AI Empowers the Construction of An Internationalization Path for Foreign Language Services to Assist the Digital Economy: The Construction of A Foreign Language Service Community Integrating Government, Industry, Academia, Research, and Application

Yingying Liu, Yanfeng Sun, Kai Fu*, Hao Yan

Harbin Cambridge College, Harbin, 150000, China
**Corresponding author: m13624515934@163.com*

Abstract: *With the spread of the digital economy around the world, many problems have arisen in the transmission and distribution of information. The old model of foreign-language services has many agents and lacks good technology, so it cannot meet the high demand for concurrency and diversity. According to the concept of a foreign language service community, the following three parts will be built in the AI-empowered architecture: First, a multi-agent ecological cooperation mechanism (niche nesting, dynamic coupling, distributed trust) will be established; Second, a digital company matrix and infrastructure integration architecture (capability pedigree classification, heterogeneous integration of the semantic computing layer and the transmission network, lightweight gateway and service bus) will be designed; and finally, a talent resource support system (skill stack iteration algorithm, generative AI training environment, cross-border knowledge distillation and transfer) will be created. The three parts of the framework are ecology, infrastructure and enterprises; together they can provide good technical support for foreign-language services and promote the internationalization of the digital economy.*

Keywords: *Foreign Language Service Community, AI Empowerment, Ecological Cooperation Architecture, Digital Company Matrix, Knowledge Distillation, Internationalization Adaptation*

Introduction

With the development of the digital economy and globalisation, production and dissemination of multilingual content have become necessary links in cross-border information flow. There are many structural problems in the current foreign-language service system, such as inefficient multi-agent collaboration, a mismatch between the technical infrastructure and business needs, outdated talent, etc.; thus, the problem of language barriers has not been effectively solved. To promote the spread of language in the internationalised digital economy, a service community with many agents, good digital infrastructure and dynamic talents need to be built. Most of the current research is focused on single-point technologies, such as neural machine translation and large language models, and there is still no general system that integrates the four dimensions of ecological cooperation, enterprise matrix, infrastructure and talent support. The Logic of Organisation in this paper is the integration of government, industry, academia, research and application, and it proposes an AI-empowered way to build a foreign-language service community, presenting innovations in the areas of ecological coordination mechanisms, integration architectures and talent adaptation algorithms.

1. Ecological Cooperation Architecture and Coordination Mechanism of the Foreign Language Service Community

1.1 Nested Structure of Multi-agent Niches and Value Symbiosis Model

All the agents in the AI-assisted foreign-language service community have different strengths and functions based on their resources and output. First, a Language Resource Provider (such as a multilingual corpus organisation) will be set up, and then this organisation will collect and annotate the initial language data. Providers of Neural Machine Translation Models are the technology companies that perform semantic encoding and decoding. Domain experts and localisation reviewers will be in the verification pool, and then fine-tuning of the generated translations will be carried out. These niches are not in a straight line but are nested; that is to say, the technology niche is located between the foundation niche and the verification niche, and it has established layered invocation and feedback loops via API interfaces. The stability of the nested structure is due to the energy flow rate among the niches; that is to say, it is the transmission efficiency and distortion-control ability of semantic feature vectors between layers[1].

The Model of Value Symbiosis is based on the mutual exchange of resources. Since the goals of the individual agents are different, we will need to use more resources (e.g., computational redundancy, annotated data, terminological assets) and obtain scarce resources from other agents to make up for the deficiencies in our own functions. The function that determines the weight of a niche is a non-linear function of the three factors of node centrality, reduction in information entropy and task completion time. A recursive value feedback loop is used to increase the gain of each agent according to the depth of its niche, reduce free-riding behaviour, and thus achieve a self-organised symbiotic stable state.

1.2 Dynamic Coupling and Resource Allocation Protocol at the Organizational Boundary

In short, the first reason for inter-organizational boundary coupling is to separate the internal operation of each organisation from its external service interface, and by means of loose coupling, achieve function cooperation. According to the concept of events, each agent will publish its own capability description file (a semantic service description in JSON-LD format) and subscribe to the state-change events of other agents in the Dynamic Coupling Mechanism. When the load of a certain agent exceeds the threshold or there is a node failure, the coupling regulator will start runtime organisation and rebinding of the service invocation chain based on the current network topology and task priorities. The two are loosely coupled, so there is little interdependence among the organisations, and only one agent needs to be upgraded or replaced without affecting the whole service.

The three kinds of scarce resources in the allocation protocol are corpora, computational nodes and translation memory banks. Protocol is a distributed negotiation algorithm that is based on double auctions; that is to say, the resource requester submits a bid with quality constraints (BLEU threshold, latency ceiling), and the resource provider broadcasts its available capacity and unit cost. To avoid gaming in the bidding stage, a sealed-bid method will be used and the second-price rule will be applied. Micro-batch processing is employed by the Allocation Execution Layer to gather a certain number of requests in a short time and then a simple consistent hashing algorithm is used for load balancing in the resource scheduling unit. A protocol can distribute the resources reasonably well and does not need a central coordination unit; therefore, it is more suitable for handling sporadic and various demands for foreign-language services in the digital economy.

1.3 Algorithmic Path for Distributed Trust Transfer and Collaborative Governance

Distributed trust transfer does not have a single authentication centre, but rather a consensus algorithm is used to build a foundation of trust among the agents. Verifiable Credentials will be used to establish the foundation of the Trust Model, and the history of service quality, response time and semantic consistency for each agent will be stored in on-chain credentials and jointly signed by several independent verification nodes. A sliding-window weighted average method is used to update the trust value, and the weight of the most recent interaction result decays exponentially; thus, the trust model can be quickly adjusted in light of any change in behaviour. The Path of trust transfer is represented by a directed edge in the trust graph. When agent A wants to know how trustworthy an unknown agent C is, it can perform trust aggregation along the path A→B→C in the trust graph. To avoid magnifying the trust, both a geometric mean and a confidence decay factor are used in the aggregation function[2].

According to the governance rules, the algorithm for joint governance will be employed to write the smart contract code. The governance parameters of the weight for resource allocation, the threshold and conditions for dispute resolution, etc., have been recorded on the blockchain. If there is a problem (e.g., the translation quality of several consecutive periods is lower than the set standard), the governance contract will automatically start an arbitration process, randomly choose some verification nodes, have them conduct a double-blind check on the disputed samples, and then, based on the reputation reward-punishment mechanism of multi-armed bandits, reduce the frequency of calls or temporarily suspend the infringing party. Algorithmic governance does not have the subjectivism of human judgment, and its decisions can be traced and audited; therefore, it is a closed-loop adaptive governance system.

2. Integration Architecture of the Foreign Language Service Digital Company Matrix and Infrastructure

2.1 Spectrum of Digital Capability and Matrix-Based Classification of Foreign Language Service Enterprises

The two kinds of dimensions used to measure the digital ability of foreign-language service enterprises are the scope of automation and the degree of specialisation. Steady progress has been made in the automation of processing, and many achievements have been made; Computer-Assisted Translation with Translation Memory (CAT+TM), full autonomy in post-editing Neural Machine Translation (NMT), and further development of context-aware generative translation based on large language models are representative cases. The three levels of granularity for domain-specific semantic adaptation are the general corpus-driven type, the vertical terminology base-constrained type, and the real-time knowledge graph-injected type. The two dimensions are combined to form a digital capability matrix, and the four ideal types obtained are: the standard batch processing type (low depth and low granularity), the vertical precision matching type (low depth and high granularity), the adaptive pipeline type (high depth and low granularity), and the full-stack cognitive type (high depth and high granularity). The enterprises are of different types and have various technology stacks and business models, so a matrix-based classification method is used to label the type for service registration and routing in the integration architecture shown below.

The framework of classification is also subject to change; that is, a company's range of digital capabilities is not fixed but will change in line with the development of its own business and changes in external technology. The Path of Migration can be shown in the capability transition probability matrix, and the migration cost is determined by the combined effect of the scale of the current data asset, the redundancy of personnel skills, and the lock-in effect of infrastructure. The first function of the entire stack for the cognitive enterprise should be to handle multimodal input parsing, cross-lingual anaphora resolution and emotional style transfer, and all of them need to work closely with the semantic computation layer in the infrastructure. Therefore, the matrix-based classification can link the features of individual enterprises with the Design of the integration architecture, and it will provide a reference for interface specifications and resource reservation strategies for all kinds of enterprises.

2.2 Heterogeneous Integration Design of the Semantic Computing Layer and Transmission Infrastructure

A semantic computing layer has been added above the old transmission channel to organise the extracted structure of meaning from the raw language signal. The three submodules at this level are: a multi-granularity word segmentation and dependency parsing module; a cross-lingual entity alignment and relation extraction module; and a context encoding module (which is a kind of Transformer-based modification of a bidirectional encoder). The microservices of the submodules are deployed at the edge node, and each module outputs a high-dimensional semantic tensor. The first transmission infrastructure (that is to say, the fibre-optic backbone network, 5G access network and data centre interconnection network) was only employed to carry the physical and network-layer transmission of data packets. The Design of Heterogeneous Integration should have a semantic-aware scheduling unit in the transmission protocol stack. Read the priority tag of the semantic tensor in this module (e.g., real-time dialogue interaction is relatively urgent compared to batch document translation), and according to this, switch to a weighted fair queue scheduling algorithm and a semantic urgency function.

Heterogeneity in the integration architecture can be seen in the coordination of the two types of resources; that is, the computation-intensive semantic parsing task and the bandwidth-sensitive raw data transmission task use the same physical link and do not interfere with each other. The Design will be to employ bidirectional computation offloading and transmission prefetching. If the computational load of the edge node is too heavy, the semantic parsing task can be offloaded to a cloud inference cluster and the next batch of unprocessed language fragments can be preloaded via the transmission link; at the same time, if the network is congested, the high-priority semantic task can be handled locally and only compressed semantic feature vectors are transmitted. Solve the online convex optimization problem in a time-varying environment to minimize the weighted sum of end-to-end language service latency and semantic distortion rate, form an adaptive scheduling strategy for deep integration of semantics and transmission[3].

2.3 Deployment Logic of the Lightweight Semantic Gateway and the Cross-Language Service Bus

The lightweight semantic gateway is at the network edge, and it connects the end-users to the service bus. The first three are to automatically recognise the source language, find breakpoints and stream segments in the source-language sentence, and check the health of all backend service nodes. The three reasons for its small size are that it is stateless (session information is stored by the upstream client or the downstream service bus), it is a small model (only the language identification and word segmentation models are deployed, not the full NMT or LLM), and it has a binary protocol parsing (avoiding the extra overhead of XML or JSON). A pipeline is set up at the gateway for parallel processing; that is to say, each incoming language package will be sequentially passed through the language identifier, breakpoint detector and routing decision module. The Routing Decision is based on a cost function that includes the current load of service nodes, the semantic cache hit rate and the coverage of the target language.

The cross-language service bus is the communication foundation for all foreign-language service digital entities in the multi-company matrix. A Bus is used to store the latest service information, and each item in the Bus consists of the service ID, the capability type label (which corresponds to the matrix classification in Section 2.1), an input-output language pair, the current endpoint address, and SLA parameters. When the lightweight semantic gateway sends a request, the bus will carry out service discovery and composition orchestration; if a single service can handle the request (e.g., English-to-Chinese document translation), it will be directly routed by the bus; if multiple services need to be linked (e.g., speech recognition → machine translation → speech synthesis), then, according to topological sorting, the bus will construct a directed acyclic graph execution plan. The bus has an idempotency check and a timeout-retry mechanism; if it cannot get a response, the failed service node will be automatically removed from the registry and replica creation will be started. Deployment Logic will be used to realise the elastic connection and functional coordination of the matrix of foreign-language service digital companies without a central coordinator.

3. Talent Resource Support System and Path to Internationalisation

3.1 Dynamic Iteration and Adaptive Update Algorithm for the Skill Stack of Foreign Language Service Talents

Skill sets of foreign-language service professionals at different levels for various foreign languages include: At the foundation is linguistic knowledge, such as syntactic, semantic and pragmatic rules; in the middle are skills in using digital tools, such as CAT, NMT, LLM invocation and debugging; and at the top are cross-cultural pragmatic strategies, such as context adaptation, style transfer and implicit information explanation. With changes in external technology and demand from the market, there have been fluctuations in the required skills for various positions, and the fixed stock of abilities of current employees may no longer meet the new demands. Therefore, an adaptive update algorithm for Bayesian Knowledge Tracing is put forward. The level of talent in a person is a multidimensional latent variable, and the observed variables are the completion quality scores and response latencies of all kinds of tasks. Forward-backward recursion is employed to calculate the mastery probabilities of all capability units. If the mastery probability of a particular capability unit is lower than the adjusted threshold (which is based on the distribution quantile of the peer talent group), then this unit will be added to the update list automatically, and the corresponding micro-learning module and verification test will be distributed.

The newness of the adaptive update algorithm is that it has added a transfer reinforcement learning mechanism and uses the skill evolution trajectories of other talents in similar positions in the foreign-language service community as the source domain knowledge. An algorithm is employed to obtain the patterns of skill co-occurrence and failure transition probabilities in the source domain, and then, with the help of maximum mean discrepancy regularization, they are transferred to the capability update model of the target talent to increase the speed of convergence and reduce the need for direct assessment. The new skill stack will be written to the distributed trust store as Verifiable Credentials and can be read in real time by the resource allocation protocol of the ecological cooperation architecture. Based on the semantics of the technical system at the level of semantic computing, periodically update the skills of talented people according to an algorithm, build a closed-loop feedback mechanism for talent resources and digital infrastructure, and prevent the decline of abilities[4].

3.2 Generative AI-driven Immersive Corpus Training Environment

Therefore, Generative AI will be employed to create many kinds of various and multi-cultural training data in an immersive manner. The three necessary parts of the environment are a Scene Generator (Multimodal Context Rendering based on a Diffusion Model), a Dialogue Agent (a virtual character driven by a large language model that can be customised according to personality and pragmatic style), and a Real-time Feedback Module (to determine the learner's grammatical accuracy and appropriateness of pragmatics at the sentence level and provide instant correction). The training environment will no longer have a fixed set of pre-recorded corpora, but will dynamically generate dialogue branches and change contexts according to the current level and training goals of the learner. There are some constraints on the generation process, such as the target language, density of domain terminology, cultural specificity index and error tolerance threshold; the difficulty gradient of the generated content should be suitable for the current skill level of the trainee.

The training effect of this environment is an increase in pragmatic competence; that is to say, it is the reduction rate of the learner's pragmatic error rate in new situations after two tests. An adversarial data augmentation method is added to the environment, a discriminator network is used to distinguish between the learner's speech and that of native speakers, and a generator network is employed to iteratively improve the realism and difficulty of generated scenes. A Head-Mounted Display and spatial audio will be employed to create an immersive world, and based on the indirect feedback obtained from the learners' paths of eye movement, speech prosody and response delays, the difficulty and speed of interaction in the next scenario will be adjusted accordingly. The training environment is closed-loop connected to the skill stack update algorithm in Section 3.1; that is to say, the capability assessment results output by the environment are directly fed into the adaptive update algorithm, and the new skill targets generated by the latter will, in turn, modify the parameters of the scene generator.

3.3 Cross-border Collaborative Knowledge Distillation and Optimisation Path for Multilingual Service Networks

Cross-border collaborative knowledge distillation can be employed to address the problem of uneven distribution of resources in multilingual service networks. High-resource languages (such as English, Chinese and Spanish) have a large number of parallel corpora and high-quality translation models, but low-resource languages (some in Southeast Asia and Africa) do not have such data and are thus prone to model underfitting. Knowledge distillation is employed to train the student model for the low-resource language with soft labels, which are the smoothed versions of the vocabulary probability distribution output by the teacher model of the high-resource language. Cross-border cooperation can also be used to distribute the execution of the distillation process; that is, multiple teacher models in different regions (each run by a local foreign-language service digital company) can generate soft labels for the same batch of Interlingua input, and the student model can combine the knowledge from all these sources using a weighted-voting method, where the weights are determined by the domain confidence of each teacher model and the linguistic proximity distance.

Another kind of meta-learning that can help a student's model do well in low-resource languages is migration optimisation path. Randomly select a small number of samples from each high-resource language in the meta-learning stage to form a support set for many pairs of high-resource languages in a low-resource setting, train the model to adapt quickly to fine-tuning in that language, and update the initial parameters of the model with a meta-optimizer. When moving to a real low-resource language, only a very small amount of labelled data is needed to reach a practical level. Therefore, we can use the lightweight semantic gateway and the cross-language service bus; when the service bus receives a request for a low-resource language, it will first call the lightweight student model distilled via cross-border collaborative knowledge distillation, and if the confidence level is below the threshold, it will fall back to multi-hop translation (relayed through an intermediate high-resource language). The entire process of knowledge distillation and migration does not need a central corpus; all that needs to be exchanged among participants are soft labels, not raw data, so it meets the requirements of distributed trust and resource allocation in the ecological cooperation architecture.

Conclusion

The four levels of the AI-enhanced foreign-language service community architecture put forward in this paper are ecological cooperation, digital company matrix, infrastructure and talent resources. The three kinds of cooperation in a loosely coupled ecology are the multi-agent niche nesting and value symbiosis model, the cross-organizational dynamic coupling protocol and the distributed trust governance algorithm. Classification of the spectrum for digital capabilities, heterogeneous integration of the semantic computing layer and transmission infrastructure, lightweight gateways and service buses, etc., all aim to achieve the integration of the enterprise matrix and infrastructure. At the same time, the iteration algorithm for skills, the generative AI training environment and the cross-border collaborative knowledge distillation and migration strategy will all be developed to promote talent and digital system training. In the future, we will carry out cross-regional model collaborative training based on federated learning, build multimodal semantic and infrastructure integration paradigms, develop a dynamic meta-learning framework for low-resource languages, and advance the theoretical foundation for application.

Fund Projects

The 2025 Key Research Project of Heilongjiang Province's Economic and Social Development (Foreign Language Special Project), with the project title: Research on the Path of Foreign Language Services Assisting the Internationalization of the Digital Economy in Heilongjiang Province, and the project number: WY2025031.

The 2025 University Scientific Research Fund Project of Harbin Cambridge College, with the project title: Construction of the AI+RCE Regional English Document Compliance Training System, and the project number: 2025JQKY06.

References

[1] Li, N. (2025). Exploration of Talent Cultivation Models for Business Foreign Language Majors in Vocational Colleges under the Background of Digital Economy. Journal of Guangxi Open University, 36(02), 57-63.

[2] Fu, R. (2025). Exploration and Practice of Bilingual Teaching of International Trade Rules under the Background of Global Digital Economy. Journal of Changchun Institute of Technology (Social Sciences Edition), 26(02), 129-132.

[3] Wang, Y. (2025). A Study on the Cultivation Model of Business English Talents in 'Foreign Language + Cross-border E-commerce' in the Digital Economy Era. English Square, (03), 79-82.

[4] Ye, Q. (2020). A Study on the Cultivation Model of Applied Interpreting Talents Serving the Local Economy: A Case Study of Guangdong Province. Science and Education Guide (Mid-term), (08), 26-27.