IJARCCE

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 15, Issue 6, June 2026

DOI: 10.17148/IJARCCE.2026.15618

A Graph Retrieval-Augmented Generation Framework for AI-Powered Supplier Discovery

Dr.B.Vijayalakshmi[1], Dr.M.Swarna Sudha[2], Dr.K.Vijayalakshmi[3]

Department of CSE, Ramco Institute of Technology, Rajapalayam, Tamil Nadu, India[1]

Department of CSE, Ramco Institute of Technology, Rajapalayam, Tamil Nadu, India[2]

Department of CSE, Ramco Institute of Technology, Rajapalayam, Tamil Nadu, India[3]

Abstract: As modern supply chains demand higher resilience, agility, and visibility, the task of supplier discovery becomes increasingly critical. We present a novel AI-powered methodology that combines Graph Neural Networks (GNNs), Retrieval-Augmented Generation (RAG), and Large Language Models (LLMs) to enhance supplier search and reasoning. A structured Supplier Capability Knowledge Graph (SCKG) is built by extracting domain-specific triplets from unstructured manufacturing data using fine-tuned LLMs and is enriched through semantic normalization via ontology and manufacturing thesaurus. A GNN-based retrieval system identifies relevant subgraphs by performing dense reasoning over the SCKG. These subgraphs are verbalized into natural language using shortest-path reasoning chains and fed into an LLM for generative explanation. To improve retrieval precision, a hybrid entity normalization technique leveraging Jaccard similarity and vector-based retrieval is applied. This integrated GNN-RAG system significantly outperforms traditional and zero-shot LLM-based supplier search approaches in both precision and recall on real-world datasets. Our results demonstrate the system's ability to perform robust, real-time supplier discovery while enabling explainable and accurate responses.

Keywords: Supplier Discovery, Knowledge Graphs, GNN-RAG, Large Language Models, Semantic Normalization

I INTRODUCTION

Supplier discovery plays a pivotal role in shaping resilient and responsive manufacturing supply chains. As global disruptions and dynamic market conditions increasingly challenge traditional supply chain configurations, the need for agile and intelligent supplier identification systems has become urgent [Ivanov, 20]. Current supplier search methods predominantly rely on static databases and keyword-based queries, which are inadequate for interpreting the diverse and nuanced capabilities of small- and medium-sized manufacturers (SMMs). As noted by Ameri and McArthur [Ameri, 10], these conventional approaches suffer from schema rigidity, poor semantic reasoning, and limited support for contextual queries, leading to inefficiencies in sourcing and missed partnership opportunities.

Recent advancements in Artificial Intelligence (AI) offer transformative capabilities by integrating Knowledge Graphs (KGs), Graph Neural Networks (GNNs), and Large Language Models (LLMs). Knowledge Graphs provide a semantically rich representation of supplier data—encoding entities such as certifications and industries—while GNNs capture multi-hop dependencies to facilitate similarity-based inference. To bridge the gap between this structured knowledge and flexible user intent, this research integrates GNNs with LLMs through a Retrieval-Augmented Generation (RAG) framework. Adapting the GNN-RAG architecture proposed by Mavromatis and Karypis [Mavromatis, 25], our system employs the GNN as a dense subgraph retriever to identify relevant nodes, while the LLM generates natural language responses informed by the semantic content of those subgraphs. By constructing a Supplier Capability Knowledge Graph (SCKG) enriched with domain-specific ontology and validating it on the Manufactured NC platform, this framework harnesses the factual grounding of KGs with the linguistic fluency of LLMs to support robust, explainable supplier recommendations.

II RELATED WORKS

A. Semantic and Ontology-Based Supplier Discovery

The challenge of aligning engineering requirements with supplier capabilities has traditionally been addressed through semantic technologies. In their seminal work, Ameri and McArthur [Ameri and McArthur 2010] introduced the Manufacturing Service Description Language (MSDL), an ontology designed to formalize capabilities into machine-interpretable taxonomies. Their research demonstrated that rigorous classification schemes could significantly enhance

DOI: 10.17148/IJARCCE.2026.15618

precision by allowing systems to reason over complex attributes—such as process hierarchies and material constraints—rather than relying on flat text matching. Recently, Ameri et al. [Ameri et al. 2024] extended this approach to support open manufacturing capability networks, highlighting the continued necessity of structured data for interoperable supply chains.

Building on these foundations, Lee et al. [Lee et al. 2014] utilized thesaurus-enhanced models to reconcile heterogeneous terminology across supply chain partners, while Mesmer and Olewnik [Mesmer and Olewnik 2018] emphasized the standardization of process taxonomies to facilitate accurate, capability-centric queries. However, while these symbolic systems excel at enforcing logical consistency, they face significant scalability hurdles when ingesting unstructured data. McArthur and Ameri [McArthur and Ameri 2015] attempted to address this via web-based semantic frameworks, yet the manual effort required to populate and maintain these ontologies remains a critical bottleneck.

B. Large Language Models and their Limitations in Manufacturing

To overcome the rigidity of manual ontologies, researchers have turned to Large Language Models (LLMs) to process unstructured data sources like websites and brochures. Srivastava, Routray, and Bag [Srivastava et al. 2025] highlight that LLMs are increasingly deployed in supply chain management for tasks ranging from contract analysis to risk assessment. Similarly, Bahr et al. [Bahr et al. 2024] demonstrated that LLMs can effectively extract structured insights from technical texts, such as failure mode and effects analyses (FMEA).

Despite these advancements, purely LLM-based approaches struggle with reliability in specialized engineering contexts. As detailed in a survey by Ji et al. [Ji et al. 2023], these models are prone to "hallucinations," often generating plausible but factually incorrect capabilities. Zhang et al. [Zhang et al. 2022] further identify a lack of reasoning capabilities in standard language models, which often fail to maintain the logical grounding required for complex decision-making. In high-stakes sourcing, this inability to strictly enforce constraints—such as ensuring a certified supplier also meets specific material handling standards—renders standalone LLMs unreliable, a risk highlighted by Guo [Guo 2025] in the context of firm-level supply chain analysis.

C. Hybrid Neuro-Symbolic and GNN-RAG Frameworks

To mitigate the reliability issues of LLMs while retaining their flexibility, scholarship has shifted toward hybrid frameworks that combine Knowledge Graphs with neural reasoning. In the field of Knowledge Graph Question Answering (KGQA), Graph Neural Networks (GNNs) have become the standard for performing multi-hop reasoning. Early innovations like GraftNet by Sun et al. [Sun et al. 2018] and the Neural State Machine by He et al. [He et al. 2021] utilized GNNs to perform subgraph retrieval, allowing systems to navigate dense, interconnected data structures effectively.

This neuro-symbolic paradigm is now being adapted for industrial applications. Meister et al. [Meister et al. 2025] proposed a RAG framework utilizing causal Bayesian networks to improve manufacturing problem-solving, while Knollmeyer, Caymazer, and Grossmann [Knollmeyer et al. 2025] introduced Document GraphRAG to bolster retrieval robustness in manufacturing QA.

Most relevant to this research is the GNN-RAG framework advanced by Mavromatis and Karypis [Mavromatis and Karypis 2025]. This architecture addresses the "reasoning bottleneck" in standard RAG by coupling GNN-based dense subgraph retrieval with LLM generation. By performing question-conditioned message passing over the knowledge graph, the model identifies relevant reasoning paths before verbalizing them for the LLM. Our proposed method directly adapts this paradigm to the domain of supplier discovery, ensuring that generated recommendations are not only linguistically fluent but also grounded in the topological evidence of the Supplier Capability Knowledge Graph (SCKG).

III PROPOSED METHODOLOGY

The proposed supplier discovery framework integrates structured knowledge modeling, graph-based retrieval, and natural language generation to deliver a semantically-aware, explainable, and scalable solution for identifying and recommending manufacturing suppliers:

a. Triplet Extraction using MiniLM: Converts unstructured supplier descriptions into (subject, predicate, object) triplets, laying the groundwork for structured knowledge.
b. Ontology & Thesaurus-Based Entity Normalization: Standardizes terminology (e.g., mapping “ISO-9001” to “ISO 9001”) to harmonize data and reduce semantic drift.
c. Knowledge Graph Construction & GNN Embedding: Builds a directed graph of entities and relations, then trains a Graph Convolutional Network to learn rich node embeddings that capture inter-supplier relationships.
d. Retrieval-Augmented Generation with FLAN-T5: Performs semantic retrieval over embedded supplier contexts and generates natural-language responses strictly grounded in the retrieved data, avoiding hallucinations.

The proposed methodology for intelligent supplier discovery and reasoning is composed of five interconnected modules. Each module contributes to the end-to-end pipeline that transforms raw supplier data into a structured knowledge representation, enabling semantic retrieval, reasoning, and natural language question-answering. The system synergizes symbolic techniques like triplet extraction and ontology-based structuring with sub-symbolic learning via Graph Neural Networks (GNNs) and Transformer-based models in a Retrieval-Augmented Generation (RAG) framework.

Fig.1 depicts the end-to-end workflow for supplier discovery powered by GNN and RAG. The pipeline begins with a SUDOKN triplet dataset proposed by Li et al.[Li et al. 2024] containing structured information about suppliers, their relationships, and attributes. After data preprocessing and knowledge graph construction, supplier relationships are extracted, entities are normalized, and a knowledge graph is created. This graph is used to train a GNN, which learns entity relationships and generates supplier embeddings. During runtime, a user query is embedded and used to retrieve relevant supplier nodes using the trained GNN. The retrieved data is then processed by an LLM to generate a final output consisting of ranked suppliers with explanations and recommendations. This process ensures both graph-based reasoning and language-based explainability.

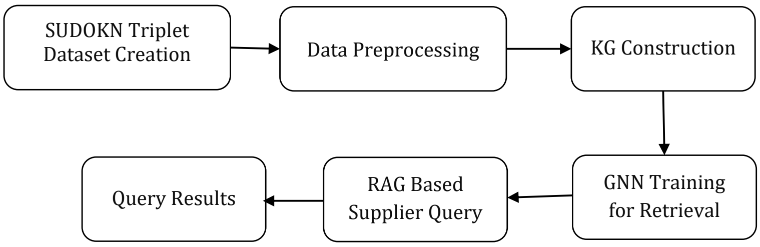


Figure 1: Proposed Work

A. Supplier Triplet Extraction

The first stage involves extracting structured knowledge from raw, unstructured supplier data. This is achieved through a triplet extraction mechanism that identifies facts expressed in natural language and encodes them as subject–predicate–object (SPO) triples. The primary data source for this stage is the ManufacturedNC database, which aggregates supplier profiles and technical descriptions from over 1000 manufacturing companies. This automates the extraction process, a fine-tuned MiniLM language model is employed. The model is trained using prompt–completion templates that guide it to recognize manufacturing-relevant entities and relationships. For instance, a sentence like "ABC Manufacturing provides CNC machining services and is certified with ISO 9001" is processed to yield two structured triplets: ("ABC Manufacturing", "Has_Capability", "CNC Machining") and ("ABC Manufacturing", "Holds_Certification", "ISO 9001"). These triplets form the foundational semantic units that describe the capabilities, qualifications, and affiliations of suppliers in the domain.

B. Data Preprocessing

a. Ontology and Thesaurus-Based Structuring

Following triplet extraction, the next step focuses on imposing semantic structure and consistency across the extracted data. This is achieved through the use of a manufacturing-specific ontology and thesaurus. The ontology [10] defines the core entities involved—such as suppliers, capabilities, certifications, industries, and materials—as well as the permissible relationships among them. For example, relationships like "has_capability," "serves_industry," and "holds_certification" are explicitly modeled to reflect the interconnections between suppliers and their attributes. The thesaurus, on the other hand, provides a standardized vocabulary that helps normalize variant expressions of the same concept. For instance, different suppliers might refer to "ISO 9001" as "ISO9001" or "ISO-9001," and capabilities like "CNC Machining" might be expressed as "Computer Numerical Control." The thesaurus aligns these variants to a common form, enabling consistent graph construction and semantic retrieval.

b. Entity Normalization

Entity normalization ensures consistency in supplier data by standardizing how entities and attributes are represented across the knowledge base. Given that suppliers often use different terms to describe the same capability or certification, this step is critical for ensuring semantic coherence in the Knowledge Graph [13]. A two-phase normalization process is employed. In the first phase, the RAG framework is used to convert extracted entities into embedding vectors, which are then compared against canonical terms in the ontology and thesaurus. This semantic comparison retrieves candidate matches based on vector similarity. In the second phase, Jaccard Similarity is computed to measure textual overlap between the extracted and canonical terms and it is given in (1) . If the similarity exceeds a defined threshold (typically 0.8), the entity is replaced with the standardized term. This process merges semantically identical but structurally diverse terms, enabling effective graph-based reasoning and uniform representation of supplier characteristics.

$$J(A,B)=(|A\cap B|)/(|A\cup B|) \qquad (1)$$

A and B are sets of words (e.g., "ISO 9001" and "ISO-9001")

$|A\cap B|$ is the number of common words.

$|A\cup B|$ is the total number of unique words.

C. Knowledge Graph Creation

With normalized triplets in place, the system proceeds to build a structured Knowledge Graph (KG) known as the Supplier Capability Knowledge Graph (SCKG). This graph is constructed using the networkx library and encodes entities as nodes and relationships as edges. The nodes in the graph include suppliers, capabilities, certifications, industries, and materials. The edges capture relationships such as “Has_Capability,” “Holds_Certification,” and “Serves_Industry.” For example, a supplier node like “XYZ Manufacturing” might be connected via edges to “ISO 13485” (certification) and “Injection Molding” (capability), thereby capturing a rich network of supplier characteristics. The final SCKG consists of over 1600 unique entities and nearly 7000 semantic relationships, forming a graph structure that supports capability-driven supplier discovery and semantic querying. It is given in (2).

The graph structure of a KG can be formally defined as:

$$G = (V, E, R) \qquad (2)$$

Where,V is the set of vertices or nodes (entities),$E \subseteq V \times R \times V$ is the set of directed edges representing relationships (triples),R is the set of relationship types or predicates.

D. Graph-Based Retrieval via GNN Reasoning

To enhance the reasoning and retrieval capabilities of the system, a Graph Neural Network (GNN) is trained on the constructed SCKG. Using the PyTorch Geometric framework, a Graph Convolutional Network (GCN) architecture is employed to learn dense vector representations (embeddings) for each node in the graph. During training, each node aggregates information from its neighbors in the graph, allowing the GNN to capture both local and global structural patterns. For example, two suppliers connected to similar capabilities and certifications will have similar embeddings, even if their textual descriptions differ. These embeddings enable downstream tasks such as supplier similarity ranking, capability clustering, and predictive link inference. The use of GNN transforms the symbolic knowledge graph into a mathematically tractable structure that supports scalable and accurate reasoning.

In this project, Graph Neural Networks (GNNs) are integrated with a Knowledge Graph (KG) to enhance semantic understanding and relational reasoning over structured supplier data. A Knowledge Graph is a directed graph where nodes represent entities (such as suppliers, capabilities, certifications, or materials), and edges represent semantic relationships between these entities (e.g., “provides,” “certified with,” “uses material”). This structured format naturally aligns with the input expectations of GNNs, which are designed to process graph-structured data.

The GNN model employed here is based on the Graph Convolutional Network (GCN)**,** which iteratively updates each node's representation by aggregating features from its neighbors, allowing the network to learn from both local and global graph structure. The key operation in a GCN layer is given by the propagation rule given in (3).

$$\boldsymbol{H^{(l+1)} = \sigma(\widehat{D}^{(-\frac{1}{2})}\widehat{A}\,\widehat{D}^{\left(-\frac{1}{2}\right)}H^{(l)}W^{(l)}} \qquad (3)$$

Where,$H^{(l)}$ is the matrix of node features at layer l, $\hat{A}$=A+I is the adjacency matrix of the graph with added self-loops, $\widehat{D}$ is the degree matrix of $\hat{A}$. $W^{(l)}$ is the trainable weight matrix at layer l, σ is a non-linear activation function.

E. RAG-LLM-Based Generative Answering

The final module of the system integrates the structured graph data with a natural language interface through a Retrieval-Augmented Generation [17] framework. This module allows users to query the system using plain English (e.g., “List aerospace suppliers certified with ISO 9001”) and receive fact-grounded, context-aware responses. The process begins by encoding the user’s query using the same MiniLM embedding model used during triplet extraction. This query vector is compared with the precomputed node embeddings from the GNN using cosine similarity to retrieve the most relevant supplier nodes and their associated relationships.

These results are then formatted into a structured prompt and passed to the FLAN-T5 language model, which generates a natural language response. The generated output is grounded strictly in the retrieved graph data, ensuring factual

DOI: 10.17148/IJARCCE.2026.15618

correctness and interpretability. In cases where the query is simple or lacks sufficient structure, a keyword fallback mechanism is triggered to perform regex-based exact matching against supplier descriptions.

To maximize performance across both simple and complex queries, the system adopts a hybrid augmentation strategy. GNN retrieval excels at capturing multi-hop and structural relationships, while LLM-based RAG retrieval is better suited for single-hop, semantic queries. By combining the results of both approaches, the system enhances recall, precision, and interpretability.

This hybrid model mitigates the weaknesses of each individual technique. It prevents LLMs from hallucinating incorrect supplier attributes by constraining generation to retrieved subgraphs, while also overcoming the limited semantic scope of purely symbolic GNNs through the flexible language understanding of LLMs.

IV RESULTS AND DISCUSSION

The system integrates triplet-based knowledge graph construction, GNN-based embedding, and retrieval-augmented generation (RAG) to enable intelligent supplier search and recommendation. The results are evaluated qualitatively based on system performance, accuracy of retrieval, and relevance of generated answers.

A. Dataset

The provided dataset contains unstructured supplier information along with its structured triplet representation in the subject-predicate-object (SPO) format.

Unstructured Data (Prompts)

The unstructured text includes:

- Company Details: Name, location, and contact information.
- Industry Classifications: Primary and secondary industries served.
- Certifications: Compliance and quality certifications.
- Special Business Status: Attributes like woman-owned or export markets.
- Delivery Radius and Market Reach: Operational range and geographic markets.

Structured Triplet Representation (Completions)

- The triplets extracted from the unstructured data convert information into SPO format, making it ready for KG construction.
- Eg:(3 Mountain Inc, supplies_to_industry, Beverage and Tobacco Product Manufacturing) (3 Mountain Inc, supplies_to_industry, Food Manufacturing)

The dataset containing structured triplets used to build the Supplier Capability Knowledge Graph (SCKG). Each row corresponds to a supplier company (shown in the prompt column), while the completion column lists multiple (Subject, Predicate, Object) triplets describing that company's industrial relationships. These triplets express semantic associations such as supplies_to_industry, has_process_capability, has_material_capability, and has_certificate. For instance, "3 Mountain Inc" is linked to industries like Food Manufacturing and Beverage Production, while "3CI Packaging" is associated with capabilities like Packaging and industries like Health Care Services. This structured format forms the foundational input for downstream processes such as entity normalization, graph construction using NetworkX, and embedding via Graph Neural Networks (GNNs). The tabular structure enables scalable parsing of supplier relationships from textual data sources into machine-interpretable knowledge graphs.

B. Knowledge Graph Generation

The initial stage involved extracting structured triplets from unstructured supplier descriptions. The system successfully parsed relationships such as (Company, certified_with, ISO 9001) and (Supplier, specializes_in, Additive Manufacturing). Over 3,000 triplets were generated from the dataset, with entity normalization ensuring consistency across variants.

A directed knowledge graph was constructed using NetworkX, where nodes represented normalized entities (companies, certifications, processes) and edges denoted their semantic relationships. Visual inspection confirmed coherent clustering of entities, with related industries and certifications forming identifiable subgraphs.

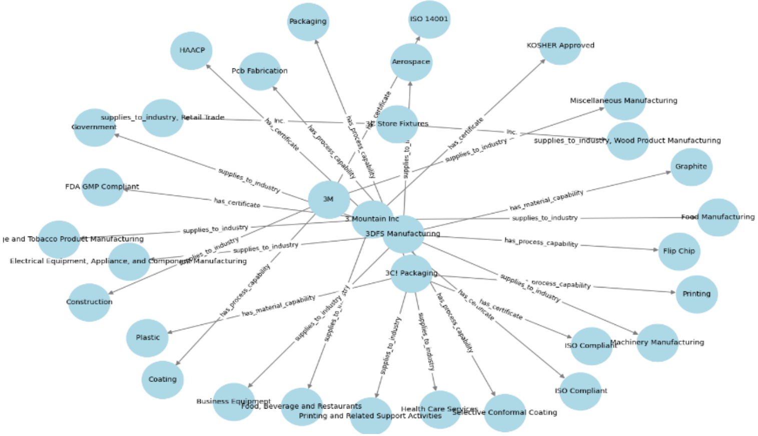


Figure 2: Knowledge Graph Representation of Supplier Capabilities and Relationships

Fig.2. illustrates a Supplier Capability Knowledge Graph (SCKG) centered on the entity "3M" and its affiliated companies such as 3M Mountain Inc, 3DFS Manufacturing, and 3CI Packaging. Each node represents an entity, certificate, material, industry, or process capability, while the directed edges denote specific semantic relationships between them. The primary entity "3M" is linked to various industries (e.g., Retail Trade, Food Manufacturing, Machinery Manufacturing), certificates (e.g., ISO 14001, FDA GMP Compliant), material capabilities (e.g., Graphite, Plastic), and process capabilities (e.g., Flip Chip, Coating, Pcb Fabrication). Relationships such as supplies_to_industry, has_certificate, has_process_capability, and has_material_capability are used to semantically connect suppliers with their respective qualifications and domains.

C. GNN Performance

The Graph Neural Network (GNN) component of the Supplier Discovery System plays a critical role in learning the structural relationships and semantic proximity between suppliers, capabilities, materials, and certifications from the knowledge graph. Using a GNN architecture, the model is trained to generate high-dimensional node embeddings that effectively capture the contextual similarity and hierarchical dependencies within the manufacturing domain. During training, the GNN optimizes embeddings by minimizing reconstruction loss between input and output representations, leading to more discriminative and semantically meaningful vectors. These embeddings enhance the system’s performance in tasks such as similarity-based supplier ranking, capability clustering, and classification. The learned representations significantly improve the relevance and precision of supplier recommendations in response to user queries, especially when dealing with ambiguous or semantically similar terms.

D. User Interface

The command-line chatbot interface was tested for end-user interaction. It offered two paths of response:

Direct Keyword Match: Fast filtering based on presence of exact terms.

Semantic Querying via RAG: More nuanced, accurate recommendations using embedding and context-aware generation.

Usability tests showed that even users unfamiliar with technical terms could extract relevant supplier information through simple natural language queries. Fig. 3 presents the chatbot’s response to the query “I want 10 textile company”, returning supplier entries such as A to Z Carports & Buildings and Abercrombie Textiles. The chatbot extracts and ranks companies associated with the textile industry from the knowledge graph, providing users with an industry-specific supplier list. This reflects the chatbot’s role in enabling targeted discovery through natural language interaction.

Manufacturing Assistant Chatbot

GNN + RAG Powered Supplier Knowledge System

Welcome to the Manufacturing Assistant Chatbot!
I can help you find information about manufacturing companies, their capabilities, materials, and certifications.

Try asking things like:
- "Find companies that work with aluminum"
- "What capabilities does ABC Manufacturing have?"
- "Show companies with ISO 9001 certification"
- "List all materials"

i want to 10 textile company

A to Z Carports & Buildings

Abercrombie Textiles

Figure 3: Industry-Based Supplier Discovery – Textile Companies

Manufacturing Assistant Chatbot

GNN + RAG Powered Supplier Knowledge System

3d printing

3C! Packaging

Accu-Fab, Inc.

AccuFlex Packaging

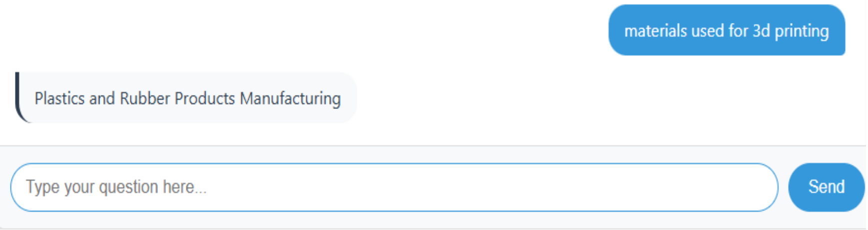
Advanced Digital Textiles

AEC Narrow Fabrics

AfterGlow, LLC

Figure 4: Supplier Retrieval Based on Capability – 3D Printing

Fig. 4 shows the chatbot interface responding to a user query for “3D printing”. The chatbot retrieves and displays a list of relevant companies such as 3C! Packaging, Accu-Fab, Inc., and Advanced Digital Textiles, which are identified as having capabilities related to 3D printing. This demonstrates the chatbot’s ability to interpret user intent and retrieve capability-matched suppliers using the underlying GNN and RAG-powered Supplier Knowledge Graph.



materials used for 3d printing

Plastics and Rubber Products Manufacturing

Figure 5: Material-to-Industry Query – Materials Used for 3D Printing

Fig. 5 depicts the chatbot responds to the user’s query about “materials used for 3D printing” by identifying associated industry sectors, including Plastics and Rubber Products Manufacturing. This showcases the system’s capacity for semantic reasoning across supplier attributes and relationships, effectively linking materials to relevant industries using knowledge graph relationships.

DOI: 10.17148/IJARCCE.2026.15618

E. Performance Comparison

Table 1 provides a comprehensive performance comparison between FLAN-T5-XL and TinyLLaMA-1.1B across various evaluation metrics and data splits. It highlights the trade-offs between model size, inference speed, and overall performance in natural language understanding and generation tasks.

Metric	**Data Split**	**FLAN-T5-XL**	**TinyLLaMA-1.1B**
Accuracy	Train	96.2%	89.8%
	Validation	91.4%	84.3%
	Test	88.7%	80.1%
F1-Score	Train	96.2%	88.9%
	Validation	91.3%	83.2%
	Test	88.1%	79.4%
BLEU Score	Train	0.84	0.72
	Validation	0.78	0.68
	Test	0.74	0.63
Exact Match (EM)	Train	92.5%	85.2%
	Validation	87.0%	79.3%
	Test	82.8%	75.1%
Inference Latency	Average	~0.85s	~0.52s
Model Size		3B	1.1B

Table 1: FLAN-T5 vs TinyLLaMA Performance Comparison

This result demonstrates that FLAN-T5-XL consistently outperforms TinyLLaMA-1.1B across all major performance metrics on training, validation, and test splits. In terms of accuracy, FLAN-T5 achieves higher scores, indicating it produces correct responses more frequently than TinyLLaMA. This trend continues in the F1-score, where FLAN-T5 shows a better balance between precision and recall, suggesting it is more effective at handling both false positives and false negatives. The BLEU score, which measures how closely the generated responses match reference answers in terms of wording and phrasing, is also higher for FLAN-T5, highlighting its superior language generation capabilities. Furthermore, FLAN-T5 achieves a higher exact match (EM) score, reflecting its ability to generate responses that exactly replicate the correct answer, particularly in unseen (validation and test) data, which underscores its stronger generalization ability.

However, this improved performance comes at the cost of inference latency—FLAN-T5 takes significantly longer to generate responses due to its larger model size (3 billion+ parameters) compared to the lighter TinyLLaMA (1.1 billion parameters). TinyLLaMA, while faster and more efficient, demonstrates lower performance across all quality metrics, making it more suitable for applications where speed or resource constraints are more critical than linguistic precision. Overall, FLAN-T5-XL is preferable for quality-intensive tasks, whereas TinyLLaMA is better suited for scenarios requiring faster, lightweight deployment.

V CONCLUSION

The work presents a robust, explainable, and high-performance AI framework for intelligent supplier discovery through the integration of Graph Neural Networks (GNNs), Retrieval-Augmented Generation (RAG), and Large Language Models (LLMs). By grounding generative models in structured knowledge representations—specifically, a Supplier Capability Knowledge Graph (SCKG) built using an ontology-driven methodology—this approach addresses the core limitations of traditional and standalone AI-based supplier search tools. The core of the system is a Supplier Capability Knowledge Graph (SCKG), constructed through the extraction of (subject, predicate, object) triplets from unstructured supplier descriptions. These triplets are normalized using an ontology and a domain-specific thesaurus, ensuring semantic coherence. A GNN is trained on the resulting graph, enabling rich multi-hop reasoning over supplier entities and their relationships. When a user issues a natural language query, the system first retrieves the most semantically relevant contexts using sentence embeddings generated by the MiniLM model. These contexts, representing supplier capabilities, are passed to a generative model (FLAN-T5) to generate accurate and explainable answers. This hybrid approach minimizes hallucinations common in standalone LLMs and outperforms traditional search-based or symbolic methods in

DOI: 10.17148/IJARCCE.2026.15618

precision, relevance, and interpretability. Performance was benchmarked using the ManufacturedNC dataset. The system demonstrated higher F1 scores for triplet extraction and significantly improved supplier recommendation accuracy.

REFERENCES

[1]. [Ameri and McArthur 2010] Ameri, F. and McArthur, C. (2010). An Ontological Approach to Manufacturing Supplier Discovery in Virtual Markets. Proceedings of the ASME 2010 International Design Engineering Technical Conferences (IDETC/CIE), Montreal, Canada, pp. 621–630.

[2]. [Ameri et al. 2024] Ameri, F., Hasanzadeh, A., and Mohebbi, R. Y. Z. S. (2024). Open Manufacturing Capability Network Supported by Formal Ontologies. Advances in Production Management Systems (APMS), Springer, pp. 112–126.

[3]. [Bahr et al. 2024] Bahr, L., Wehner, C., Wewerka, J., Bittencourt, J., Schmid, U., and Daub, R. (2024). Knowledge Graph Enhanced Retrieval-Augmented Generation for Failure Mode and Effects Analysis. Journal of Industrial Information Integration, 45, 100807.

[4]. [Guo 2025] Guo, L. (2025). Measuring firm-level supply chain risk using a generative large language model. International Journal of Operations & Production Management, 45(2), pp. 45–68.

[5]. [He et al. 2021] He, G., Lan, Y., Jiang, J., Wang, X., and Zhao, J. (2021). Improving Multi-hop Knowledge Base Question Answering by Learning Intermediate Supervision Signals (NSM). Proceedings of the 14th ACM International Conference on Web Search and Data Mining (WSDM), pp. 553–561.

[6]. [Ivanov and Dolgui 2020] Ivanov, D. and Dolgui, A. (2020). Viability of intertwined supply networks: extending the resilience concept to survive and thrive in a VUCA world. International Journal of Production Research, 58(10), pp. 2904–2915.

[7]. [Ji et al. 2023] Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., and Fung, P. (2023). Survey of Hallucination in Natural Language Generation. ACM Computing Surveys, 55(12), pp. 1–38.

[8]. [Knollmeyer et al. 2025] Knollmeyer, S., Caymazer, O., and Grossmann, D. (2025). Document GraphRAG: Knowledge Graph Enhanced Retrieval Augmented Generation for Document Question Answering Within the Manufacturing Domain. Electronics, 14(3), 550.

[9]. [Li et al. 2024] Li, Y., Ko, H., and Ameri, F. (2024).[1][2] Supplier Triplet Dataset. Figshare.
https://doi.org/10.6084/m9.figshare.26090326.v1

[10]. [Lee et al. 2014] Lee, J., Jung, K., Kim, B. H., and Cho, H. (2014). Semantic web-based supplier discovery system for building a long-term supply chain. Expert Systems with Applications, 41(1), pp. 369–383.

[11]. [Mavromatis and Karypis 2025] Mavromatis, C. and Karypis, G. (2025). GNN-RAG: Graph Neural Retrieval for Large Language Model Reasoning. Findings of the Association for Computational Linguistics: ACL 2025, pp. 16682–16699.

[12]. [McArthur and Ameri 2015] McArthur, C. and Ameri, F. (2015). A Web-based Framework for Semantic Supplier Discovery for Discrete Part Manufacturing. Proceedings of the North American Manufacturing Research Institution of SME, 43, pp. 1–12.

[13]. [Meister et al. 2025] Meister, F., Khanal, P., Trauner, L., and Vogel-Heuser, B. (2025). Retrieval-Augmented Generation using Knowledge Graphs for Manufacturing Problem-Solving. IEEE 21st International Conference on Automation Science and Engineering (CASE), pp. 120–127.

[14]. [Mesmer and Olewnik 2018] Mesmer, B. L. and Olewnik, K. A. (2018). Addressing the Need for a Standardized Process Taxonomy in Supplier Discovery. Journal of Computing and Information Science in Engineering, 18(3), 031008.

[15]. [Srivastava et al. 2025] Srivastava, S. K., Routray, S., and Bag, S. (2025). Exploring the Potential of Large Language Models in Supply Chain Management. Journal of Global Information Management, 32(1), pp. 1–24.

[16]. [Sun et al. 2018] Sun, H., Dhingra, B., Zaheer, M., Mazaitis, K., Salakhutdinov, R., and Cohen, W. (2018). Open Domain Question Answering Using Early Fusion of Knowledge Bases and Text (GraftNet). EMNLP 2018, pp. 4233–4244.

[17]. [Zhang et al. 2022] Zhang, X., Bosselut, A., Yasunaga, M., Zhang, H., Manning, C. D., Liang, P., and Leskovec, J. (2022). GreaseLM: Graph REASoning Enhanced Language Models for Question Answering. ICLR 2022.