture long-range dependencies, laying the architectural foundation for subsequent large-scale pre-training. The multi-head attention mechanism, by operating multiple attention heads in parallel, allows the model to capture semantic relationships [3-4] across different subspaces, with its mathematical formulation shown in Equation (1).

$$\begin{gathered} Mhead(Q,K,V) = Concat(h_1, h_2, \cdots, h_i)W^o \\ h_i = Attenetion(QW_i^Q, KW_i^K, VW_i^V) \end{gathered} \tag{1}$$

In 2018, the release of BERT and GPT-1 officially ushered in the era of pre-trained language models. Since then, model parameter sizes have continued to grow, with models such as GPT-3 and PaLM demonstrating outstanding few-shot and in-context learning capabilities. In 2022, ChatGPT emerged, leveraging RLHF technology to significantly improve model instruction following and conversational interaction, propelling large-scale models into a phase of widespread industrial application.

2.2. Method for Building Vertical Domain Models

(1) Supervised Fine-Tuning: Supervised fine-tuning (SFT) trains models using domain-labeled data to acquire industry-specific knowledge and task paradigms, forming the foundation for building vertical models. Medium-scale open-source models optimized through SFT can achieve performance comparable to large models on domain-specific tasks; in automotive marketing, they can learn vehicle specifications, sales scripts, and customer purchase decision pathways.

(2) Parameter-efficient fine-tuning: To reduce the computational cost of full-parameter fine-tuning, methods such as LoRA introduce low-rank adapter modules, enabling near-full-parameter fine-tuning performance with only a small number of trainable parameters. The core idea of LoRA is to decompose the weight update matrix into the product of two low-rank matrices, as shown in Equation (2).

$$W^{'} = W + \Delta W = W + BA \tag{2}$$

Formula (2) reduces the number of trainable parameters to one ten-thousandth of the original level without introducing inference latency, making it particularly suitable for model customization by automotive marketing companies under limited computational resources. Here, W represents the pre-trained weights, while B and A are low-rank matrices that are trainable.

Retrieval-Augmented Generation: Retrieval-Augmented Generation (RAG) leverages external knowledge bases to expand a model's domain expertise, making it particularly suitable for business scenarios where knowledge evolves rapidly. This approach decouples the knowledge retrieval process from content generation, enabling seamless updates to the knowledge base without requiring retraining of the model. The generated outputs are traceable and verifiable, effectively addressing issues such as knowledge lag in large models and inaccuracies in industry-specific information. It provides a lightweight knowledge-enhancement pathway [5-6] tailored for vertical automotive marketing applications. The mathematical formulation of Retrieval-Augmented Generation is shown in Equation (3).

$$f = LLM(x \oplus \mathrm{Re}\,trieve(x, \mathrm{K})) \tag{3}$$

Here, x represents user input, K is the domain knowledge base, Retrieve is the retrieval function, and $\oplus$ denotes concatenation. In automotive marketing scenarios, integrating a RAG Agent with continuously accumulated business data can significantly enhance the professional capabilities and accuracy of vertical marketing models.

2.3. Automotive Digital Marketing AI Technology

User Intent Understanding and Precision Marketing: Accurately understanding user intent is a prerequisite for personalized automotive precision marketing. By using named entity recognition (NER) to extract vehicle model entities from text, and then matching these entities with an automotive knowledge graph, we can achieve structured analysis of users' car purchase needs. The mathematical expression for extracting vehicle parameter features through BERT-based text feature extraction is given in Equation (4) [7-8].

$$P_i = \frac{1}{m} \sum_{i=1}^{m} BERT(w_{ij}) \tag{4}$$

Here, P_i is the feature vector of automotive parameters for the i-th annotated entity, w_{ij} is the text sequence obtained from the j-th random walk, and m is a random number. On this basis, the marketing feature similarity is calculated to achieve precise matching from user needs to recommended marketing solutions.

3. AUTOMOTIVE DIGITAL MARKETING CORPUS

3.1. Corpus Scope and Collection Strategy

The dedicated automotive digital marketing corpus serves as the foundation for training vertical large models. This study follows the principles of covering the entire marketing lifecycle, emphasizing scenario-specific expertise, and ensuring high-quality content, thereby defining six key directions for data collection.

First, vehicle specifications and configuration data, including dimensions, powertrain, range, optional features, and technical highlights, are collected from sources such as automaker websites, automotive vertical media, and Ministry of Industry and Information Technology announcements, providing a foundation for the model to understand product capabilities.

Second, marketing content and creative materials, including official copy, advertising scripts, short video scripts, live-streaming talking points, review articles, and more, cover styles across multiple platforms such as Douyin, Xiaohongshu, and WeChat. In the trend of content-driven marketing, the quality of this type of material determines the upper limit of a model's content generation capability.

Third, user behavior and anonymized profile data—including car purchase inquiry records, community comments, lead information, as well as profile tags such as purchase stage, budget, and vehicle preferences—are key resources for enabling intent recognition and precision marketing.

Fourth, competitive analysis and market intelligence involve collecting comparative data on competitors, sales reports, price changes, and promotional policies to support competitor analysis and market trend forecasting.

Fifth, the marketing scripts and customer service knowledge base integrates sales scripts, Q&A libraries, objection-handling strategies, and after-sales support information, supplemented by the AMCK public dataset to enrich maintenance content, thereby enhancing intelligent sales assistant scenarios.

Sixth, industry trends and policies and regulations, including dynamic materials such as new energy support policies, dual-credit regulations, and smart connected vehicle-related laws, ensure compliance of marketing content and provide a solid basis for market analysis.

Data collection is conducted through automated scraping and API interfaces to obtain publicly available marketing data; at a deeper level, we collaborate with automakers, dealer groups, and digital marketing service providers to access non-public business data and experiential knowledge. All data collection and usage strictly comply with privacy regulations.

3.2. Knowledge Annotation and Structuring

To enhance the retrievability of data and the relevance of model training, this study constructs a multi-layered knowledge annotation system tailored for marketing scenarios.

(1) Marketing entity annotation: Identify key marketing entities in the text, including brands and models, product specifications, marketing concepts, channels and platforms, and marketing campaigns.

(2) Marketing relationship labeling: Label semantic relationships between entities, such as "vehicle model—competitive comparison—vehicle model," "user needs—matched with—vehicle features," and "marketing content—published on—channel," laying the foundation for constructing a marketing knowledge graph.

(3) User Intent Labeling: For user query data, label the vehicle purchase intent category and the stage of the purchasing process.

This study employs an iterative annotation process combining "model pre-labeling with manual refinement." First, a general large model is used to perform initial labeling on the raw data. Then, domain experts in automotive marketing conduct sampling verification and bias correction on the labeled results, continuously refining the annotation guidelines based on feedback. This approach ensures high-quality dataset annotation and provides high-quality labeled samples for subsequent training of specialized large models.

3.3. Data Cleaning and Preprocessing

The original corpus suffers from diverse formats, high noise levels, and inconsistent quality, requiring a systematic cleaning and preprocessing workflow. The specific data cleaning and preprocessing flowchart is shown in Figure 1.

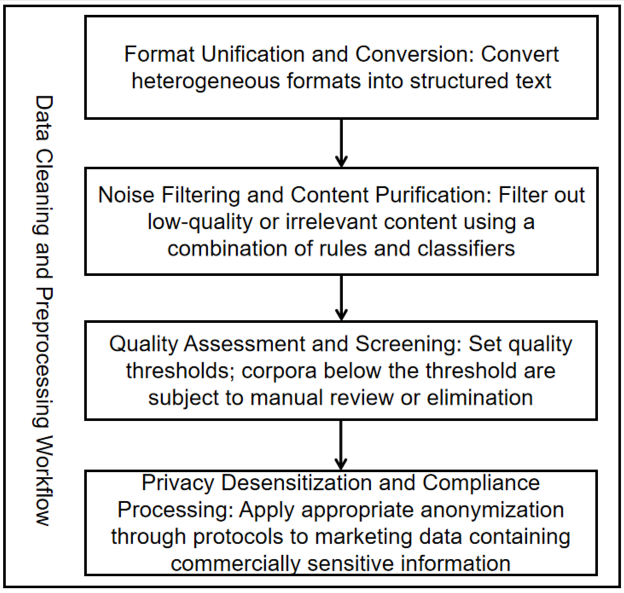

Figure 1. Detailed workflow diagram of data cleaning and preprocessing

As shown in Figure 1, the data preprocessing consists of four steps: First, multi-source and heterogeneous format conversion is performed, using ASR to extract text from short videos while preserving hierarchical structures and metadata. Second, noise such as advertisements and duplicate content is filtered out to purify UGC materials. Third, quality screening is conducted based on dimensions like term density and relevance to eliminate low-quality data. Fourth, user privacy is anonymized and commercially sensitive information is de-identified to ensure data compliance and usability. The preprocessed data is stored in a structured format, with each record containing fields such as document ID, source type, brand/model tags, content category, main text, and metadata.

4. SUPERVISED TRAINING METHODS IN DIGITAL MARKETING

4.1. Supervised Fine-Tuning Training Strategy

Supervised fine-tuning is a core step for vertical models to acquire domain knowledge in automotive marketing. This paper proposes an SFT training strategy divided into two stages: knowledge injection in the marketing domain and instruction fine-tuning for marketing tasks, as detailed below.

Phase 1: Knowledge Injection in Marketing Domain. Use high-quality professional documents from the automotive marketing corpus as training data. The training objective is to maximize the following likelihood function, whose mathematical expression is given in Formula (5).

$$L_{sft} = -\sum_{t=1}^{T} \log P(d_t \mid d_{<t};\theta) \tag{5}$$

Here, d_t represents the t-th token, θ denotes the model parameters, and T is the sequence length. This process aims to reduce the semantic distribution gap between the general-purpose model and the automotive marketing domain, enabling the model to acquire the fundamental knowledge structure and expression patterns specific to marketing scenarios.

Phase 2: Fine-tuning Marketing Task Instructions. Train the model using data in the "instruction-output" format to enable it to understand user intent in marketing scenarios and generate responses that meet professional standards. The training data covers typical marketing tasks in the automotive industry, including content generation, competitive product analysis, user intent recognition, and marketing strategy recommendations, as detailed below:

(1) Marketing Content Creation: Please write a promotional copy for XX suitable for the YY platform;

(2) Competitive Product Comparison Analysis: Please compare the differences between "Model" and "Model B";

(3) User Intent Recognition: Please analyze the stage of car purchase the user's query falls into;

(4) Marketing Strategy Recommendations: Based on XXX, recommend which vehicle model and sales pitch to use.

4.2. Retrieval-Augmented Generation Architecture

Given the frequent updates in automotive marketing information—such as new model launches, price adjustments, and changes in promotional policies—relying solely on SFT struggles to cover all time-sensitive knowledge. This study introduces an RAG architecture on top of SFT, forming a dual-track knowledge enhancement mechanism of "SFT+RAG," which significantly improves the professional competence and accuracy of vertical marketing models. The core components of the RAG system include building a marketing knowledge base, vector retrieval, and fusion-based generation.

Marketing Knowledge Base Construction: Using an automotive marketing corpus as the knowledge source, documents are segmented into semantically complete paragraphs, and each paragraph is converted into a vector representation via an embedding model.

Vector retrieval: After encoding the user query with the same embedding model, the k most relevant paragraphs are retrieved from the knowledge base via approximate nearest neighbor search. The mathematical expression of the retrieval process is shown in Formula (6).

$$D_{re} = TopK_{d \in K} \frac{E(q) \cdot E(d)}{\|E(q)\| \cdot \|E(d)\|} \tag{6}$$

where q is the user query, d is a document paragraph from the knowledge base, E() is the embedding function, and the similarity is measured by cosine similarity.

Fusion generation: Concatenate the retrieved marketing knowledge paragraphs with the user query to form an enhanced prompt, then input it into the fine-tuned vertical model to generate a response.

4.3. Marketing Knowledge Graph Integration

To enhance the model's structured reasoning capability and improve marketing decision accuracy, this paper constructs an automotive marketing knowledge graph and establishes a collaborative mechanism between the graph and large models. The graph is built using a combination of top-down and bottom-up approaches, defining entities, relationships, and attribute ontologies based on the automotive marketing business system. Vertical models can leverage the knowledge graph through two pathways—enhanced retrieval and reasoning enhancement—to achieve deep integration with the graph.

Graph-enhanced retrieval: When user queries involve entity relationship searches, the system prioritizes retrieving answers from the knowledge graph and injects the query results as context into the model [9-10]. The mathematical expression for knowledge graph-based marketing recommendations is shown in Formula (7).

$$S(u,v) = sim(F(u), P(v)) = \frac{F(u) \cdot P(v)}{\|F(u)\| \cdot \|P(v)\|} \tag{7}$$

Here, u represents user requirements, v denotes the target vehicle model, F(u) is the feature vector of user requirements, and P(v) is the feature vector of vehicle parameters. Precise matching between users and vehicles is achieved through similarity calculation.

Model Inference Enhancement: This paper leverages triplets from the automotive marketing knowledge graph as auxiliary training signals, incorporating contrastive learning into the optimization objective to strengthen large models' ability to encode structured marketing knowledge. By training on triplets such as (vehicle model, competitive benchmarking, competing model), the model learns to deeply understand competitive relationships among vehicle models. Meanwhile, using triplets like (user tag, recommendation, vehicle model) enables the model to grasp the matching logic between user profiles and products, thereby improving reasoning and decision-making capabilities in marketing scenarios.

5. MULTI-AGENT COLLABORATIVE ARCHITECTURE

Automotive marketing operations encompass multiple tasks such as content generation, lead management, user interaction, and data analysis, making it difficult for a single large model to independently adapt to all business scenarios. To address this, this study draws on the design concept of the Deep+Agent multi-agent platform to build a collaborative multi-agent architecture tailored for automotive marketing, enabling end-to-end business support through functional decomposition.

This paper designs four functionally independent agents. The first is the Content Creativity Agent, which focuses on generating marketing content using SFT+RAG technology to produce various promotional materials such as short video scripts,text-image copy, and live-streaming scripts. It can also access automotive marketing knowledge graphs to retrieve professional information like vehicle specifications and product highlights, ensuring the professionalism of its output. The second is the Lead Management Agent, responsible for managing leads throughout their entire lifecycle, including lead cleaning, grading, and long-term nurturing. By integrating intelligent outbound calling and semantic analysis of user conversations to identify purchase intent, it dynamically tags users and automatically generates targeted lead nurturing strategies. The third is the User Understanding Agent, which primarily handles intent parsing and user profiling by extracting key information—such as purchasing needs, budget ranges, and preferred configurations—from user inquiries, delivering structured outputs that support decision-making for other agents. The fourth is the Knowledge Management Agent, serving as the system's central orchestrator, responsible for task distribution and knowledge retrieval coordination. It matches relevant knowledge resources—including vehicle databases, script libraries, and marketing case studies—to different business scenarios and selects appropriate retrieval strategies accordingly.

6. CONCLUSION

This study centers on building a vertical AI foundation for digital marketing in the automotive industry by integrating open-source large models with domain-specific automotive corpora. A systematic investigation was conducted covering corpus construction, model training, performance validation, and engineering deployment, leading to three key conclusions: First, a specialized automotive marketing corpus defines the upper limit of vertical model performance; high-quality data spanning vehicle models, marketing materials, user profiles, and competitive intelligence forms the foundational basis for model capabilities. Second, the optimal framework combines SFT (Supervised Fine-Tuning), RAG (Retrieval-Augmented Generation), and knowledge graphs—when integrated, this tripartite approach significantly improves model accuracy on marketing tasks compared to the base model. Third, large-scale deployment must balance efficiency, iterative evolution, and data compliance, relying on model quantization, data flywheels, and safety filtering mechanisms to ensure successful implementation.

REFERENCES

[1] DeepZero. (n.d.). Deep Agent. https://www.deepzero.com/product/alphadata/deepagent/

[2] Sohu. (2025, March 7). KuKa Technology announces strategic transformation: "KuaiChao" integrates DeepSeek for upgraded automotive marketing intelligence. https://www.sohu.com/a/868122895_120389468

[3] Ma, W., Li, Y., & Gao, J. (2026). Research on robot target recognition based on multimodal visual self-attention mechanism. Computer Measurement & Control, 34(7), 261-267. https://doi.org/10.16526/j.cnki.11-4762/tp.2026.07.032

[4] Jia, S. (2026). Chinese character relationship extraction based on BiLSTM-CNN and attention mechanism. Research and Application of Digital Intelligence Technology. Advance online publication. https://doi.org/10.26917/j.cnki.issn.2097-597X.2026.02.001

[5] Guo, X., Li, B., Wu, Y., Liu, J., & Xu, M. (2026). A visual generation system for ammunition transfer and supply schemes based on semantic graph retrieval enhancement. Journal of Computer-Aided Design & Graphics. Advance online publication. https://link.cnki.net/urlid/11.2925.tp.20260806.1649.002

[6] Chen, L., Lei, K., & Liu, C. (2026). Research on question answering method in power grid domain based on graph retrieval-augmented generation. Computer Applications and Software. Advance online publication. https://link.cnki.net/urlid/31.1260.TP.20260806.1406.002

[7] Guan, S., Xi, J., & Song, H. (2026). A bidirectional LSTM sentiment analysis method for reviews integrating BERT and attention mechanism. Software Engineering, 29(8), 19-22+28. https://doi.org/10.19644/j.cnki.issn2096-1472.2026.008.004

[8] Zhu, W., Su, B., & Li, H. (2026). A multi-classification method for civil aviation news based on BERT and BiGRU. Computer and Modernization, (7), 91-96.

[9] Dou, J., Wang, X., Qiao, J., et al. (2026). An intelligent decision support framework for distribution network maintenance based on knowledge graph retrieval-augmented generation and large language models. Engineering Science and Technology, 58(4), 67-79.

[10] Liu, C. (2026). Research on key technologies for retrieval-augmented generation of geographical knowledge graphs based on multimodal data [Master's thesis, Chongqing Jiaotong University]. https://doi.org/10.27671/d.cnki.gcjtc.2026.000508