# Large Multi-Modal Models (LMMs) as Universal Foundation Models for AI-Native Wireless Systems

Shengzhe Xu, Christo Kurisummoottil Thomas, Omar Hashash, Nikhil Muralidhar, Walid Saad, and Naren Ramakrishnan


## Abstract

Large language models (LLMs) and foundation models have been recently touted as a game-changer for 6 G systems. However, recent efforts on LLMs for wireless networks are limited to a direct application of existing language models that were designed for natural language processing (NLP) applications. To address this challenge and create wireless-centric foundation models, this paper presents a comprehensive vision on how to design *universal foundation models* that are tailored towards the unique needs of next-generation wireless systems, thereby paving the way towards the deployment of *artificial intelligence (AI)-native networks*. Diverging from NLP-based foundation models, the proposed framework promotes the design of *large multi-modal models (LMMs)* fostered by three key capabilities: 1) processing of *multi-modal sensing* data, 2) *grounding* of physical symbol representations in real-world wireless systems using causal reasoning and retrieval-augmented generation (RAG), and 3) enabling *instructibility* from the wireless environment feedback to facilitate dynamic network adaptation thanks to logical and mathematical reasoning facilitated by neuro-symbolic AI. In essence, these properties enable the proposed LMM framework to build universal capabilities that cater to various cross-layer networking tasks and *alignment* of intents across different domains. Preliminary results from experimental evaluation demonstrate the efficacy of grounding using RAG in LMMs, and showcase the *alignment* of LMMs with wireless system designs. Furthermore, the enhanced rationale exhibited in the responses to mathematical questions by LMMs, compared to vanilla LLMs, demonstrates the logical and mathematical reasoning capabilities inherent in LMMs. Building on those results, we present a sequel of open questions and challenges for LMMs. We then conclude with a set of recommendations that ignite the path towards LMM-empowered AI-native systems.


## Introduction

Future artificial intelligence (AI)-native wireless systems (e.g., 6 G and beyond) must leverage machine learning (ML) and AI algorithms to design, optimize, and operate various facets of the network, including resource allocation, transceiver design, and others [1]. Consequently, cross-layer network functionalities implemented by AI models could enable advanced network capabilities, including: 1) *resilience*, enabling 6 G networks to withstand disruptions and maintain connectivity even in challenging scenarios; 2) *intent management*, allowing networks to autonomously translate high-level business intents into closed-loop network configurations; 3) *big-data analytics*, enabling diagnostics using historical wireless data, addressing software or hardware failures, improving communication and computing resource usage, and predicting future user and network behavior; and 4) *non-linear signal processing*, allowing networks to process multi-modal signal characteristics.

To achieve the above goals, a promising avenue is to explore generative AI's universal knowledge retrieval and generation capabilities, particularly foundation models such as large language models (LLMs). Trained on diverse datasets, LLMs can discern intricate patterns and offer insights for optimizing end-user experience in future wireless applications.

### Related Works and Limitations

LLMs for wireless networks have been studied in [2], [3], and [4]. However, the LLMs of [2] and [3] are confined to processing a single mode of textual data, which restricts their role to network chatbots. Accordingly, such LLMs cannot capture the multi-modal data arising from the multiple functions (e.g., sensing, communication, etc.) of future wireless networks. Although [4] focuses on utilizing *multi-modal* LLMs, their approach relies on LLMs like GPT-x, LLaMA, or Falcon tailored for natural language processing (NLP) tasks. To become effective, such multi-modal LLMs must be fine-tuned as wireless tasks change, and, thus, they cannot act as a *universal* solution to different interrelated, cross-layer tasks in AI-native networks.

In addition, the works in [2], [3], and [4] overlook how AI-native networks can fuse, at scale, the environmental *sensing* data that drive their multi-modal LLMs. Moreover, state-of-art solutions

Digital Object Identifier: 10.1109/MNET.2024.3427313
Date of Current Version: 16 September 2024
Date of Publication: 15 July 2024

*Shengzhe Xu and Naren Ramakrishnan are with the Department of Computer Science, Virginia Tech, Arlington, VA 22203 USA; Christo Kurisummoottil Thomas (corresponding author), Omar Hashash, and Walid Saad are with the Bradley Department of Electrical and Computer Engineering, Virginia Tech, Arlington, VA 22203 USA; Nikhil Muralidhar is with the Department of Computer Science, Stevens Institute of Technology, Hoboken, NJ 07030 USA.*

like [4] neglect the fact that even textual wireless data can be structured in specific formats such as tables and network performance evaluations can be presented in the form of graphs or images. This limitation in handling structure and modality, restricts the range of functionalities (e.g., wireless chatbots, text to image conversion) for which LLMs can be effectively adopted.

Moreover, current LLMs [2], [3], [4] lack the essential *grounding* abilities that connect their abstract, language-based knowledge to real-world experiences. In fact, LLMs majorly gain their knowledge upon being trained on extensive corpuses of text data. Hence, these LLMs cannot capture the complex physics governing the wireless environment, such as the propagation of wireless signals, thereby leading to potentially inconsistent decisions and predictions. The absence of grounding impedes AI-native networks from carrying out logical, causal, and mathematical reasoning operations, necessary for achieving goals such as resilience, intent management, non-linear signal processing, and others. Therefore, it is necessary to ensure that the representations acquired by LLMs accurately interpret information from the real world and adhere to the network goals.

Another persistent challenge of LLMs is their tendency to hallucinate by generating "human-like" outputs that do not connect to reality, essentially fabricating false information [5]. If such hallucinations occur, AI-native networks driven by LLMs may generate inaccurate information. For example, LLMs may propose power allocations that violate the regulated thresholds of transmit powers of base stations, leading to alignment problems. *Alignment* here pertains to fulfilling network objectives, adhering to physical constraints like radiated power or environmental sustainability goals, and ensuring compliance with governmental regulations. Finally, existing LLMs [2], [3], [4] lack precise instructibility from the environment and, hence, they cannot perform dynamic problem-solving. *Instructibility* is the capability of LLMs to dynamically adjust their behavior based on explicit feedback from user equipment, system engineers, or operators. This adaptability should also ensure that LLM decisions are explainable.

The absence of grounding impedes AI-native networks from carrying out logical, causal, and mathematical reasoning operations, necessary for achieving goals such as resilience, intent management, nonlinear signal processing, and others.

## Contributions

The main contribution of this paper is the introduction of *grounded and instructible large multi-modal models (LMMs)* that are *universal* and have *alignment* capabilities, as shown in Fig. 1. Here, *universal foundation model* for wireless systems are AI models tailored to handle a wide array of tasks and applications within the wireless domain, irrespective of the network architecture and standards. Our key contributions include:

- We propose a novel framework for universal, wireless-centric foundation models, that goes beyond [2], [3], [4] by integrating the following capabilities into LMMs: 1) *Multi-modal data fusion:* fusing multi-modal sensing information to a shared semantic space thus enabling efficient training of universal foundation models, 2) *Grounding:* involving the creation of a wireless-specific language through retrieval augmented generation (RAG) [6] and leveraging causal reasoning, and 3) *Instructibility:* facilitating transparent interactions between the wireless environment and LMMs through online reinforcement learning (RL) and neuro-symbolic AI to perform logical and mathematical reasoning. This approach ensures *alignment* by developing

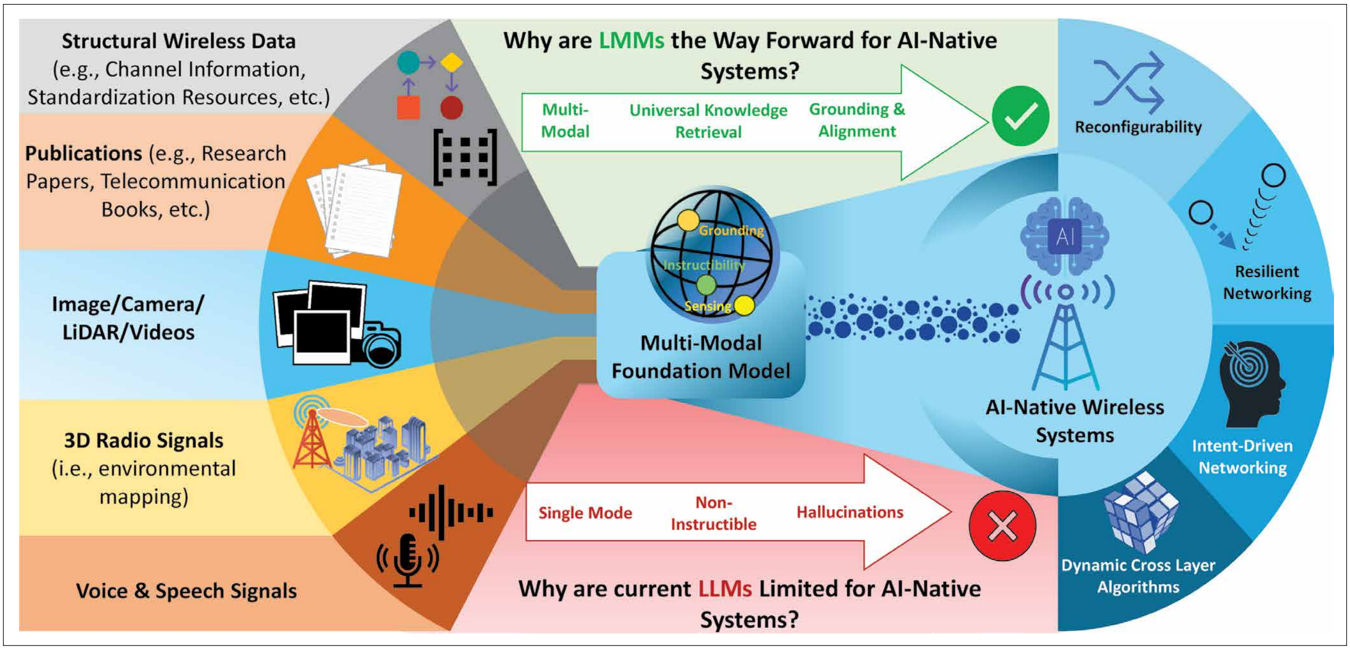


FIGURE 1. Illustrative figure of the proposed framework for LMM-empowered AI-native wireless systems.

trustworthy LMMs that can explain the reasons behind wireless data, propose cross-layer network actions aligned with network goals, and accommodate physical constraints.
- Initial experimental results using RAG demonstrate that infusing more wireless context improves the accuracy of LMM responses, thereby reducing hallucinations compared to responses generated without wireless context. Furthermore, for mathematical questions, an LMM delivers accurate responses with proper rationale, showcasing its ability to reason effectively when grounded in the right context.
- We present a use case of intent-based management employing LMMs, comprising problem formulation, intent assurance, and a validation phase. The improved logical and mathematical reasoning capabilities (shown in experiments) enable LMMs to function as dynamic problem solvers. We demonstrate that logical and mathematical reasoning capabilities enable continuous monitoring of network performance–a critical aspect of building resilient networks. In contrast to deep RL-based methods, which may be constrained to specific domains, LMMs can accelerate network service recovery during failures by proposing a sequence of remediation actions.
- We highlight challenges in constructing universal foundation models, covering aspects such as network planning, acquiring diverse datasets, and adapting to evolving standards.

## LMM-Empowered AI-Native Wireless Systems: Proposed Framework

To build universal foundation models, as shown in Fig. 2, the proposed multi-modal LMM framework is built upon the principles of multi-modal data fusion, grounding, and instructibility. The components of the proposed universal foundation models framework are discussed next.

### Fusion of Multi-Modal Sensing Information: A Tradeoff Between Minimality and Redundancy

For capturing the real-world wireless environment, there is a need for precise sensing and mapping of its diverse surroundings. Prior works like [4] (and references therein) discussed exploiting visual

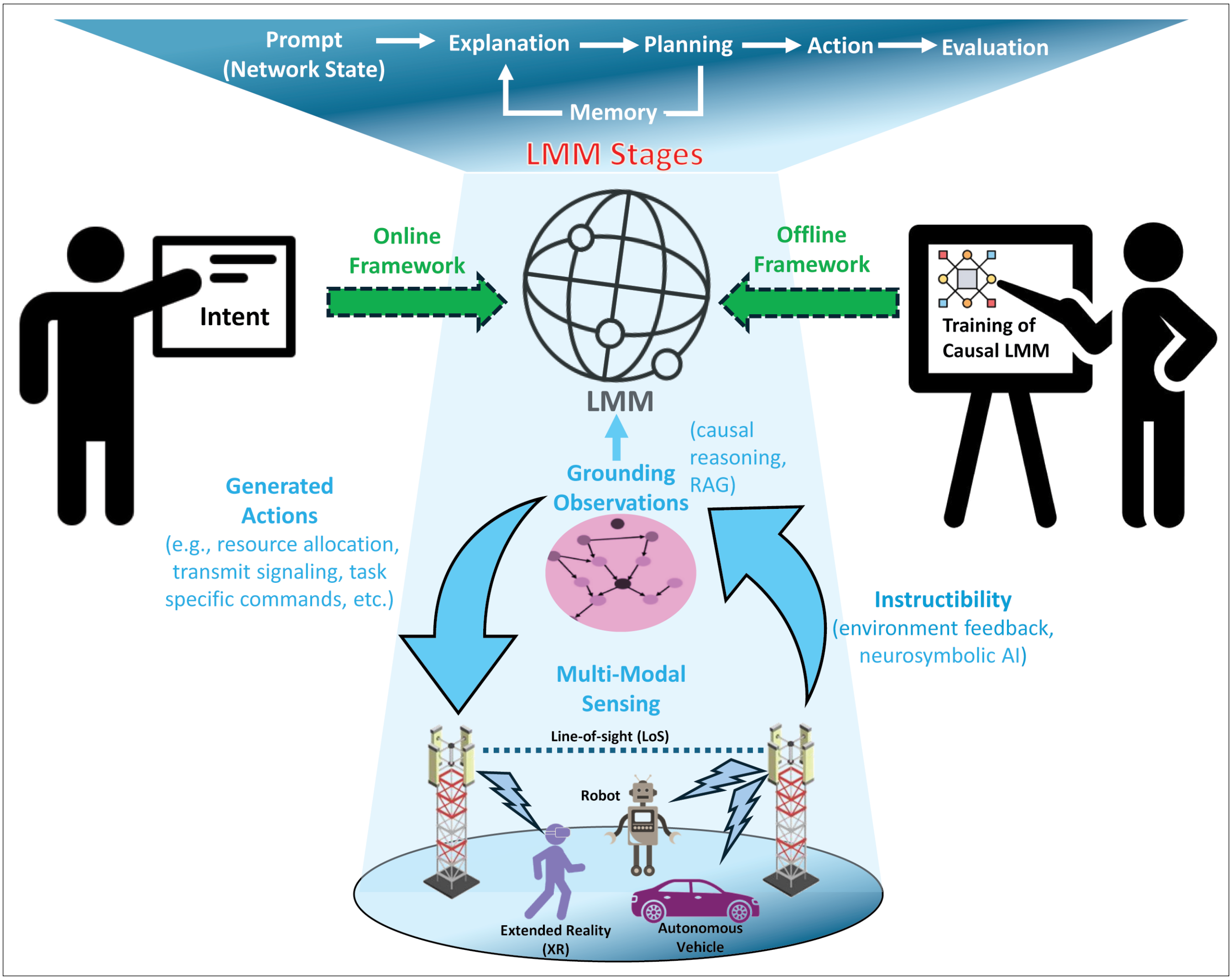


FIGURE 2. Proposed framework of wireless-centric LMMs with capabilities of grounding, instructibility, and alignment.

generative AI models such as meta-transformers to map multi-modal wireless sensing information to a semantic latent space. By capturing the characteristics of the wireless environment, such models can enhance contextual and situational awareness for sensing applications in AI-native networks. However, providing the entire mapped sensing information (received from diverse sources) as input to train the LMM is resource-intensive and requires substantial time for retraining. This hinders the timely execution of dynamic updates essential for maintaining seamless connectivity. To address this limitation of meta-transformers inherent in [4], we propose to convey a compressed sequence of pertinent information to the LMM using dimensionality reduction. To achieve this, we start with the identification of physical symbols present across the multi-modal data. *Physical symbols* refer to abstract entities present in the data that have relevant semantics with respect to the wireless network. For example, this may involve associating symbols with various extrinsic and intrinsic elements. *Extrinsic elements* encompass dynamic objects, such as scattering elements in the environment and users. Meanwhile, *intrinsic elements* involve static network features like network addresses and signal processing methods underlying wireless transmission or reception, among others. There could be redundancy among the information conveyed by the symbols across multiple modalities of data. To mitigate redundancy, the shared semantic latent space's construction should adhere to the information bottleneck principle [7]. In this framework, it is assumed that a dominant mode of information, called *prime modality*, exists within a dataset, serving as the primary source of information. Other modalities complement or enhance the information provided by this prime modality. Here, the compact representation for any modality should convey as little information as possible about the raw data, and, simultaneously, prime modality representation should convey maximum information about other modalities. This ensures that the resulting semantic latent space is of *minimal dimension while avoiding redundant information*. Further, for LMM training, we advocate using this filtered representations in the shared semantic space as inputs. Additionally, filtering also determines when to perform dynamic updates of the neural network (NN) parameters of the LMM, taking into account the nature of the captured data, which can be either static (e.g., 3GPP standards) or dynamic (e.g., wireless channel information).

While the fusion and filtering of multi-modal information and the training of LMMs is important, on its own, simply identifying symbols is not enough if LMMs are to be universal. While a vanilla LLM can effectively predict the events following an observed sequence of sensing information, it lacks precise understanding of what causes the event and its implications from a wireless system perspective. For example, translating images of trees in a wireless environment into a set of angles of arrival or departure that describe the RF signal propagation environment requires associating meaning from a wireless perspective with each extracted physical symbol. This aspect, called *grounding*, is detailed next.

## Causal Reasoning for Grounding in LMMs: Reducing Hallucinations and Bolstering Trustworthiness

Traditional grounding approaches typically entail creating a knowledge base, which represents an instance of symbolic AI. The knowledge base captures the possible logical relations among physical symbols, such as scattering objects, users, network topology, and transmission or reception parameters. Nevertheless, the use of knowledge base methods faces scalability challenges as the number of relations and physical symbols expands. To overcome this limitations in conventional grounding methods, we propose that LMMs infer the relations among various physical symbols identified using *causal reasoning* [8], as discussed next.

While specific experiments demonstrate that language models might exhibit causality, it is predominantly attributed to the causal knowledge ingrained in the training data, rather than indicative of LLMs possessing inherent causal understanding. In [9], a gradient-based, transformer-type algorithm for zero-shot optimal covariate balancing for causal treatment effect is introduced. We propose to advance [9] by incorporating theoretical methods to construct causal foundation models, focusing on wireless concepts as the relevant physical symbols. Here, one may ask: *how to identify the causal relations among physical symbols and how to ensure that the learned relations are aligned with the wireless concepts in standards and textbooks?* One common approach is to perform finetuning [4] that takes a pre-trained language model trained on large amounts of general text and then continue to train it on a small-scale task-specific text. Fine-tuning is appropriate if the user specifically knows the ground-truth causal relations. For wireless scenarios, fine-tuning can be beneficial for constructing a wireless specific chatbot capable of extracting valuable information from its knowledge base. However, fine-tuning LLMs may impose limitations on the wireless applications supported. This limitation arises from the narrow set of NN parameters that are tuned during the fine-tuning process (limited degrees of freedom). This, in turn, requires re-tuning as the wireless environment or task change. To address these limitations, we suggest the use of RAG coupled with causal discovery, as discussed next and in Fig. 3.

**1) How to Perform Causal Discovery Through RAG?:** Through querying from a wireless-specific database that includes wireless textbooks, research papers, 3GPP standard, or any device instruction handbook, RAG [10] enables the LMM to understand the domain-knowledge context. Once this information is retrieved, the generation component of RAG can help formulate new content that infers or expresses causal relations among the identified physical symbols. For example, when the LMM is tasked with deducing causal relationships between scattering objects in the environment and channel parameters like angle of arrival (AoA) or angle of departure (AoD), RAG can map these wireless observations to the underlying physical concepts from the database.

Through an evolvable external knowledge component and multi-agent cooperation, RAG can allow the implementation of emerging

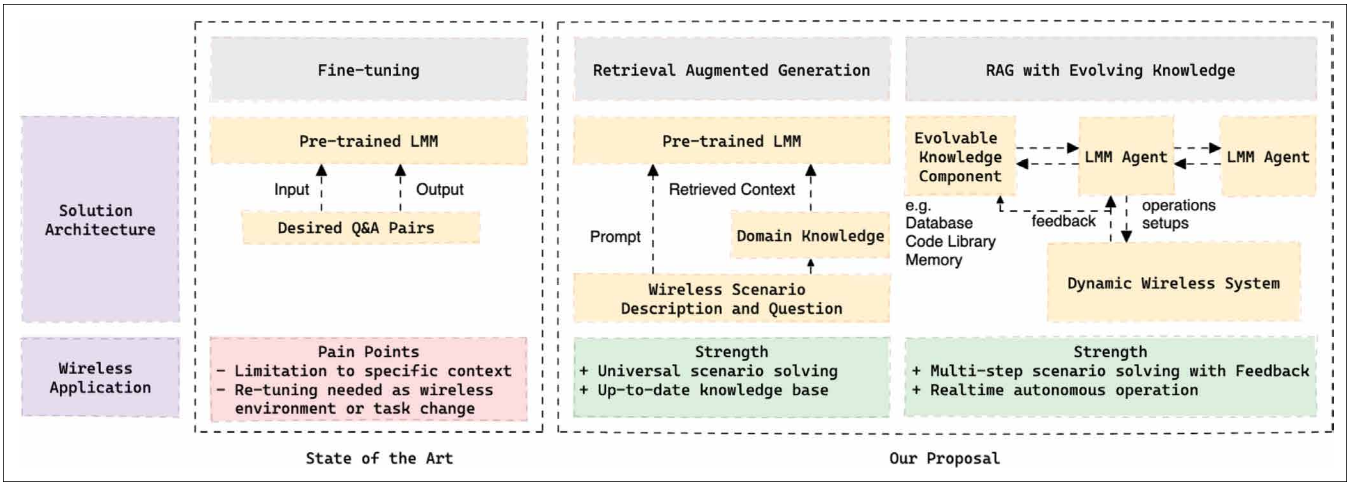


FIGURE 3. Applying LMM solutions for wireless applications: 1) Fine-tuning, 2) RAG, and 3) RAG with evolving knowledge.

applications that require multi-tasking, like in connected homes or industrial robots. Moreover, performing retrieval from an evolving knowledge base can enable universal knowledge retrieval for semantic communications [11], and intent management. With its continuous learning capability, RAG, with evolving knowledge, enables continually updating the wireless algorithms across all open systems interconnection (OSI) layers while ensuring compatibility with the advancements in the semiconductor industry and software solutions. This proves advantageous, particularly for intent management and resilience (see the sections "LMMs for Intent Management" and "LMMs for Resilient Networking"). Apart from the continual learning capability, RAG with evolving knowledge enables the knowledge retrieved to be dynamically adjusted to cater to the specific application demands. For example, in a multi-user communication system, the retrieved literature on signal processing algorithms must differ from what might be required in a single-user scenario.

**2) What Does Causal Discovery Through RAG Entail for LMMs?:** Grounding via causal discovery entails endowing LMMs with the capability to comprehend the causal relationships among physical symbols and subsequently engage in causal inference through interventions and counterfactuals [8]. Through interventions and counterfactuals, the LMM can indulge in chain-of-thought kind of reasoning, where it analyzes a sequence of causal state-action pairs $(\mathbf{s}_t, \mathbf{a}_t)$ and their effects, $\mathbf{s}_0 \xrightarrow{\mathbf{a}_0} \mathbf{s}_1 \xrightarrow{\mathbf{a}_1} \mathbf{s}_2 \cdots \xrightarrow{\mathbf{a}_{N-1}} \mathbf{s}_N$. This ability facilitates long-term planning for wireless resource allocation, signaling schemes for transmission and reception (and may include beamforming, modulation, coding, and control signaling, among others), and quality-of-service (QoS) management, thereby contributing to the establishment of robust and resilient wireless systems.

## Instructiblity From Environmental Feedback

For instructibility, LMMs should be able to dynamically adjust resource allocation, signaling policies, and many other cross-layer network functionalities in real time, catering to diverse tasks, environments, and optimization objectives. Additionally, they should be able to continuously monitor wireless observations to identify and address any unforeseen issues that might impede seamless network connectivity. A standard approach for dynamically adjusting wireless resource management and signaling schemes based on user feedback using deep RL. However, deep RL techniques are task-specific and require retraining when the wireless environment and optimization objectives change. While multi-task RL solutions exist, they are mostly limited to specific domains or OSI layers. Moreover, they lack the ability to continuously evolve their state and action space to cope with changes in standards or advancements in wireless technology. Conventional multi-task RL solutions also cannot perform abductive reasoning, a crucial aspect for making inferences about missing data or determining the best explanations for observed data. These features are essential for achieving dynamic adaptability and reconfigurability, necessary for resilience and intent management. They are also crucial for supporting abductive reasoning capabilities required for semantic communications and other related tasks. We next discuss the key components needed to provide instructibility to LMMs. We begin by detailing the framework incorporating communication context, prompting, and an online LMM with wireless environment feedback, contributing to establishing an instructible system. Subsequently, we explore how to instill LMMs with logical and mathematical reasoning, that are essential for constructing self-evolving and dynamically adaptable systems, thus achieving instructibility.

**1) Communication Context:** Causal representations that are used to represent the physical symbols, similar to tokens in NLP-based LLMs, form the *communication context* for an LMM. This context encapsulates critical aspects of the wireless communication scenario. The components of the LMM context include:

- **Network setup** including details about 1) *communicating devices*, 2) *communication link* (downlink or uplink), and 3) *physical topology* that describes the antenna configuration, as well as any miscellaneous network architecture.
- **Communication constraints** including constraints on the total power and shared communication/computation resources arross frequency, time, and other dimensions.

- **Wireless standards/text snippets** read using RAG (the section "How to Perform Causal Discovery Through RAG?"), that include excerpts from relevant wireless communication standards or documents, providing a contextual basis for the communication scenario.
- **End-to-end optimization objectives** that may include quality-of-service (QoS) measures such as average throughput, delay, reliability/quality-of-experience.
- **Historical wireless data** that may involve diverse measurements such as uplink pilots, user feedback on channel quality indication, various sensing measurements, and received uplink signal measurements, among other relevant parameters.

Next, we explain how to construct an online LMM by instructing it with environmental feedback.

**2) Online LMM With Wireless Environment Feedback Using Neuro-Symbolic AI:** One common apporoach to instill instructiblity is to use an iterative prompting mechanism in which an LMM is guided through multiple rounds of interaction with human prompts. In each iteration, the model refines and improves subsequent responses using the feedback (e.g., the QoS results based on the wireless policy of the LMM) from the previous round. However, iterative prompting requires human intervention. We propose building an online LMM framework to address this limitation and enable the development of autonomous wireless systems. In this setup, LMM functions as the wireless policy and is operationally embedded within an interactive setting using online RL. This entails utilizing gathered wireless observations and feedback from the environment to iteratively enhance its functionality, aligning with goals expressed in wireless language. The formulation of the LMM-powered cross-layer network functionalities can be represented as a partially observable Markov decision process. Here, the states are defined by the communication context and prompts, actions are represented by the wireless policy suggested by the LMM, and rewards are determined using performance metrics obtained from the wireless environment. If available, the network goal or intent can be articulated in natural language by the network operator. To ensure continuous operation without disrupting connectivity, online LMMs should possess the ability to explain wireless observations and infer any missing data, necessitating logical reasoning capabilities. Furthermore, given that many wireless concepts can be expressed mathematically, LMMs must inevitably be capable of performing mathematical reasoning. This includes tasks such as channel predictions, beamforming vector computations, channel quality measurements, and many other cross-layer network computations.

Here, multi-task RL can be an alternative approach to perform diverse wireless tasks. However, since they lack logical and mathematical reasoning capabilities, we advocate incorporating them through the use of *neuro-symbolic AI* [11]. In our setting, symbolic AI serves to evaluate diverse logical and mathematical formulas, while the neural component is responsible for learning the logical and mathematical equations from wireless observations and context information. When prompted with communication context and grounded wireless observations using causal discovery (the section "How to Perform Causal Discovery Through RAG?"), symbolic AI connects facts and data through rules and algorithms, resembling the cognitive operations of the human brain in storing high-level concepts and engaging in nuanced inference. To prevent hallucination, LMMs must have the ability to explain wireless observations and infer any missing data. Additionally, they should understand the connections between various physical symbols through symbolic AI. Here, a viable approach is to develop a formal logical language that can encapsulate the exhaustive ontology of wireless concepts and articulate rules governing the functioning of wireless systems (and is the symbolic part here). This strategy is reminiscent of the Cyc concept [12], which serves a similar purpose for web-based data. Beyond their lack of logical reasoning abilities, existing LLMs face challenges in accurately capturing mathematical formulas and executing mathematical derivations. To overcome this limitation, a promising approach involves leveraging a neuro-symbolic problem solver [13], having three main components. First, a *problem reader* encodes math word problems, presented as textual prompts, into vector representations. Second, a *programmer* generates symbolic grounded equations, which are executed to produce answers. Lastly, a *symbolic executor* obtains final results. In this setup, the programmer learns the weights (neural part) that establish connections between various mathematical symbols. The resulting neuro-symbolic problem solver enables the construction of dynamic problem solvers, a critical component for intent management and resilience, as discussed in sections "LMMs for Intent Management" and "LMMs for Resilient Networking."

## Experimental Validation: A Use Case Study for Intent Management and Resilient Networking

Here, we first demonstrate illustrative experiments conducted on a dataset specific to wireless scenarios using RAG. We highlight the enhanced performance of LMM compared to vanilla LLMs which does not have any wireless context, characterized by succinct explanations (resulting in *reduced hallucinations*), precise answers (demonstrating *grounding in wireless concepts*), and well-founded rationales (illustrating *mathematical reasoning* capabilities). Additionally, we discuss how the results indicate the potential application of the proposed LMM in addressing specific challenges in future wireless networks.

### Experimental Validation

To evaluate the efficacy of RAG in wireless contexts, we conducted certain Q/A experiments, where a sample question from the dataset is shown in Fig. 4. In the RAG process, relevant paragraphs are extracted from [14] to serve as wireless context information. This information encompasses a combination of textual content and mathematical symbols and equations. This

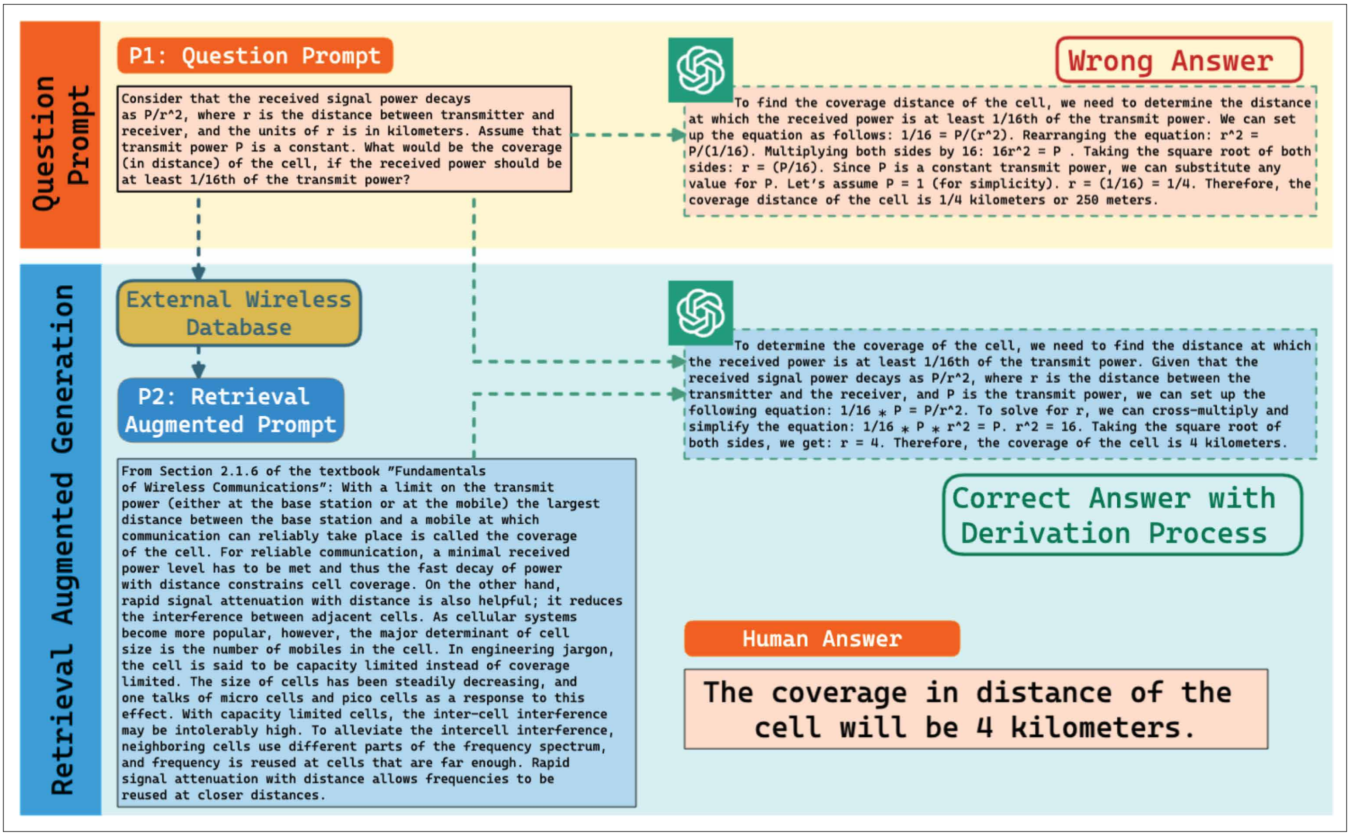


FIGURE 4. A sample mathematical Q/A pair from the dataset.

| Evaluation Measure | Question Prompt | Retrieval Augmented Prompt |
|---|---|---|
| Precision (↑) | 0.06 | **0.08** |
| Recall (↑) | 0.59 | **0.65** |
| F1 Score (↑) | 0.11 | **0.14** |
| ROUGE-L (F-measure) (↑) | 0.17 | **0.20** |
| Over Explaining (↓) | 0.34 | **0.12** |
| Conceptual Question Ratertion (↑) | **0.89** | 0.84 |
| Conceptual Question Asserte (↑) | 0.88 | **0.95** |
| Mathematical Question Rationale (↑) | 0.77 | **0.94** |
| Mathematical Question Assertion (↑) | 0.76 | **0.97** |
| Mathematical Question Derivative Steps (↑) | 0.48 | **0.87** |

TABLE 1. Prompting GPT-3.5 Turbo with retrieval-augmented context shows a general advancement over purely prompting with questions in 4 quantitative measurements (upper section) and 6 human-evaluated measurements (lower section). For each metric (row), the symbol (↑) indicates that higher scores are better, and better results are highlighted in **bold** for the two prompting methods. **Precision:** the number of shared words to the total number of words in the generated answers; **Recall:** the number of shared words to the total number of words in the human answers; **F1 score:** $\frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$; **ROUGE-L (F-measure):** based on the longest common subsequence (LCS) between the generated answer and human answer, which indicates that a longer shared sequence should indicate more similarity between the two sequences. **Human Evaluated Score:** Participants are asked to rate each Q/A sample without knowing the source of it. They will give score 0 for no and score 1 for yes for each of the Rationale, Assertion, and Over Explaining items. Mathematical questions require an additional derivative step to be scored.

structured wireless information represents a simple multi-modality case. Table 1 shows an evaluation using 16 human participants on responses from different prompting methods of 4 conceptual wireless questions and 7 mathematical wireless questions. With the help of retrieved knowledge context, common LMM evaluation metrics (see Table 1 for definitions), including precision, recall, F1 score, and ROUGE-L measure, indicate a performance improvement over vanilla LLMs ranging between 15% to 30%. We also interpret the human evaluation result in the following way: 1) For conceptual questions, the standard *Question Prompt* can retrieve reasonable rationale, while RAG can refine the assertion, leading to an 8% improvement in the assertion metric. Furthermore, the over-explaining metric, which gauges the alignment of responses with human expectations, shows a remarkable almost threefold improvement for RAG compared to vanilla LLMs. Improved assertions imply that LMMs can mitigate hallucinations in their responses, thereby aligning more effectively with the goals of the network. 2) In the context of mathematical questions, *Retrieval Augmented Prompt* consistently provides the correct answer and can offer more detailed mathematical derivatives. Specifically, the rationale exhibits a 22% improvement with RAG, while the derivative steps are 81% more detailed than vanilla LLMs. This indicates that LMMs exhibit enhanced logical and mathematical reasoning abilities.

As discussed earlier, the demonstrated logical and mathematical reasoning capabilities, evidenced by improved rationale measures, enable LMMs to map wireless observations and context information fed as input to them into a mathematical problem formulation. This empowers LMMs to function as dynamic problem solvers (defined in the section "Online LMM With Wireless Environment Feedback Using Neuro-Symbolic AI"). To exemplify the application of these capabilities in future wireless systems, we next discuss a few use cases, including intent management and resilience.

### LMMs for Intent Management

A recent work that exploits LLMs for intent management appeared in [3]. However, this prior work is limited to using an LLM as a chatbot to convert human specified intent in natural language to infrastructure level intents as network service descriptors. Furthermore, the authors in [3] utilize human feedback to enhance the configurations generated by the LLM, hindering their ability to ensure intent assurance autonomously. In contrast to [3], we propose to employ LMMs for various phases in intent management spanned across the OSI layers:

- *Problem Formulation Phase:* The network must autonomously translates the operator specified intents into an optimization problem, considering multiple objectives and physical constraints. An LMM can facilitate dynamic problem formulation without human intervention since they possess better logical and mathematical reasoning abilities. Moreover, LMM responses are grounded in wireless physics, as discussed in the section "Causal Reasoning for Grounding in LMMs: Reducing Hallucinations and Bolstering Trustworthiness," and clarified in Fig. 4. It can thus formulate precisely wireless optimization problems, for resource allocation, signaling schemes, or a combination of cross-layer objectives.
- *Intent Assurance Phase:* Leveraging a continuous stream of wireless measurements, LMMs can function as intent *assurance agents*. Using neuro-symbolic AI capabilities, these agents can assess logical formulas representing the desired intent (typically defined by QoS targets). If the intent is not fulfilled, LMMs can identify and articulate the specific issues that need resolution, guiding efforts towards achieving intent assurance within a specified timeframe.
- *Validator Agent:* For solving the LMM-designed problem formulation provided we can use multiobjective RL with causal reasoning games building on [8]. Validating solutions against regulatory norms and physical constraints for long-term intent fulfillment is crucial. This is the *alignment goal* described previously in the section "Introduction." To autonomously manage intent, a validator role can be fulfilled by LMM, possessing a solid understanding of wireless concepts.

Next, we discuss the impact of minimizing hallucinations through improved assertions and providing precise answers (as reflected in over-explaining metric). This capability is essential for swiftly recovering from network service disruptions and thereby ensuring resilience.

Improved assertions imply that LMMs can mitigate hallucinations in their responses, thereby aligning more effectively with the goals of the network.

### LMMs for Resilient Networking

*Resilience* is the ability of wireless networks to: a) detect or predict in advance any failures or performance disruptions arising due to network functionality issues across any OSI layer, changing wireless environment, user dynamics, or external malicious influences; and b) recover back to their normal functionality within a stipulated time frame, thereby ensuring seamless connectivity for all connected devices. In [8], we proposed a robust framework for building resilient wireless networks causal Bayesian optimization. However, the application of our solution in [8] is limited, because it mainly focuses around quickly recovering from QoS deviations in the network. However, network service disruptions can stem from changes in the wireless environment or malfunctions in hardware or software functionalities across diverse edge devices. To address this challenge, we suggest leveraging LMMs equipped with causal knowledge, not only pertaining to the wireless environment but also grounded in wireless standards and cross-layer network functionalities. Such a universal foundation model can handle service disruptions across multiple domains and tasks by operating in a closed loop fashion as discussed next.

- *Continuous Monitoring of Service Disruptions:* To detect network service disruptions, the LMM should continuously monitor and predict potential issues across OSI layers. For example, consider a situation where the software code representing functionality at any OSI layer on an edge device becomes corrupted due to processor malfunctions. Alternatively, critical information intended for storage on an edge server might face corruption due to jamming attempts. Herein, since LMMs are grounded in wireless concepts, they can adeptly analyze error messages and descriptions of software malfunctions or data corruptions, offering suggestions aligned with network standards to rectify the issues, without any human intervention. This approach enhances the model's capability to provide context-aware solutions across diverse tasks or domains or environments. Furthermore, LMMs can be consistently prompted to check for potential wireless environment issues that might lead to performance deviations in the near future. Such network issues can be formulated as either logical formulas or mathematical equations. For example, a potential logical formula could be $p : X \rightarrow (Y < \tau)$, signifying that with probability $p$, the wireless observation $X$ results in a performance $Y$ below the expected target $\tau$. However, vanilla LLMs face challenges in handling such logical and mathematical problems, as discussed in the section "Instructiblity From Environmental Feedback." In this context, logical and mathematical reasoning capabilities using neuro-symbolic problem solvers

Instructibility allows LMMs to generate a sequence of network actions in response to the feedback from the wireless environment.

play a crucial role in assessing performance quality. In contrast to [8], which necessitates the construction of specific causal models for monitoring particular tasks or QoS targets, the universal nature of LMMs allows a single model to be used for monitoring performance deviations and software or hardware malfunctions across any OSI layer. Given that a failure is detected, we next look at how the LMM can help the network functionalities quickly (within a stipulated time) recover back to the expected performance.

- *Network Service Recovery via LMM:* As detailed in the section "How to Perform Causal Discovery Through RAG?," RAG enables the model to comprehend the causal implications of network actions by grounding wireless observations to the extracted knowledge. This enables LMMs to execute the minimal interventions required to restore the network to a normal functioning state. Further, as discussed in the section "Instructiblity From Environmental Feedback," instructibility allows LMMs to generate a sequence of network actions in response to the feedback from the wireless environment. These actions can involve repairing malfunctioning code, adjusting resource allocation, or refining signaling schemes to restore the network to normal functioning. In contrast to [8], which might require separate causal AI models to monitor diverse functionalities across OSI layers, the universality of LMMs can possible enable faster switching between tasks requiring repair or refinement, utilizing a single AI model.

## Challenges and Open Questions for LMM-Empowered AI-Native Wireless Systems

### How Can We Enable LMMs To Do Planning?

Meeting future wireless network goals requires the capability of LMMs to provide recommendations across various OSI layers. This includes resource allocation policies, waveforms under non-linear signal models, network slicing policies, and more. To ensure that these recommendations align with long-term network goals in terms of performance, sustainability, or seamless connectivity, LMMs must possess the ability to perform planning. In this context, planning entails the ability of LMMs to propose a sequence of multi-dimensional network actions, enabling the network to optimize performance objectives. The term "multi-dimensional" reflects that these network actions are not restricted to a specific task or a single layer but can extend across multiple OSI layers. Standard methods like deep RL optimize actions for specific tasks but lack universality, as discussed in the section "LMM-Empowered AI-Native Wireless Systems: Proposed Framework." Here, we look at possible approaches to incorporate planning in LMMs. The first approach, uses fine-tuning, and involves taking a pretrained LLM and refining it using planning problems—consisting of instances and their solutions. While additional fine-tuning data and efforts might result in improved empirical performance, we must recognize that fine-tuning essentially transforms the planning task into a memory-based (approximate) retrieval process. This, however, falls short of providing conclusive evidence regarding the inherent planning capabilities of LMMs.

The second approach to enhance planning performance involves prompting the LMM with hints or suggestions to improve its initial plan guess. Crucial considerations in this context include whether the back prompting is manual or automated, the entity certifying the correctness of the final answer, and whether the prompts offer additional problem knowledge or merely encourage the LMM to reconsider its approach. A more popular methodology, here is "chain of thought prompting (CoT)," that involves having a human (a system engineer) in the loop prompt the LMM. However, CoT is susceptible to the Clever Hans effect, where the LMM generates wireless policies, and the human in the loop, aware of right vs. wrong solutions, inadvertently guides the LMM. The responsibility for accuracy, if achieved, lies with the human in the loop. This framework raises concerns when the human cannot verify the answer to the planning problem themselves. However, in a wireless network, to ensure automated network operation, we cannot afford to rely on human intervention. Therefore, a promising approach entails the LMM critiquing its predictions through self-reflection capability and iteratively self-improving. This self-reflection capability can be achieved by incorporating causal reasoning. With the LMM being aware of each network action's causal effects, it can store these experiences in a short-term memory. Subsequently, when presented with new wireless observations, the LMM can self-reflect based on its previous experiences and engage in planning to ensure that network objectives are met.

### What Are the Challenges Associated With Training a Universal Foundation Model?

Training a foundation model for wireless communication enables domain-specific optimization, efficiency gains through reduced NN weights, and enhanced performance. However, achieving this long-term goal requires collaboration among stakeholders in wireless communication and computer science. The associated challenges include the need for seamless interdisciplinary cooperation, addressing diverse communication standards, and incorporating evolving technologies to ensure the model's adaptability and effectiveness.

- *Diverse and Representative Datasets:* To ensure seamless connectivity across different wireless environments and diverse applications, an LMM should be trained under diverse signal conditions, interference patterns and fading scenarios.
- *Adaptability to Evolving Standards and Technologies:* The incorporation of evolving 3GPP standards into the foundation models is crucial. This integration ensures that the decisions made by the LMM comply with

both the network and the unified 3GPP standards. This alignment contributes significantly to enhancing the trustworthiness of the language model by ensuring its compatibility and compliance with the latest industry standards. In addition to leveraging LMMs for understanding and adhering to existing standards, they can also play a crucial role in the creation of standards, especially in scenarios where technologies are not standardized. LMMs, with their capacity for natural language processing and generation grounded in wireless concepts, can contribute to the formulation and documentation of wireless standards, fostering innovation and clarity in technology development.

## Conclusion and Recommendations

This article developed a new framework for designing AI-native wireless systems (6G and beyond) for multiple tasks using foundation models built on multi-modality, grounding, and instructibility principles. We conclude with three key recommendations:

- **Speeding Up Next-G Standardization to System Design:** LMMs can assist in the swift prototyping of diverse system design scenarios. Leveraging the capabilities of RAG, LMMs can retrieve pertinent text-based descriptions and specifications by considering the provided input, whether it be a network intent or system design goals. This enables LMMs to actively contribute to rapidly exploring design alternatives and their associated implications.
- **Building a Repository of Wireless Datasets:** While the results based on RAG offer unique insights into LMMs' capabilities, it is essential to acknowledge the challenges associated with generating a comprehensive dataset. To address this, we recommend the creation of an opensource ontology for wireless concepts and algorithms, sourced from a curated selection of textbooks and wireless literature, encompassing 3GPP standards. This approach ensures the quality, reliability, and trustworthiness of the dataset, making it applicable for research and development across the entire wireless community.
- **Compositions of Short Language Models and Distributed Architecture:** Creating a universal foundation model at each wireless base station may prove impractical due to the substantial energy consumption and computational resources involved. Thus, we recommend a distributed architecture that involves constructing distinct short language models learned at edge servers. These edge servers might require only a condensed language model, as the applications or tasks they handle are limited in scope. This distributed architecture facilitates collaborative reasoning based on the principle of compositionality [15], through the combination of representations from multiple smaller models. This compositional approach empowers the LMM-based network with the capacity to acquire diverse skills and functionalities over time, and thus gaining universal capabilities in a distributed fashion.
- **Sustainable Foundation Models Using Next-Generation AI:** Existing LLMs are power-hungry due to billions of NN parameters in their architecture. However, to achieve the goal of sustainability in future wireless networks, it is critical to build foundation models that can work with less data and are smaller models but still able to meet the capabilities of LMMs. Herein, we recommend to construct universal foundation models with the next-generation AI capabilities of explainability (using causal reasoning), reasoning (using neurosymbolic AI), planning, and common sense (using world models) as envisioned in [1]. Such next-generation AI capabilities allow the foundation models to understand how the world works (using world models), form better future predictions, and thereby gain universal capabilities with limited training overhead.

## Acknowledgment

The work of Shengzhe Xu and Naren Ramakrishnan was supported in part by the U.S. Department of Defense through the Office of the Assistant Secretary of Defense for Research and Engineering (ASD(R&E)) under Contract HQ003419D0003. The work of Christo Kurisummoottil Thomas, Omar Hashash, and Walid Saad was supported by the U.S. National Science Foundation under Grant NSF CNS-2225511. Any views, opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the U.S. Department of Defense nor ASD(R&E).

## References

[1] W. Saad et al., "Artificial general intelligence (AGI)-native wireless systems: A journey beyond 6G," 2024, *arXiv:2405.02336*.
[2] A. Maatouk et al., "TeleQnA: A benchmark dataset to assess large language models telecommunications knowledge," 2023, *arXiv:2310.15051*.
[3] S. Tarkoma, R. Morabito, and J. Sauvola, "AI-native interconnect framework for integration of large language model technologies in 6G systems," 2023, *arXiv:2311.05842*.
[4] L. Bariah et al., "Large generative AI models for telecom: The next big thing?" 2023, *arXiv:2306.10249*.
[5] V. Rawte, A. Sheth, and A. Das, "A survey of hallucination in large foundation models," 2023, *arXiv:2309.05922*.
[6] T. Carta et al., "Grounding large language models in interactive environments with online reinforcement learning," 2023, *arXiv:2302.02662*.
[7] X. Xiao et al., "Neuro-inspired information-theoretic hierarchical perception for multimodal learning," in *Proc. 12th Int. Conf. Learn. Represent.*, 2024, pp. 1–29.
[8] C. K. Thomas et al., "Causal reasoning: Charting a revolutionary course for next-generation AI-native wireless networks," *IEEE Veh. Technol. Mag.*, vol. 19, no. 1, pp. 16–31, Mar. 2024.
[9] J. Zhang et al., "Towards causal foundation model: On duality between causal inference and attention," 2023, *arXiv:2310.00809*.
[10] A. Asai, M. Gardner, and H. Hajishirzi, "Evidentiality-guided generation for knowledge-intensive NLP tasks," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Human Lang. Technol.*, 2022, pp. 9459–9474.
[11] C. K. Thomas and W. Saad, "Neuro-symbolic causal reasoning meets signaling game for emergent semantic communications," *IEEE Trans. Wireless Commun.*, vol. 23, no. 5, pp. 4546–4563, May 2024.
[12] D. Lenat and G. Marcus, "Getting from generative AI to trustworthy AI: What LLMs might learn from Cyc," 2023, *arXiv:2308.04445*.
[13] J. Qin et al., "Neural-symbolic solver for math word problems with auxiliary tasks," 2021, *arXiv:2107.01431*.

[14] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[15] R. Bansal et al., "LLM augmented LLMs: Expanding capabilities through composition," 2024, *arXiv:2401.02412*.

## Biographies

Shengzhe Xu (shengzx@vt.edu) received the B.S. degree in telecommunication engineering with management from the Beijing University of Posts and Telecommunications in 2016. He is currently pursuing the Ph.D. degree with the Department of Computer Science, Virginia Tech. He has published research articles in the fields of deep generative models, time-series analysis, large language models, semantic analysis of program code, and wireless communications. He has completed applied machine learning internships at Facebook, Microsoft Research Asia, Google, and Tsinghua National Laboratory.

Christo Kurisummoottil Thomas (Member, IEEE) (christokt@vt.edu) received the B.S. degree in electronics and communication engineering from the National Institute of Technology Calicut, India, in 2010, the M.S. degree in telecommunication engineering from the Indian Institute of Science, Bengaluru, India, in 2012, and the Ph.D. degree from EURECOM, France, in 2020. From 2012 to 2014, he was a Staff Design Engineer of 4G LTE with Broadcom Communications, Bengaluru. From 2014 to 2017, he was a Design Engineer with Intel Corporation, Bengaluru. From November 2020 to June 2022, he was a Staff Engineer with the Wireless Research and Development Division, Qualcomm Inc., Espoo, Finland. He is currently a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, Virginia Tech. His current research interests include semantic communications, statistical signal processing, and machine learning for wireless communications. He was a recipient of the Best Student Paper Award from the 2018 IEEE SPAWC, Kalamata, Greece.

Omar Hashash (Graduate Student Member, IEEE) (omarnh@vt.edu) received the B.E. degree in communications and electronics engineering from Beirut Arab University, Lebanon, in 2019 and the M.E. degree in electrical and computer engineering from the American University of Beirut, Lebanon, in 2021. He is currently pursuing the Ph.D. degree with the Electrical and Computer Engineering Department, Virginia Tech. His research interests include 6G, artificial intelligence (AI), digital twins, metaverse, and edge intelligence. He is a finalist for the 2023 Qualcomm Innovation Fellowship in North America. He served as a technical program committee member in multiple flagship IEEE conferences and a frequent reviewer for several IEEE journals. He is a member of the Wireless@VT Research Laboratory.

Nikhil Muralidhar (nmurali1@stevens.edu) received the B.S. degree in computer science from Virginia Tech in 2012, the M.S. degree in computer science from George Mason University in 2015, and the Ph.D. degree in computer science from Virginia Tech in 2022. During his Ph.D. degree, he was with NEC Laboratories America Inc., in 2017 and Los-Alamos National Laboratories in 2021. He is currently an Assistant professor with the Computer Science Department, Stevens Institute of Technology. He leads the Scientific Artificial Intelligence (ScAI) Laboratory, Stevens Institute of Technology. His research is focused on scientific machine learning. His research efforts have garnered awards in global competitions. He has also served extensively on the program and review committees for conferences like NeurIPS, ICLR, ICML, IEEE Big Data, and SIAM SDM. He received the Academic Excellence Award for his M.S. degree. He received the Two-Year NSF Urban Computing Fellowship for his Ph.D. degree.

Walid Saad (Fellow, IEEE) (walids@vt.edu) received the Ph.D. degree from the University of Oslo, Norway, in 2010. He is currently a Professor with the Department of Electrical and Computer Engineering, Virginia Tech, where he leads the Network Science, Wireless, and Security (NEWS) Laboratory. His research interests include wireless networks (5G/6G/beyond), machine learning, game theory, security, UAVs, semantic communications, cyber-physical systems, and network science. He was a recipient of the NSF CAREER Award in 2013, the AFOSR Summer Faculty Fellowship in 2014, and the Young Investigator Award from the Office of Naval Research (ONR) in 2015. He was a recipient of the 2015 and 2022 Fred W. Ellersick Prize from the IEEE Communications Society, the IEEE Communications Society Marconi Prize Award in 2023, and the IEEE Communications Society Award for Advances in Communication in 2023. He was also the co-author of the articles that received the IEEE Communications Society Young Author Best Paper Award in 2019, 2021, and 2023. He also received the 2017 IEEE ComSoc Best Young Professional in Academia Award, the 2018 IEEE ComSoc Radio Communications Committee Early Achievement Award, and the 2019 IEEE ComSoc Communication Theory Technical Committee Early Achievement Award.

Naren Ramakrishnan (Fellow, IEEE) (naren@cs.vt.edu) is currently the Thomas L. Phillips Professor of engineering with Virginia Tech. His work has been featured in venues, such as the NIH outreach publication *Biomedical Computation Review* and the National Science Foundation's Discoveries series, *Wall Street Journal, Newsweek, Smithsonian Magazine, Popular Science, Chronicle of Higher Education, and ACM TechNews*. His research interests include data science, urban analytics, recommender systems, and computational epidemiology. In these research areas, his articles has garnered more than ten best paper awards. His research has been supported by DHS, DARPA, DOE, DOT&E, DTRA, IARPA, ICWERX, NIH, NEH, NSF, ODNI, ONR, U.S. Army Research Office, U.S. Naval Surface Warfare Center (NSWC), AIRC, Boeing, Ford, General Motors, General Dynamics, HP Labs, L3 Communications, Lockheed Martin, Northrop Grumman, NEC Labs, and The Washington Post. He is a fellow of the Asia-Pacific AI Association (AAIA) and the American Association for Advancement of Science (AAAS).