ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 3 - 2026

Inference-Driven Intelligent Manufacturing Decision Support Combining Retrieval-Augmented Generation and Industrial Knowledge Graphs

Hailong Yang[1,*], *Yonggang Zhang*[2], *Yu Bai*[1], *Renzhi Gao*[1], *Yi Qi*[1] *and Baorui Du*[1]

[1] Institute of Engineering Thermophysics, Chinese Academy of Sciences, Beijing, 100190, China
[2] Xi'an XAE Flying Aviation Manufacturing Technology Co., Ltd.

SUMMARY: *Considering that the challenges of manufacturing demands are becoming increasingly more complicated, particularly in aviation machining, decision-making regarding the choice of the most appropriate process parameters remains a daunting task. This work primarily aims at exploring how Knowledge Graphs and Retrieval-Augmented Generation can be used as a suggested architecture of intelligent and explainable decision-making systems. The literature analysis was through content analysis of recent refereed journal articles as a technique of secondary research methodology applied in the study to determine the existence of technological complementarities as well as identify gaps in research. The findings confirm the fact that the KGs enhance the context and audit trails whereas RAG provides the dynamic and context-based searches of information that have a positive influence on the machining decisions. This integration will surely be of great help since it would reduce the trial and error, improve accuracy, and sustainability of critical applications. This work can make a contribution through the combination of semantic technologies and generative AI in making decisions in the manufacture industry of aircraft. This piece of work is applicable to the future development of smart and dynamic supporting systems in line with the technological requirements of Industry 4.0.*

KEYWORDS: *Inference-Driven; Intelligent Manufacturing; Decision Support; Retrieval-Augmented Generation (RAG); Industrial Knowledge Graphs (KGs); Aviation Machining*

1 Introduction

The intelligent decision-making systems have become even more central in the context of the aviation engines manufacturing industry, specifically, the tolerances are extremely small and the machining parameters are so complicated that intelligent systems may become valuable tools [1]. Even the tiniest elements of the engine are under the control of an aviation machinist and it has to tweak a multiplicity of tied up factors, such as cutting speed, feed rate, tool geometrical, and material properties, all of which must be fined-tuned to the point of performance and safety serviceability standards [2]. The older method that placed much emphasis on expert intuition and manual control are no longer sufficient in the management of the amount and complexity of process variables. The difficulty has driven advanced computational schemes, specifically knowledge graphs and generative AI systems, to the realm of aerospace machining [3].

*yanghailong22@iet.cn
https://doi.org/10.65102/is20261076
Published: 10 June, 2026

One of the most important in this direction suggested an integrated resource assigning tool aimed at aerospace workplaces with the addition of machining data into a real-time graph neural network [4]. This method integrated semantic manufacturing knowledge into a graph, which was structured and facilitated optimal use of equipment and saved a lot of processing time. The analysis showed that knowledge graphs have the ability to learn and reason on multidimensional manufacturing data and provide a basis to intelligent and data-driven decision support systems in the aviation manufacturing settings [5].

Based on this, the domain knowledge graph application has been on the increase in order to enhance the conception of intelligent design and automated process system. As an example, [6] created a knowledge graph based on a data system as a way of understanding the complexity of mould designing by extracting insights out of huge engineering records. It allowed automatic selection of the process parameters in the process of injection moulding and it was no longer necessary to rely on human expertise, but rather structured semantic retrieval of the relevant knowledge could assist in supporting dynamic and real-time decision-making processes in the industry. The work of Figure 1 depicts the fusion of Knowledge Graphs and Retrieval-Augmented Generation (RAG) to smart machining decisions via AI. The illustration compares the classic manual machining (where there is lack of efficiency, errors and time-consuming decision making) to AI-enabled machining, which provides adaptive, accurate, real-time decision making. It is done through the combination of RAG Implementation (inference-driven query responses) and Knowledge Graph Integration (structuring semantic manufacturing knowledge) and permits better returns.

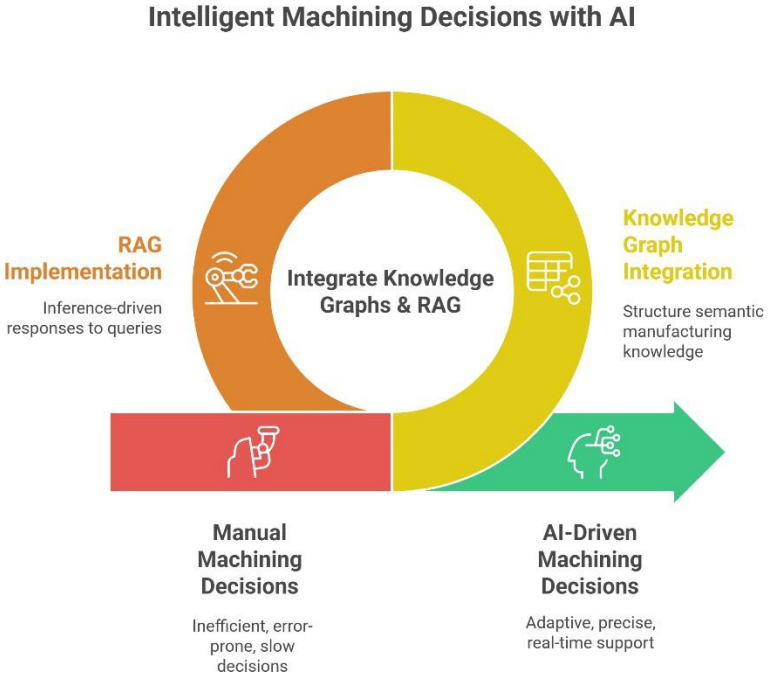

Figure 1: Integration of Knowledge Graphs and Retrieval-Augmented Generation (RAG) for intelligent machining decisions with AI

This was further applied to the optimization of machining processes by building a process knowledge graph employing cosine similarity matching to process the proposed graph with feature-specific machining recommendations. Their model represented geometries of parts, machining features, and other parameters in a semantic format which could easily be retrieved to construct mature machining schemes on complex geometries such as holes and pockets. As discussed in the paper, the prospective of knowledge graphs to simulate situational

manufacturing knowledge, enhance the precision of retrieval as well as better designing-to-manufacture pipeline in aerospace and related sectors is significant [7].

In spite of these developments, the application of generative AI, in particular, retrieval augmented generation (RAG) to machining decision systems is a poorly explored field, particularly when structural aviation parts and engine components are concerned. RAG RAG, together with knowledge graphs, has the potential to push the limits of existing decision support systems because it allows inference-driven responses using not only the structured knowledge but also the dynamic query resolution. [8] emphasizes that this combination of retrieval and generation is a good opportunity to provide real-time and situation-specific assistance. However, its application in machining is very rudimentary.

The new convergence of the knowledge graph reasoning and RAG is one solution to the inflexibility of the traditional rule-based or heuristics machine learning methods. With the ability to access the organized knowledge of the domain quickly, and by providing the specific suggestions applicable in the situation at hand, such systems promise to improve the decision-making process in aerospace workshops. According to [9] this fusion can decrease production errors and improve throughput and benefits on response times that are particularly important in high-value and low-tolerance sectors such as the aerospace industry. On the same note, [10] observe that production cycles can be automated through graph-augmented decision-making to enhance the accuracy in machining quality consistency.

This literature review critically examines the integration of industrial knowledge graphs with RAG-based generative AI for decision-making in the machining of high-precision aviation parts and engine systems. The analysis synthesizes current methodologies for knowledge graph–based decision support, evaluates the application of RAG for machining parameter recommendations, and identifies key limitations and challenges in current approaches. The review further proposes directions for the development of inference-driven, intelligent decision systems that can adaptively support manufacturing operations.

Drawing on the experience of case studies and methodological understanding, this paper develops a conceptual framework of structured-reasoning systems in aviation machining that can mediate between domain-specific knowledge and AI-inspired inference. It highlights the possibilities of transformation of semantic knowledge representation coupled with generative models in order to facilitate intelligent, adaptive, and accurate decision support. This not only makes the machinists navigate through complicated decision-making choices, but also provides solutions that can be scaled to suit the entire aerospace manufacturing ecosystem. Finally, the study prepares the ground to develop the next generation AI systems that will result in a better cognitive load and improved contextual reasoning in aircraft component manufacturing, which is a high-stakes field.

Bibliometric analysis of 252 selected references was done to position this review in the changing academic literature. This effort illuminates thematic focal points, patterns of publication, and new directions of research in this field. The most common words used in titles converge as in Figure 2 around knowledge, based, manufacturing and process, indicating that there has always been interest in intelligent systems and data driven machining. The cloud itself can be a brief visual overview of these high-impact topics whose visual representation shows not the depth of the techniques but the conceptual trends as well (e.g. retrieval, graph, cognitive).

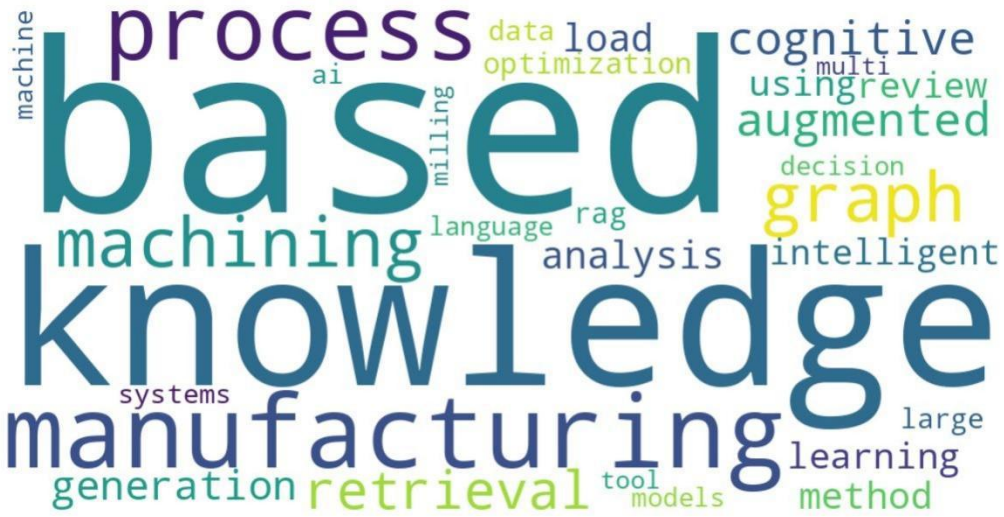

Figure 2: Word cloud of frequently occurring title terms in the reviewed literature

Figure 3 captures the temporal trend in publications, showing a marked acceleration after 2018. This sharp rise aligns with the increasing integration of AI and knowledge engineering into manufacturing domains. Notably, publication volume peaks in 2025, suggesting that this is an especially fertile year for research contributions.

Regarding the dissemination channels, Figure 4 illustrates the most commonly mentioned journals in this review. Among the unnecessary ones, one can distinguish arXiv, Applied Sciences, and IOP Conference Series: Materials Science and Engineering, which means that there is a good balance between open-access preprints and peer-reviewed conference papers. Such diversity implies a rapidly developing multidisciplinary research field where contributions occur specifically in the area of theory development as well as in the area of application technologies.

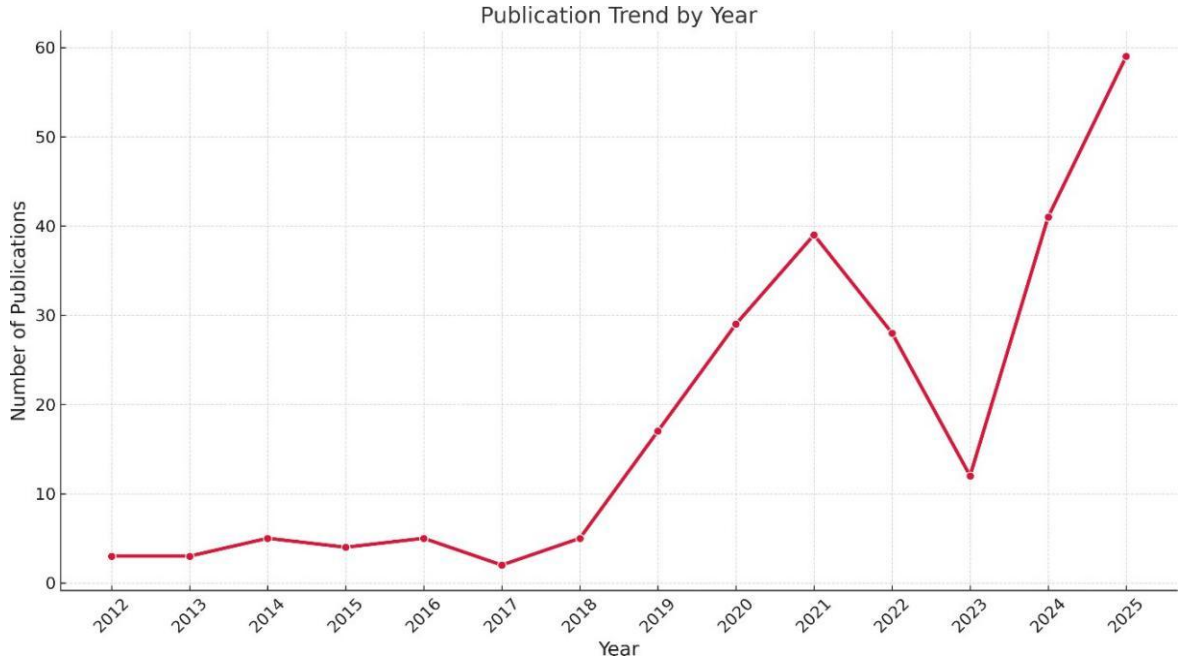

Figure 3: Annual publication trend from 2012 to 2025.

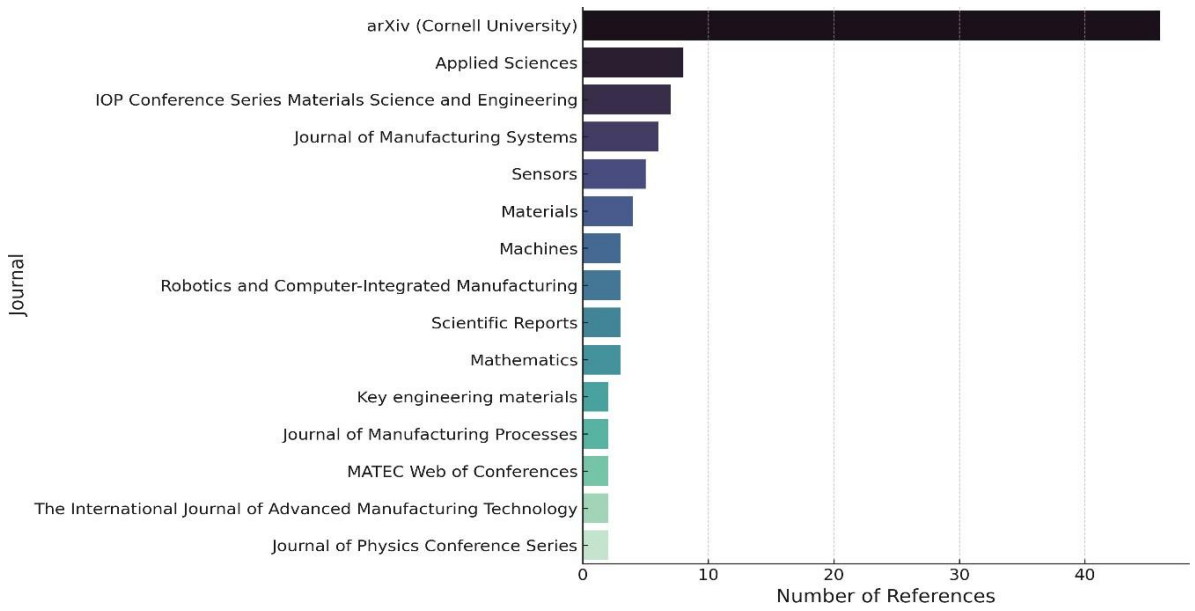

Figure 4: Top journals and conference proceedings represented in the reference dataset.

Lastly, Figure 5 presents keyword co-occurrence network graph constructed from the top 50 most frequent title terms in the reviewed literature. Nodes represent individual keywords, and edges reflect their co-occurrence within titles. The layout illustrates conceptual proximity and research clusters across themes like manufacturing, knowledge systems, cognitive graphs, and process optimization.

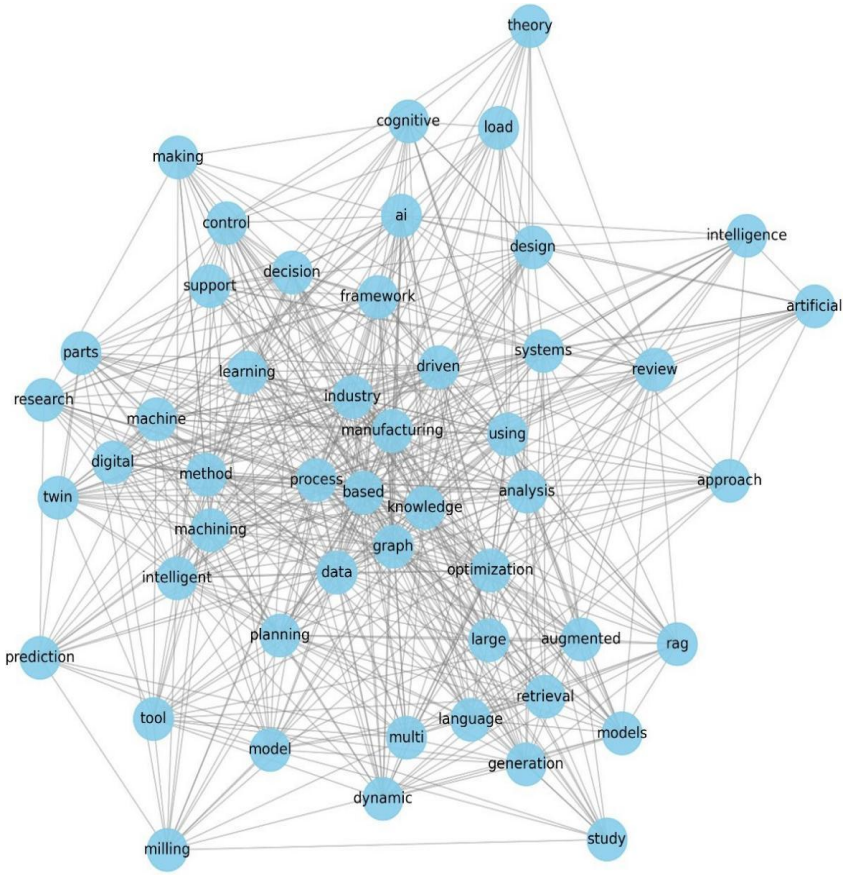

Figure 5: Keyword co-occurrence network graph.

2 Theoretical Framework

CLT has been instrumental in the development of the "cognitive manufacturing" frameworks, which aim at matching human-machine cooperation to the capacity of the working memory. [11] put forward a conceptual model for Industry 4.0 contexts that directly aims to minimize cognitive load through semantic integration and HCI technologies. Through the lens of cognitive manufacturing, CLT explains why engineers, who have to choose from dozens of parameters affecting machining outcomes (feed rate, spindle speed, tool path) experience overload and how structured AI can assist in this process. Figure 6 gives the enhanced Cognitive Load Theory framework showing the connections between intrinsic load (caused by information complexity), germane load (shaped by mental effort and motivation), and extraneous load (affected by how information is presented). These factors impact schema acquisition, ultimately influencing outcomes such as the effectiveness of valuation methods, presentation issues, and decision-making. As it stands, CLT differentiates between intrinsic load, or the complexity resulting from aviation component machining, extraneous load, which is the inefficiency resulting from the presentation of information, and germane load, which is the effort that is devoted to constructing a schema. In precision machining of engine parts, the intrinsic load is high: every property of material, the strategy of cutting and the tolerance levels available are in a very large decision space. Exposing these relationships in an industrial knowledge graph helps to offload important relationships from the minds of engineers and designers who otherwise must constantly switch between mental models of tool-workpiece coupling. In this way, by providing only contextually relevant nodes and edges instead of tables or manuals, the load that can be considered as irrelevant according to Sweller is reduced and the resources are freed to be used for processing of new configurations.

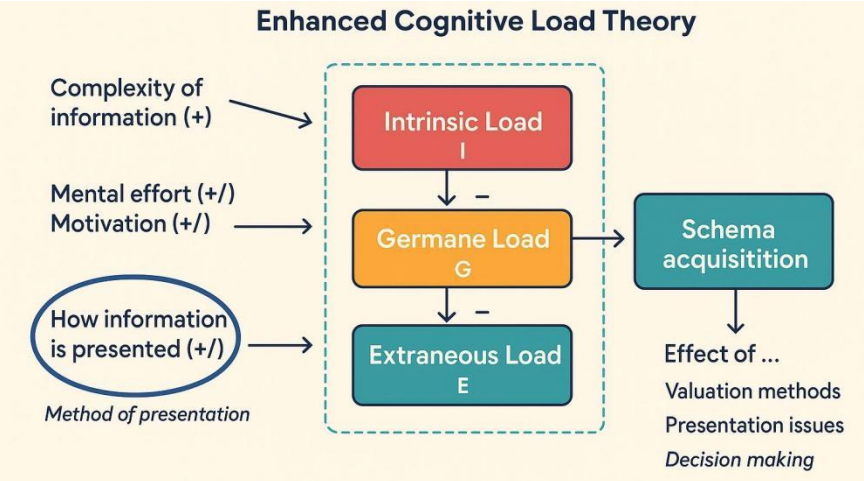

Figure 6: Enhanced Cognitive Load Theory Framework

Building upon this, Retrieval Augmented Generation (RAG) not only retrieves and synthesises the knowledge into graphs but also provides a concise, human-interpretable textual explanation. [12] showed that using the conceptual-level understanding of the AI explainability methods developed with CLT, it is possible to greatly reduce the users' effort when interpreting the model outputs. The RAG retrieval filters out information overload for engineers so that they do not waste time on excessive data input and provides immediate feedback for the refinement of the schema in terms of what is useful and valuable. Based on these concepts, the theoretical framework places CLT at the centre of the architectural support of inference-based decision-making. The knowledge graph represents the system knowledge

and can contain machining parameters materials' characteristics and tools constraints (Guo). The RAG layer works as the cognitive layer, and it interacts with the graph to generate specific and contextually relevant information. A cognitive load monitor module determines query complexity and length of the result page to either reduce or expand the results returned. In this way, the proposed framework aims to change high-precision aviation machining from a cognitively demanding task into an efficient human-machine performance. It not only shows how AI can enhance the decision-making of experts in manufacturing but also how future systems should be assessed: success is defined not only by the accuracy of the recommendations made but by the ability to decrease cognitive load and thus increase throughput and error rates [13].

3 Methodology

This literature review follows a qualitative research approach, which is a secondary approach because it utilizes a content analysis technique that will synthesize and assess the recent scholarly discussion in the domain of Knowledge Graphs (KGs) and Retrieval-Augmented Generation (RAG) model application to manufacturing situations. It particularly examines the way they are used in decision support, parameters optimization of the process, and easy identification of any implementation challenges. The content analysis was selected because it is an appropriate method of interpreting patterns, themes, and relationships in qualitative textual data that are based on peer-reviewed literature. The relevance, credibility, and recency criteria helped to select the sources, but some exceptions were made. Keywords like knowledge graph, RAG model, manufacturing decision support, machining optimization, semantic reasoning, and intelligent process planning were used in the search of databases like Scopus, IEEE Xplore, SpringerLink, ScienceDirect, and Web of Science. Peer-reviewed articles and books concerning technical and engineering applications were only considered and no materials that were not in English or peer reviewed to ensure academic rigour were taken. To structure the literature, a thematic coding framework was utilised in order to have three categories based on the objectives that the study is going to achieve: knowledge graph-based decision support, RAG applications in machining, and integration challenges. A similarity was observed in the findings in many articles to bring out convergences and differences between the findings. It was specifically focused on frameworks, case studies, system architectures and statistics benchmarks that can aid in critical interpretation. Through the use of content analysis, such a methodology enables subtle synthesis in the interdisciplinary studies within semantic technologies and AI across intersections of conceptual and technological understanding between semantic technologies and deep learning in manufacturing [14]. It is also transparent and replicable, which contributes to the creation of informed recommendations and the future study direction where gaps and trends are identified.

Figure 7 shows the key benefits and challenges of integrating Knowledge Graphs and Retrieval-Augmented Generation (RAG) models in manufacturing. The figure highlights improved decision-making processes and enhanced machining precision as major advantages, while illustrating data integration complexities and technical skill gaps as primary challenges associated with this integration.

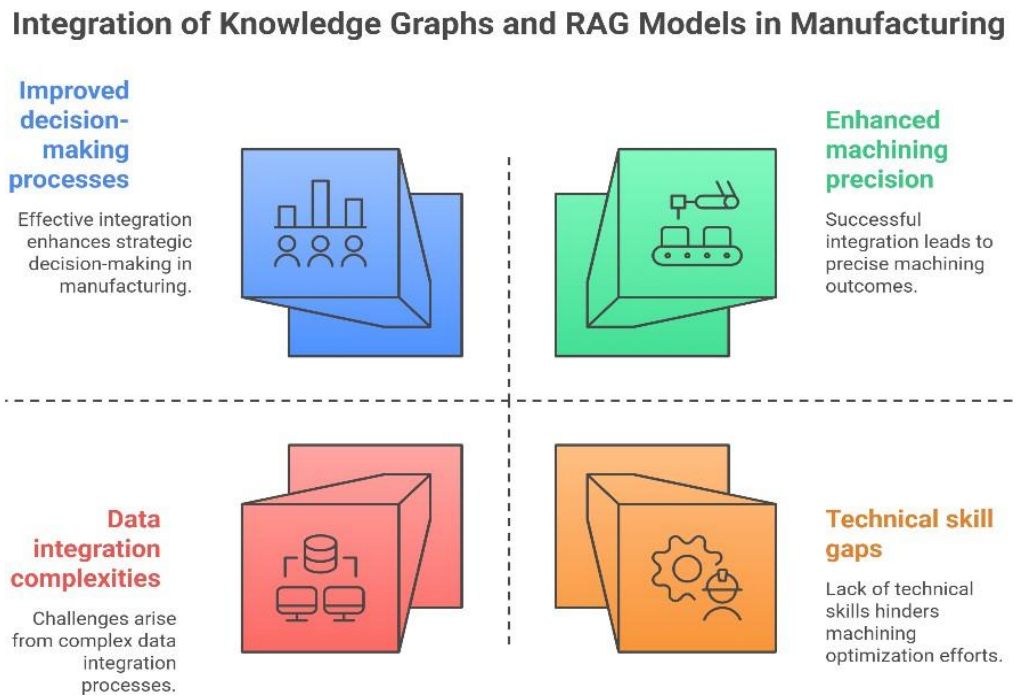

Figure 7: Key benefits and challenges of integrating Knowledge Graphs and Retrieval-Augmented Generation (RAG) models in manufacturing.

3.1 Foundations of Intelligent Manufacturing Decision Support Definition and Evolution of Intelligent Decision Systems in Manufacturing

Based on its application, the IDS in manufacturing has grown from the traditional rule-based expert system to an intelligence system supported by real-time analytics, a semantic knowledge base, and human-computer interaction. Initially, rules in the expert systems were programmed manually; thus, there was a problem of scalability and flexibility. Thus, Industry 4.0 utilises advanced decision aids such as deep learning, the digital twin concept or knowledge graphs to gather information that requires context awareness and prediction support seen in Figure 8. This recent and comprehensive review also points out how these architectures of AI-based DSSs use multiple-layer models of the flow of streams of data or real-time sensors and enterprise databases as well as semantic reasoning, to enable production and maintenance schedules and quality control in industries [15].

In an example of digitalisation in DSS application, a detailed analysis of how process modelling using digital twins can help in the tactical and operational level decisions, in terms of lead time as well as resource allocation across different manufacturing lines. These developments can be said to have seen a major shift away from the traditional models of rule-based systems, where once a decision has been made, there is no possibility of change to the learning-enabled platforms, where decisions change with the arrival of new data.

In the aviation sector especially in structural parts and engine machining, the decision-making environment involves close tolerances, difficult and intricate shapes such as thin walls, and varying material properties. Thin-walled aerospace components have many problems: changes in thickness, spring back, and vibration modes of the thin-walled aerospace components have critical impacts on the surface and dimensional control. Many discuss intelligent monitoring technologies for thin-walled machining and identify the lack of adaptive decision support for high-frequency sensor data to alert the operator about emerging

anomalies [16]. In the same way, surface quality in the machining of an engine part requires consideration of interrelated effects of cutting speed, feed rate, and tool wear have been used in AI models to predict the surface finish based on these parameters, thus showing the drawback of adjusting the parameters traditionally due to the complexity of the interrelated conditions. Altogether, these works reveal the challenges of the cognitive and the technical aspects that engineers experience when they try to meet high performance in the face of the rising and large amount of data.

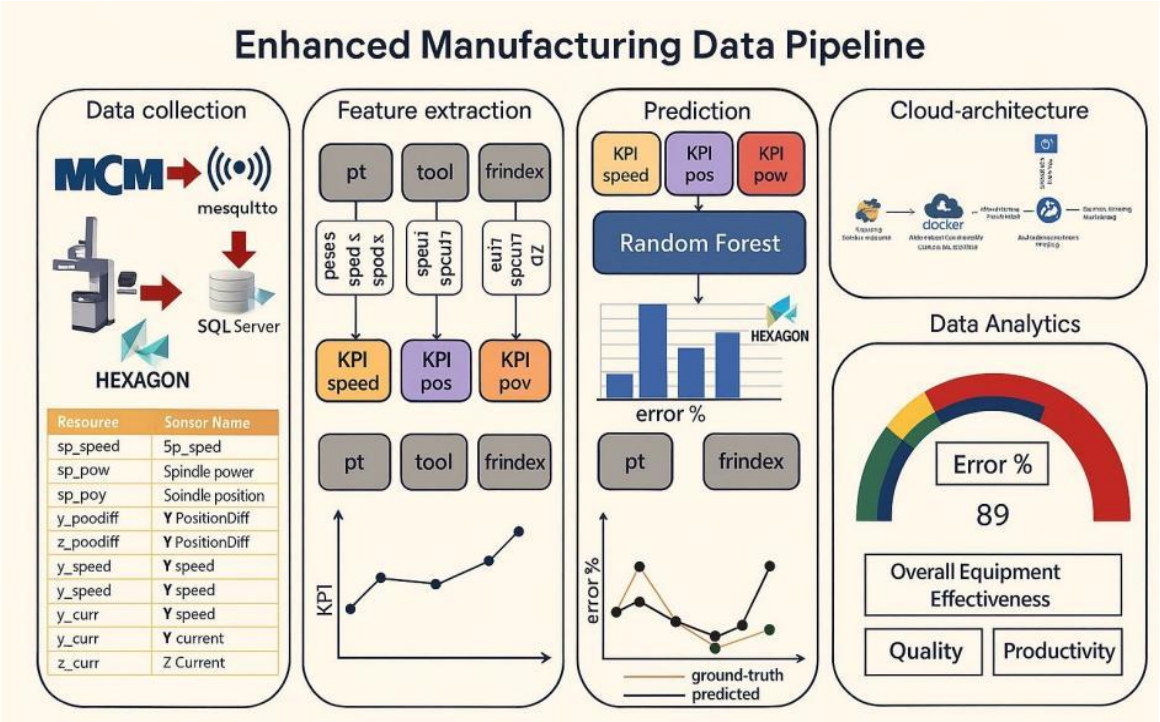

Figure 8: AI-Based Decision Support Systems in Industry 4.0

Conventional methods of choosing the machining process parameters include Taguchi's method, DoE, and guidelines based on the expert's experience. Even though these methods have brought changes gradually, they are not effective in learning the changing production environments, the new grades of the material, or the new types of faults. On the other hand, intelligent decision support includes knowledge graphs to encode manufacturing ontologies such as the materials, tools, and machines and uses the retrieval augmented generation (RAG) to fetch and generate the required parameter recommendations when needed. AI-based DSS frameworks can therefore provide the feed rates or spindle speeds that are most beneficial based on structured knowledge and historical data in a short time, and the context of the environment [17]. As pointed out in the review of the AI DSS, this moves from fixed rule sets to dynamic learning systems will go a long way in cutting down on setup time, taking out the guesswork and increasing the resilience of the process.

3.2 Industrial Knowledge Graphs in Machining Applications Definition and Role of Knowledge Graphs in Manufacturing

Industrial KGs are formal models of domain objects and their relationships in a format that can be processed by machines as seen in Figure 9. In the manufacturing industry, KGs are centralised knowledge storage that aggregates information from various sources, including CAD models, logs, and experts' guidelines. This allows for a higher level of reasoning on

relationships, performs automated inference and decision-making, and helps to trace decisions made during the production process. That is why, when using KGs as a knowledge base for decision support systems, it is possible to enforce ontological definitions and logical rules so that the system can make semantic queries and obtain the required knowledge that is relevant to the context without the need for additional curation [18].

It is in this context that the modelling of machining knowledge within these graphs entails certain aspects of the manufacturing process, including material properties, tool paths, and cutting parameters. For example, a materials terminology KG built mechanically from engineering literature has attributes like hardness, thermal conductivity, and microstructure associated with the recommended machining conditions. In parallel, there is a machining process route generation method that formulates specific operations (drilling, pocketing) as graph entities, where the edges represent the allowed sequences and tooling. It is such a graph that enables the generation of process plans through relations that link part geometries with tool path templates.

Altogether, these approaches show that KGs involve both domain knowledge and procedural logic for the accurate manufacturing of components. Various examples demonstrate the use of KGs in the context of processing parameter optimisation in machining. [19] proposed a process knowledge graph with an improved cosine similarity matching to present feature-specific machining schemes. It encodes part features, tool libraries, and cutting conditions into a single graph to get the mature machining sequences to reduce the trial-and-error setup and enhance surface quality in aerospace parts. The cosine similarity matching calculation model is as follows:

$$CosineSimilarity = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \cdot \|\mathbf{B}\|} \tag{1}$$

where, $\mathbf{A}$ and $\mathbf{B}$ are two vectors that represent embedded representations of two machining features or parameters (such as tool geometry, material properties, etc.). By calculating cosine similarity, the matching degree between different processing schemes can be quantified, thereby recommending the optimal parameters.

Similarly, an intelligent numerical control (NC) programming system uses a KG to translate high-level manufacturing intents such as "roughing pass for titanium alloy" to NC code templates and control feed rates and spindle speeds based on linked material and tool entities. These cases illustrate how KGs can be applied in practice for automating decision-making and incorporating expertise rules into knowledge-driven business processes.

The use of knowledge graphs in machining applications has many advantages. It provides indirect connections, for example, between tool wear rates and cooling strategies, which enables it to provide better decision-making in conditions of risk. These filters help in achieving semantic filtering that enables users to get recommendations that suit specific part geometries and material grades. Furthermore, because of traceability, graph provenance is traceable, and the engineers can verify the decision-making process and meet the high standards of aerospace quality. Altogether, these advantages place the industrial KGs based on intelligent, transparent, and adaptable manufacturing systems.

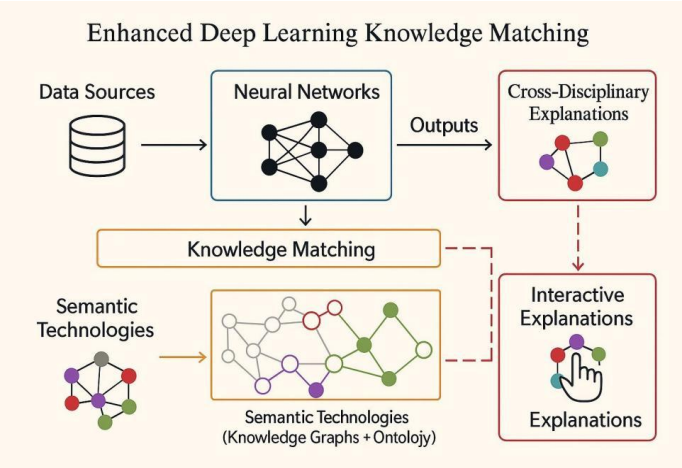

Figure 9: Schematic of an explainable AI system integrating semantic technologies with deep learning.

3.3 Retrieval-Augmented Generation (RAG) in Engineering Context Overview of RAG: Combining Pre-Trained Language Models with External Knowledge Retrieval

Retrieval Augmented Generation (RAG) is a type of AI system that combines features from large language models that are pre-trained, and an efficient knowledge retrieval system as seen in Figure 10. In a conventional RAG, a user query is passed through a retrieval component either via sparse vector techniques such as BM25 or dense embedding to obtain the relevant documents or knowledge snippets from a predetermined knowledge source. These retrieved contexts are then appended to the original context and passed through a generative model to produce well-formed responses that are semantically connected to the specific context [20]. This architecture also helps to reduce the hallucination of generative models by fixing the output to factual sources and can also adapt to domain-specific updates dynamically. Researchers proposed a dynamic selection-based RAG approach that improves the multi-document QA for commercial purposes by adjusting the weights of the retrieved passages dynamically, which achieved a 15% improvement in answer accuracy and the coherence of the responses with the static retrieval models.

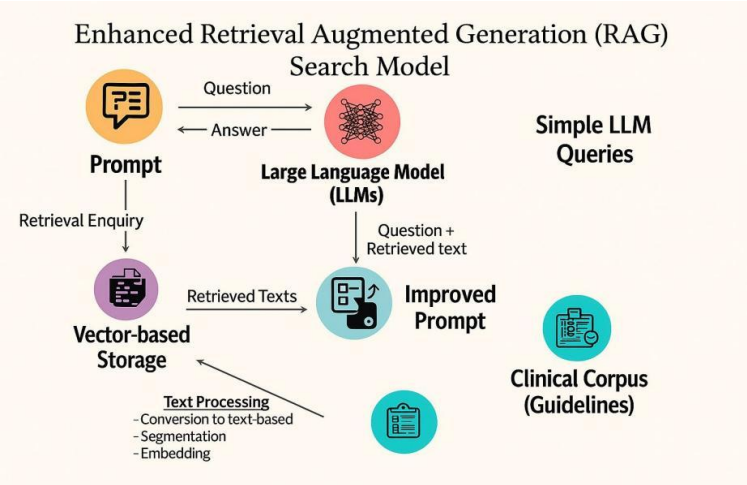

Figure 10: Architecture of a RAG model integrating a retrieval module with a language model.

RAG has been applied in various engineering fields in the recent past, making it a versatile tool. According to Niv, in systems engineering, RAG can incorporate domain ontologies like component specification and functional requirements in the retrieval process to help LLMs generate design rationale and trade-off analysis that meets strict engineering standards. The retrieval part guarantees that the generated content is compliant with formal requirements, and the generative layer creates more elaborate storylines for further communication with stakeholders. [21] proposed an intelligent fault information retrieval system for new energy vehicles for the automotive sector where an LLM was connected to a database of maintenance records and sensor logs. Their RAG-based system obtained a retrieval accuracy of more than 90% and generated diagnostic reports that reduced the average time to repair in pilot studies. These applications demonstrate how RAG can provide detail and context in technical areas and how it can be used to provide an understanding of the subject matter.

When it comes to high-precision manufacturing, there are various opportunities for using RAG frameworks to improve decision-making in the following areas. For the recommendation of machining parameters, RAG systems can access process logs and material databases to pull cutting conditions such as feed rate, spindle speeds, and tool path approaches and provide brief, justified recommendations based on part geometry. In fault diagnosis, the real-time sensors' anomalies can be matched with failure cases and a maintenance procedures database, the generative component will then generate the root causes and the detailed steps of the rectification. RAG is advantageous in intelligent process planning as it integrates detailed operation sequences: the best practice workflows from the digital twin repositories are retrieved, and then an instruction that considers the mechanics of the machine, the availability of the tool, and the quality requirements. Through retrieval for factual grounding and generative synthesis for human-friendly guidance, RAG systems have the potential to shift conventional manufacturing decision-aiding tools into contextually aware and self-learning tools that can effectively handle the diverse and dynamic nature of current manufacturing systems [22].

3.4 Synergistic Integration: Knowledge Graphs + RAG for Machining

In this context, the KG is considered the semantic core that stores materials, tools, and process parameters and their related ontologies, while the RAG layer is the cognitive layer that queries specific subgraphs for machining decisions as well as preparing concise context-aware advice. The framework starts with KG construction where domain experts transform workpiece alloys (e.g. stainless steel, aluminium) and tools (e.g. end mill, drill) into concepts while relations such as suitable for and requires cooling into relation in a KG that represents best practice in machining as identified. At runtime, a query module interprets the engineer's demand ("roughing titanium pocket") to a graph-structured query, which retrieves subgraphs that store analogous past operations and results. The obtained subgraph context is then joined with the original prompt and passed to an LLM to obtain parameter recommendations such as feed rate, and spindle speed which are semantically aligned with KG and free of hallucinations [23].

RAG's strength is in the usage of structured manufacturing knowledge to provide improved explanatory responses. Instead of using raw corpora, the system acquires KG-derived embeddings and textual annotations, for example, tool wear correlations and coolant strategies, and asks the LLM to generate insights with links to the corresponding nodes in the KG. This helps in having a better contextual understanding; for instance, when the system is machining a deep and thin-walled aerofoil pocket, it can look at the KG and come up with a plan that would have multiple steps addressing how to remove material while at the same time

maintaining stability.

In practice, this synergy allows for a dynamic recommendation of these parameters, critically. For feed rate tuning, the KG incorporates previous research studies that relate cutting velocities to surface finishes, while RAG translates these into straightforward instructions such as "use 20m.s−1 to 30m.s−1 for roughing of Ti-6Al-4V". It also applies to spindle speed optimization: the ability to obtain subgraphs with success in certain tool coating pairs lets the LLM adapt RPM suggestions depending on the current sensor data [24]. For the selection of the tool material, the framework can query the KG for information on tool hardness, thermal stability and cost, which can then be matched with the geometry of the component and the number of components to be produced.

Emerging industrial architectures showcase this integration. A recent research work of ASME has proposed a microservice-based RAG engine integrated with a hosted Neo4j KG that can retrieve aerospace workshop machining schemas within a sub-second time. Another prototype implements Apache Jena for KG storage and Lang Chain for orchestration and scalability of the solution in the CNC shop floor with the continuous learning loop to update the KG from the sensor logs and operators' feedback.

Altogether, these examples demonstrate how KGs + RAG build a solid decision support framework – one that is open, dynamic, and based on the existing machining knowledge and the state-of-the-art generative AI approaches. Figure 11 shows the conceptual Framework of an Inference-Driven Decision Support System Integrating Knowledge Graphs and Retrieval-Augmented Generation.

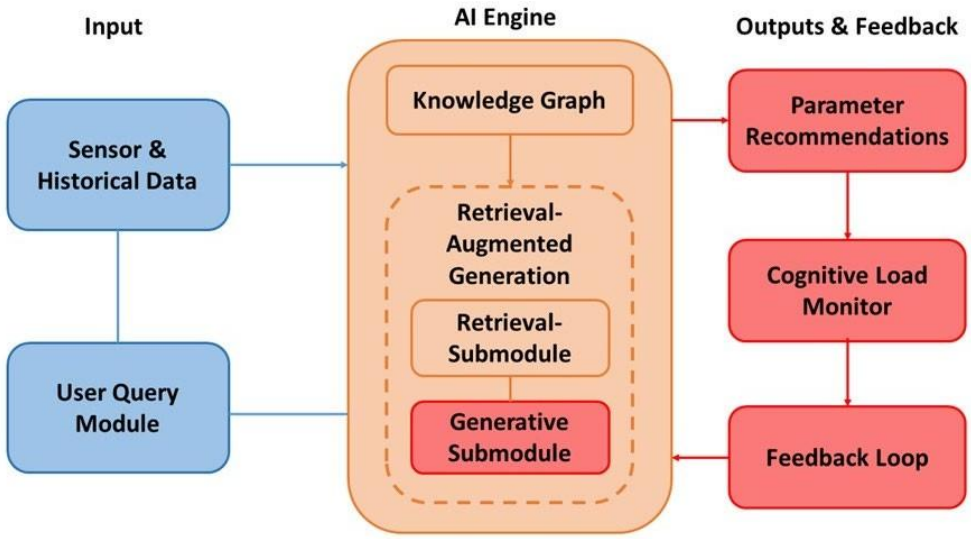

Figure 11: Conceptual Framework of an Inference-Driven Decision Support System Integrating Knowledge Graphs and Retrieval-Augmented Generation

4 Challenges and Research Gaps

The current developments of inference-driven decision support for machining have revealed the following main challenges that need to be overcome for the technology to be adopted widely in the industry: One of the major challenges is the ability to exchange and incorporate data in existing manufacturing systems. Most aerospace and engine machining shops use a mixture of CNC controllers, PLM, and MES systems, which results in distributed and separate access to material properties, tool logs, and processes. In a poll conducted with more than 50 aerospace workshops, it was established that more than 70% of critical process data

were locked into vendor-proprietary formats, which limited real-time knowledge graph population and dynamic querying. Many pointed out the absence of a shared exchange protocol (MTConnect vs. OPC UA) as a problem that hinders KG updates and RAG retrieval lines. Table 1 shows the challenges faced by intelligent decision-making systems in aviation manufacturing.

Table 1: Challenges Faced by Intelligent Decision Systems in Aviation Manufacturing

Challenge	Details
Data exchange and integration	Most aerospace and engine processing workshops use multiple CNC controllers, PLM, and MES systems, and over 70% of critical process data is in supplier proprietary formats, limiting real-time knowledge graph filling and dynamic queries; Lack of shared exchange protocols (such as the MTConnect and oPCUA dispute) hinders knowledge graph updates and RAG retrieval
Real time constraints and computational costs	RAG needs to perform high-speed semantic search on large knowledge graphs, which are then generated by complex Transformer LLM. For medium-sized knowledge graphs (over 100K nodes), the query latency exceeds 2 seconds, making it unsuitable for real-time adaptive control loops and requiring a large number of GPUs and infrastructure, resulting in high energy consumption
Specific contexts and general intelligence	Although general LLM can predict parameter settings, it lacks specialized knowledge in specific fields, such as reducing vibration when processing thin-walled titanium parts; The RAG system adapted to the domain has lower errors in feed rate, but lacks the process of integrating professional process heuristic methods into the hybrid RAG architecture
Knowledge Graph Pattern	The existing knowledge graph uses specific materials, tools, and process planning ontology, which cannot be shared and benchmarked across factories; Recent literature has proposed various competition models, but they have not received widespread attention. There is a lack of standard language and relationship types, which hinders knowledge sharing among aerospace manufacturers

The second problem relates to the effects of real-time constraints and computational costs of RAG implementations on the shop floor. In Retrieval Augmented Generation, a high-speed semantic search needs to be conducted on large KGs and then followed by computation of complex transformer-based LLMs for generation [25]. Further showed that the end-to-end RAG latency may take more than 2 seconds per query in mid-sized KGs with more than 100K nodes, which is not suitable for real-time adaptive control loops. However, it is quite challenging to maintain such workloads that demand not only a massive amount of GPU but also more infrastructure and energy consumption in 24/7 operations. Another research gap is context-specific vs. general intelligence regarding machining decisions. Even though generalist LLMs can learn to foresee practical she/binary parameter settings, it suffices to say that they hardly are endowed with that specificity of expertise that is always helpful when dealing with specialities such as minimizing chatter while machining thin-walled titanium parts. [26] compared a general-purpose RAG system with a domain-adapted variant fine-tuned on aerospace process data; the latter had a lower error rate in feed rate, implying the need to develop more specialized reasoning models. However, there are few procedures for the integration of highly specialized process heuristics into hybrid RAG architectures systematically. The query latency calculation model of RAG is as follows:

$$Latency = T_{\text{retrieval}} + T_{\text{generation}} \tag{2}$$

where, $T_{\text{retrieval}}$ is the knowledge graph retrieval time, and $T_{\text{generation}}$ is the language model generation time. For large-scale knowledge graphs, retrieval time may become a bottleneck and latency needs to be reduced through index optimization or distributed computing.

Finally, the absence of well-defined knowledge graph schemas for manufacturing knowledge is an issue that does not allow for sharing and reusability across organizations. The current KGs use specific ontologies for materials, tooling, and process plans, which do not allow cross-plant sharing and benchmarking. [27] identified more than a dozen competing schema proposals in the recent literature, but none of them has received considerable attention. So long as there is no standard language and standard relationship types, sharing and combining graph-based knowledge between different aerospace producers will remain challenging. Figure 12 shows the key challenges hindering inference-driven machining adoption. The figure highlights data silos due to proprietary formats, slow retrieval-augmented generation (RAG) with latency issues, lack of machining expertise in general large language models (LLMs), and the absence of knowledge graph (KG) schemas which inhibits knowledge sharing.

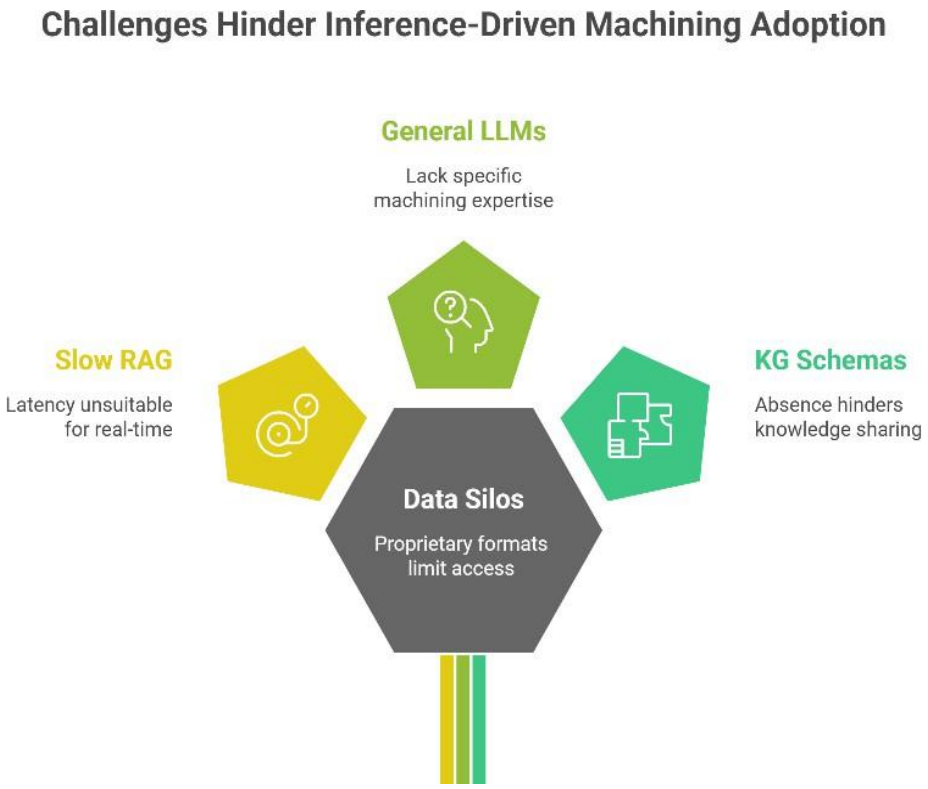

Figure 12: Key challenges hindering inference-driven machining adoption.

5 Discussion

The studies demonstrated that, in the context of machining aerospace impeller components, the use of a process knowledge graph and better cosine similarity matching helped in reducing the feature-specific machining error rates and proved the efficacy of KG-driven parameter recommendations. However, these systems are based on static graphs updated offline, and they do not have a real-time connection with the shop floor data streams and hence cannot

learn from the tool wear or variability in material.

Researchers proposed an M-KG model to support the manufacturing decision-making for casing machining in real-time, with real-time parameter tuning to reduce cycle time. One of the interesting observations made during the literature review was that although KGs have been applied for encoding the expert knowledge for the parameter selection and still, most of the existing systems are not dynamic and do not possess the ability to update themselves in real-time. It has been shown that the application of the KG-driven system for machining the aerospace impellers helped to decrease the feature-specific errors because of the more accurate choice of parameters. However, the system only used pre-defined graphs updated offline and hence, it was not sensitive to the real production conditions.

Another study proposed the real-time KG-enhanced machining advisor, which resulted in a decrease in the cycle time through the dynamism in feed rate and spindle speed through the use of sensors. This indicates that it is possible to significantly improve the efficiency of operations by integrating KGs with live data. On the other hand, [28] presented a case-based reasoning framework with KG for welding parameter recommendation that enhanced the consistency aspect, but it is not dynamic. Such inconsistency across studies suggests that there is a need to have better integration of KGs into closed-loop manufacturing systems.

One of the major issues that were pointed out was that there were no established templates for knowledge graphs in manufacturing. From the literature reviewed during this study, it was realized that with no schema consistency, even the best graphs cannot interoperate or support scalable RAG pipelines. This was further described in the works done where OMPKG (Open Manufacturing Process Knowledge Graph) was built to solve the problem of semantic fragmentation, having the key-value and entity-relationship paradigms [29, 30]. The other challenge is how to handle data from the legacy system. Therefore, the idea of integrating RAG-KG systems will remain undefined if there is no significant enhancement of system compatibility, schema matching, and hybrid model selection.

6 Future Trends and Research Opportunities

There’s growing momentum around using domain-specific language models in the manufacturing sector, particularly in high-precision machining. As access to large-scale industrial knowledge graphs becomes easier, and as learning systems for real-time machine control and smart factory integrations evolve, the idea of inference-based decision-making is getting stronger. AI isn’t just assisting—it’s starting to understand the deeper needs of machining environments.

The more language model is trained on manufacturing specific content, the more relevant the results are. These models start speaking engineer language and machine language. The accuracy is enhanced and the amount of inaccuracy in process parameter prescribing reduces significantly. Indeed, case-based learning with machining data has already begun to demonstrate obvious benefits as compared to general-purpose AI systems in terms of generating meaningful information.

Industrial knowledge graphs are also moving in the direction of the objectives of smart manufacturing. They are being designed to be FAIR (Findable, Accessible, Interoperable, and Reusable) and are prepared to large-scale applications of AI. Even some of the largest technology firms are already experimenting on how such graphs can be made inter-operable among plants. Even though the data schema gap between different systems still exists, such as how to fit data schema into ontology-based approaches, modular data approaches are proving useful in bridging those disjunctions and enabling systems to communicate at scale.

Knowledge graph based reinforcement learning is assisting in achieving quicker more

responsive scheduling in more dynamic aspects of production. This is important in the adaptive process control where timing and sequence decisions must vary in real-time. Developments in the development of capturing spindle power feedback in machining is also aiding these adaptive controls so that even in case of process variability, cycle times are more balanced.

The digital twins are introducing a new set of intelligence. They provide an interface between real machines and the virtual ones, allowing one to modify operations according to expected results or detect faults before they occur. The use of digital twins in offloading real-time performance in CNC machining is already underway, and virtual performance approaches the actual performance by a thin band of error. Such a simulation does not only raise the confidence in automated decision-making, but it also paves the way to completely closed-loop systems, which are able to learn and self-correct errors in course of operations.

These developments suggest that the integration of AI, knowledge graphs, and digital twins isn't just theoretical anymore—it's happening on the factory floor. The precision, adaptability, and decision support they offer are transforming how manufacturers think about machining, quality control, and process efficiency.

7 Implications and Conclusion: Towards Intelligent Decision Support in Aviation Machining

The adoption of intelligent systems in machining of the airline offers a ground-breaking change in the decision making shop floor. In a business where accuracy, dependability and effectiveness are the key elements, reducing dependence on human implication and increasing the use of facts can be far-reaching. The predictive abilities of intelligent systems, especially those based on the methods of artificial intelligence, prevent the unintended downtime, because the equipment failures can be predicted. This reduces the time wasted in maintenance and also operational budgets are reduced considerably.

In addition, they facilitate high-quality scheduling, minimize variation in process, and inspire good use of resources, hence, leading to productivity. These intelligent frameworks are further complemented by additive manufacturing (AM) technologies as the demand of complex components is on the rise and hard-to-machine materials prevail in the aviation industry. The capacity of AM to form elaborate geometries and still retain the structural integrity is in line with the strict performance requirements in the aviation industry. Intelligent systems and AM can also support environmental practices that are sustainable by facilitating fuel-efficient design and reduce the use of materials, which is the voice of aviation regarding green manufacture.

An overview of the current literature supports the idea that the knowledge graphs (KGs) and retrieval-augmented generation (RAG) represent an attractive direction in the creation of the advanced decision support systems in the field of aviation machining. KGs give a systematic presentation of the domain specific information and this better leads to traceability, transparency and contextual meaning to the process planning. This is quite important especially in aviation situations where regulatory adherence, safety provisions, and quality assurance require very responsible decision-making systems.

RAG also add to the capabilities of the knowledge graphs as it allows access to context specific information in real-time. This exoset mode is more effective in accelerating the decision-making process by memory access of the pertinent insights in real time and development of dynamic suggestions, particularly in situations where complex and intolerant machining procedures are characteristic of aerospace pieces. The KGs and RAG combination

therefore diminishes human addictions, minimize mistakes and facilitates more stable and precise production result.

These innovations are not simply technically well-founded, but also well-calculated along with more general industry transformations under Industry 4.0. In continuation of the development of digital manufacturing, the further studies should consider how KG-RAG systems can be made interoperable with digital twins, intelligent scheduling algorithms, and self-learning systems. This kind of integration will play a pivotal role in the achievement of resilient manufacturing ecosystems, adaptive and sustainable manufacturing ecosystems.

To conclude, the merging of knowledge graphs and retrieval based generation in aviation machining can be seen as a strong answer to such old issues in in-precision manufacturing. Their joint use improves precision, responsibility, and flexibility. Although the industry has been on the frontline to push the limits of innovativeness, the future of high-performance, sustainable, and intelligent aviation manufacturing settings will be underpinned by intelligent and inference-based systems.

References

[1] Buchgeher G, Gabauer D, Martinez-Gil J, et al. Knowledge graphs in manufacturing and production: a systematic literature review[J]. IEEE Access, 2021, 9: 55537-55554.

[2] Yahya M, Breslin J G, Ali M I. Semantic web and knowledge graphs for industry 4.0[J]. Applied Sciences, 2021, 11(11): 5110.

[3] Rožanec J M, Lu J, Rupnik J, et al. Actionable cognitive twins for decision making in manufacturing[J]. International Journal of Production Research, 2022, 60(2): 452-478.

[4] El Kalach F, Yousif I, Wuest T, et al. Cognitive manufacturing: definition and current trends[J]. Journal of Intelligent Manufacturing, 2025, 36(6): 3695-3715.

[5] Rožanec J M, Fortuna B, Mladenić D. Knowledge graph-based rich and confidentiality preserving Explainable Artificial Intelligence (XAI)[J]. Information fusion, 2022, 81: 91-102.

[6] Kumar A, Starly B. “FabNER”: information extraction from manufacturing process science domain literature using named entity recognition[J]. Journal of Intelligent Manufacturing, 2022, 33(8): 2393-2407.

[7] Farbiz F, Habibullah M S, Hamadicharef B, et al. Knowledge-embedded machine learning and its applications in smart manufacturing[J]. Journal of Intelligent Manufacturing, 2023, 34(7): 2889-2906.

[8] Melluso N, Grangel-González I, Fantoni G. Enhancing industry 4.0 standards interoperability via knowledge graphs with natural language processing[J]. Computers in Industry, 2022, 140: 103676.

[9] Kosasih E E, Margaroli F, Gelli S, et al. Towards knowledge graph reasoning for supply chain risk management using graph neural networks[J]. International Journal of Production Research, 2024, 62(15): 5596-5612.

[10] Bharadwaj A G, Starly B. Knowledge graph construction for product designs from large

CAD model repositories[J]. Advanced Engineering Informatics, 2022, 53: 101680.

[11] Abu-Rasheed H, Weber C, Zenkert J, et al. Transferrable framework based on knowledge graphs for generating explainable results in domain-specific, intelligent information retrieval[C]//Informatics. MDPI, 2022, 9(1): 6.

[12] Ali M I, Patel P, Breslin J G, et al. Cognitive digital twins for smart manufacturing[J]. IEEE Intelligent Systems, 2021, 36(2): 96-100.

[13] Stan L, Nicolescu A F, Pupăză C, et al. Digital Twin and web services for robotic deburring in intelligent manufacturing[J]. Journal of Intelligent Manufacturing, 2023, 34(6): 2765-2781.

[14] Rajabi E, Etminani K. Knowledge-graph-based explainable AI: A systematic review[J]. Journal of information science, 2024, 50(4): 1019-1029.

[15] Selvam A. Knowledge Graphs for Integrating Multi-Omics Data in Pharmaceutical Research[J]. Newark Journal of Human-Centric AI and Robotics Interaction, 2022, 2: 512-550.

[16] Joshi P, Masilamani V, Mukherjee A. A knowledge graph embedding based approach to predict the adverse drug reactions using a deep neural network[J]. Journal of biomedical informatics, 2022, 132: 104122.

[17] Castañé G, Dolgui A, Kousi N, et al. The ASSISTANT project: AI for high level decisions in manufacturing[J]. International Journal of Production Research, 2023, 61(7): 2288-2306.

[18] Rosati R, Romeo L, Cecchini G, et al. From knowledge-based to big data analytic model: a novel IoT and machine learning based decision support system for predictive maintenance in Industry 4.0[J]. Journal of Intelligent Manufacturing, 2023, 34(1): 107-121.

[19] Ibrahim N, Aboulela S, Ibrahim A, et al. A survey on augmenting knowledge graphs (KGs) with large language models (LLMs): models, evaluation metrics, benchmarks, and challenges[J]. Discover Artificial Intelligence, 2024, 4(1): 76.

[20] Buehler M J. Accelerating scientific discovery with generative knowledge extraction, graph-based representation, and multimodal intelligent graph reasoning[J]. Machine Learning: Science and Technology, 2024, 5(3): 035083.

[21] Ramonell C, Chacón R, Posada H. Knowledge graph-based data integration system for digital twins of built assets[J]. Automation in Construction, 2023, 156: 105109.

[22] Tamašauskaitė G, Groth P. Defining a knowledge graph development process through a systematic review[J]. ACM Transactions on Software Engineering and Methodology, 2023, 32(1): 1-40.

[23] Verma S, Bhatia R, Harit S, et al. Scholarly knowledge graphs through structuring scholarly communication: a review[J]. Complex & intelligent systems, 2023, 9(1): 1059-1095.

[24] Huet A, Pinquié R, Véron P, et al. CACDA: A knowledge graph for a context-aware cognitive design assistant[J]. Computers in Industry, 2021, 125: 103377.

[25] Kejriwal M. Knowledge graphs: A practical review of the research landscape[J]. Information, 2022, 13(4): 161.

[26] Meloni A, Angioni S, Salatino A, et al. Integrating conversational agents and knowledge graphs within the scholarly domain[J]. IEEE Access, 2023, 11: 22468-22489.

[27] Chhetri T R, Kurteva A, Adigun J G, et al. Knowledge graph based hard drive failure prediction[J]. Sensors, 2022, 22(3): 985.

[28] Jaimini U, Sheth A. Causalkg: Causal knowledge graph explainability using interventional and counterfactual reasoning[J]. IEEE Internet Computing, 2022, 26(1): 43-50.

[29] Kusiak A. Predictive models in digital manufacturing: research, applications, and future outlook[J]. International Journal of Production Research, 2023, 61(17): 6052-6062.

[30] Mattera G, Nele L, Paolella D. Monitoring and control the Wire Arc Additive Manufacturing process using artificial intelligence techniques: a review[J]. Journal of Intelligent Manufacturing, 2024, 35(2): 467-497.