Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 17, No. 2, 2026

Key Technologies and Applications of Multimodal Cognitive Agent Construction in the Vertical Domain of Inspection and Testing

Zhiruo Li [1], Wenlin You [1], Xinrong Zhang [1], Shiyong Zheng [1, 2, *]

[1] School of Business, Guilin University of Electronic Technology, Guilin, Guangxi, China
[2] College of Digital Economics, Nanning University, Nanning, Guangxi, China
*** Corresponding author:** Shiyong Zheng (Email: shiyongzheng123@whu.edu.cn)

Abstract: Inspection and testing underpin product safety and regulatory compliance across industries such as manufacturing, healthcare and food and beverage. However, conventional engineering test processes that mainly depend on manual, offline actions, failed to adapt to next generation, automated, digital and environmental-sensitive testing, resulting in very low productivity and high cost for most firms. The automation and digitalization of inspection and testing processes have become a research hotspot in both academia and industry. The natural language processing and computer interpretation of test results have been a focus of AI research as well. Nonetheless, due to the lack of real data, the verification and simulation of real inspection and testing environments are still difficult for researchers. As a solution, we manage to develop a multimodal cognitive test system that fuses the textual regulatory documents and instrument data, through a series of modules and processing flows. Our test system is based on a so-called multimodal cognitive agent, which includes large language model, vision module, knowledge graph and retrieval-augmented generation. We introduce the design, development and application of our test system, which used for rubber heater in a glass factory, and layout the future challenges for the exploration of multimodal agent technology in the test engineering. The project is supported by the Guangxi Key Research and Development Program. This paper was partially presented at the IEEE 2023 International Conference on Intelligent Commerce (ICIC).

Keywords: Multimodal Agent; Large Language Model; Inspection and Testing; Cognitive Computing; Knowledge Graph.

1. Introduction

Whenever you walk into any accredited testing lab, you see, dramatically, something you don't want to see: sophisticated instrumentation surrounding a handedly performed workflow. Mass spectrometers detecting parts per billion sit next to technicians flipping through a hundred-twenty-page standards handbook, and workers typing a multi-million-yuan testing machine's result into a report template. Prime ministerial regulations arrive every quarter, each cross-referencing others. The divide between analytical capability and information management has never been wider. This motivates, and we believe will enable us to serve, the research we present on multimodal cognitive agents for inspection and testing.

The inspection and testing industry has a distinct secondary role in the quality infrastructure. The observer and authority that ensures that food is safe to eat, bridges are sufficiently robust, pharmaceuticals contain stated ingredients provides for a bottleneck. The increasing workloads coupled with ever more complex operations and tighter audit regimes aggravate existing challenges. The data landscape is a large and growing environment of register entries, regulatory documents, instrument spectra and images, calibration data, chain-of-custody documentation, and formatted tabular datasets; the data is diverse, it is stored in multiple locations and in multiple formats, each with its own meaningful structure. For a large part, the quality assurance industry relies on a combination of spreadsheets, LIMS databases, and field knowledge that dies with the staff.

Recent progress in artificial intelligence has fueled maniacal enthusiasm in this area. Large language models now understand regulatory text reasonably well[1]. Multimodal learning lets us connect the text of a standard to the image of what an instrument observes[2]. Multi-agent architectures allow multiple AI modules to collaborate on complex analytical challenges[3]. But enthusiasm is not deployment. The chasm between exciting proof-of-concept and a laboratory director trusting it with real samples is big, and to be fair, many of the hard challenges are not algorithmic.

We discuss what we learned developing multimodal cognitive agents supporting inspections, testing, and repair activities, for a project funded by the Guangxi Key Research and Development Program. The projects is a collaboration between researchers at Guilin University of Electronic Technology, scholars from the industry in China and Europe. The collaboration was fruitful. We do not want to claim that we solved everything. Instead, we are providing an honest perspective of the relevant technologies that matter, the architectural choices that mattered, and the applications that made a difference in laboratory settings. The partners from industry and academia have helped us shape what follows.

2. Where the Real Challenges Lurk

Before we get into how we might solve these problems, it's worth reminding ourselves as a group about what makes a domain like this especially difficult for AI. Some of these challenges are obvious to anyone who's actually worked in a lab; others are only revealed when you try to solve the problem from scratch.

One of the main issues is that the data in this domain is impossible to neatly separate into categories. One workflow could involve a PDF of a regulatory standard, a proprietary instrument format mass spectrum, some sample preparation notes scrawled in a paper form, and a LIMS schema to hold the numerical results even though it was designed twenty years ago. It's not just that they look like such different things; even a peak in a chromatogram doesn't 'mean' the same thing a clause in a standard 'means'. One of the hardest technical issues is to solve this semantic gap instead of taking all the data and knuckling it into text and hoping for the best [4][5].

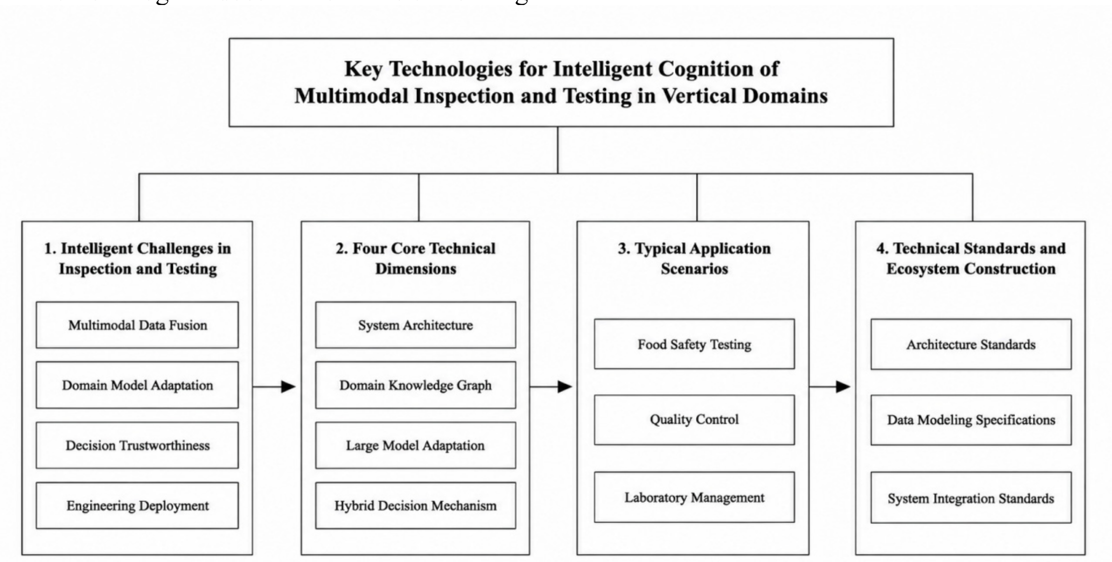

Figure 1. Technical Roadmap

Second, domain knowledge is more important than model size. Regardless of how big a general-purpose language model is, it does not know what a 'matrix spike recovery' is, or why a detection limit calculated by one method is different from a detection limit calculated by another method, or which version of GB 2760 applies to a food additive in a particular year. This knowledge is not rare. It is written down in the standards and method documents. However, they are too numerous, and have to be updated frequently[6]. It takes more than fine-tuning to teach a system this type of knowledge. It requires constructing a knowledge infrastructure that is as easy to maintain as the regulations are to update[7].

Third, the consequences of being wrong are not abstract. If an AI system tells a technician that a batch of infant formula passes aflatoxin screening when it does not, people get hurt. This might sound like an argument for keeping humans in the loop, and it is, but it also means that airborne interface between analytics and people must be explainable in terms people can understand. Simply saying 'the model is 94% confident' is not enough; they need to see which standard clause was checked, which spectral feature triggered the flag, and what the historical baseline looks like[8][9].

Fourth, and most practically, laboratories are not software companies. They cannot afford to hire a machine learning engineer in the summer, they run on tight budgets to begin with, and their IT infrastructure is conservative for good reasons. Auditors do not like experiments, and running tests in a lab on a lifetime of data is risky business. The only gantry you can afford is one that decouples and is load-balanced. Any AI system sitting in the middle of that environment must be deployable and usable on modest hardware, integrate with existing LIMS/instrument software, degrade gracefully when components fail [10][11][12].

3. Key Technologies for Constructing Domain-Specific Cognitive Agents

3.1. System Architecture: What the Four Layers Do

Four layers ultimately defined our architecture. It did not originally, we explored it over time as we encountered obstacles, but the final four have proven to be stable over our deployments

The bottom of the stack is what we call the Multimodal Model Access Layer. Its job is not exciting, but it is a crucial piece of infrastructure: it gives the rest of the system a simple, uniform interface to a small queue of foundation models (OpenAI, DeepSeek, Qwen, Doubao, Kimi, and simply named Ollama instances), so that control is not required as to which service does what query. It also maintains API keys, tally usage and cost, deals with rate limits, and supports swapping models without having to restart the system. The last of these proved more important than we anticipated: inspection tasks truly find different models more suited to their particular needs, and the ability to route a food safety compliance check to one model while routing ahead a mathematical uncertainty estimation to another proved an inconvenient but real requirement[3][14].

Above is the Data and Knowledge Layer. It ingests data from relational databases, document repositories, web crawlers, and instrument APIs. A non-trivial amount of engineering went into the document parsing pipeline alone, regulatory PDFs and Word documents come in bewildering variations of formatting, and extracting structured information from them required a combination of layout-aware parsing and fallback OCR. The layer also handles vector embedding generation, indexing, and chunking strategies that respect document structure rather than blindly splitting at token boundaries.

Above is the Intelligent Workflow Layer, which includes the platform's core wizard-based workflow authoring tools. The set of available nodes is modular and configurable, including natural language understanding, conditional branching, loops, knowledge and data retrieval, parameter extraction, and various template transformations. The entire set of nodes were designed to empower domain experts to define what is essentially a script of steps to be used by the system without writing code[29].

Here at the top is the Agent Development and Application Layer, which is where individual agents are composed and deployed. We support three different interaction paradigms: conversational (the user asks questions and they get answered), automated (an agent monitors a stream of data and reacts when it meets a certain condition), and autonomous (an agent plans and executes a multi-step workflow). Developers can compare the performance of different foundation models with our multi-model comparison tools, and we trace every decision so that you can audit it later.

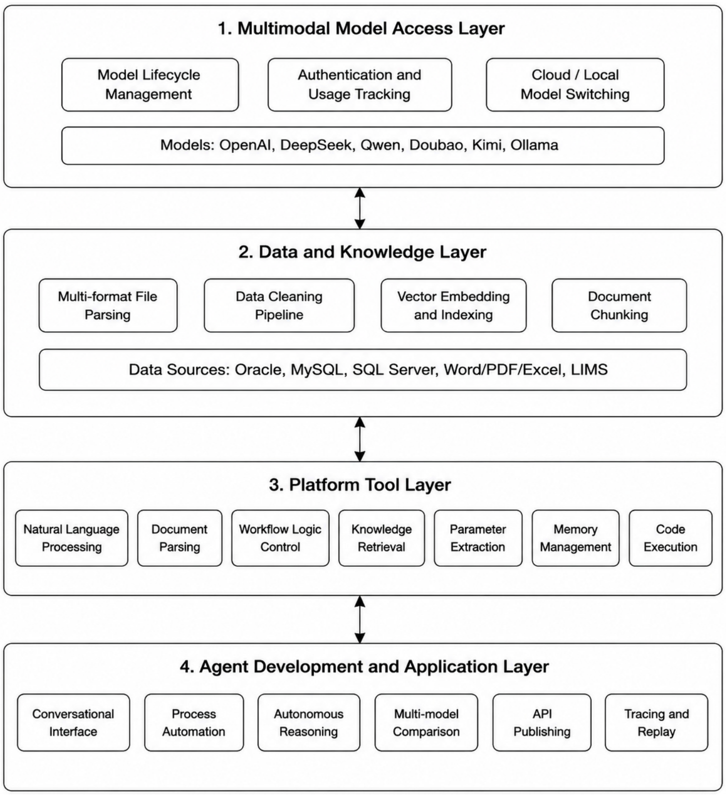

Figure 2. Architecture of the Multimodal Cognitive Agent System

3.2. Building a Knowledge Graph That Actually Helps

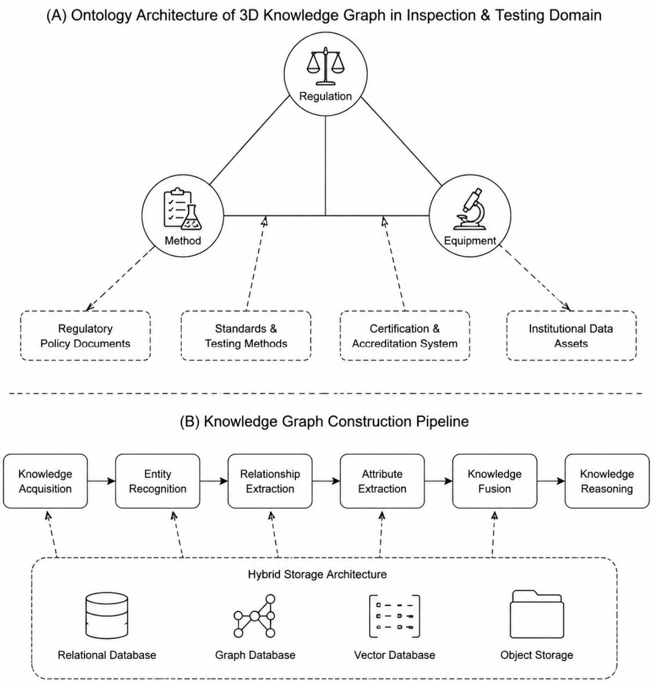

Figure 3. Knowledge Graph Ontology Architecture & Construction Pipeline Combination Diagram

Building a knowledge graph is a good achievement, but only as useful as the questions it can answer. Because regulations and methods are too numerous to categorize in a single dimension, we organize our knowledge graph along three axes: regulations, methods, and equipment. That is to say, any inspection activity can be represented in a three dimensional space in which each dimension captures an aspect of the activity: which regulation applies, what method is required, and what instrument is used. In our experience, the vast majority of relationships that arise in everyday lab work can be represented within these three dimensions [13] [14] [15].

The knowledge acquisition pipeline pulls from many sources. We use automated crawlers that alert us to new developments in various standards repositories - a daunting task when no single organization uses a standard format for publication and review[8]. APIs with LIMS systems allow us to pull in operational data from the equipment itself. We use OCR to digitize existing paper documents which are still common in many Chinese laboratories, followed by a post-correction pipeline using language models trained to the domain (extraction of data from metrology documents). As a starting point we partnered with Beijing Sunway World to access the YiBiaozhun platform, which already included over one million standard documents in more than sixty subdomains[8].

Our choice of storage is not driven by architectural desiderata, but rather by the strengths of individual database technologies. Structured metadata is stored in a relational database. Entity relationships are stored in a graph database due to its flexibility in multi-hop queries, which are ubiquitous in the extraction of relationship information from regulatory documents. Semantic search is supported by a vector database of text and image embeddings. Finally, we store raw documents and images in an object store, which is cheap and easy to use. As of now our knowledge graph includes more than one hundred thousand typed relationships and semantic cross-modal queries return in less than two hundred milliseconds, making it practical for interactive use[15][24].

3.3. Making Language Models Work in a Specialized Domain

General-purpose language models know a great deal about many things, but their knowledge how inspection and testing is shallow at best. We found that three adaptation strategies, used together, made a meaningful difference.

First, we fine-tune embedding models on domain corpora. Standard embedding models like BERT and CLIP, trained on general web text and images, do not capture the semantic similarity between, say, two standard methods that describe the same analytical procedure in different wording. Fine-tuning on a corpus of inspection standards, test reports, and technical manuals substantially improves retrieval quality for domain-specific queries[6][17]. The preprocessing pipeline for this corpus, cleaning, deduplication, terminology normalization, and augmentation through back-translation, was more work than the fine-tuning itself.

Second, we use retrieval-augmented generation as the default inference mode rather than an optional add-on. Whenever the user asks a question, we first retrieve relevant entities from the knowledge graph and relevant passages from the document store, and then the language model generates an answer conditioned on this evidence, with citations. This obviously does not prevent hallucination (nothing can), yet it certainly grounds hallucination in some factual basis and thus gives the user a way to check the sources of what the system claims[19][20][21][22][23].

Third, we employ parameter-efficient fine-tuning using Low-Rank Adaptation, which allows us to adapt large models to domain-specific tasks without full fine-tuning, in terms of training time and memory. Empirically, training only about 0.1 percent of the parameters of the model can lead to considerable gains in professional text understanding, and the tiny adapter weights can be replaced in the future as regulations evolve over time. Adapters can also be loaded for different inspection subdomains—for example, for laboratories that provide inspection services for a range of product categories[16][18].

3.4. Hybrid Decision-Making: Rules and Models

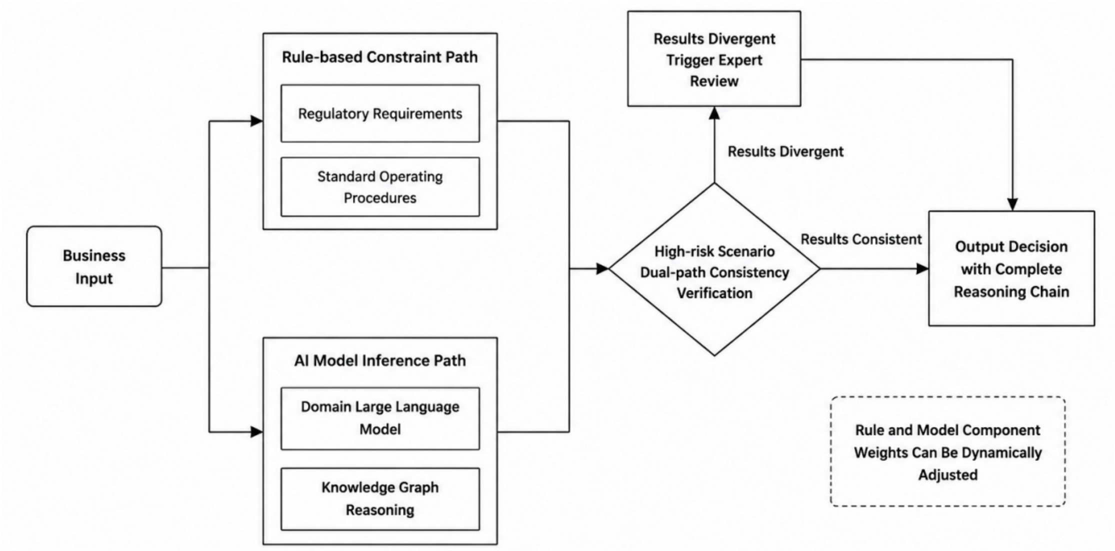

Figure 4. Schematic diagram of the dual-path hybrid decision-making mechanism

Based on the understanding of the decision-making processes in achieving compliance to the regulations and on the modeling approaches that can be used to mimic the decision-making process, we present a hybrid approach that uses different decision-making mechanisms that run in parallel and reconcile at the end. The hybrid architecture is designed in such a way that it can be used for any high-risk decision with and without human-in-the-loop. The verification of the interpretations and the model-based predictive inference can be performed by either automating the compliance monitoring in the presence of new data or by providing a reasoning chain that can be used for tracking and cross-checking compliance statements and predictions with the regulators. We showcase the hybrid decision architecture using the food safety regulatory compliance problem as an example. The architecture is designed such that if rule-based checks agree on the compliance status, and the model-based predictive inference also agrees with the rule-based approach, a recommendation for compliance status is made. In the cases of a disagreement between the two approaches, a high-risk case is flagged and sent for a new human decision-making instance. Each decision made using the hybrid decision architecture traces the reasoning and the data interpretation that led to the decision, providing the transparency needed for compliance with the regulation and the ability to perform root-cause analysis if the system eventually malfunctions. [9] [11] [28]

The weighting between rule-based and model-based components adjusts to the situation. For routine, well-defined checks where the regulatory guidance is clear and unambiguous, the rule-based pathway dominates. For ambiguous cases that require interpretation, the model-based pathway receives more influence. However, there is no point asking a neural network whether a food additive exceeds its maximum use level when the regulation states a precise numerical limit. In the case of a novel ingredient, the regulatory guidance is perhaps less clear, and its utilization might need to be interpreted: does it fall under an existing category, for instance?

4. What This Looks Like in Practice

4.1. Food Safety: The Hardest Test

We began with a deployment in food safety inspection for good reasons. The regulatory environment is especially dense, China's sheer number of documents on food safety standards is in the thousands, and the cost of error is publicly visible at once. Three specific capabilities were deployed.

The label review system is the component that sees the most use. By taking an image of a product label, recognizing and reading the text on the label, matching the claimed ingredients and nutrition information to whatever standard applies, and then automatically highlighting potential discrepancies, the system is able to reduce the time a label—one that previously required a human reviewer for a full twenty minutes—from nearly riright onto under three seconds and in the process, the system writes down the specific items for a human to verify. The standard label review uses throughput is approximately one thousand per hour on a modest hardware system[22][23]. The system does not replace the human but instead triages, allowing the human to work only on those labels where something appears to be wrong.

In contrast, the illegal additive screening module adopts a different approach. Rather than referencing fixed lists of illegal additives, it integrates mass spectrometry with kernel-based knowledge graph matching to flag compounds that are not allowed in the target product body. The knowledge graph captures known relationships among product bodies, allowed additives, and disallowed substances, so that the system can detect not only known illegal additives, but also potentially harmful analogues [2] [10].

The food safety suite also supports automated draft report generation from LIMS and instrument data. The engineer selects the product body, which is used to determine the compliance rules to apply. The system pulls the raw data from LIMS and instruments, performs the relevant statistical checks, and drafts a CNAS-compliant test report, including methods, raw data, uncertainty, and pass/fail status. The engineer then reviews the auto-draft report and signs it off – the system is not signatory at all – but it brings the drafting time down from about forty minutes, or so, to less than five.

4.2. Quality Control: More Than Pass/Fail

Quality control in manufacturing is a different beast from regulatory compliance. The scale is larger, the cycle time is shorter, there is a cost to a false positive, i.e. we would have stopped a production line unnecessarily, and so on.

Product compliance is a complex problem that involves matching test results against multiple industry and national standards. Standards compliance assessment is time consuming, with multiple interfaces to inventory test data, test instrument data, multiple versions of potentially hundreds of standards and continuous updates to test methods and standard requirements. Our standards module performs curation of more than two hundred food, pharmaceutical and consumer goods standards, includes semantic interpretation such that 'heavy metals' is understood to be a group of elements with individual limits; and knowledge about whether or not compliance to an older version of a standard satisfies an updated standard. The curation is used by a standards matcher to provide product compliance assessment. An anomaly detection tool is used to automatically search for outliers against data stored in a data warehouse. For an outlier, relevant data is automatically associated and presented to the analyst in his or her MIS workspace for investigation. The matching and anomaly detection allows for a new case to be created in less than one second, where a human analyst would require 20 minutes of searches against multiple systems. The anomaly detection uses isolation forest combined with one-class support vector machine (SVM) algorithms.

4.3. Managing the Laboratory Itself

There are many challenges in the laboratory that are less helped by statistical methods but highly important in their own right. Sample tracking, equipment scheduling, workload balancing, quality monitoring. These consume staff time and are highly error-prone when people are busy and juggling several things at once. We have just finished a new workflow management module that will track a sample from reception through to reporting, flag bottlenecks such as an instrument queue backing up or a test that is due for review, and automatically bring the issue to the attention of a manager in a timely manner. We are also working on a sophisticated genetic algorithm for scheduling people, instruments and reagents[5][11]. These are not glamorous uses of AI, but we are found to save them a lot of time more than any "analytical" feature they offer us.

5. Standards, Trust, and What Needs to Happen Next

Building the technology is necessary but not sufficient. For multimodal cognitive agents to be adopted at scale in inspection and testing, an industry that is, rightly, conservative about new technology, there needs to be a framework of standards that addresses interoperability, reliability, and security. Our project contributed to the development of three categories of standards, and the process of drafting them was instructive.

The first category covers cognitive agent architecture: what functional components a system must have, how interfaces should be defined, and what data flows are expected. These standards aim to ensure that agents from different vendors can interoperate and that laboratories are not locked into a single provider's ecosystem[26]. The second addresses multimodal data processing and knowledge modeling: data formats for standard documents, test results, and analytical spectra, as well as schemas for knowledge graphs and ontology definitions. Getting different stakeholders to agree on data formats is never easy, but it is the kind of unglamorous work that determines whether a technology ecosystem actually forms or remains a collection of isolated point solutions.

The third group, system integration and data exchange, is literally the most interesting. This group defines the interfaces and protocols that enable the cognitive agents to interface with the LIMS, instrument software and enterprise systems. And it is exactly this group which will fail the industry[29][30] if left to our imagination.

I should make a more general point here. From what the industry needs fast, a digital transformation that is enabled kAI will not be driven by AI capability alone. It will require an ecosystem, one in which vendors, standards bodies, accreditation bodies and regulators all have a parallel working group. We contributed to a number of these standards, and there is a lot left to do, for instance on the validation and certification of AI-assisted inspection decisions, which poses some new questions that the current systems of conformity assessment were not designed to handle [12][18].

6. Conclusion

In this paper we have described what we learned while building multimodal cognitive agents for the domain of inspection and testing. This work sits at the intersection of several active research areas such as multimodal learning, knowledge graphs, retrieval-augmented generation and hybrid AI systems, but the primary contribution is an integration story. We found that making these technologies work together in a real laboratory setting exposed a set of engineering problems that would not be encountered in benchmark evaluations: how to keep a knowledge graph up-to-date as regulations change, how to make laboratory staff feel confident in AI-generated recommendations, and how to run complex AI systems on the infrastructure that laboratories actually use.

Several directions deserve more attention than they currently receive. Causal reasoning, the ability to trace an anomaly not just to a data point but to a root cause, would move these systems from detection to diagnosis. Better few-shot and zero-shot learning would reduce the burden of creating labeled training data for every new standard or product category. And perhaps most importantly, we need better ways of measuring whether these systems actually make laboratories more effective, not just whether they score well on technical benchmarks, but whether they reduce errors, save time, and help people make better decisions.

The inspection and testing industry is not going to be transformed by AI overnight, and anyone who promises otherwise is selling something. But steady, well-engineered integration of cognitive agents into laboratory workflows can make a real difference, reducing the time spent searching for information, catching errors before they reach a report, and freeing skilled analysts to focus on the cases that genuinely require human judgment. That, rather than any particular algorithmic breakthrough, is what we consider the real contribution of this work.

Acknowledgments

Guangxi Key Research and Development Program Project: Research and Application of Key Technologies for Constructing Multimodal Cognitive Agents in Vertical Inspection and Testing Fields (Grant No. Guike FN2504240021)

National Natural Science Foundation of China: Research on the Behavioral Mechanism of Doctors' Knowledge Sharing in Online Medical Communities—From the Perspective of Network Effects (Project Number: 72462011)

2023 Key Special Project of the 14th Five-Year Plan for Educational Science Planning in Guangxi: Research on Innovation and Entrepreneurship Education in Local Application-Oriented Universities Serving the Development of Local Regional Economy (No.: 2023ZJY1456).

Graduate Education Innovation Program Project, Guilin University of Electronic Technology (No. YCSW2024337 & 2024YCXS091).

2026 Graduate Innovation Project, Guilin University of Electronic Technology: Research on the Impact of Supply Chain Digitalization on Corporate Greenwashing.

References

[1] Sun, H., Wang, T., Yuan, L., et al. (2026). From structure to synergy: A survey of vision-language perception paradigm evolution in multimodal large language models. *Information Fusion*, 133, 104285. https://doi.org/ 10.1016/j. inffus. 2026. 104285.

[2] López, V. S., Olivas, S. E., & Montañés, S. M. (2026). RGB ensemble strategies for unsupervised industrial anomaly detection on the AutoVI dataset. *Electronics*, 15(10), 2077. https://doi.org/10.3390/electronics15102077.

[3] Qu, J., Wang, Y., Fu, Y., et al. (2026). A multi-agent AI framework for explainable battery system maintenance. *Cell Reports Physical Science*, 7(6), 103388. https://doi. org/10. 1016/ j. xcrp.2026.103388.

[4] Geetika, D., Vanita, A., & Ravibabu, M. (2023). Artificial neural network based sub-surface defect detection in glass fiber reinforced polymers: Nondestructive evaluation 4.0. *Sensing and Imaging*, 24(1). https://doi.org/10.1007/s12217-023-00987-9.

[5] Liu, S., Sun, Y., Zhang, L., et al. (2021). Fault diagnosis of shipboard medium-voltage DC power system based on machine learning. *International Journal of Electrical Power and Energy Systems*, 124. https://doi.org/10.1016/ j.ijepes. 2020. 106345.

[6] Yao, L., Fang, Z., Xiao, Y., et al. (2021). An intelligent fault diagnosis method for lithium battery systems based on grid search support vector machine. *Energy*, 214, 118866. https:// doi. org/10.1016/j.energy.2020.118866.

[7] Meng, A., Zhang, W., & Qi, M. (2026). A semi-supervised multi-stage pipeline for low-resource industrial condition monitoring by injecting domain knowledge via cross-modal multimodal large language model fine-tuning. *Engineering Applications of Artificial Intelligence*, 181, 115250. https:// doi. org/ 10.1016/j.engappai.2026.115250.

[8] Qi, W., Yongsheng, H., & Feng, C. (2021). Deepening the IDA* algorithm for knowledge graph reasoning through neural network architecture. *Neurocomputing*, 429, 101–109. https:// doi. org/10.1016/j.neucom.2021.01.024.

[9] Baptista, L. M., Henriques, M. E., & Goebel, K. (2021). More effective prognostics with elbow point detection and deep learning. *Mechanical Systems and Signal Processing*, 146. https:// doi.org/10.1016/j.ymssp.2020.106989.

[10] Mao, J., Liu, J., & Shi, T. (2026). Task-guided dynamic attention mechanism for industrial anomaly detection with large vision-language models. *Concurrency and Computation: Practice and Experience*, 38(10), e70774. https://doi.org/ 10. 1002/ cpe.70774.

[11] Zhang, X., Zhang, L., Wang, P., et al. (2026). KPIs anomaly detection through dual-branch spatio-temporal-frequency feature extraction and enhanced transformer. *International Journal of Machine Learning and Cybernetics*, 17(7), 341. https: // doi.org/10.1007/s13048-026-02189-6.

[12] Haha, B. A. M., Bohli, A., Haddour, N., et al. (2026). A systematic review of industrial IoT anomaly detection and the forensic interpretability gap. *Electronics*, 15(11), 2240. https:// doi. org/10.3390/electronics15112240.

[13] Zhang, C., Zhang, Q. M., Luo, Y., et al. (2026). Hard constraints and soft learning dual-graph anomaly detection for industrial processes. *Engineering Applications of Artificial Intelligence*, 177, 114927. https://doi.org/10. 1016/j. engappai. 2026. 114927.

[14] Xu, R. H., Zhang, N., Yin, Y. Z., et al. (2025). Multimodal framework integrating multiple large language model agents for intelligent geotechnical design. *Automation in Construction*, 176, 106257. https://doi.org/10.1016/j.autcon.2025.106257.

[15] Guo, Q., Jiang, L., Dong, K., et al. (2026). KG-Anchored RAG: Retrieval-augmented generation for power system professional documents integrating topic modeling and knowledge graphs. *Electronics*, 15(11), 2362. https://doi.org/ 10.3390/ electronics 15112362.

[16] Songping, L., Feifei, L., Yusen, Y., et al. (2020). Nondestructive evaluation 4.0: Ultrasonic intelligent nondestructive testing and evaluation for composites. *Research in Nondestructive Evaluation*, 31(5–6), 370–388. https://doi. org/ 10.1080/0934984X.2020.1784964.

[17] Alhazmi, H. M. (2026). An explainable AI framework (XAI-CoffeeNet) with deep learning and vision transformers for Saudi coffee disease recognition and severity assessment. *Smart Agricultural Technology*, 14, 102259. https://doi.org/ 10. 1016/j.smartat.2026.102259.

[18] Brunaccioni, L., Fusco, D. F., & Vallerio, M. (2026). Dynamic SHAP: Time resolved explainable AI for batch processes. *Computers and Chemical Engineering*, 213, 109717. https:// doi. org/10.1016/j.compchemeng.2026.109717.

[19] Xiao, Y., Yin, K., & Mostafavi, A. (2026). CrisiSense-RAG: Crisis sensing multimodal retrieval-augmented generation for rapid disaster impact assessment. *Computer-Aided Civil and Infrastructure Engineering*, 48, 100096. https:// doi. org/ 10. 1111/ mice.13096.

[20] Chen, X., Li, Y., Bi, Y., et al. (2026). ReflectRAG: Enhancing retrieval-augmented generation with GRPO-optimized iterative reflection. *Neurocomputing*, 695, 134047. https://doi. org/ 10. 1016/j.neucom.2026.134047.

[21] Karabowicz, P., Charkiewicz, R., Charkiewicz, A., et al. (2026). Target-aware molecule SMILES generation using a large language model with retrieval-augmented generation, multi-turn memory, and a predictive model. *Journal of Computer-Aided Molecular Design*, 40(1), 134. https://doi.org/ 10.1007/ s10822-025-00489-9.

[22] Kasireddy, E., Chow, C., & Pourrahmat, M. M. (2025). Evaluating the application of GPT-4o and retrieval-augmented generation (RAG) for assessing risk of bias and study quality in systematic literature reviews (SLRs). *Value in Health*, 28(6S1), S278. https://doi.org/10.1016/j.jval.2025.04.018.

[23] Raghavan, P., Vaidhyanathan, K., & Blumen, R. (2025). Abhinav Kimothi on retrieval-augmented generation. *IEEE Software*, 42(6), 116–119. https://doi.org/10. 1109/MS. 2025. 3412476.

[24] Hai, W., Jinrui, L., & Jianfeng, D. (2020). Iterative visual relationship detection via commonsense knowledge graph. *Big Data Research*, 23, 100175. https://doi.org/10. 1016/j.bdr. 2020. 100175.

[25] Wang, Z. W., Li, L. B., & Wei, L. S. (2026). An integrated multimodal sensing and data fusion system for rapid, community-scale inspection of self-built housing safety hazards. *Advanced Engineering Informatics*, 76, 104864. https://doi.org/10.1016/j.aei.2026.104864.

[26] Ding, M., & Dong, S. (2026). Digentity (Digital-Entity) human dyad: A perspective on human-AI collaboration and decision-making. *Australasian Marketing Journal*, 34(2), 120–125. https: //doi.org/10.1016/j.ausmj.2026.03.002.

[27] Pan, X., Wang, D., & Tsung, F. (2026). Empowering intelligent quality control with large models: A comprehensive survey of methods, challenges, and perspectives. *Frontiers of Engineering Management*, 13(1), 42–64. https://doi. org/ 10. 1016/ j.fem.2026.01.003.

[28] Dong, L., Chen, Y., Ke, W., et al. (2026). Multimodal fake news detection via evidence retrieval and visual forensics with large vision-language models. *Information*, 17(4), 317. https:// doi. org/10.3390/info17040317.

[29] Rashidin, S. M., Taheri, B., & Javed, S. (2026). Frugal AI as a catalyst for smart hospitality: Rethinking digital transformation in resource-constrained settings. *International Journal of Hospitality Management*, 139, 104785. https://doi.org/ 10. 1016/ j.ijhm.2026.104785.

[30] Fernández, D. M., Prieto, R. C. J., & Cabrales, L. Á. (2026). Digital transformation and social sustainability: The mediating role of employees' digital competencies. *Technovation*, 156, 103607. https://doi.org/10.1016/j.technovation.2026.103607.