AI-Based Supply Chain Mapping System: A Prototype Development Roadmap

Introduction

The Indian Army faces critical challenges in maintaining operational readiness due to vulnerabilities inherent in its complex, multi-tier supply chains. Recent global conflicts and the COVID-19 pandemic have starkly highlighted how disruptions in the availability of essential equipment, components, and sub-components can severely impede military capabilities. This strategic imperative underscores the need for advanced solutions that can provide deep visibility into supply networks, identify points of failure, and enable proactive mitigation. The "AI Based Supply Chain Mapping System" initiative, outlined within the VII Edition Compendium of Problem Definition Statements (CPDS) 2025, represents a significant step by the Indian Army to leverage innovation from academia and industry to enhance defense capabilities and foster self-reliance.[1]

This report presents a comprehensive and actionable roadmap for developing a working prototype of an AI-based supply chain mapping system. The proposed solution is designed to meet the Indian Army's specific requirements, focusing on core functionalities achievable within a challenging one-month timeframe. The approach prioritizes the use of open-source information, leverages cutting-edge Artificial Intelligence and Machine Learning techniques, and outlines pragmatic strategies for data collection, solution formulation, and prototype development. This document aims to provide a structured guide, enabling the successful delivery of a submittable and impactful solution.

I. Deep Analysis of the Problem Statement & Submission Criteria

Problem Genesis and Importance: Why Multi-Tier Supply Chain Mapping is Critical

The core problem articulated by the Indian Army stems from the inherent fragility of its multi-tier supply chains. The problem statement explicitly notes that "Recent wars and the pandemic have shown that disruption of supply chains can hamper the availability of equipment, components and sub-components, thus affecting the readiness levels". This directly links supply chain resilience to national security, highlighting how logistical vulnerabilities can translate into operational deficiencies. The Indian Army procures a vast array of items, including complex equipment and spares, which are assembled from components sourced from numerous vendors across multiple tiers. This intricate web of dependencies means that the "surge capability for production... will be dependent on the capability of the least denominator amongst the suppliers". This observation points to a critical systemic risk: a single point of failure deep within the supply chain, potentially at the second, third, or even fourth tier, can have cascading and disproportionate impacts on the overall readiness and operational continuity.

The profound significance of this problem extends beyond mere logistical efficiency; it is a strategic imperative for national security. The capacity to maintain "readiness levels" during "a state of heightened conflict" necessitates a shift from reactive problem-solving to proactive vulnerability identification and predictive intelligence. The objective is not simply to know who the suppliers are, but to anticipate *when* and *how* disruptions might occur and to have pre-planned mitigation strategies. This aligns with approaches taken by other defense organizations, such as the US DoD's Defense Logistics Agency (DLA), which leverages AI for supply chain risk management to predict bottlenecks, forecast demand, and recommend alternative suppliers.[2] Furthermore, the aerospace and defense sectors globally operate within highly complex and regulated supply chains, facing unique challenges like factory fires, labor disruptions, and business sales, which AI-powered solutions are increasingly used to monitor and mitigate.[3] The Army's initiative, therefore, represents a forward-thinking move to address these challenges and ensure a robust defense posture.

Key Deliverables Breakdown: Detailed Interpretation of Each Requirement (a-e)

The problem statement meticulously outlines five key deliverables, each critical for the functionality and utility of the proposed AI-based supply chain mapping system:

- **(a) Map the multi-tiered supply chain (second, third, and fourth tier suppliers) using open-source information, such as financial transaction record details.**
  - This deliverable forms the foundational mapping capability. The emphasis on "open source information" is a crucial constraint, particularly concerning "financial transaction record details." While commercial solutions like Altana.ai and Resilinc utilize vast proprietary data networks and supplier-validated information for multi-tier visibility [4], the directive for open-source data suggests that direct, granular financial transaction data for private entities will likely not be readily available in bulk.[6] Consequently, the approach must pivot towards inferring relationships and financial health from publicly available, less direct sources. This will involve extensive use of Natural Language Processing (NLP) to extract information from unstructured public documents, news articles, and public company registries.[9] The system's ability to identify and link companies across multiple tiers will be a core demonstration of its mapping prowess.
- **(b) Sense risks across the supply chain based on the identification of ownership up to 1% holdings, including the Board of Directors, State Owned Enterprises (SOEs), and country of origin.**
  - This requirement demands a deep dive into corporate structures and geopolitical affiliations. Identifying ownership down to a 1% holding is an exceptionally granular level of detail, which is typically challenging to obtain from open sources for all entities. This necessitates sophisticated data extraction and inference techniques, likely involving NLP to parse public documents and cross-reference information about individuals and entities.[10] The inclusion of "Board of Directors, State Owned Enterprises, and country of origin" points directly to the need for assessing control points and geopolitical vulnerabilities, which is paramount in a defense context. AI models for

dynamic risk scoring, capable of integrating diverse data points, are directly applicable here.[12]

- **(c) Suggest alternate suppliers of components and sub-components based on their function and role.**
  - This deliverable focuses on the prescriptive aspect of the AI system – providing actionable mitigation strategies. It requires a robust recommendation system that can identify suitable alternative suppliers not just by part number, but by understanding their core capabilities, functional expertise, and role within the broader supply chain ecosystem. AI-powered supplier recommendation systems are a well-established field, capable of matching criteria and surfacing relevant data to suggest new or alternative suppliers.[14] The system will need to leverage a rich, attribute-based representation of suppliers and components within its knowledge base to make intelligent, function- and role-based recommendations.
- **(d) The algorithm should be able to furnish data as desired above with effect from a mutually agreed time stamp.**
  - This implies a historical data retention and querying capability. The system must be able to provide a snapshot of the supply chain's state, vulnerabilities, and recommendations at a specific point in time. While real-time updates are ideal for a production system, for a prototype within a one-month timeframe, this could involve demonstrating the ability to process data up to a certain historical point or simulating data changes over time to show temporal analysis capabilities. The underlying data model, particularly a knowledge graph, would need to support versioning or temporal queries to fulfill this requirement.
- **(e) The software/algorithm developed should be able to work on a standalone mode, updated in offline mode, and be upgradable with latest updates from the developer.**
  - This is a critical operational and deployment constraint for a defense application. It mandates that the solution must be self-contained and function without continuous internet access for its core operations after initial setup. This significantly impacts architectural choices, favoring local processing and pre-trained models. The ability to update in "offline mode" implies a mechanism for periodic data synchronization or model redeployment via local file transfer or a secure, pre-packaged update bundle. The "upgradable" aspect suggests a modular design that allows for future enhancements and maintenance.

Submission Guidelines & Evaluation Criteria: Navigating CPDS 2025 Requirements, TRL 3-7 Implications, and Selection Process

Understanding the CPDS 2025 guidelines and evaluation criteria is paramount for tailoring the prototype and the accompanying proposal to maximize its chances of selection.

- **Eligibility and Classification:** As engineering students, the proposers fall under "Academic Institutions," making them explicitly eligible to submit solutions. A crucial requirement is that proposals must be "unclassified in nature" and include an undertaking for sharing with technical experts for evaluation, with confidentiality assured by the Army Design Bureau (ADB). This means the prototype must strictly rely on publicly available, unclassified data, and no sensitive military information can be used or assumed.

- **Technology Readiness Level (TRL):** Solutions are expected to range from TRL 3 to TRL 7.
  - **TRL 3 (Proof-of-Concept):** This level signifies that an analytical and experimental critical function or characteristic proof-of-concept has been established, often involving the construction of a proof-of-concept model.[18]
  - **TRL 6 (Developed Prototype):** At this stage, a fully functional prototype or representational model has been demonstrated in a relevant environment.[18]
  - **TRL 7 (Operational Environment):** This is the highest level sought, where a prototype is demonstrated in an operational environment.[18]
  - Given the one-month timeline and the "from scratch" development, achieving TRL 7 (operational environment demonstration) is highly improbable. The most pragmatic and achievable target for a "working prototype" within this timeframe is a robust TRL 3, demonstrating the core AI/ML logic, with a clear pathway towards TRL 6. This involves showcasing the methodology and feasibility of mapping, vulnerability detection, and recommendation using a curated, representative dataset, rather than attempting comprehensive real-world data integration at scale. The emphasis will be on the algorithmic and architectural soundness.
- **Proposal Format and Content:** The submission requires a "comprehensive solution in the prescribed format", detailing "well defined objectives, scope, outcomes and quantifiable deliverables with specific time lines". This report's structure and detailed timeline are designed to address these requirements.
- **Evaluation Criteria:** Proposals will be assessed based on several key criteria:
  - **Relevance:** Direct alignment with the problem statement and operational needs.
  - **Innovation:** Novelty of the solution and use of advanced technologies.
  - **Feasibility:** Practicality and ease of implementation within the Army's environment, especially crucial given the tight development timeline.
  - **Scalability:** Potential for widespread adoption and future expansion.
  - **Cost-Effectiveness:** Budget viability, naturally aligned with open-source solutions.
  - **Development Timeline:** Realistic and achievable plan.
  - **Indigenous Content:** A stated preference for solutions maximizing indigenization. This criterion strongly favors solutions developed in India and leveraging Indian data sources. By focusing on publicly available Indian company data (e.g., from the Ministry of Corporate Affairs) and open-source AI/ML frameworks, the proposed solution can naturally align with this preference.
- **Selection Process:** The selection process involves initial screening by Domain Specialists and technical experts, followed by an expert panel review, and potentially a "Product Demo". This highlights the critical importance of a clear, demonstrable prototype and a compelling proposal that explicitly addresses all assessment criteria. The process also indicates a long-term vision, as successful proposals can lead to "Technical Consultancy" and selection as a "Developing Agency" for future R&D and procurement. This suggests that the prototype should be framed as a foundational

step towards a deployable, scalable, and sustainable solution.

II. Solution Formulation: Architectural Design & AI/ML Strategy

The proposed AI-based supply chain mapping system will adopt a modular architecture, centered around a robust Knowledge Graph (KG). This design choice is critical for representing the complex, multi-tiered relationships and ownership structures inherent in supply chains, while also enabling the integration of various AI/ML models for analysis and recommendation.

Core System Architecture: Proposing a Modular, Scalable Design

For rapid prototyping within a one-month timeframe, a layered, modular architecture is most appropriate. This approach allows for parallel development by the two students, simplifies debugging, and ensures clear interfaces between components, which is vital for future upgradability and scalability, as required by the problem statement. The system will consist of distinct, loosely coupled modules:

- **Data Ingestion & Pre-processing Module:** This module will be responsible for collecting raw data from various open-source channels, performing initial cleaning, normalization, and structuring.
- **Knowledge Graph Construction & Management Module:** This is the heart of the system, responsible for populating and maintaining the Knowledge Graph, which serves as the central repository for all supply chain and ownership information.
- **Vulnerability Assessment & Risk Scoring Module:** This module will apply advanced AI/ML models to the Knowledge Graph to identify potential risks and calculate risk scores for entities and linkages.
- **Alternate Supplier Recommendation Module:** This module will leverage the insights from the Knowledge Graph and AI models to suggest suitable alternative suppliers.
- **User Interface (UI) / Interaction Layer:** This front-end component will allow users to interact with the system, input queries, visualize the mapped supply chain, and view risk assessments and recommendations.

This modular design inherently supports the "standalone mode" and "offline mode" requirements, as each module can be developed and packaged for local execution, minimizing external dependencies during operation.

Knowledge Graph as the Foundation for Multi-Tier Mapping and Ownership

A Knowledge Graph (KG) is the optimal data structure for representing the intricate web of entities and relationships within a multi-tier supply chain. Its graph-based nature allows for intuitive modeling of complex connections that are difficult to manage efficiently in traditional relational databases.[20]

- **Representation:** In this KG, nodes will represent various entities such as companies, specific components, sub-components, manufacturing sites, geographical locations, individuals (e.g., Board of Directors), and countries. Edges will define the relationships between these entities, such as "supplies," "manufactures," "owned_by," "located_in," "board_member_of," "procured_by," or "is_a_component_of." Attributes can be associated with both nodes and edges to store properties like company financial health, component function, supplier role, risk scores, or country of origin.

- **Multi-Tier Mapping:** The KG will explicitly model the relationships between Tier 1, 2, 3, and 4 suppliers and their respective components/sub-components, directly addressing Deliverable (a). This is achieved by creating interconnected nodes and edges that trace the flow of materials and components through successive layers of the supply chain. Leading commercial solutions like Altana.ai also build their core platform on a "shared knowledge graph" to provide a "dynamic view of 530 million companies and their ownership networks".[4]
- **Ownership Mapping:** The KG is uniquely suited to represent granular ownership information, including "up to 1% holdings," Board of Directors, State Owned Enterprises (SOEs), and country of origin, fulfilling Deliverable (b). For instance, an "owned_by" edge can include a "percentage" attribute, and individuals can be linked to companies via "board_member_of" relationships. This structure allows for easy traversal to identify ultimate beneficial ownership or control, which is critical for assessing geopolitical risks.
- **Addressing Data Challenges:** A significant advantage of KGs in this context is their ability to address data sparsity and enhance context understanding.[20] Given that open-source data may be incomplete or fragmented, the KG can help infer missing relationships by leveraging existing connections and attributes. For example, if a company's direct Tier 1 suppliers are known, but their Tier 2 suppliers are not explicitly stated in public records, the KG can infer potential Tier 2 relationships based on industry commonalities or shared board members, effectively enriching the map. This capability is crucial for overcoming the inherent limitations of open-source data and building a comprehensive map beyond readily available direct data.

AI/ML Models for Core Functionalities

The AI/ML strategy will integrate various models to address the specific deliverables of the problem statement, with a strong emphasis on techniques suitable for graph data and unstructured text.

- **Mapping & Entity Resolution (Tier 2, 3, 4 Suppliers, Ownership):**
  - **Technique: Natural Language Processing (NLP)** will be the primary tool for extracting structured information from unstructured open-source text. This involves **Named Entity Recognition (NER)** to identify companies, individuals, locations, and component types, and **Relationship Extraction (RE)** to determine their connections (e.g., "Company X supplies Component Y to Company Z," "Person A is a director of Company B," "Company C owns X% of Company D").[10] Large Language Models (LLMs) can enhance this process by understanding context and inferring relationships from global media and public reports.[9]
  - **Integration with KG:** The extracted entities and relationships will directly populate the Knowledge Graph. **Graph Neural Networks (GNNs)** can then be applied to the constructed graph

for tasks like link prediction (inferring missing supplier relationships) or node classification (categorizing companies by tier or type based on their connections).[22]

- **Vulnerability & Risk Prediction (Disruption Likelihood, Risk Scoring):**
  - **Technique: Graph Neural Networks (GNNs)** are exceptionally well-suited for analyzing complex network structures to identify critical nodes, bottlenecks, and propagate risk through the supply chain.[22] They can leverage the graph's topology and node/edge attributes (e.g., financial health indicators, geopolitical stability of a country, supplier performance data) to predict vulnerabilities and calculate dynamic risk scores. GNNs have been shown to outperform traditional ML models in supply chain risk assessment and uncovering hidden risks.[22]
  - **Anomaly Detection:** Machine Learning models can establish baselines of "normal" supply chain behavior and flag subtle deviations (e.g., unusual transaction patterns, sudden changes in supplier status, or unexpected logistical delays) as potential threats.[13]
  - **Predictive Analytics:** AI models will analyze historical data and external factors such as weather patterns, economic indicators, and geopolitical events (extracted from news feeds) to forecast potential disruptions.[3] This allows for proactive risk management, anticipating challenges before they escalate.
  - **Dynamic Risk Scoring:** The system will generate continuous risk scores for each supplier or component based on real-time data ingestion and analysis, moving beyond static assessments to a dynamic risk landscape.[12] The ability of GNNs to understand the *interconnectivity* of the supply chain means they can not only identify a vulnerable supplier but also predict the *ripple effect* of that vulnerability across the entire multi-tier network, providing a more holistic risk assessment.[23]
- **Alternate Supplier Recommendation (Function- and Role-Based):**
  - **Technique:** A **Knowledge-Based Recommendation System** or **Content-Based Filtering** approach will be employed, leveraging the rich attributes stored in the Knowledge Graph.[16] When a disruption is identified, the system will query the KG for suppliers that match the "function and role" of the disrupted entity. This will involve matching based on component specifications, manufacturing capabilities, geographical location, capacity, and risk profile.
  - **Context-Aware Recommendations:** The system can incorporate real-time conditions (e.g., current disruptions, geopolitical stability of a region, transport availability) to refine recommendations, ensuring the suggested alternatives are viable and resilient.[16] AI-powered supplier discovery tools already exist in commercial procurement, demonstrating the feasibility of this approach.[14]

Standalone & Offline Capability: Design Considerations for Deployment

The requirement for the software to work in "standalone mode" and be "updated in offline mode" is a critical design constraint for a defense application where continuous internet connectivity may not be guaranteed.

- **Local Deployment:** The prototype will be designed as a desktop application or a locally hosted web application (e.g., using Python Flask/Streamlit with a local graph database instance). All necessary data, pre-trained models, and dependencies will be bundled with the application, allowing it to run on a local machine without requiring constant internet access for its core functionality.

- **Offline Updates:** The system will support periodic data and model updates via a local file transfer or a pre-packaged update bundle. This means the data ingestion pipeline will be designed to process batch updates of downloaded open-source data.
- **Upgradability:** The modular architecture facilitates future upgrades and enhancements. New data sources, improved AI models, or expanded functionalities can be integrated by updating specific modules without requiring a complete system overhaul.

Table: Key Deliverables and Proposed AI/ML Approach

This table succinctly maps the Indian Army's specific deliverables to the proposed technical solutions, providing a clear overview of how each requirement will be addressed within the prototype.

Deliverable (from PDS)	Core Functionality Addressed	Proposed AI/ML Approach/Model Types	Key Data Sources (Open Source)	Expected Prototype Output
(a) Map multi-tier suppliers (2nd, 3rd, 4th tier) using open-source information.	Multi-tier Supply Chain Mapping	NLP (NER, Relationship Extraction), Knowledge Graph Construction, GNNs (Link Prediction)	MCA Website (Master Data, Public Documents), Data.gov.in, Public News/Business Media, Industry Reports, Geospatial Data [9]	Interactive graph visualization of mapped supply chain (nodes: companies, components; edges: relationships, tiers) for selected equipment.
(b) Sense risks based on ownership (up to 1% holdings), BoD, SOEs, country of origin.	Vulnerability & Risk Assessment	GNNs (Node/Graph Classification, Anomaly Detection), Predictive Analytics, NLP (Sentiment Analysis, Entity Resolution)	MCA Public Documents, Public News/Business Media, Geopolitical Risk Indices (publicly available), Synthetic Data [3]	Highlighted vulnerable nodes/paths on the graph, risk scores for entities, identification of high-risk ownership/country ties.
(c) Suggest alternate suppliers based on function and role.	Alternate Supplier Recommendation	Knowledge-Based Recommendation Systems, Content-Based Filtering, KG Traversal	Mapped KG data (supplier capabilities, component attributes, location, risk profiles), Industry Standards/Taxonomies (inferred/public) [14]	List of alternate suppliers with their profiles (function, role, location, risk score) for selected components/sub-components.
(d) Furnish data with effect from a mutually agreed time stamp.	Temporal Data Handling	Temporal Graph Data Model, Data Versioning/Snapshotting	Archived versions of open-source data (if available), Synthetic Data with timestamps [4]	Ability to query and display the supply chain map and risk assessment as it existed at a specified past date.
(e) Work on standalone mode, updated offline, upgradable.	Deployment & Maintenance	Modular Application Design, Local Database, Packaged Executable/Containerization	N/A (System Design Constraint)	Self-contained software package, clear update instructions for new data/models.

III. Data Collection & Preparation Techniques

The success of the AI-based supply chain mapping system hinges on a robust and pragmatic data collection strategy, particularly given the reliance on open-source information and the inherent challenges of obtaining granular multi-tier data.

Open-Source Data Strategy

- **Publicly Available Company Registries (India-Specific):**
  - The primary source for identifying Indian entities will be the Ministry of Corporate Affairs (MCA) website (www.mca.gov.in) and the Open Government Data (OGD) Platform India (data.gov.in).[24] These portals provide essential master data such as Corporate Identification Number (CIN), company name, status, authorized/paid-up capital, date of registration, registered state, principal business activity, and registered office address. Crucially, the "Public Documents" section of the MCA website allows access to filed documents like annual returns and financial statements, which, while not directly providing raw transaction data, contain valuable unstructured text that can be processed for deeper insights into company structure and financial health.[24] These structured and semi-structured data points will form the foundational nodes and initial attributes within the Knowledge Graph.
- **Financial Transaction Records (Proxy Data, Publicly Available Reports):**
  - A significant challenge lies in obtaining "open source financial transaction record details". Direct, granular transaction data between private companies is proprietary and generally not publicly available.[6] Therefore, the strategy must pivot to inferring financial relationships and dependencies from proxy data. This involves leveraging publicly available *financial statements and annual reports*.[24] NLP will be used to extract high-level financial health indicators (e.g., revenue, profit, debt, major investments, mergers, acquisitions, significant supplier/customer mentions) that can serve as proxies for financial stability and business relationships. News articles reporting on major business deals, partnerships, or supply agreements will also be critical for inferring financial ties.[9] This adaptive approach is necessary to overcome the inherent limitations of open-source financial data.
- **News, Social Media, and Market Intelligence (NLP for Insights):**
  - Reputable news agencies, business journals, and industry-specific news feeds will be vital sources. Generative AI and NLP tools are highly effective at analyzing vast amounts of unstructured text from "news articles, social media posts and market reports" to identify potential risks.[12] This includes detecting emerging risks [23], identifying specific disruption types like factory fires, labor disruptions, or extreme weather events [3], and sensing geopolitical tensions or economic shifts that could impact supply chains.[13] Semantic Visions, for example, maps supplier networks using AI and real-time data from global media in multiple languages, extracting business relationships and sentiment (threat/opportunity).[9] This continuous monitoring of public discourse provides crucial, dynamic insights for vulnerability assessment.
- **Geospatial Data for Location Intelligence:**
  - Integrating geographical context is essential for understanding location-based risks. Data

sources like OpenStreetMap, public mapping APIs (e.g., OpenLayers/Leaflet for visualization), and government geospatial data portals can provide locations of companies, manufacturing sites, and transportation hubs. This enables the visualization of supply chain routes and the identification of geographical concentrations of suppliers, which are critical for assessing vulnerabilities related to natural disaster zones, geopolitical hotspots, or proximity to conflict areas.[27] Geospatial AI combines GIS with AI to automate analytics and elevate predictive capabilities, making it highly relevant for enhancing the vulnerability mapping component.[27]

- **Other Public Data:**
  - Where available, industry-specific databases, government reports on particular sectors (e.g., defense manufacturing trends), and aggregated trade statistics can provide additional context and relationships.

Data Acquisition Challenges & Mitigation: Addressing Data Sparsity and Privacy

The reliance on open-source data presents several challenges that must be addressed:

- **Granularity and Completeness:** Obtaining highly granular data, such as "1% ownership details" or comprehensive transaction records for all tiers, from purely open sources is exceptionally difficult. Public information is often incomplete or aggregated.
- **Data Quality and Noise:** Open-source data can be inconsistent, contain errors, or include significant irrelevant information, particularly from news and social media.
- **Entity Resolution:** A major challenge will be accurately identifying and linking different mentions of the same company or individual across various disparate data sources (e.g., "Company X" in a news article versus "X Corp Pvt Ltd" in a company registry).
- **Privacy and Confidentiality:** While the problem specifies "open source," strict adherence to data usage ethics and ensuring no sensitive or classified information is inadvertently processed or stored is paramount, especially given the defense context.

To mitigate these challenges, a multi-pronged approach is necessary:

- **Inference via NLP:** As discussed, NLP will be extensively used to infer relationships and attributes from unstructured text where direct structured data is unavailable.[10] This means the system will be designed to deduce connections rather than solely relying on explicit declarations.
- **Focus on Key Entities:** For the prototype, the scope of data collection will be strategically limited. Instead of attempting to map the entire Indian Army's supply chain, the focus will be on a few selected "major weapon systems/equipment" and tracing their supply chains to Tier 3 or 4 for a limited number of critical components. This allows for demonstrating the methodology and depth of analysis without being overwhelmed by data volume.
- **Synthetic Data Generation (Crucial Mitigation):** Given the inherent limitations of real open-source data, especially for granular details like 1% ownership or specific financial transactions, **synthetic data generation** will be a crucial supplementary approach. This allows for creating

controlled, realistic datasets that mimic complex multi-tier relationships and disruption scenarios.[28] Tools like SupplySim, available in the snap-stanford/supply-chains GitHub repository, can generate synthetic supply chain network data, which is invaluable for training and testing AI/ML models and demonstrating the system's full capabilities.[30] This pragmatic solution enables robust algorithm development even when real-world data is sparse or sensitive.

Data Pre-processing & Feature Engineering: Preparing Data for AI/ML Models

Once data is acquired, rigorous pre-processing and feature engineering are essential to transform raw information into a structured format suitable for the Knowledge Graph and subsequent AI/ML analysis.

- **Data Cleaning and Normalization:** This involves handling missing values, correcting inconsistencies, standardizing data formats (e.g., company names, addresses, dates), and removing duplicate entries.[4]
- **Entity Resolution and Linking:** A critical step is to accurately identify and link different mentions of the same company, individual, or product across various disparate data sources. This ensures that the Knowledge Graph maintains a coherent and accurate representation of real-world entities.
- **Relationship Extraction:** Formalizing the inferred or explicitly stated connections into a structured format of nodes and edges for the Knowledge Graph is paramount.
- **Feature Engineering for GNNs:** For the Graph Neural Networks, meaningful attributes will be derived for both nodes and edges. For example, company nodes might have attributes like a calculated financial health score (inferred from public reports), industry sector, or a geopolitical risk index based on country of origin. Edges could have attributes like the criticality of the supplied component or the volume of a transaction (if inferred). These features will be crucial inputs for the GNNs to perform risk assessment and prediction.[20]
- **Temporal Features:** Incorporating timestamps for data points and relationships will enable the system to analyze the supply chain's state "with effect from a mutually agreed time stamp". This allows for historical analysis and understanding how vulnerabilities evolve over time.

Table: Open-Source Data Sources for Multi-Tier Supply Chain Mapping

This table outlines key open-source data sources relevant to the project, highlighting their utility for different deliverables and noting associated challenges. This serves as a practical guide for data acquisition.

Data Type	Relevance to Deliverables	Potential Open-Source	Key Information/Notes

		Sources (Indian Context)	
Company Registries	Multi-tier Mapping (a), Ownership (b)	Ministry of Corporate Affairs (MCA) website (www.mca.gov.in), Open Government Data (OGD) Platform India (data.gov.in) [24]	Company Name, CIN, Registration Date, Address, Directors, Principal Business Activity. Public Documents (Annual Returns, Financial Statements) for unstructured text.
Financial/Business News	Multi-tier Mapping (a), Ownership (b), Risk Sensing (b), Alternate Suppliers (c)	Major Indian Business News Outlets (e.g., Economic Times, Livemint), Global Business News (e.g., Reuters, Bloomberg - for publicly available articles) [9]	Mentions of partnerships, contracts, mergers/acquisitions, investments, financial distress, operational disruptions. Requires NLP for extraction.
Geospatial Data	Risk Sensing (b)	OpenStreetMap, Public Mapping APIs (e.g., OpenLayers/Leaflet for visualization), Government geospatial portals (if available) [27]	Locations of facilities, transport routes. Essential for visualizing geographical risks (e.g., natural disasters, geopolitical hotspots).
Public Reports/Filings	Multi-tier Mapping (a), Ownership (b), Risk Sensing (b)	MCA Public Documents (Annual Reports, Director Reports), Industry Association Reports, Government Policy Documents [24]	High-level financial indicators, major shareholder mentions, board changes, strategic partnerships. Unstructured text requiring NLP.
Synthetic Data	All Deliverables (a-d) (for demonstration/training)	snap-stanford/supply-chains GitHub repository (SupplySim simulator), Gretel.ai resources, DataCamp tutorials [28]	Customizable multi-tier network structures, simulated financial transactions, ownership percentages, disruption events. Crucial for overcoming real data sparsity and testing.

IV. Prototype Development Process: From Concept to Working Model

The prototype development will follow an agile, iterative approach, with a strong focus on delivering core functionalities within the one-month timeframe. The choice of technology stack will prioritize open-source tools that facilitate rapid development and align with the "standalone" and "offline" requirements.

Technology Stack Recommendation: Programming Languages, Libraries, and Frameworks

The following technology stack is recommended for building the prototype:

ry	logy/Tool	ale & Snippet Relevance
mming Language		y standard for AI/ML and data s ve libraries and community support
Database	**Graph** or **Neo4j (Community Editi**	al for storing the Knowledge raph is fully open-source (Apac and scalable for massive graphs.[2 nity Edition is popular with native and Cypher query language, tho is GPLv3.[21]
Neural Network (GNN) Libraries	**h Geometric (PyG)** or **Deep (DGL)**	Python libraries for deep learn Both are high-performance, sc ll-documented, suitable for implem or risk assessment and link predicti
l Language Processing works	or **Hugging Face Transformers**	Named Entity Recognition nship Extraction (RE), and text proc nstructured data. spaCy is effici tion, while Transformers provides e-of-the-art LLMs for advance anding.[10]
l AI/ML Libraries	**earn, NumPy, Pandas**	nental libraries for data manip al analysis, and implementing tra dels if needed for baseline comp ific sub-tasks.
terface (UI) Framework	**lit** or **Flask/Dash**	pid development of a simple, inte sed UI. Streamlit is excellent fo pp creation, while Flask/Dash offe ization for a locally hosted web ser
nerization (Optional mended)		ackaging the application in ntained unit, facilitating "stan deployment and ensuring con ments.

+

Category	Technology/Tool	Rationale & Snippet Relevance
Programming Language	Python	Industry standard for AI/ML and data science; extensive libraries and community support.
Graph Database	JanusGraph or Neo4j (Community Edition)	Essential for storing the Knowledge Graph. JanusGraph is fully open-source (Apache 2.0 license) and scalable for massive graphs.[21] Neo4j Community Edition is popular with native graph storage and Cypher query language, though its license is GPLv3.[21]
Graph Neural Network (GNN) Libraries	PyTorch Geometric (PyG) or Deep Graph Library (DGL)	Leading Python libraries for deep learning on graphs. Both are high-performance, scalable, and well-documented, suitable for implementing GNNs for risk assessment and link prediction.[32]
Natural Language Processing (NLP) Frameworks	spaCy or Hugging Face Transformers	For Named Entity Recognition (NER), Relationship Extraction (RE), and text processing from unstructured data. spaCy is efficient for production, while Transformers provides access to state-of-the-art LLMs for advanced text understanding.[10]
General AI/ML Libraries	Scikit-learn, NumPy, Pandas	Fundamental libraries for data manipulation, statistical analysis, and implementing traditional ML models if needed for baseline comparisons or specific sub-tasks.
User Interface (UI) Framework	Streamlit or Flask/Dash	For rapid development of a simple, interactive web-based UI. Streamlit is excellent for quick data app creation, while Flask/Dash offer more customization for a locally hosted web service.
Containerization (Optional but Recommended)	Docker	For packaging the application into a self-contained unit, facilitating "standalone mode" deployment and ensuring consistent environments.

The selection of these tools balances performance with ease of use and rapid prototyping capabilities. While some GNN libraries offer high data throughput with GPU acceleration [32], the immediate priority for a one-month prototype is demonstrating core functionality. Therefore, libraries that are easy-to-use and have strong community support are favored.

Iterative Development Approach: Agile Sprints for Rapid Prototyping

Given the aggressive one-month deadline, an agile development methodology with short, focused sprints (e.g., weekly) is crucial. This allows for continuous progress tracking, early identification of roadblocks, and iterative refinement of the prototype.

- **Sprint 1: Foundation & Data Strategy (Week 1)**
    - **Focus:** Initial data collection, defining the Knowledge Graph schema, environment setup.
    - **Tasks:** Research and identify specific open-source data APIs/scraping targets (e.g., MCA website structure). Define the initial KG schema (nodes, relationships, attributes) based on deliverables. Set up the development environment (Python, chosen graph database, Git repository, virtual environment). Begin initial data scraping for a small sample of companies/components.
    - **Milestone:** Initial data sources identified, prototype scope (e.g., 1-2 major equipment types, focus on Tier 1-3 for a few entities) defined, basic KG schema drafted, development environment configured.
- **Sprint 2: Core Mapping & Ownership (Week 2)**
    - **Focus:** Populating the Knowledge Graph with multi-tier supplier and ownership data using NLP.
    - **Tasks:** Implement web scraping and data ingestion scripts to extract raw data. Develop and fine-tune NLP models (NER, Relationship Extraction) to identify company names, ownership links, Board of Directors, and country of origin from scraped news articles and public documents. Populate the Knowledge Graph with the extracted entities and relationships for the selected prototype scope. Begin initial data validation.
    - **Milestone:** Functional data ingestion pipeline, KG populated with initial multi-tier supplier and ownership data for the defined scope.
- **Sprint 3: Vulnerability & Risk + Recommendation (Week 3)**
    - **Focus:** Implementing the core AI logic for risk assessment and alternate supplier recommendation.
    - **Tasks:** Define specific vulnerability criteria (e.g., concentration risk, geopolitical risk indicators, single points of failure). Implement initial graph algorithms or GNNs to analyze the KG for these vulnerabilities and calculate basic risk scores. Develop the alternate supplier recommendation algorithm (e.g., content-based matching based on product attributes and supplier roles from the KG). Test AI logic with synthetic data to ensure functionality.
    - **Milestone:** Core AI logic for risk assessment and supplier recommendation implemented and tested on sample/synthetic data.
- **Sprint 4: Prototype Integration, UI & Documentation (Week 4)**
    - **Focus:** Integrating all modules, developing the user interface, preparing for submission.
    - **Tasks:** Integrate the data ingestion, KG, risk assessment, and recommendation modules. Develop a simple UI (using Streamlit or Flask/Dash) to allow user input (e.g., "Show supply chain for X

equipment," "Find alternate suppliers for Y component"), visualize the mapped supply chain, and display risk scores and recommendations. Implement the "standalone" and "offline update" capabilities. Package the prototype for submission. Prepare the final submission document, including technical details, scope, and future enhancements. Practice the prototype demonstration.
  - **Milestone:** Working prototype ready, all deliverables addressed for the defined scope, comprehensive submission document prepared.

Module-wise Development Plan

- **Data Ingestion & Knowledge Graph Construction Module:**
  - **Student A (AI/ML Focus):** Lead development of NLP models for entity and relationship extraction from unstructured text; design and implement the KG schema; develop scripts for populating the KG from processed data.
  - **Student B (Supply Chain Focus):** Identify and curate specific open-source data sources relevant to the chosen equipment/components; define the initial set of relevant suppliers and components for mapping; provide domain expertise for validating extracted relationships and data quality.
  - **Collaboration Points:** Regular synchronization on data schema, validation of extracted entities and relationships, ensuring data consistency for KG population.
- **Vulnerability Assessment & Risk Scoring Module:**
  - **Student A (AI/ML Focus):** Develop and implement GNNs or other graph-based ML algorithms for vulnerability detection and risk scoring (e.g., identifying critical paths, single points of failure, tracing ownership to high-risk countries); integrate external risk factors (e.g., publicly available country risk data).
  - **Student B (Supply Chain Focus):** Define specific vulnerability criteria and risk parameters relevant to defense supply chains; provide input on the interpretation of risk scores; validate identified vulnerabilities against domain knowledge.
  - **Collaboration Points:** Joint design of risk assessment logic, interpretation of AI model outputs, and refinement of risk parameters.
- **Alternate Supplier Recommendation Module:**
  - **Student A (AI/ML Focus):** Implement the recommendation algorithm (e.g., content-based matching) based on attributes from the Knowledge Graph; develop the logic for filtering and ranking potential alternatives.
  - **Student B (Supply Chain Focus):** Define the criteria for "alternate supplier" based on function, role, and other relevant attributes; provide feedback on the relevance and practicality of recommended suppliers.
  - **Collaboration Points:** Jointly define the attributes and criteria for supplier matching, iterative refinement of recommendation logic.
- **User Interface (UI) / Interaction Layer:**
  - **Both Students:** Jointly design the user interface to ensure it is intuitive and meets the Army's

potential interaction needs. Student A will primarily implement the backend API that serves data to the UI. Student B will focus on the user experience, data visualization, and presentation of insights (e.g., interactive graph visualization, clear display of risk scores and recommendations).
○ **Collaboration Points:** Continuous feedback loop to ensure the UI effectively communicates the system's capabilities and insights.

Testing & Validation Strategy: Ensuring Functionality and Adherence to Deliverables

A rigorous testing strategy is crucial to ensure the prototype is functional, reliable, and adheres to all specified deliverables within the tight timeline.

- **Unit Testing:** Individual functions and components of each module will be tested to ensure they perform as expected.
- **Integration Testing:** Verify seamless interaction and data flow between different modules (e.g., data ingestion to KG, KG to risk assessment).
- **Functional Testing:** Test the prototype against each of the five key deliverables (mapping, risk sensing, recommendations, temporal query, standalone/offline capability) using a small, controlled dataset (both real-world samples and synthetic data).
- **Performance Testing (Basic):** Assess the prototype's responsiveness and processing time for the limited dataset to ensure it is practical for demonstration.
- **Deliverable Compliance Check:** A meticulous checklist will be used to ensure every point in Deliverables (a) through (e) is addressed by the prototype's functionality and the accompanying documentation. This also includes verifying adherence to TRL requirements.
- **User Acceptance Testing (Informal):** If possible, a peer or mentor not directly involved in the development can test the UI for usability and clarity, providing valuable external feedback.

V. Project Execution: One-Month Timeline & Work Distribution

The following detailed 4-week timeline outlines the key focus areas, deliverables, and work distribution for each student, ensuring a structured approach to meet the ambitious deadline.

Detailed 4-Week Timeline: Breakdown of Tasks Per WeekKey Milestones & Checkpoints: Ensuring Progress Towards a Working Prototype

- **End of Week 1:** Initial Data Sources & KG Schema defined.
- **End of Week 2:** Functional Data Ingestion & Populated Knowledge Graph for selected scope.
- **End of Week 3:** Core AI Logic (Risk Assessment & Recommendation) implemented and tested.
- **End of Week 4:** Working Prototype & Submission Package complete.

Contingency Planning: Addressing Potential Roadblocks in a Tight Timeline

Given the aggressive timeline, proactive contingency planning is essential:

Week	Key Focus Areas	Deliverables/Milestones for the Week	Student A (AI/ML Focus) Responsibilities	Student B (Supply Chain Focus) Responsibilities
Week 1	Foundation & Data Strategy	Initial data sources identified, prototype scope defined, basic KG schema drafted, development environment configured.	Research open-source data APIs/scraping techniques. Initial NLP model selection for entity extraction. Define initial KG schema (nodes, relationships, attributes). Set up Python environment.	Deep dive into Indian company registry data (MCA, data.gov.in). Identify specific "major weapon systems/equipment" for prototype scope. Define initial set of relevant suppliers/components for mapping. Research proxy financial data sources.
Week 2	Core Mapping & Ownership	Functional data ingestion pipeline, KG populated with initial multi-tier supplier and ownership data for selected scope.	Implement web scraping/data ingestion scripts. Develop/fine-tune NLP models for entity (company, person, location) and relationship (supplier-of, owned-by, board-member-of) extraction from news/public reports. Populate initial KG.	Validate extracted entities and relationships against known information for selected equipment/suppliers. Refine data cleaning rules. Identify key ownership structures (SOE, private holdings) for target companies.
Week 3	Vulnerability & Risk + Recommendation	Core AI logic for risk assessment and supplier recommendation implemented and tested on sample/synthetic data.	Implement GNN/graph algorithms for vulnerability detection and risk scoring (e.g., identifying single points of failure, tracing ownership to high-risk countries). Develop initial alternate supplier recommendation algorithm (e.g., content-based matching).	Define specific vulnerability criteria (e.g., concentration risk, geopolitical risk indicators). Provide input for risk scoring logic. Define criteria for "alternate supplier" based on function/role. Validate initial risk scores and recommendations.
Week 4	Prototype Integration, UI & Documentation	Working prototype ready, all deliverables addressed, comprehensive submission document prepared.	Integrate all modules. Develop simple UI (Streamlit/Flask) for demonstration. Implement standalone/offline capability.	Prepare demonstration scenarios. Gather/curate sample input data for demo. Draft sections of the

		Package the prototype (e.g., Docker container or executable).	proposal (problem analysis, deliverables, importance, indigenous content). Prepare for potential Q&A on supply chain aspects.

- **Data Availability:** If comprehensive open-source data proves too sparse for certain granularities (e.g., 1% ownership), increase reliance on synthetic data for demonstration, clearly stating this limitation and outlining future plans for real data integration. The prototype's value will then lie in demonstrating the *methodology* with realistic data.
- **Model Complexity:** Start with simpler, more interpretable ML models (e.g., rule-based systems, traditional graph algorithms) for initial implementation. Only transition to more complex GNNs if time permits and the performance benefits are clearly demonstrable within the remaining schedule.
- **Scope Creep:** Rigorously adhere to the defined Minimum Viable Product (MVP) scope for the prototype. Resist the temptation to add advanced features that are not explicitly required by the deliverables. Postpone such enhancements for future development phases.
- **Technical Issues:** Allocate buffer time within each sprint for debugging and troubleshooting. Leverage online communities (e.g., Stack Overflow, GitHub issues) and documentation for rapid problem resolution.

VI. Useful Online Resources & Learning Paths

To effectively execute this project, leveraging a combination of academic, technical, and domain-specific online resources will be invaluable.

AI/ML & GNN Resources

- **Courses:** For a strong theoretical foundation in Graph Neural Networks (GNNs), the "Stanford Course: CS224W Machine Learning with Graphs" is highly recommended.[32]
- **Books:** "Network Science" by Albert-László Barabási and "Graph Representation Learning Book" by William L. Hamilton provide in-depth knowledge on graph theory and GNNs.[32]
- **Libraries:** For practical implementation of GNNs, **PyTorch Geometric (PyG)** and **Deep Graph Library (DGL)** are leading Python libraries known for their performance, scalability, and ease of use.[32]
- **NLP:** For text processing, **spaCy** offers efficient tools for production-ready NLP pipelines, while **Hugging Face Transformers** provides access to a vast array of pre-trained Language Models for advanced tasks like Named Entity Recognition and Relationship Extraction.
- **General ML: Scikit-learn** is indispensable for traditional machine learning tasks, and the

official documentation for **TensorFlow** and **PyTorch** will be crucial for deep learning model development.

- **Synthetic Data:** Resources from **Gretel.ai** [28] and **DataCamp tutorials** [29] offer insights into various synthetic data generation methods. Critically, the **snap-stanford/supply-chains GitHub repository** includes code for "Learning Production Functions for Supply Chains with Graph Neural Networks" and provides a synthetic data simulator called SupplySim, which can be directly used to generate realistic supply chain network data for training and testing.[30]

Supply Chain Analytics Resources

- **Optimization/Simulation: Optilogic's resources on supply chain simulation** provide valuable insights into how simulation can be used to test risk mitigation strategies and improve network design.[26]
- **Risk Management:** Articles such as those on **Forbes Tech Council discussing Generative AI in supply chain risk assessment and mitigation** [12] and the **Emerald article on AI-driven methods for supply chain risk management** [23] offer contemporary perspectives and specific AI techniques.
- **General Concepts:** Reviewing foundational concepts from reputable academic institutions like **MIT OpenCourseWare on Supply Chain Management** or professional certifications like **APICS** can solidify understanding of supply chain principles.

Open-Source Data Portals (India-Specific)

- **Company Registries:** The **Ministry of Corporate Affairs (MCA) website** (www.mca.gov.in) [24] and the **Open Government Data (OGD) Platform India** (data.gov.in) [25] are the authoritative sources for Indian company master data and public filings.
- **Public News Archives:** Major Indian business news portals (e.g., Economic Times, Livemint) and global news agencies (e.g., Reuters, for publicly available articles) will be key for extracting unstructured textual data.

Relevant Research Papers & Case Studies

- **GNNs in Supply Chain:** Explore academic papers specifically on **Graph Neural Networks in Supply Chain** to understand their application in risk assessment and relationship detection.[22]
- **AI in Defense Supply Chain:** Review case studies and articles on **AI in defense supply chain management** from organizations like the DLA [2] and analyses of aerospace and defense supply chain challenges.[3] These provide real-world context and validation for the problem.
- **Commercial Solutions:** While proprietary, studying the features and approaches of commercial solutions like **Altana.ai** [4], **Resilinc** [5], **Semantic Visions** [9], and **TealBook** [14] can offer inspiration for functionality and system design, particularly regarding multi-tier visibility, risk sensing, and supplier recommendations.

VII. Conclusion & Recommendations

The development of an AI-based supply chain mapping system for the Indian Army presents a critical opportunity to enhance national defense readiness by mitigating vulnerabilities exposed in complex, multi-tier supply chains. This report has outlined a comprehensive roadmap for developing a working prototype within a challenging one-month timeframe, emphasizing a pragmatic, open-source, and AI-driven approach.

The analysis indicates that a **Knowledge Graph** serves as the optimal foundational data model for representing the intricate relationships within multi-tier supply chains and granular ownership structures. This choice is critical for addressing data sparsity and enhancing contextual understanding from disparate open-source information. The reliance on **Natural Language Processing (NLP)** and **Large Language Models (LLMs)** is paramount for extracting structured entities and relationships from unstructured public documents and news, especially given the difficulty in obtaining direct, granular financial transaction data. **Graph Neural Networks (GNNs)** are identified as the most suitable AI models for analyzing the interconnectedness of the Knowledge Graph, enabling sophisticated vulnerability assessment, dynamic risk scoring, and the prediction of ripple effects across the supply chain.

For the prototype, a **realistic Technology Readiness Level (TRL) target of TRL 3-6** is recommended, focusing on demonstrating the core AI/ML logic and system capabilities with a curated, representative dataset, potentially augmented by **synthetic data generation** to overcome real-world data limitations. The prototype's design will prioritize **standalone operation and offline update capabilities**, aligning with critical defense deployment requirements. Furthermore, maximizing **indigenous content** by leveraging Indian public data sources and open-source technologies will be a significant advantage in the evaluation process.

Key Recommendations for Prototype Development:

1. **Prioritize Knowledge Graph Construction:** Dedicate significant effort to building a robust Knowledge Graph schema and populating it with accurately extracted entities and relationships from diverse open-source Indian data sources. The quality of the KG will directly determine the effectiveness of subsequent AI models.
2. **Focus on Core Deliverables (MVP):** Given the one-month deadline, strictly adhere to an MVP approach. Demonstrate the core functionalities of multi-tier mapping, risk sensing, and alternate supplier recommendation for a limited scope (e.g., one or two major equipment systems and their critical components).
3. **Leverage NLP Extensively:** Develop strong NLP pipelines for Named Entity Recognition and Relationship Extraction from unstructured text (news, public filings, annual reports) to infer financial health, ownership, and supply chain connections where direct structured data is unavailable.

4. **Integrate Synthetic Data:** Utilize synthetic data generation tools (e.g., SupplySim from the snap-stanford/supply-chains repository) to train and test GNNs and other AI models, especially for demonstrating granular ownership analysis (e.g., 1% holdings) and various disruption scenarios where real data is scarce or sensitive.
5. **Design for Standalone and Offline:** Ensure the chosen technology stack and architectural design allow for the prototype to run locally without continuous internet connectivity and to be updated periodically in an offline mode. Containerization (e.g., Docker) can facilitate this.
6. **Emphasize Explainability:** While not explicitly requested, consider incorporating elements of explainable AI (XAI) where feasible. Providing some transparency into *why* a risk is flagged or a supplier is recommended can build trust and facilitate adoption in a defense context.
7. **Document Thoroughly:** Prepare a comprehensive proposal document that clearly articulates the problem understanding, the proposed solution's architecture, the data strategy, the AI/ML models used, the achieved TRL, and a clear plan for future enhancements and scalability.

By meticulously following this roadmap, the engineering students can successfully develop a compelling AI-based supply chain mapping system prototype that not only meets the Indian Army's requirements but also demonstrates significant innovation and a clear path toward enhancing national defense readiness.

Works cited

1. India moving towards becoming a global leader in defence ... - PIB, accessed on June 5, 2025, https://www.pib.gov.in/PressReleasePage.aspx?PRID=2102439
2. Utilization of Artificial Intelligence (AI) to Illuminate Supply Chain ..., accessed on June 5, 2025, https://www.dla.mil/About-DLA/News/News-Article-View/Article/4186367/utilization-of-artificial-intelligence-ai-to-illuminate-supply-chain-risk/
3. Top 5 Aerospace and Defense Supply Chain Challenges of 2024 ..., accessed on June 5, 2025, https://resilinc.ai/blog/aerospace-and-defense-supply-chain-challenges/
4. Platform | Altana, accessed on June 5, 2025, https://altana.ai/platform
5. Multi-Tier & Agentic Supply Chain Mapping | Resilinc, accessed on June 5, 2025, https://resilinc.ai/products/multi-tier-mapping/
6. B2B Integration Platform & Supply Chain Solutions | OpenText, accessed on June 5, 2025, https://www.opentext.com/products/business-network-cloud
7. Mine - The Future of Data Ownership, accessed on June 5, 2025, https://www.saymine.com/
8. Company Data - LSEG, accessed on June 5, 2025, https://www.lseg.com/en/data-analytics/financial-data/company-data
9. Multi-Tier Supply Chain Mapping | Visibility Beyond the First Tier, accessed on June 5, 2025, https://www.semantic-visions.com/solutions/multi-tier-supply-chain-mapping
10. Interpreting Land Titles in Land Registry using Natural Language Processing, accessed on June 5, 2025, https://fastdatascience.com/nlp-land-titles/
11. Legal NLP - Natural Language Processing for Legal Documents - John Snow Labs, accessed on June 5, 2025, https://www.johnsnowlabs.com/legal-nlp/
12. Leveraging Generative AI In Supply Chain Risk Assessment - Forbes, accessed on June 5, 2025,

https://www.forbes.com/councils/forbestechcouncil/2025/02/13/leveraging-generative-ai-in-supply-chain-risk-assessment-and-mitigation/
13. The Critical Role of AI in Strengthening Supply Chain Security | NeuralTrust, accessed on June 5, 2025, https://neuraltrust.ai/blog/ai-supply-chain-security
14. Aime | Gen AI Assistant - TealBook, accessed on June 5, 2025, https://www.tealbook.com/solutions/ai-supplier-recommendations/
15. Top 10 AI Procurement Tools in 2025 – Use Cases, Examples & Key ..., accessed on June 5, 2025, https://suplari.com/top-10-ai-procurement-tools/
16. How to Implement AI-Based Recommendation System Step-by-Step - PixelPlex, accessed on June 5, 2025, https://pixelplex.io/blog/ai-based-recommendation-system/
17. What Are Recommender Systems? Use Cases, Types & Techniques - Coralogix, accessed on June 5, 2025, https://coralogix.com/ai-blog/what-are-recommender-systems-use-cases-types-and-techniques/
18. Eligibility of technology readiness levels (TRL) – UKRI, accessed on June 5, 2025, https://www.ukri.org/councils/stfc/guidance-for-applicants/check-if-youre-eligible-for-funding/eligibility-of-technology-readiness-levels-trl/
19. Technology Readiness Levels - NASA, accessed on June 5, 2025, https://www.nasa.gov/directorates/somd/space-communications-navigation-program/technology-readiness-levels/
20. Knowledge Graph in Machine Learning: All You Need to Know, accessed on June 5, 2025, https://www.puppygraph.com/blog/knowledge-graph-in-machine-learning
21. 7 Best Open Source Graph Databases - PuppyGraph, accessed on June 5, 2025, https://www.puppygraph.com/blog/open-source-graph-databases
22. Graph Neural Networks in Supply Chain Analytics and Optimization: Concepts, Perspectives, Dataset and Benchmarks - arXiv, accessed on June 5, 2025, https://arxiv.org/html/2411.08550v1
23. Artificial intelligence applications for supply chain risk management ..., accessed on June 5, 2025, https://www.emerald.com/insight/content/doi/10.1108/mscra-10-2024-0041/full/html
24. Unlocking Business Transparency: A Step-by-Step Guide to Checking Company Registration in India - Vakilkaro, accessed on June 5, 2025, https://www.vakilkaro.com/blogs/check-company-registration-in-india-guide
25. Companies |Open Government Data (OGD) Platform India, accessed on June 5, 2025, https://www.data.gov.in/dataset-group-name/Companies
26. Supply Chain Simulation Explained - Optilogic, accessed on June 5, 2025, https://optilogic.com/resources/blog/supply-chain-simulation-explained/
27. Artificial Intelligence in GIS | Geospatial AI - Esri, accessed on June 5, 2025, https://www.esri.com/en-us/geospatial-artificial-intelligence/overview
28. What is Synthetic Data Generation? Definition and FAQs - Gretel.ai, accessed on June 5, 2025, https://gretel.ai/technical-glossary/what-is-synthetic-data-generation
29. Synthetic Data Generation: A Hands-On Guide in Python - DataCamp, accessed on June 5, 2025, https://www.datacamp.com/tutorial/synthetic-data-generation
30. snap-stanford/supply-chains - GitHub, accessed on June 5, 2025, https://github.com/snap-stanford/supply-chains
31. Top 10 Open Source Graph Databases in 2025 | GeeksforGeeks, accessed on June 5, 2025, https://www.geeksforgeeks.org/open-source-graph-databases/
32. Graph Neural Networks: Libraries, Tools, and Learning Resources - neptune.ai, accessed on June

5, 2025, https://neptune.ai/blog/graph-neural-networks-libraries-tools-learning-resources
33. Avoid defense dollar waste with industrial AI supply chain ... - Avathon, accessed on June 5, 2025,
https://avathon.com/blog/avoid-defense-dollar-waste-with-industrial-ai-supply-chain-management/