

An Evaluation Survey of Knowledge-Based Approaches in Telecommunication Applications

Georgios P. Koudouridis ^{1,2,*} , Serveh Shalmashi ¹ and Reza Moosavi ¹

¹ Systems Management, Global AI Accelerator, Ericsson, 16483 Stockholm, Sweden; serveh.shalmashi@ericsson.com (S.S.); reza.moosavi@ericsson.com (R.M.)

² Radio Communications Laboratory, Department of Physics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

* Correspondence: george.koudouridis@ericsson.com

Abstract: The purpose of this survey study is to shed light on the importance of knowledge usage and knowledge-driven applications in telecommunication systems and businesses. To this end, we first define a classification of the different knowledge-based approaches in terms of knowledge representations and reasoning formalisms. Further, we define a set of qualitative criteria and evaluate the different categories for their suitability and usefulness in telecommunications. From the evaluation results, we could conclude that different use cases are better served by different knowledge-based approaches. Further, we elaborate and showcase our findings on three different knowledge-based approaches and their applicability to three operational aspects of telecommunication networks. More specifically, we study the utilization of large language models in network operation and management, the automation of the network based on knowledge-graphs and intent-based networking, and the optimization of the network based on machine learning-based distributed intelligence. The article concludes with challenges, limitations, and future steps toward knowledge-driven telecommunications.

Keywords: network automation; generative AI; large language models; knowledge-based systems; knowledge representation; machine reasoning; machine learning; intent-based networking



Citation: Koudouridis, G.P.; Shalmashi, S.; Moosavi, R. An Evaluation Survey of Knowledge-Based Approaches in Telecommunication Applications. *Telecom* **2024**, *5*, 98–121. <https://doi.org/10.3390/telecom5010006>

Academic Editor: Alessandro Polo

Received: 13 December 2023

Revised: 11 January 2024

Accepted: 18 January 2024

Published: 26 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the ever-evolving landscape of telecommunications, the relentless pursuit of innovation and efficiency continues to drive transformative advancements [1,2]. Within this dynamic realm, knowledge-based approaches represent a promising shift in how we design, manage, and optimize telecommunications networks. These approaches use advanced techniques like knowledge representation and machine learning/reasoning, along with domain expertise, to turn raw data into valuable operational insights. By analyzing vast and diverse datasets, knowledge-based systems enable informed decision-making, predictive analytics, and autonomous network orchestration and management. The potential impact on telecommunications is significant. Imagine networks that autonomously adapt to changing conditions, predict and mitigate anomalies, and optimize performance in real time [3]. This could lead to reduced signaling overhead, reduced operational costs, better resource allocation, and proactive identification of problems and vulnerabilities before they affect service quality. These intelligent systems have the potential to greatly improve the efficiency, reliability, and responsiveness of telecommunications networks [4,5].

This paper is an evaluation survey that aims to shed light on the importance of knowledge usage and knowledge management in telecommunication systems and telecommunications businesses. Firstly, based on a set of knowledge-related definitions, various forms of knowledge representations and reasoning (KRR) are classified and presented. The classification is used to associate the various representation and reasoning forms with the applications they are mostly suitable for supporting.

Secondly, a set of assessment criteria is defined for the evaluation of the different approaches. These criteria correspond to KRR features characterizing the quality and usefulness of each approach given the tasks they are majorly used for. The evaluation according to the quality criteria is of assistance when determining the suitability of a KRR approach in different automation processes' demands.

Finally, we consider the usage of KRR for the automation of telecommunication networks in three different levels of business automation abstraction. To this end, we define and reason on the usefulness and importance of the knowledge automation processes on three use-case categories including automation of network operation centers, telecommunication system solutions towards intent-driven operation, and distributed radio access networks (RAN) optimization functions. The paper concludes with a forward strategy and next steps.

2. Knowledge-Based Approaches

Throughout the years, knowledge-based approaches have been accommodated by data warehouses tailored for telecommunications [6–8]. When considering the automated utilization of knowledge in telecommunication systems and in general, there is a set of requirements that are put forward by the applications. In the first subsection, we first provide a categorization of the knowledge representation approaches. For the evaluation, we further suggest a set of assessment criteria for comparison purposes. Finally, in the last subsection, we present the evaluation in tabular form based on the assessment criteria.

2.1. Categories of Knowledge Representation and Reasoning Approaches

In general, there are many knowledge representation forms, each with an associated reasoning approach. In this paper, we divide them into three major categories as depicted in Figure 1. The first category corresponds to logic-based and mathematical formalisms that depart from standard logic and mathematics. Due to practical limitations in the applications of logic under uncertainty, these formalisms have been extended to non-standard logic approaches of reasoning to cope with incomplete and uncertain knowledge. To this end, various forms of reasoning have been developed including abductive, inductive, fuzzy, probabilistic, and many others. The second category consists of structural approaches, typically based on graphs. Reasoning in graphs corresponds to traversing through the graph from a root or a set of nodes, that correspond to the premises, to another non-empty set of nodes, that correspond to the conclusions. The third category refers to data-driven formalisms that accommodate reasoning based on data and/or data-derived models as in digital twin, regression, neural network, and generative AI. It has to be noted that the two last categories, in the description that follows, correspond to the connectionist approach, as depicted in Figure 1. However, they will be defined and evaluated in the sequel separately as they can be used for different purposes. Common to all these different categories is the existence of a knowledge base where predicates and facts, graphs, data, and models that characterize and describe the state of the system are stored. In this section, these different representation and reasoning forms will be described and their applicability in telecommunications systems will be discussed.

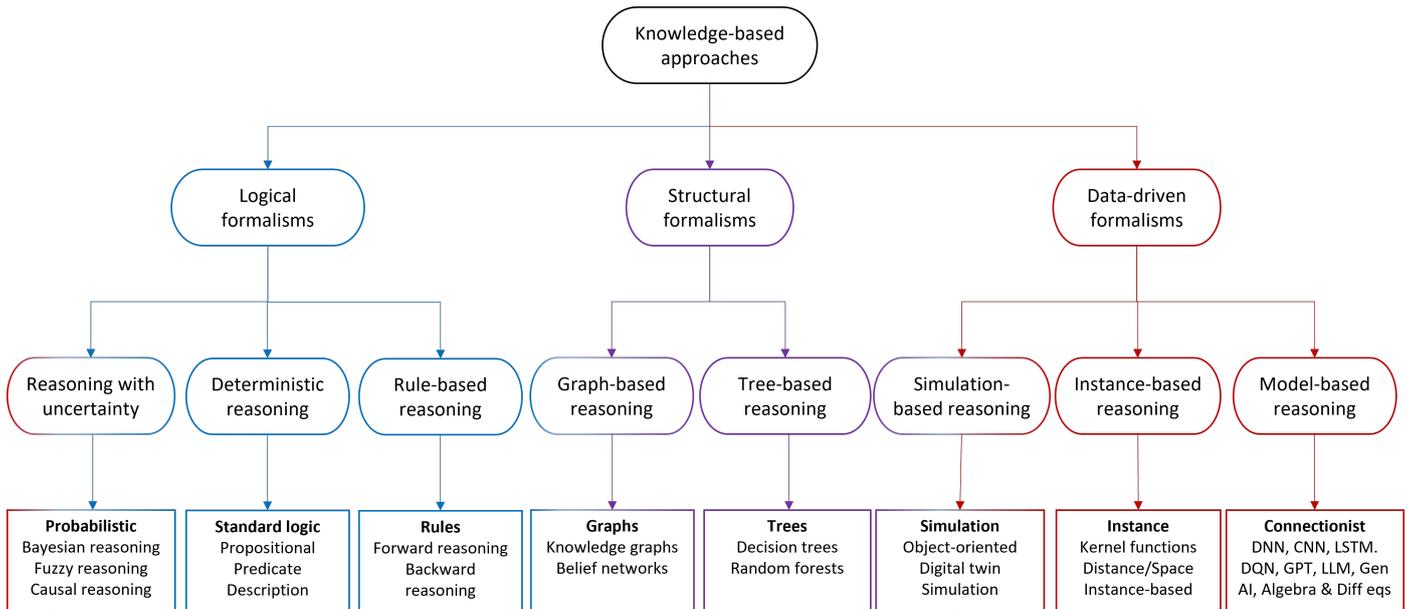


Figure 1. Knowledge representation and reasoning categories.

2.1.1. Probabilistic Formalisms

These formalisms employ probabilistic models including non-standard logic models to represent uncertainty and causal relationships between variables. They enable reasoning under uncertainty by quantifying the likelihood of different events and updating beliefs based on evidence. Uncertainties may be due to incomplete, imprecise, vague knowledge or knowledge that is probabilistic and/or causally inferred. The latter corresponds to Bayesian and/or causal reasoning that is governed by the conditional probabilities and/or causal relationships represented by probabilistic and causal models, respectively. The former can be alternatively addressed by approaches such as non-monotonic logic, fuzzy logic, and temporal logic. These reasoning methods are suitable for application scenarios with varying levels of certainty and requirements for timely adapting to dynamic and uncertain environments [9–11].

2.1.2. Standard Logic-Based Formalisms

These formalisms use logical rules and symbols to represent knowledge and perform deductive reasoning. They provide a precise and rigorous framework for expressing relationships and making inferences based on logical rules and constraints. Examples of such formalisms are the first-order logic and the propositional logic [12].

2.1.3. Rule-Based Formalisms

Rule-based formalisms utilize a collection of if-then rules to represent knowledge and make decisions. They allow explicit representation of domain-specific knowledge and provide a flexible mechanism for capturing complex relationships and conditions. This category includes formalisms, such as expert systems, policy-based systems, and production systems [13].

2.1.4. Graph-Based Formalisms

Graph-based formalisms represent knowledge using nodes and edges to represent knowledge entities and their relationships. They provide a versatile structure for organizing and representing complex knowledge, allowing for efficient traversal and exploration of interconnected information. Graph-based reasoning involves traversing nodes and edges in the graph to explore relationships between entities. Examples of such formalisms are knowledge graphs, ontologies, and semantic networks [14,15].

2.1.5. Tree-Based Formalisms

Tree-based formalisms use hierarchical structures to represent knowledge in a sequential decision-making fashion. They allow for the classification or regression of input data by partitioning it into smaller subsets based on specific attributes or conditions. A conclusion is made by traversing a path through the tree nodes until the path reaches a conclusion represented by a unique leaf node. Examples of tree-based formalisms are decision trees [16] and random forests [17].

2.1.6. Simulation-Based Formalisms

Simulation-based formalisms involve modeling a system or phenomenon to study its behavior over time. They allow for the representation of dynamic processes, enabling the examination of different scenarios and the prediction of outcomes based on the defined rules and constraints. Reasoning is performed based on emulating the behavior of networks' components to simulate scenario outcomes and make predictions on taken actions. Simulators, frames, and digital twins [18] are examples of such formalism realizations.

2.1.7. Instance-Based Formalisms

Instance-based formalisms use past instances or cases to make predictions or decisions about new situations. They rely on similarity measures and retrieve relevant past instances to infer likely outcomes or solutions for new cases. Representative approaches constitute tensor/space, regression, instance-based reasoning, case-based reasoning, k-nearest neighbors, and interpolation methods [19], which have wide use in telecommunication applications. These representations are effective in handling high-dimensional and complex data, allowing for automatic feature extraction and pattern recognition, but lack interpretable and explicit knowledge representation.

2.1.8. Neural Network-Based Formalisms

Neural network-based formalisms, and other similar deep connectionist formalisms, employ interconnected layers of artificial neurons to learn patterns and relationships from data. They excel at handling complex, high-dimensional data and are capable of automatic feature extraction, making them suitable for tasks such as classification, regression, and pattern recognition [20,21]. These formalisms correspond mainly to approaches that are realized by artificial neural networks (ANN) approaches, including Deep Neural Networks (DNN), Deep Q-learning Networks (DQN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Deep Reinforcement Learning (DRL) [22].

2.1.9. Generative AI-Based Formalisms

Generative AI-based formalisms utilize models that can generate new data or simulate human-like responses. They are trained on large amounts of data and can comprehend and generate human language, making them valuable for tasks like language translation, text generation, summarization, and natural language understanding. These formalisms rely on very deep neural network approaches and include Variational Autoencoders (VAE) and Generative Adversarial Networks (GANs), where two neural networks contest with each other in the form of a zero-sum game and where one agent's gain is another agent's loss. They also include approaches implementing Large Language Models (LLMs) and General Process Transformation (GPT) for Natural Language Understanding/Processing (NLU/NLP), of which their impact on the telecommunications industry is highly anticipated [23–25]. It is important to understand that GenAI does not only mean Large Language Models (LLMs). GenAI is a broad field that covers numerous modalities such as text, images, audio, and videos. Also, not all LLMs are capable of generating content. Only decoder-only and encoder-decoder models are generative, and for simplicity, we will refer to these as generative LLMs, while the rest will be referred to as encoder-based LLMs. Additionally, LLMs are built using various architectures and training methodologies.

2.2. Knowledge-Based Applications

In this paper, we elaborate on different operational aspects corresponding to distinct ways in which networks benefit from knowledge-based approaches.

- *Network operation*—Different use cases may require different forms of knowledge representation and reasoning approaches. Network operation and management use cases that provide operation support based on human interaction greatly benefit from Generative AI. In Section 4.1, a description of such use cases is provided.
- *Network automation*—There are different approaches to realize automation in networks. One promising approach is intent-based networking automation. In Section 4.2, a description of a knowledge-assisted solution for intent-based automation is provided.
- *Network optimization*—Improving network performance for different use cases requires different network optimization approaches. For optimization in RAN, which is an inherently distributed network, one promising approach is distributed learning. In Section 4.3, a description of such approaches is provided.

3. Evaluation of Different Knowledge-Based Approaches

For the evaluation, we first define a set of assessment criteria for comparison purposes. We use the criteria to derive relevant measurements for the qualitative evaluation, the results of which are discussed in the last subsection and summarized in tabular form.

3.1. Assessment Criteria

The selection of knowledge representation and reasoning to implement a use case is based on relevant characteristics that are generally or specifically defined. These definitions collectively provide a framework for evaluating knowledge representation formalisms based on key criteria, while ensuring that the chosen formalism aligns with specific requirements and objectives. Different criteria are of interest for different use cases. Herein, the following list of criteria may be considered:

- *Knowledge formalism*—Knowledge formalism refers to the specific language, representation framework, or set of rules used to express and structure knowledge within a knowledge representation system. It defines how information is encoded, stored, and manipulated to enable reasoning and inference.
- *Knowledge accuracy*—Accuracy refers to how well the knowledge representation reflects the true state of the world or domain it is intended to model. It manifests the degree to which the information encoded in the knowledge representation is correct and free from errors. High accuracy indicates that the knowledge representation provides a faithful reflection of reality, while low accuracy implies that it may contain inaccuracies or inaccurately represent aspects of the domain.
- *Knowledge precision*—Precision relates to the level of detail and specificity in the knowledge representation. It manifests how well the representation captures fine-grained distinctions, nuances, or specific aspects of the domain. High precision implies that the knowledge representation provides detailed and specific information, while low precision may result in a more abstract or generalized representation.
- *Knowledge coverage*—Coverage refers to the extent to which the knowledge representation includes relevant information about the domain. It manifests the comprehensiveness of the representation, indicating whether it covers all the necessary concepts, relationships, or facts related to the domain. High coverage implies that the representation includes a broad range of domain knowledge, while low coverage may result in gaps or missing information.
- *Knowledge generalizability*—Generalizability considers how well the knowledge or conclusions derived from the representation can apply to broader contexts or domains beyond the specific use case. It manifests the extent to which results can be generalized or adapted to different scenarios.
- *Knowledge explainability*—Explainability gauges how transparent and understandable the reasoning process and the conclusions drawn from the knowledge base are. It

manifests the degree to which the representation facilitates mechanisms for generating explanations, justifications, or evidence to support the derived conclusions.

- *Knowledge certainty*—Certainty refers to the degree of confidence or trustworthiness associated with the knowledge encoded in the representation. It manifests how certain or reliable the information is in the reasoning process. High certainty implies a high degree of confidence in the knowledge, while low certainty indicates uncertainty or ambiguity.
- *Knowledge consistency*—Consistency pertains to the ability of the knowledge representation to handle conflicting or contradictory information and ensure coherence. It manifests how well the representation deals with conflicting facts or conclusions and resolves inconsistencies within the knowledge base.
- *Knowledge confidence*—Confidence assesses the levels of trust or belief in the conclusions drawn from the knowledge base. It manifests how confident one can be in the validity of the results obtained through reasoning with the represented knowledge.
- *Knowledge expandability*—Expandability refers to the capability of the knowledge representation to accommodate the derivation of new knowledge from existing facts or rules. It also manifests how easily new knowledge can be added to the representation.
- *Knowledge scalability*—Scalability relates to how well the knowledge representation can handle increasing complexity and size, particularly when dealing with large amounts of knowledge, complex relationships, or diverse domains. It manifests the ability to efficiently store, retrieve, and reason with knowledge.

Although further criteria could be identified, the listed criteria cover a whole range of aspects related to knowledge quality characteristics and can be broadly divided into two categories. The first category refers more to the *declarative* quality characteristics of a formalism that are pertinent to the knowledge representation. The representation-centric quality category consists mainly of the formalism, accuracy, precision, coverage, generalizability, and explainability criteria. The second category refers more to the *procedural* quality characteristics of a formalism that relate to the knowledge reasoning process. The reasoning-centric quality category consists mainly of the certainty, consistency, confidence, expandability, and scalability criteria.

It should be noted that different representations and reasoning formalisms address the above criteria differently. Consequently, the selection of one approach over the other may depend on the use-case requirements for these criteria. Nonetheless, the above criteria can help evaluate and compare different knowledge representation approaches based on their capabilities and limitations in addressing the various aspects of knowledge representation and reasoning.

3.2. Assessment Analysis Measurements

For each one of the criteria, the following qualitative measurements are principally defined:

- *Knowledge formalism*—The choice of knowledge formalism can be assessed based on its expressive power, syntactic and semantic clarity, and adherence to the established standards. Evaluation involves considering how well the formalism captures the complexities of the domain, supports precise representation, and aligns with the requirements of the application. An indication of the level of knowledge formalism would be given by the percentage of relevant standards, like expressiveness, clarity, etc., met by the knowledge formalism.
- *Knowledge accuracy*—Accuracy in a formalism is measured by assessing the correspondence between the represented knowledge and the actual state of affairs. This can be performed via continued evaluations of the formalism's performance in making correct inferences or predictions within the domain. An indication of knowledge accuracy would be given by the percentage of predictions or inferences that align with real-world data or expert assessments.
- *Knowledge precision*—Precision can be evaluated by examining the level of detail in the representation, such as the granularity of concepts, relationships, or constraints. A

highly precise representation contains specific rules, attributes, or axioms that leave little room for ambiguity. An indication of knowledge precision would be given by the proportion of highly detailed rules or attributes in the knowledge base.

- *Knowledge coverage*—Coverage can be evaluated by comparing the knowledge representation to a comprehensive domain ontology or by assessing whether it includes all relevant concepts, entities, and relationships needed to address specific use cases or queries within the domain. An indication of knowledge coverage would be given by the extent to which the knowledge base covers the domain ontology's concepts.
- *Knowledge generalizability*—Generalizability can be assessed by examining the performance and applicability of the knowledge representation across a range of related domains or use cases. It involves testing whether the knowledge and reasoning methods can be extended to diverse contexts effectively. An indication of knowledge generalizability would be given by the representation's effectiveness when applied to different domains.
- *Knowledge explainability*—Explainability is measured by examining the degree to which formalism facilitates methods for generating human-readable explanations of the reasoning process, making it comprehensible to users or stakeholders. An indication of knowledge explainability would be given by the normalized length of the path from premises to conclusions of the reasoning processes.
- *Knowledge certainty*—Certainty can be quantified using probabilistic measures or certainty factors that assign degrees of belief or confidence to individual pieces of knowledge within the representation. An indication of knowledge certainty would be given by the probability of correct/incorrect individual knowledge pieces over the total of knowledge pieces.
- *Knowledge consistency*—Consistency can be assessed by checking whether the representation provides mechanisms to detect and resolve conflicts or contradictions, ensuring that the knowledge remains internally coherent. An indication of knowledge consistency would be given by the consistency of knowledge when comparing coherent pieces, i.e., non-conflicting knowledge pieces against the total.
- *Knowledge confidence*—Confidence in a formalism can be evaluated by examining its provisions for attaching confidence scores or measures to conclusions and assessing the mechanisms for assessing the reliability of knowledge sources. An indication of knowledge confidence would be given by statistical confidence or the confidence ratings given by users or experts.
- *Knowledge expandability*—Expandability can be evaluated by examining whether the representation supports mechanisms for inference, induction, or the incorporation of new facts, rules, or domain-specific knowledge over time. An indication of knowledge expandability would be given by the efficiency of deriving new insights relative to the addition of new knowledge.
- *Knowledge scalability*—Scalability can be assessed by evaluating the performance of the knowledge representation as the size and complexity of the knowledge base grow. It involves examining factors such as response times, resource utilization, and efficiency of storage. An indication of knowledge scalability would be given by the system's performance under maximum load over its baseline performance.

These criteria collectively help evaluate the suitability and quality of a knowledge representation formalism for various applications and domains. The choice of a representation should align with specific objectives and requirements related to the criteria. Based on the above description of measurements, Table 1 lists a set of defining questions to be answered for the qualitative assessment. These questions constitute a methodology that is used to assess conformance to the criteria in the range of high, moderate, and low conformance. It should be emphasized that the comparison of the related work in this survey relies on assessments based on these qualitative measurements. For further clarity of the criteria, some indicators have been also presented. However, a rigorous formalization of the criteria

with well-defined quantitative metrics is an interesting direction for future research toward a unified knowledge framework.

Table 1. Questions defining the assessment criteria.

Criteria	Relevant Assessment Defining Questions
<i>Formalism</i>	What formal language or representation framework is used to express knowledge? What is its expressive power and semantic clarity? Is knowledge represented in a clear, unambiguous, and straightforward manner?
<i>Accuracy</i>	How accurate is the representation of knowledge in reflecting the true state of the world? Can the knowledge base be evaluated for its overall correctness or accuracy? How many times is the knowledge overall correct?
<i>Precision</i>	Does the representation allow for precise and detailed domain descriptions? Does the same input give the same output? Can it capture fine-grained distinctions and nuances? Does it support the representation of explicit constraints to enforce precision?
<i>Coverage</i>	Is everything covered? How comprehensive is the representation in covering the relevant aspects of the domain? Does it handle missing or incomplete knowledge gracefully? Are there mechanisms to handle uncertainty or partial knowledge gaps?
<i>Generalizability</i>	Does it generalize properly or stand for a specialized objective? To what degree results can apply to a broader context?
<i>Explainability</i>	How transparent and explainable is the reasoning process and the conclusions drawn from the knowledge base? Does the representation provide mechanisms for generating explanations, justification, or evidence to support the derived conclusions?
<i>Certainty</i>	Does the representation support the encoding of uncertainty or degrees of belief? How certain is the representation of knowledge in the reasoning process? Is there a mechanism to quantify and reason with uncertain or probabilistic knowledge?
<i>Consistency</i>	Does the representation allow for the detection and resolution of conflicts between facts/conclusions? How sufficient are conflicting facts/conclusions resolved? Is there a mechanism to handle inconsistencies and ensure coherence in the knowledge base?
<i>Confidence</i>	Can the representation assign confidence levels to the conclusions drawn from the knowledge base? What are the levels of confidence? Are there mechanisms to assess the reliability or trustworthiness of the knowledge sources and inference results?
<i>Expandability</i>	Does the representation support the derivation of new knowledge from existing facts or rules? Are there mechanisms for logical inference or induction to generate new knowledge? Can the representation accommodate the evolution of knowledge over time?
<i>Scalability</i>	Can the representation handle complexity and scalability in time and space? Can it effectively handle large amounts of knowledge, complex relationships, or domains? Does it provide mechanisms for efficient storage, retrieval, and reasoning with knowledge?

3.3. Assessment Results

Table 2 provides a comparative overview of how the listed formalisms generally align with the given criteria. It is important to note that each formalism has its own strengths and weaknesses, and its suitability varies depending on the specific use case, problem complexity, available data, and the expertise of practitioners. The evaluation should consider specific requirements and context when choosing an appropriate knowledge representation approach. Consequently, the ratings in the table are generalized and the assessment is based on a general perspective on how some of the formalisms might align with the criteria based on their typical characteristics and may not reflect the nuances of individual applications. Different implementations or variations in each formalism may have different levels of support for the listed criteria.

Table 2. Evaluation Results of Knowledge-based Approaches.

Criteria Category	Quality Criteria ¹	Bayesian/ Causal ²	Standard Logic	Rules	Graphs ³	Trees ³	Simulation	Instance	Neural Networks	LLMs/ Generative
Declarative	Formalism	High	High	High	High	Medium	Medium	Low	Low	Low
	Accuracy	High	High	High	High	High	High	High	High	Medium
	Precision	High	High	High	High	Medium	High	Medium	Medium	Medium
	Coverage	Medium	Medium	Medium	High	Medium	High	Medium	Medium	Medium
	Generalizability	Medium	Medium	Medium	Medium	Medium	Medium	Low	Medium	High
	Explainability	Medium	High	High	High	High	Medium	Medium	Low	Low
Procedural	Certainty	High	Medium	Medium	Medium	Low	Medium	Low	Low	Medium
	Consistency	High	Medium	Medium	High	High	High	Medium	Medium	Medium
	Confidence	High	Low	Medium	Medium	Low	Medium	Low	Medium	Medium
	Expandability	High	High	High	High	Medium	Low	Low	Low	Low
	Scalability	Medium	Low	Low	High	Medium	Medium	High	Medium	Medium

¹ It should be noted that knowledge representation is strongly associated with and inherent to the reasoning process. For example, symbolic reasoning requires a symbolic representation; therefore, symbolic reasoning based on knowledge graphs requires first transferring the graph to a symbolic form. ² Belief networks are a subcategory of graphs that allow for Bayesian reasoning. However, graphs are not a prerequisite; rather, the main knowledge representation is the joint probability tensor. ³ Reasoning in graphs and tree structures implies traversing through the graph or traversing through the tree from root to leaf nodes.

For the analysis, we assume that knowledge presentations are not allowed to be excessive in size. Rather it is assumed that the storage is limited; therefore, the entities and facts of the world cannot be exhaustive. Wherever required, tradeoffs will be explained and the assessment justified. It is important to note that the suitability of each formalism varies depending on the specific use case, problem complexity, available data, and the expertise of practitioners. The assigned grades are provided based on general characteristics and may not reflect the nuances of individual applications.

3.3.1. Bayesian/Causal Networks

In terms of declarative aspects, Bayesian/Causal Networks excel in providing robust frameworks for representing knowledge via probabilistic relationships, offering precise and formal representations aligned with real-world uncertainties and complexities. They accurately model complex probabilistic dependencies and causal relationships within the domain, but might face challenges in handling missing or incomplete knowledge beyond encoded probabilities or causality. While these models generalize within encoded probabilistic relationships, they might struggle with entirely new scenarios.

With regard to their procedural qualities, Bayesian/causal reasoning approaches effectively encode uncertainty and probabilities, ensuring high consistency and reliability in the reasoning process. However, due to their probabilistic nature, their explainability might be less intuitive, impacting the explainability of the reasoning process.

In summary, probability-based formalisms excel in representing uncertainty and offer high accuracy, consistency, and confidence. They provide precise and comprehensive descriptions but might encounter challenges in scalability and intuitive explainability due to their probabilistic nature. Additionally, while expandable, generalizing beyond encoded relationships might be a limitation.

3.3.2. Standard Logic-Based Formalisms

In terms of the declarative aspects, standard logic-based formalisms are high as they offer rigorous, well-defined formal languages to express knowledge, providing precise and unambiguous representation. These formalisms guarantee accuracy, ensuring that conclusions drawn from the knowledge base are logically correct and allow for precise and detailed descriptions of the domain, capturing fine-grained distinctions and enforcing precision via explicit constraints or rules. Nonetheless, while comprehensive within the

defined logical constraints, they might struggle to generalize beyond the scope and/or handle incomplete knowledge beyond the logic of the defined rules. On the other hand, they inherently provide transparent reasoning processes, offering clear and explainable justifications via logical rules.

In terms of procedural qualities, standard logic ensures strict adherence to formal rules, allowing for the detection and resolution of conflicts, and ensuring coherence in the knowledge base. However, it provides no mechanisms to deal with inconsistencies due to non-monotonicity where contradicting knowledge may exist. They represent knowledge deterministically, without encoding uncertainty and without assigning confidence levels to conclusions or assessing the reliability of inference results, which can be both an advantage and a limitation in handling uncertain domains. Standard logic formalisms are highly expandable, deriving new knowledge from existing facts or rules, but may fail with scalability when dealing with vast amounts of knowledge or complex relationships due to computational complexity.

In summary, standard logic-based formalisms excel in providing precise, and accurate, representations within their defined scope and rules. They offer excellent explainability and expandability, but may face challenges in handling uncertainty, scalability, and generalizing beyond explicitly defined logical rules.

3.3.3. Rule-Based Formalisms

Rule-based systems provide a structured and explicit representation of knowledge, leveraging if-then rules to offer a clear and formal language to express domain-specific information. These systems stand out for their accuracy in reflecting domain-specific knowledge and rules, ensuring that the conclusions drawn align well with the established rule set. Additionally, they offer a high degree of precision by offering detailed descriptions and constraints via explicit rules, enabling fine-grained control and specific reasoning within the domain.

While these systems comprehensively cover the knowledge explicitly represented in rules, they might encounter challenges in handling missing or incomplete information not explicitly encoded within the rule set, generalizing beyond the defined rule set. The reasoning process is based on explicit rules, allowing for traceability and justifications for conclusions. This clarity in the reasoning process makes them highly explainable and suitable for presenting transparent decision-making pathways.

Nonetheless, these systems may encounter limitations in handling certainty and consistency, especially in dealing with uncertainty or encoding probabilistic beliefs as effectively as probabilistic models. While they follow a structured order of execution (forward or backward chaining) to resolve conflicts or inconsistencies in a deterministic manner, they lack mechanisms to prevent contradictory rules. Furthermore, while they allow for the incorporation of new rules or modifications to existing ones, expanding to entirely new knowledge domains might necessitate significant effort and expertise. Managing a large number of rules might lead to scalability issues, impacting performance and complexity as the rule base grows.

In summary, rule-based systems offer structured and explicit representations with high accuracy and precision. They excel in explainability and expandability, but might face challenges in dealing with uncertainty, scaling to large rule bases, and generalizing to new domains.

3.3.4. Graph-Based Formalisms

Graph-based representations are adept at reflecting real-world entities and relationships precisely, maintaining accuracy by directly mirroring the connections and attributes present in the domain. They facilitate nuanced and detailed modeling of relationships, supporting precise descriptions within the domain by allowing for a granular depiction of associations and attributes. They effectively accommodate diverse information and multiple facets of a domain and provide a visually intuitive representation of the domain's generic relationships.

However, the extent of their generalizability may differ depending on the domain, and in some cases, adaptations or extensions might be necessary for broader applicability.

Despite their strengths, they are not directly suitable to explicitly capture uncertainty in reasoning within the graph structure. It necessitates additional mechanisms or annotations to express degrees of belief or uncertainty explicitly. Furthermore, while graph structures inherently ensure consistency by facilitating conflict detection and resolution, expressing absolute confidence levels can be challenging without specific scoring mechanisms. On the other hand, knowledge graphs are highly expandable. They allow for the incorporation of new nodes and edges, facilitating the integration of evolving domain knowledge and the derivation of novel insights. Additionally, they handle scalability efficiently, managing large volumes of data, complex relationships, and diverse domains effectively.

In summary, graph-based representations, particularly knowledge graphs, excel in various aspects of knowledge representation, but might have certain limitations in expressing certainty, confidence, and generalizability across diverse domains.

3.3.5. Tree-Based Formalisms

Tree-based representations, typified by decision trees and random forests, organize information hierarchically, although they might be less expressive in capturing intricate relationships compared to graph-based approaches. They strive for high accuracy by creating straightforward yet effective models to precisely classify or predict outcomes based on given attributes. These models can represent domain-specific details, but might not explicitly capture nuanced relationships as graph-based structures, due to their hierarchical nature. While they can cover various attributes, their hierarchical organization may limit the coverage of complex relationships found in certain domains. In terms of generalizability, decision trees may perform well within specific domains, but might struggle to adapt to significantly different contexts without retraining or modification. However, they excel in explainability by providing transparent rules and paths, making it easy to interpret the decision-making process.

With regards to their procedural qualities, tree-based structures might lack inherent mechanisms to explicitly represent uncertainty or probability. They ensure consistency by partitioning data based on specific rules, ensuring coherent classification or prediction within each node. However, they might not inherently provide confidence levels for predictions, often requiring additional techniques to estimate reliability. Regarding expandability, expanding decision trees might necessitate restructuring, and adding new attributes might not seamlessly integrate without altering the entire structure. Scalability challenges might arise when handling extensive and diverse knowledge, potentially impacting storage and efficiency.

In summary, tree-based representation formalisms, like decision trees and random forests, effectively provide consistent and accurate results with high explainability. However, they may face limitations in handling uncertainty, capturing complex relationships, and scaling to diverse domains.

3.3.6. Simulation-Based Formalisms

Simulation-based models, often using mathematical or computational models, may not offer the expressiveness found in logic-based or graph-based representations. However, they aim for high accuracy by reflecting real-world behavior and providing precise predictions based on well-defined models. Simulations excel in precision, capturing fine-grained nuances and behaviors of the modeled systems comprehensively. These representations strive to cover a broad spectrum of behaviors and scenarios, although achieving this may require thorough modeling of each aspect. While they generally generalize within modeled domains, they might face limitations when applied to vastly different scenarios.

Simulations generate results according to certain distributions and input parameters, maintaining consistency by adhering to defined physical laws or constraints within simulated environments. However, they might not explicitly capture uncertainty inherent in real-world scenarios, hindering certainty representation. Interpreting complex interrela-

tionships within simulations might require specific expertise, impacting their explainability. Confidence estimation relies on input data accuracy and model validation, but lacks inherent methods to quantify confidence levels directly. Expanding simulations often involves modifying underlying models or parameters, posing challenges in accommodating additional or new knowledge seamlessly. Scaling simulations might be computationally complex, especially for real-time, large-scale, or dynamic systems.

In summary, simulation-based representations offer accurate, precise, and comprehensive representations of modeled systems. However, challenges in scalability, interpretability, and adaptability to new knowledge may limit their expandability and generalizability across diverse domains.

3.3.7. Instance-Based Regression Formalisms

Instance-based methods, encompassing case-based, regression, and kernel function models, lack a formalized language for explicit knowledge expression, relying on storing and retrieving instances or patterns. However, they excel in accuracy by leveraging similarity among instances or patterns for predictions or classifications. While they can handle a wide variety of instances, these methods may suffer from data sparsity or biased instance representation, impacting their coverage. They generalize well within their trained instance spaces, but face limitations in adapting to unseen or significantly different instances. These methods provide moderate explainability based on instance similarity, but might struggle with complex explanations in high-dimensional or non-linear spaces.

Lacking explicit mechanisms for uncertainty or confidence levels associated with instances or predictions, they may not adequately encode uncertainty. Handling inconsistencies via proximity measures for instance retrieval might not inherently resolve conflicts in contradictory instances. Confidence estimation relies on proximity measures or statistical evaluations, but it is not inherent to these methods. Although accommodating new knowledge might be challenging without extensive retraining or model modification, they scale well by relying on efficient instance retrieval and comparison techniques.

In summary, instance-based methods excel in accuracy and scalability, effectively capturing similarities within instance spaces. However, they may lack explicit mechanisms for dealing with uncertainty, inconsistent knowledge, or accommodating new knowledge without significant retraining. Their generalizability might be limited to the scope of the training instances.

3.3.8. Neural Network-Based Formalisms

Neural network-based approaches, including DNN, CNN, RNN, LSTM, DRL, DQN, etc., encode knowledge implicitly in their weights and activations, lacking explicit representation in a formalized language. These models often achieve high accuracy in predictions or classifications, especially with large datasets. However, while they capture data patterns and features well, they might lack precise definitions of explicit rules or constraints, impacting precision. Their coverage spans a wide range of data patterns, but may suffer from biases in training data, affecting coverage of less frequent patterns. Despite good generalization within trained data distributions, they might lack robustness when applied to unseen or significantly different data distributions. Neural networks are often considered black-box models, posing challenges in explaining decisions, particularly in deep and complex architectures.

Concerning procedural qualities, they lack mechanisms to encode uncertainty or confidence levels explicitly, resulting in non-probabilistic outputs. With a lack of mechanisms dealing with conflicting knowledge, these models might overfit or underfit data, leading to inconsistencies in predictions. While techniques like softmax probabilities estimate confidence, accuracy in reflecting true confidence might vary. Furthermore, via scalable with increased data sizes and handling complex relationships, integrating new knowledge often demands retraining or fine-tuning, posing challenges in incorporating new information without compromising prior learning.

In summary, neural networks showcase high accuracy and scalability, being potent tools for pattern recognition and complex data processing. However, their limitations lie in the explicit handling of uncertainty, explainability, expandability, and generalization to new or diverse data.

3.3.9. Generative AI-Based Formalisms

Generative models, including Large Language Models (LLMs), Variational Autoencoders (VAEs), and Generative Adversarial Networks (GANs), learn implicit knowledge from data without an explicit formalized structure. While they achieve high-quality outputs, accuracy may vary, particularly with limited or biased training data. They also lack precise encoding of rules or constraints found in other formalisms to capture data patterns and nuances. Covering a wide range of data patterns, generative models might struggle with rare or outlier patterns due to biases in training data. On the other hand, their generalizability surpasses neural-based approaches due to a two-step learning approach, initially gathering general knowledge from vast data and fine-tuning for specific use cases.

In addition to limited explainability, they also lack explicit certainty or confidence expression in generated outputs, requiring proper prompt-based context to improve output quality and resolve conflicting or ambiguous information. Generative models might offer confidence scores, yet their accuracy in reflecting true confidence levels varies. Integrating new knowledge requires retraining or fine-tuning, challenging incremental learning without forgetting previous knowledge.

In summary, these models excel in generating content, displaying scalability, coverage across patterns, and generalization. However, they lack explicit handling of uncertainty, struggle with explainability and expandability, and are susceptible to data biases.

4. Knowledge-Driven Applications in Telecommunications

Knowledge-based applications in telecommunications in general, and in wireless communications in particular, offer several advantages. The advantages refer to all aspects of telecommunications, including the efficient operation of the network, controlled utilization of resources, provision of advanced and personalized services, and improved security and maintenance of the telecommunication system. More specifically, some of the advantages of knowledge-based applications for telecommunication use cases are as follows.

- *Efficient resource management*—Knowledge-based systems aid in optimizing resource allocation, such as spectrum utilization, by intelligently predicting and adapting to network demands based on historical data and patterns. This improves overall network efficiency.
- *Enhanced network intelligence*—By leveraging knowledge bases, wireless networks can make informed decisions in real time. This intelligence enables networks to adapt to changing conditions, predict potential issues, and proactively address them, leading to improved performance and reliability.
- *Dynamic network configuration*—Knowledge-based applications facilitate dynamic re-configuration of network parameters based on learned patterns, allowing networks to self-optimize and adapt to varying conditions, such as changes in user behavior or environmental factors.
- *Context-aware services*—Knowledge-based systems enable the provision of context-aware services. By understanding user preferences, locations, and behaviors, wireless networks can deliver personalized and location-specific services, enhancing user experience.
- *Optimized quality of service (QoS)*—Knowledge-based applications can predict network performance based on historical data and adjust parameters to ensure consistent QoS. This ensures that critical services receive adequate resources and prioritization.
- *Predictive Maintenance*—Knowledge-based systems can predict potential failures or maintenance needs within wireless network components by analyzing historical performance data. This predictive capability minimizes downtime and prevents unexpected failures.

- *Improved security and anomaly detection*—Knowledge-based approaches help in identifying abnormal network behavior by recognizing patterns that deviate from the norm. This aids in detecting anomalies and potential security threats, allowing for quicker responses to security breaches.
- *Resource optimization in dedicated networks*—In dedicated environments, such as Internet-of-Things (IoT) and factory environments, knowledge-based systems can optimize device connectivity, manage data flows, and coordinate communication among devices, while ensuring efficient utilization of resources and minimizing energy consumption and communication delays.

Overall, knowledge-based applications empower wireless networks to become more adaptive, intelligent, and responsive to dynamic conditions, leading to enhanced efficiency, improved services, and better overall performance. They do so for all different levels of telecommunication systems abstraction, including system, service, and network levels. The different levels operate at different data levels, making certain KRRs more suitable than others. There are various forms of representing and using knowledge corresponding to different levels of abstraction. Figure 2 illustrates three different levels of abstraction: (i) system and network operation at the Operation and Management (OAM) level, (ii) service and network orchestration at the service management and orchestration (SMO) level, and (iii) network infrastructure and transmission medium at the radio access/core network (RAN/CN) level. At the RAN level, physical phenomena, such as electromagnetic and wave propagation, are typically expressed in mathematical equations that concretely relate to the physical parameters. They are majorly derived from empirical observations and theoretical foundations of many sciences including physics, mathematics, information, and communication. Equations about physical entities are the lowest level of abstraction in the knowledge representation line and have been used to develop algorithms that control the communication infrastructure to optimally use the resources of various media, such as the radio environment. Moving from physics towards the digitalization line, more abstract levels of information are to be identified that are closer to human cognition capabilities. The capability of the human mind to think at higher levels of abstraction also requires the movement from equation representations to a more human language-like form of representing knowledge. The representation of knowledge in telecommunication systems is related to the application it is aimed at facilitating. Between the infrastructure that benefits from the environment, e.g., in the RAN domain, and the operation center benefiting from the users and operators, e.g., in the network operation center (NOC) domain, from customer relationship managers (CRMs) to operations support systems (OSSs), there is the service operation domain where the telecommunication system should also generally benefit from to improve its operations.

As illustrated in Figure 2, the knowledge in the different domains is associated with different reasoning processes pertinent to operations in the corresponding domains. The NOC domain implies user interaction and human reasoning based on natural language understanding and understanding capabilities, while the operation of the system requires knowledge representations that allow for the automation of the reasoning processes within the system. Getting closer to the RAN domain, the reasoning process is based on representations that can express numerical expressions. Furthermore, RAN is naturally distributed: the antennas, as in distributed antenna systems (DAS) and/or the base stations, are always geographically spread out. As a consequence of that, it is natural to study how intelligence, i.e., AI algorithms, should be distributed. Do we bring data to the algorithm, or do we distribute the algorithms? If we distribute, then how is this efficiently performed, and which technologies are suitable including transfer learning, federated learning, and split learning? The different categories of KRR, including learning and inferring, are described in the next section, and their application is further elaborated in the use-case description section.

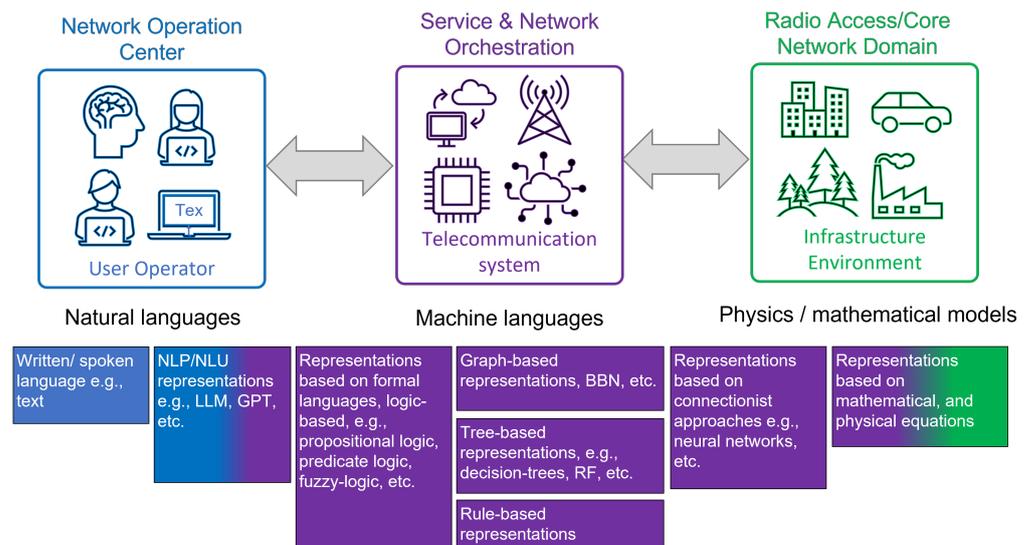


Figure 2. Knowledge representation and reasoning categories.

4.1. Network Operations

Large Language Models (LLMs), such as encoder-based models like BERT and causal/Generative types like GPTs, have a significant impact on knowledge-intensive tasks in development and operations in the telecommunications domain. LLMs are large and deep neural networks whose underlying architecture is in the form of transformer, self-attention, convolutions, etc. The knowledge is implicitly captured in a high-dimensional vector space via the weights of the model. The models learn from vast volumes of textual data via a process called pre-training, where the input data plays a significant role in creating an efficient model. Furthermore, one of the key benefits of LLMs is the ease of interacting and querying the knowledge they contain. Unlike many other knowledge representation systems, LLMs do not necessitate the use of any specialized querying language or formalisms. Rather, the input/query usually involves a natural language expression in plain text.

LLMs can be directly used for telecom purposes. However, for better performance in the applications that are heavily using the telecom terminologies in order to understand those terminologies and the relation among them, general LLMs need to be domain adapted, meaning the domain knowledge needs to be encapsulated into the models in the pre-training stage. The requirement of domain adaptation is creating a large telecom dataset [26]. The telecom domain, being on the frontiers of digital transformation, has the knowledge already documented in textual format in standard format, such as specifications of the 3rd Generation Partnership Project (3GPP) or company-wide documentation of product specifications handcrafted by subject matter experts or any technical reports in the field.

Figure 3 shows two methods for building telecom-specific language models. In one method, existing pre-trained models in the general domain are used as a starting point, and the model is further pre-trained on a telecom dataset. In the second method, the models are trained from scratch using a mix of general and telecom domain datasets. The two approaches result in similar models that have a representation of general and in-domain knowledge. The key difference is that the second approach has a more tailored vocabulary and requires more training time and resources; as a result, it becomes more expensive [26].

The authors in [27] explored domain adaptation strategies for BERT-like models in the telecom domain. They investigated three approaches demonstrating that continued pre-training on telecom-domain text starting from a general-domain checkpoint, pre-training from scratch on telecom-domain text, or a combination of both yielded varying levels of performance and generalizability.

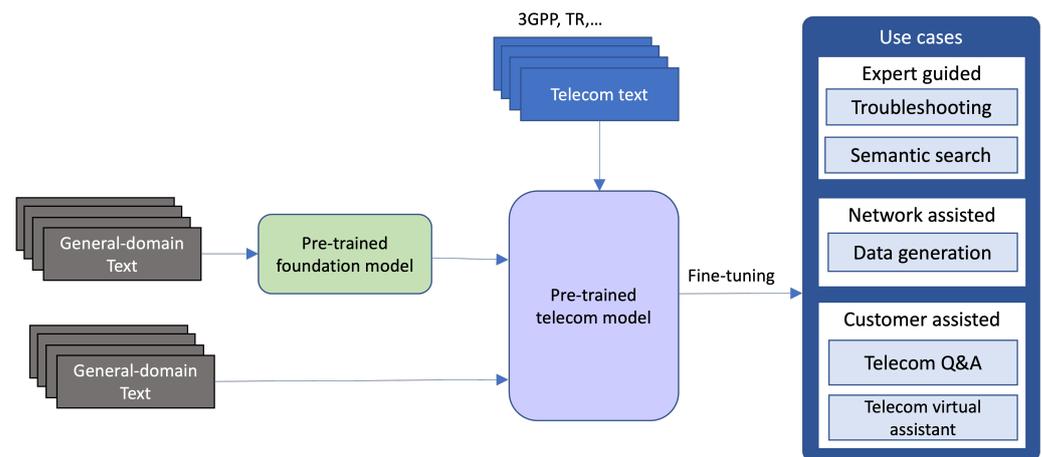


Figure 3. Pre-Training Telecom Language Models (LMs)—Incorporating telecom-domain knowledge into the models in the pre-training phase.

LLM use cases towards research and development can be leveraged in three directions: (i) *expert-guided* use cases, (ii) *network-assisted* use cases, and (iii) *customer-assisted* use cases. The intended user of the first category can be an expert in the field, a NOC operator, a product developer, a support service, a researcher, a 3GPP delegate, or any other individual involved in telecommunications seeking to address their information needs. In this first category, LLMs can be used for semantic search such as troubleshooting and test case recommendation. Troubleshooting is a very exhausting and time-consuming task that requires lots of expertise. One application of troubleshooting is answering the trouble reports (TRs) that correspond to some observed fault in a product. To this end, the knowledge source can be text descriptions from tickets. The authors in [28] present a BERT-based model designed for troubleshooting in the telecommunication industry. This approach utilizes a two-stage process to retrieve and re-rank candidate solutions for reported problems, outperforming non-BERT models like BM25. The efficacy of LLMs in enhancing accuracy in recommending solutions within a telecom-specific context is proven. Some trouble reports might be written differently, but in essence, they may refer to the same problem; therefore, the same response would apply. Duplicate trouble tickets lead to extra effort and cost. Domain-adapted LLM can be applied to identify duplicates. In [29], a BERT-based retrieval system is introduced for trouble reports (TRs) in large software-intensive organizations. The work aimed to identify duplicate TRs efficiently by leveraging domain-specific knowledge, specifically in telecommunications. It addressed the challenge of generalization to out-of-domain TRs and demonstrated that incorporating domain-specific data into fine-tuning improved overall model performance and generalizability.

Furthermore, in [30], an end-to-end recommendation system is proposed for troubleshooting failed test cases using NLP techniques. The study explores various models, including language models, for information retrieval and re-ranking to address the challenges posed by massive data in the troubleshooting process. This work demonstrated the effectiveness of a two-stage BERT model fine-tuned on domain-specific data in generating recommendations.

In network-assisted use cases, generating synthetic data is a very important usage. Rare events that are difficult to capture in the field can be generated with this data collection process technology, which impacts the model training phase or safe online exploration. For example, the work in [31] focuses on synthetic data generation in the telecom domain for extractive question answering (QA). It develops a pipeline using BERT-like and T5 models to generate synthetic data and achieves state-of-the-art results in telecom QA. The findings suggest that synthetic data, in addition to human-annotated data, can improve model performance.

In the case of customer-assisted use cases, two use cases stand out. The first use case is the question and answering within the domain, and the second use case is the

telecom virtual assistance designed to support field technicians in their daily work. For such purposes, the knowledge source can be text descriptions from any telecom-related document, such as product manuals describing a system's operation. The authors in [32] conducted research on adapting question-answering models to the telecommunication domain. By fine-tuning the pre-trained LLMs with a relatively small dataset, a significant domain adaptation is achieved, as a result reducing the need for extensive annotation efforts. This work emphasized the potential for constructing domain-specialized question-answering systems using LLMs.

4.2. Network Automation

Intent-based network management has been introduced in telecommunications to offer automated control and intelligent optimization of complex networks [33,34]. Intent-based networking (IBN) [35] and Intent-driven Networks (IDNs) have evolved from SDNs and NFVs management disciplines [36] as a successive evolution of policy-based networks (PBNs) [37]. Central to these management disciplines are the network policies governing the network operation. Intent-based operation extends previous approaches in that it first offers an abstraction layer expressing the desired outcome of the network service, and second, it automatically determines how the network should be configured in an efficient and reliable manner [38].

As shown in Figure 4, IBN is characterized by a closed-loop cycle comprising four processes: (i) intent translation, where the abstracted intent is mapped to a network configuration; (ii) intent resolution, where configuration conflicts are resolved; (iii) intent actuation, where the requested service is provisioned; and (iv) intent assurance, where provisioned services are resilient to network changes. In its simplest form, IBNs require a translation of intents to network policies, and finally, to detailed network configurations that govern the infrastructure and the resources. More flexible and dynamic forms require the employment of intent-driven systems with associated reasoning engines.

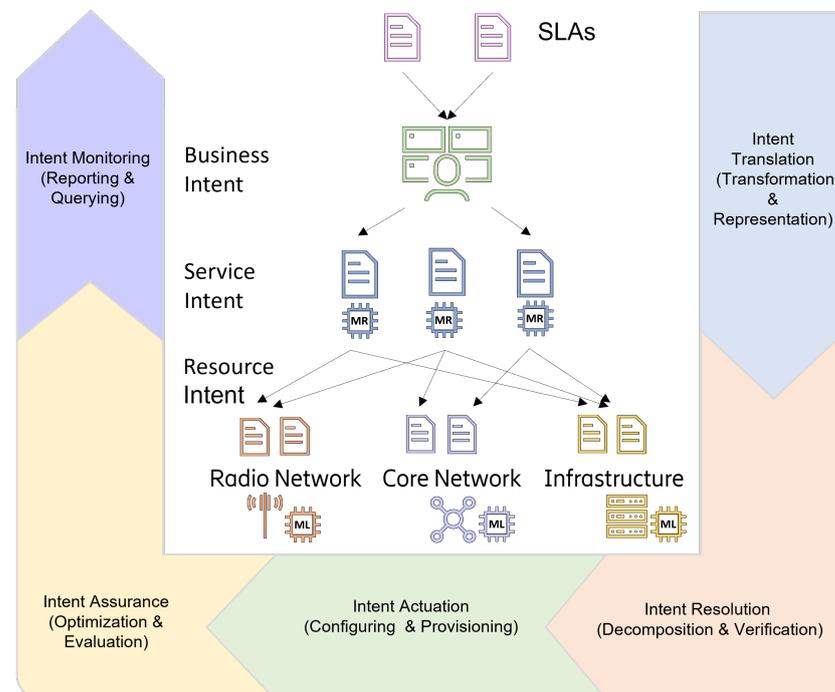


Figure 4. Intent-based automation process cycle.

At the core of all flexible and dynamic intent-driven systems lies the crucial element of knowledge representation. This involves modeling intents in a way that is easily understood by humans while also being executable by machines. In this context, knowledge graphs have emerged as a widely adopted method for representing intents. These

knowledge graphs encompass entities, relationships, and constraints, serving as structured repositories for intent definitions, essentially defining an ontology [39].

However, it is essential to note that knowledge representation is just one aspect of the broader picture. Intent-driven systems rely on reasoning engines as their key components, which can take the form of rule-based logic, advanced semantic reasoning capabilities, or large language models. These engines play a pivotal role in interpreting intents, translating them, and breaking them down into specific network actions. They are specifically designed for tasks such as inferring intent decomposition, resolving conflicts, and optimizing configurations to ensure the provision of services aligned with the intended goals [40].

Another distinctive characteristic of intent-driven systems is their adaptability. Combining knowledge-based systems with machine learning techniques allows these systems to evolve alongside changing network conditions and user requirements [41]. They can be designed to learn from historical data and user feedback, continuously enhancing their ability to comprehend and fulfill intents. Moreover, intent-driven systems contribute significantly to ongoing network improvements. They achieve this by automating routine tasks, optimizing resource allocation, and ensuring compliance with established policies. Their reliability is further bolstered by their ability to anticipate potential consequences of actions, foresee issues, and proactively mitigate them by selecting alternative network configuration solutions that align with the intended objectives.

Advances in natural language understanding, explainable AI, and distributed knowledge graphs are promising future directions to further augment the capabilities of these intent-driven systems, making them indispensable in the face of growing network complexity.

4.3. Network Optimization

Apart from dealing with network demands in terms of operations and automation, telecommunication systems, especially the RAN domain, are expected to support an increased number of differentiated businesses and a massive amount of devices and services. In such a scenario, the complexity of resource allocation decisions will increase beyond existing RAN computational and storage capacity, hence making it difficult to efficiently optimize network configuration and planning. To address the increasing complexity, AI-driven network optimization based on machine learning (ML) techniques has been proposed. Trained on large datasets, generalized ML models are generated with relatively high accuracy and precision. Although the training can be performed centrally, in many cases within RAN, a centralized approach is prohibitive due to requirements on cost, latency, and bandwidth. In the more general case, privacy preservation concerns and/or regulatory limitations advocate in favor of a distributed learning solution such as federated learning, split learning, or distributed learning based on DNN, DQN, DRL, etc. In the case of RAN, distributed learning is also mandated by the nature of RAN, which is inherently distributed. Furthermore, with the recent advances of distributed learning [42], it is now possible to address these challenges and exploit the knowledge that is distributed in many network nodes. As depicted in Figure 5, decentralized learning can be divided into three main categories:

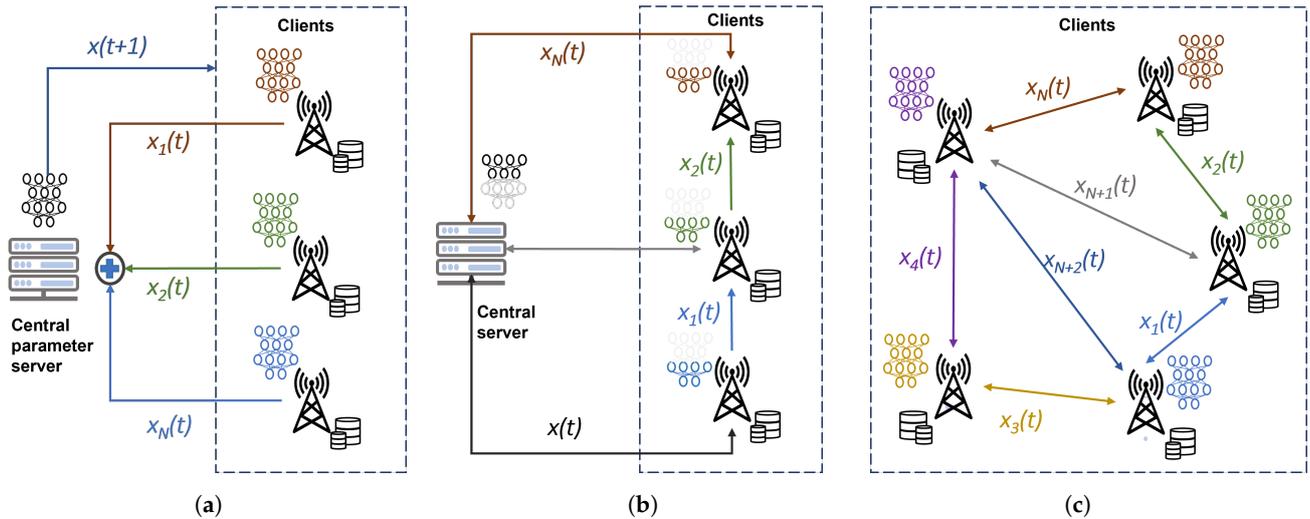


Figure 5. Distributed learning approaches for network optimization. (a) Federated learning setup, with a central parameter server and a number of clients that participate in the training of a machine learning model located in the server; (b) split learning setup, with a central parameter server. Here, all clients have a copy of the partial model and cooperate in training the model in a sequential order; (c) fully distributed learning setup consisting only of clients each having a local copy of the model and cooperating in training the model via direct communication.

- Federated learning**—In federated learning (FL), multiple (possibly large numbers of) clients participate in the training of a shared global model by exchanging model updates with a central parameter server (PS) [43]. This is performed by using an iterative process where each global iteration, often referred to as *communication round*, is divided into three phases:
 1. The PS sends the current model parameter vector x (an M -dimensional vector) to all participating clients. Either the model is sent directly, or a differential compared to the previous round is transmitted.
 2. The clients, the total number of which is N , perform one or more steps of stochastic gradient descent on their own (private) training data and obtain a model update. Specifically, the n th client computes a gradient update x_n (M -dimensional vector).
 3. The M -dimensional model updates x_1, \dots, x_N from the N clients are then sent back to the PS. The PS aggregates these updates by adding them together to obtain the updated model x . A typical aggregation rule is to compute a weighted sum of the updates, that is, compute the M -dimensional total update $\sum_{n=1}^N \alpha_n x_n$ for some weights α_n . Other non-linear aggregation rules are also possible, for example, to safeguard against malicious or malfunctioning agents [44].

The above procedure is repeated until a termination criterion is met. The termination can, for example, be when a desirable training accuracy is achieved, when a predefined number of iteration rounds have been reached, or when a global loss function meets the minimum requirements. The loss function depends on the computational problem intended to be solved using the FL system, and typically, it represents a prediction error on a test dataset. The above procedures are depicted in Figure 5a.

- Split learning**—Split learning (SL) is another cooperative ML technique used to train a neural network model between a client and a central server by splitting the model at a cut layer into a client-side model and a server-side model [45]. The existing SL approach conducts the training process sequentially across clients, where each client calculates the intermediate results and transmits them to the next client in order. Each client passes the intermediate gradient in backward order when casting backward propagation. Furthermore, relevant features from the data are then transmitted to the

central server which processes these features, performs the necessary computations, and sends back the updated parameters. This approach allows the sensitive data to remain on the client's side, ensuring privacy, while still benefiting from the centralized server's computational power. Split learning maintains a partition between data and computations, safeguarding user privacy during model training [46]. A depiction of SL is illustrated in Figure 5b;

- *Fully distributed learning*—In fully distributed learning setups, clients collaboratively perform learning without transmitting model updates to a central parameter server [47]. Instead, all clients have their own instance of the learning model. Each client has its own private training data. The clients are connected according to a connectivity graph such that each client has one or more neighbors. This scenario is illustrated in Figure 5c. The training proceeds by exchanging updates between the clients and their neighbors using, for instance, gossip-based averaging, where nodes iteratively share and merge information to collectively refine their local models. Fully distributed learning operates in a decentralized manner, promoting data locality and fostering collaboration among neighboring nodes while avoiding a central server's reliance [48].

Wireless networks often have limited bandwidth and high latency; therefore, distributed learning is beneficial. Firstly, it minimizes data transmission by processing data locally on sites, which reduces the risk of exposing sensitive information. Instead of transmitting raw data to a central server, only model updates or aggregated information are shared. Secondly, utilizing multiple sites allows for scaling machine learning tasks while reducing latency in decision-making and increasing system efficiency and robustness. If a site in the network fails or has no available computational resources, the overall learning process can continue without significant interruptions. Finally, transferring learned models creates a smaller network footprint than transferring raw data, thus offloading the user plane. A combination of a low network footprint and better resource utilization yields energy savings.

Future research in distributed learning needs to address data, system, and model heterogeneity. Data heterogeneity relates to challenges with domain adaptation and identical and independent distributions, (i.i.d). System heterogeneity refers to system-related inconsistencies encountered between different sites that may skew the overall process in terms of robustness and resilience. Whilst, model heterogeneity refers to FL with different model architectures.

4.4. Limitations, Challenges, and Future Research

Knowledge-based applications in wireless networks offer several advantages and prove themselves invaluable in addressing the complexity of telecommunications use cases. Despite the advantages, knowledge-based applications are limited in many aspects as they introduce new demands to be met. The following list, which is by no means exhaustive, summarizes some of the limitations of knowledge-based applications in wireless communications. It should be also noted that the magnitude probability of occurrence and ease of mitigation of what is listed below is different in each category of KRR.

- *Data quality*—Knowledge-based applications heavily rely on historical and real-time data and data quality. If data is inaccurate in case of data poisoning and harmful content or is incomplete or outdated, it can lead to incorrect decisions and predictions.
- *Initial training*—Training certain knowledge-based models such as LLMs initially requires a substantial amount of data and computational resources. This can be a barrier to adoption, especially for smaller networks.
- *Data privacy*—The collection and analysis of user data or exposure/training on a company's proprietary data for knowledge-based applications raise privacy concerns. Striking a balance between data-driven insights and user privacy is a significant challenge in the case of private information exfiltration. There is a chance that the data being used for training is exposed in the output in the case of Generative AI type of use cases; therefore, data source and proper curating of data is very important.

- *Security*—Protecting the knowledge base from intervention threats and ensuring the security of data used for training and decision-making is a significant challenge.
- *Trustworthiness*—resilience, fairness/biases, and hallucinations are important challenges in Generative AI-based formalisms, where the research in this area is still fairly new.
- *Scalability*—As wireless networks grow in size and complexity, managing and analyzing vast amounts of data can become challenging in wireless communication, and optimizing spectrum allocation and management in real-time is complex. Knowledge-based systems must adapt to changing spectrum conditions and demands. Complexity and scalability issues may arise in knowledge-based systems.
- *Resource consumption*—Knowledge-based applications can be computationally intensive and costly, requiring significant processing power and memory. This can strain network resources, especially in resource-constrained environments.
- *Interoperability*—Integrating knowledge-based systems with existing network infrastructure can be complex. Modern wireless networks encompass various technologies (e.g., 5G, IoT, and mobile edge computing) and require handling diverse types of data. Ensuring compatibility and smooth integration is a challenge.
- *Real-time decision making*—Wireless networks are inherently dynamic, and adapting knowledge-based models to rapidly changing conditions can be difficult. Ensuring that knowledge-based systems can make timely decisions in dynamic environments is challenging.

The limitations listed above have a major impact, mainly on the technology aspects of knowledge-based applications. Further, other partly non-technical challenges governing the usage of advanced AI technologies in telecommunications also comprise

- *Ethical considerations*—The use of user data and AI algorithms in wireless networks raises ethical concerns, including bias in decision-making and the responsible use of AI technology.
- *Regulatory compliance*—Compliance with data protection and privacy regulations, e.g., General Data Protection Regulation (GDPR), poses challenges for knowledge-based applications that involve user data.
- *Human-machine interaction*—Ensuring that knowledge-based systems can effectively interact with human operators and users is a challenge, particularly in critical applications.
- *Cyber-security threats*—In the age of digitization, protecting the knowledge base from cyber threats is a significant challenge for the protection of telecommunications businesses. Access to knowledge base contents implies access to critical business processes and technology know-how. Cyber attacks may enable the infection of knowledge bases with malicious knowledge and misbehaving models.

The rich range of limitations and challenges and the use-case discussions reveal that knowledge-based applications are still in their infancy. In addition to the future work suggested above for the top three technologies, some further future research and development directions can be outlined based on the general challenges and limitations discussed. To this end, we envisage different research areas on how to use, experiment, and develop knowledge-driven solutions for wireless networks:

- i. *Optimization* of network operation including advanced self-optimized, self-configured, and self-healed networks where resource allocation is performed intelligently and in real-time based on context awareness and accurate data;
- ii. *Security* of knowledge utilization to improve privacy-preserving techniques and enhance security protocols in wireless networks. The former should aim at using knowledge representation to understand data sensitivity and reasoning to enforce data protection policies, while the latter should aim to understand complex attack patterns and to implement proactive defense mechanisms;
- iii. *Automation* where KRR and emerging technologies like 5G/6G, IoT, and edge computing are integrated to develop cross-domain applications enabling network automation;

- iv. *User experience* where KRR is utilized to model user preferences or behavior and to adapt network services accordingly for enhanced user experience. In addition, KRR can be also utilized to implement QoS-aware protocols that can dynamically adjust to meet the varying service quality requirements of different applications;
- v. *Sustainability* where KRR is applied to reduce the carbon footprint of wireless infrastructure and improve the energy-efficient operation of the network, as well as to model renewable energy sources while integrating them into the network's energy management strategies;
- vi. *Interoperability* that require working towards standardizing KRR approaches for wireless networks to ensure interoperability between different systems and vendors, as well as collaborating with international bodies to develop guidelines and best practices for the introduction of KRR in telecommunications.

Each of these research areas not only presents opportunities for technological advancement, but also indicates the key research directions, such as the need for robust data management, advanced algorithmic development, and the balancing of efficiency with user privacy and security concerns. Collaboration between academia, industry, and regulatory bodies will be crucial to address these challenges and harness the full potential of KRR in wireless communications.

5. Conclusions

In this article, we have identified different categories of knowledge representation and reasoning formalisms that can be used in telecommunication systems. We further defined a methodology to evaluate the identified categories based on a set of qualitative criteria. The criteria have been defined and the conformance of the different approaches have been discussed. From the evaluation, we could conclude that different use cases are better served by different knowledge-based approaches. Further, we have also identified the application domains and discussed the knowledge-driven evolution of telecommunication concerning use cases targeting network operation, automation, and optimization objectives. Overall, knowledge-driven applications empower wireless networks to become more adaptive, intelligent, and responsive to dynamic conditions, leading to enhanced efficiency, improved services, and better overall performance.

Generally, current knowledge representation and reasoning applications in wireless communications struggle with dynamic environmental adaptation, real-time decision-making, and efficient resource management, while also facing challenges in ensuring security and user privacy. Future research directions should focus on developing sophisticated, context-aware models for dynamic networks, enhancing self-optimization, improving security protocols, and integrating with emerging technologies like 5G/6G and IoT, all while maintaining energy efficiency and global interoperability. Within the key AI technologies, future studies should focus on advances with LLMs and generative approaches, extensions of knowledge graphs, and enhancements of deep AI technologies for distributed intelligence. To this end, future standards and applications strengthen the importance of knowledge in AI, and its deployment in telecommunication businesses and services is expected to only amplify.

Author Contributions: Conceptualization, G.P.K.; methodology, G.P.K.; validation, G.P.K., and S.S.; formal analysis, G.P.K.; investigation, G.P.K., S.S. and R.M.; resources, G.P.K., S.S. and R.M.; data curation, G.P.K., S.S. and R.M.; writing—original draft preparation, G.P.K.; writing—review and editing, G.P.K., S.S. and R.M.; visualization, G.P.K., and S.S.; supervision, R.M.; project administration, R.M.; funding acquisition, R.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Acknowledgments: The authors like to thank Fitsum Gaim Gebre, Jyotirmoy Banerjee, Nikhil Korati Prasanna, Pratyush Kiran Uppuluri, Sarbashis Das, and Sathiyanyanan Sampath, all from Global AI Accelerator, Ericsson, for fruitful discussions.

Conflicts of Interest: The authors of this research study were employees of Ericsson. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Creswell, A.; White, T.; Dumoulin, V.; Arulkumaran, K.; Sengupta, B.; Bharath, A.A. Generative adversarial networks: An overview. *IEEE Signal Process. Mag.* **2018**, *35*, 53–65. [[CrossRef](#)]
2. Gozalo-Brizuela, R.; Garrido-Merchan, E.C. A survey of Generative AI Applications, *arXiv* **2023**, arXiv:2306.02781.
3. Niemöller, J.; Sarmonikas, G.; Washington, N. Generating actionable insights from customer experience awareness. *Ericsson Technol. Rev.* **2016**, *2020*, 2–11.
4. Niemöller, J.; Mokrushin, L. Cognitive technologies in network and business automation. *Ericsson Technol. Rev.* **2018**, *6*, 5.
5. Inam, R.; Karapantelakis, A.; Vandikas, K.; Mokrushin, L.; Feljan, A.V.; Fersman, E. Towards automated service-oriented lifecycle management for 5G networks. In Proceedings of the 2015 IEEE 20th Conference on Emerging Technologies & Factory Automation (ETFA), Luxembourg, 8–11 September 2015; pp. 1–8.
6. Conine, R. The data warehouse in the telecommunication industry. In Proceedings of the NOMS 98 1998 IEEE Network Operations and Management Symposium, New Orleans, LA, USA, 15–20 February 1998; Volume 1, pp. 205–209.
7. Hossain, M.M.; Azim, T.; Karim, M.Y.; Hoque, A.S.M.L. Integrated data warehousing for telecommunication industries. In Proceedings of the 2009 12th International Conference on Computers and Information Technology, Dhaka, Bangladesh, 21–23 December 2009; pp. 657–662.
8. Chaudhuri, B.G.; Rani, S. Future's Backbone Network Monitoring With Metadata in Data Warehouse for Telecom Industry. In Proceedings of the 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India, 28–30 April 2023; pp. 949–954.
9. Haddad, M.; Altman, Z.; Elayoubi, S.E.; Altman, E. A Nash-Stackelberg Fuzzy Q-Learning Decision Approach in Heterogeneous Cognitive Networks. In Proceedings of the 2010 IEEE Global Telecommunications Conference GLOBECOM, Miami, FL, USA, 6–10 December 2010; pp. 1–6.
10. Dirani, M.; Altman, Z. A cooperative Reinforcement Learning approach for Inter-Cell Interference Coordination in OFDMA cellular networks. In Proceedings of the 8th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, Avignon, France, 31 May–4 June 2010; pp. 170–176.
11. Zhang, Z.; Lin, H.; Liu, K.; Wu, D. A hybrid fuzzy-based personalized recommender system for telecom products/services. *Inf. Sci.* **2013**, *235*, 117–129. [[CrossRef](#)]
12. Abdallah, A.; Maarof, M.A.; Zainal, A. Fraud detection system: A survey. *J. Netw. Comput. Appl.* **2016**, *68*, 90–113. [[CrossRef](#)]
13. Hilas, C.S. Designing an expert system for fraud detection in private telecommunications networks. *Expert Syst. Appl.* **2009**, *36*, 11559–11569. [[CrossRef](#)]
14. Yoshinov, R.; Kulikov, I.; Zhukova, N. Methods of Composing Hierarchical Knowledge Graphs of Telecommunication Networks. *Probl. Eng. Cybern. Robot.* **2020**, *72*, 69–78. [[CrossRef](#)]
15. Kulikov, I.; Vodyaho, A.; Stankova, E.; Zhukova, N. Ontology for Knowledge Graphs of Telecommunication Network Monitoring Systems. *Comput. Sci. Appl.* **2021**, *21*, 432–446.
16. Quinlan, J.R. C4.5: Induction of decision trees. *Mach. Learn.* **1986**, *1*, 81–106. [[CrossRef](#)]
17. Imtiaz, S.; Schiessl, S.; Koudouridis, G.P.; Gross, J. Coordinates-Based Resource Allocation Through Supervised Machine Learning. *IEEE Trans. Cogn. Commun. Netw.* **2021**, *7*, 1347–1362. [[CrossRef](#)]
18. Öhlen, P.; Johnston, C.; Olofsson, H.; Terrill, S.; Chernogorov, F. Network digital twins—Outlook and opportunities. *Ericsson Technol. Rev.* **2023**, *1*, 11.
19. Koudouridis, G.P.; Qvarfordt, C. A Method for the Generation of Radio Signal Coverage Maps for Dense Networks. In Proceedings of the 2018 IEEE International Conference on Communications (ICC), Kansas City, MO, USA, 20–24 May 2018; pp. 1–6.
20. Zhou, Y.; Zhang, X.; Xing, Y.; Xie, Z. GAN-based Signal-to-Noise Ratio Prediction for Wireless Communication Systems. In Proceedings of the IEEE 91st Vehicular Technology Conference (VTC Spring), Antwerp, Belgium, 25–28 May 2020.
21. Lin, F.; Zhang, Y.; Liu, X.; Dai, B.; Zhang, L.; Chen, X. GAN-based Channel Estimation for Wireless Communication Systems with Large-Scale Antenna Arrays. *arXiv* **2021**, arXiv:2108.03881.
22. Koudouridis, G.P.; He, Q.; Dán, G. An architecture and performance evaluation framework for artificial intelligence solutions in beyond 5G radio access networks. *Wirel. Commun. Netw.* **2022**, *2022*, 94. [[CrossRef](#)]
23. Maatouk, A.; Piovesan, N.; Ayed, F.; Domenico, A.D.; Debbah, M. Large Language Models for Telecom: Forthcoming Impact on the Industry. *arXiv* **2023**, arXiv:2308.06013.
24. Zou, H.; Zhao, Q.; Bariah, L.; Bennis, M.; Debbah, M. Wireless Multi-Agent Generative AI: From Connected Intelligence to Collective Intelligence. *arXiv* **2023**, arXiv:2307.02757.

25. Bariah, L.; Zhao, Q.; Zou, H.; Tian, Y.; Bader, F.; Debbah, M. Large Language Models for Telecom: The Next Big Thing? *arXiv* **2023**, arXiv:2306.10249.
26. Ericsson Blog. 2021. Available online: <https://www.ericsson.com/en/blog/2022/1/neural-language-models-telecom-domain> (accessed on 17 January 2024).
27. Holm, H. Bidirectional Encoder Representations from Transformers (BERT) for Question Answering in the Telecom Domain: Adapting a BERT-like Language Model to the Telecom Domain Using the ELECTRA Pre-Training Approach. Ph.D. Thesis, Royal Institute of Technology (KTH), Stockholm, Sweden, 2021.
28. Grimalt, N.M.I.; Shalmashi, S.; Yaghoubi, F.; Jonsson, L.; Payberah, A.H. BERTicsson: A Recommender System For Troubleshooting. In Proceedings of the SDU@AAAI, Virtual, 1 March
29. Bosch, N.; Shalmashi, S.; Yaghoubi, F.; Holm, H.; Gaim, F.; Payberah, A.H. Fine-Tuning BERT-based Language Models for Duplicate Trouble Report Retrieval. In Proceedings of the 2022 IEEE International Conference on Big Data, Osaka, Japan, 17–20 December 2022; pp. 4737–4745.
30. Sun, X.; Holm, H.; Molavipour, S.; Gebre, F.G.; Pawar, K.R.Y.; Shalmashi, S. Recommendation System for Product Test Failures Using BERT. In Proceedings of the International Conference on Knowledge Discovery and Information Retrieval, KDIR, Rome, Italy, 13–15 November 2023.
31. Bissessar, D.; Bois, A. *Evaluation of Methods for Question Answering Data Generation: Using Large Language Models*; Linköping University, Department of Computer: Linköping, Sweden, 2022.
32. Gunnarsson, M. *Multi-Hop Neural Question Answering in the Telecom Domain*; LTH, Lund University: Lund, Sweden, 2021.
33. ETSI. *Zero Touch Network and Service Management (ZSM) Means of Automation*; ETSI: Sophia Antipolis, France, 2018.
34. Saha, B.K.; Tandur, D.; Haab, L.; Podleski, L. Intent-based networks: An industrial perspective. In Proceedings of the 1st International Workshop on Future Industrial Communication Networks, New Delhi, India, 2 November 2018; pp. 35–40.
35. Clemm, A.; Ciavaglia, L.; Granville, L.; Tantsura, J. *Intent-Based Networking—Concepts and Definitions*; ITU: Geneva, Switzerland, 2021.
36. Wei, Y.; Peng, M.; Liu, Y. Intent-based networks for 6G: Insights and challenges. *Digit. Commun. Netw.* **2020**, *6*, 270–280 [[CrossRef](#)]
37. Du, Z.; Jiang, S.; Nobre, J.; Ciavaglia, L.; Behringer, M. *ANIMA Intent Policy and Format*, ANIMA WG; IETF: Fremont, CA, USA, 2016.
38. Leivadreas, A.; Falkner, M. A Survey on Intent-Based Networking. *IEEE Commun. Surv. Tutor.* **2023**, *25*, 625–655. [[CrossRef](#)]
39. Niemöller, J.; Silvander, J.; Stjernholm, P.; Angelin, L.; Eriksson, U. Autonomous networks with multi-layer intent-based operation. *Ericsson Technol. Rev.* **2023**, *2023*, 2–13. [[CrossRef](#)]
40. Niemöller, J.; Szabó, R.; Zahemszky, A.; Roeland, D. Creating autonomous networks with intent-based closed loops. *Ericsson Technol. Rev.* **2022**, *2022*, 2–11. [[CrossRef](#)]
41. Niemöller, J.; Mokrushin, L.; Mohalik, S.K.; Vlachou-Konchylaki, M.; Sarmonikas, G. Cognitive processes for adaptive intent-based networking. *Ericsson Technol. Rev.* **2020**, *2020*, 2–11. [[CrossRef](#)]
42. Chen, M.; Gunduz, D.; Huang, K.; Saad, W.; Bennis, M.; Feljan, A.V.; Poor, H.V. Distributed learning in wireless networks: Recent progress and future challenges. *IEEE J. Sel. Areas Commun.* **2021**, *39*, 3579–3605. [[CrossRef](#)]
43. Amiri, M.M.; Gunduz, D. Federated learning over wireless fading channels. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 3546–3557. [[CrossRef](#)]
44. Blanchard, P.; Mhamdi, E.E.; Guerraoui, R.; Stainer, J. Machine learning with adversaries: Byzantine tolerant gradient descent. In *Advances in Neural Information Processing Systems*; Springer: Berlin/Heidelberg, Germany, 2017; Volume 30.
45. Thapa, C.; Chamikara, M.A.P.; Camtepe, S. SplitFed: When federated learning meets split learning. *arXiv* **2020**, arXiv:2004.12088.
46. Wu, W.; Li, M.; Qu, K.; Zhou, C.; Shen, X.; Zhuang, W.; Li, X.; Shi, W. Split Learning over Wireless Networks: Parallel Design and Resource Management. *arXiv* **2022**, arXiv:2204.08119v2.
47. Liang, L.; Ye, H.; Li, G.Y. Spectrum sharing in vehicular networks based on multi-agent reinforcement learning. *IEEE J. Sel. Areas Commun.* **2019**, *37*, 2282–2292. [[CrossRef](#)]
48. Assran, M.; Loizou, N.; Ballas, N.; Rabbat, M. Stochastic gradient push for distributed deep learning. In Proceedings of the 36th International Conference on Machine Learning, PMLR, Long Beach, CA, USA, 9–15 June 2019; Volume 97, pp. 344–353.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.