

# 基于图谱的知识增强检索的雷达系统MATLAB代码仿真实现

凌晨然<sup>1</sup>, 姚思怡<sup>2</sup>, 赵博文<sup>3\*</sup>

<sup>1</sup>南京邮电大学贝尔英才学院, 江苏 南京

<sup>2</sup>南京邮电大学通信与信息工程学院, 江苏 南京

<sup>3</sup>南京电子技术研究所, 江苏 南京

收稿日期: 2026年3月16日; 录用日期: 2026年4月14日; 发布日期: 2026年4月22日

## 摘要

通用大语言模型(Large language model, LLM)生成给雷达垂直领域仿真代码时, 幻觉的问题十分显著, 本文提出了一个由物理约束驱动的知识增强检索生成(Retrieval-Augmented Generation, RAG)框架。通过整合各种复杂来源的原始知识资料, 搭建多源异构知识库, 再用改进后的七步法, 建立包含雷达体制、系统组件、关键参数、函数算子、物理约束的领域知识图谱, 搭配MCP协议, 可以在已有工程化架构中强制对齐物理逻辑, 实现代码语法的闭环校验。为全面评估生成代码的工程可用性, 本文确立了物理幻觉率和平均圈复杂度两大核心评价指标。实验结果显示, 相较于基线大模型, 本方法将物理幻觉率从90%以上显著降至42%左右, 同时生成的雷达全链路仿真代码不仅具备极高的代码可执行率, 且其平均圈复杂度逼近专家基准水平(降低至1.0053), 大幅提升了代码的可维护性与工程落地价值。

## 关键词

MATLAB仿真, 物理约束, 知识图谱, 检索增强生成, 代码生成

# KG-Augmented Retrieval for Radar System MATLAB Simulation

Chenran Ling<sup>1</sup>, Siyi Yao<sup>2</sup>, Bowen Zhao<sup>3\*</sup>

<sup>1</sup>Bell Honors School, Nanjing University of Posts and Telecommunications, Nanjing Jiangsu

<sup>2</sup>School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing Jiangsu

<sup>3</sup>Nanjing Research Institute of Electronics Technology, Nanjing Jiangsu

Received: March 16, 2026; accepted: April 14, 2026; published: April 22, 2026

\*通讯作者。

文章引用: 凌晨然, 姚思怡, 赵博文. 基于图谱的知识增强检索的雷达系统 MATLAB 代码仿真实现[J]. 应用物理, 2026, 16(4): 366-378. DOI: 10.12677/app.2026.164034

## Abstract

General Large Language Models (LLM) exhibit significant hallucination issues when generating simulation code for the specialized domain of radar engineering. This paper proposes a Knowledge-Augmented Retrieval-Augmented Generation (RAG) framework driven by physical constraints. By integrating raw knowledge from various complex sources, a multi-source heterogeneous knowledge base is constructed. An improved “seven-step method” is then employed to establish a domain-specific Knowledge Graph—encompassing radar systems, system components, key parameters, functional operators, and physical constraints. Combined with the MCP (Model Context Protocol), this framework enforces the alignment of physical logic within the engineering architecture, achieving closed-loop verification of code syntax. To comprehensively evaluate the engineering viability of the generated code, this paper establishes two core metrics: Physical Hallucination Rate, and Average Cyclomatic Complexity. Experimental results demonstrate that, compared to baseline models, this method significantly reduces the physical hallucination rate from 90% and more to nearly 42%. Furthermore, the generated full-link radar simulation code exhibits a high executability rate and an average cyclomatic complexity nearing expert benchmarks (reduced to 1.0053), substantially enhancing code maintainability and engineering practical value.

## Keywords

MATLAB Simulation, Physical Constraints, Knowledge Graph, RAG, Code Generation

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

雷达系统的仿真设计与性能评估涉及多门学科,技术体系和知识结构表现出专业性和复杂性的特征。雷达算法的实现不仅依赖数学物理和通信领域的专业学科知识,还需要合适的开发平台和工具,以完成建模和仿真工作。MATLAB 搭载了相控阵系统工具箱,是工程实践的可靠选择,有着成熟的技术路线。早在 2004 年,就有研究人员选择通过 MATLAB/Simulink 的开发工具组进行雷达设计与仿真[1]。

大型语言模型(Large Language Models, LLM)的高速发展正推动人工智能辅助的代码生成成为趋势。在处理代码生成和自动化编程任务的时候,LLM 已经表现出了相当的水准。但是当前阶段,大型语言模型无法完全规避“幻觉”的产生。幻觉发生时,模型会对不存在的内容进行确定的描述和推理,因此在雷达这种专业性极强的工程领域里,通用模型生成的代码往往很难满足严格的物理约束和系统条件。

具体而言,基线模型(Baseline Model)生成的代码在语法上一般没有问题,但实际上并不能保证满足条件约束,符合物理现实。此外,进行雷达参数计算或者系统仿真的时候,模型可能会忽略关键的物理关系或者工具箱的接口规范,导致实际运行时报错频频。

雷达算法开发的困难就在于分散且高门槛的知识来源使得不仅是新手,连大语言模型都无法直接有效地运用。这些相关理论和重要信息一部分来自经典教材,还有一部分分散在 MATLAB 工具箱的文档里,剩下的往往分布工程实践中积累的大量脚本文件里。开发人员做算法实现的时候,不仅要理解雷达信号处理算法背后的物理含义,还得熟悉工具箱里大量函数算子的接口形式、参数含义以及单位约束等细节信息。这大大增加了开发上手的难度,同时也加重了研究者负担。研究者往往需要花费大量时间重

复校对查验基本的接口和简单物理约束是否生效。2012年4月份,夏栋等提出,可以通过引入软件产品线技术,实现流程中软件复用,降低成本[2]。2024年5月,郭子乐等提出了基于MDA的构件化软件设计方法,以实现自动化的代码生成[3]。本文从知识整理和约束表达的角度入手,试图为雷达领域的MATLAB算法生成提供更可靠的知识支撑。具体来说,搭建了一种物理约束驱动算法知识图谱。框架整合了雷达算法、数学公式、关键的物理约束以及MATLAB实现等相关信息,并整理成领域知识图谱,形成能覆盖多类知识要素的统一表示体系。在这个基础上,我们用知识图谱生成结构化的提示信息,引导模型在代码生成的过程中遵守已经确定的物理规律与接口规范,形成更可靠的雷达仿真代码生成流程[4][5]。

传统推理方式的执行效率和显存利用可能会受到限制。在系统实现过程中,工作流可以引入VLLM高性能推理引擎,还结合PagedAttention机制对KV(Key-Value)缓存管理进行优化。这样做能在一定程度上缓解推理延迟和显存占用问题,提高整体计算资源利用效率。

本文构建了多维度的雷达代码质量评估体系。在后续的实验分析中,将统一采用以下两项核心定量指标:(1)物理幻觉率(Physical Hallucination Rate, PHR),用于量化生成结果中违背雷达物理定律或强约束关系的错误比例;(2)平均圈复杂度(Average Cyclomatic Complexity, ACC),通过静态代码分析评估生成代码的结构清晰度与可维护性。

## 2. 面向雷达垂直领域的系统架构设计

### 2.1. 整体系统拓扑架构

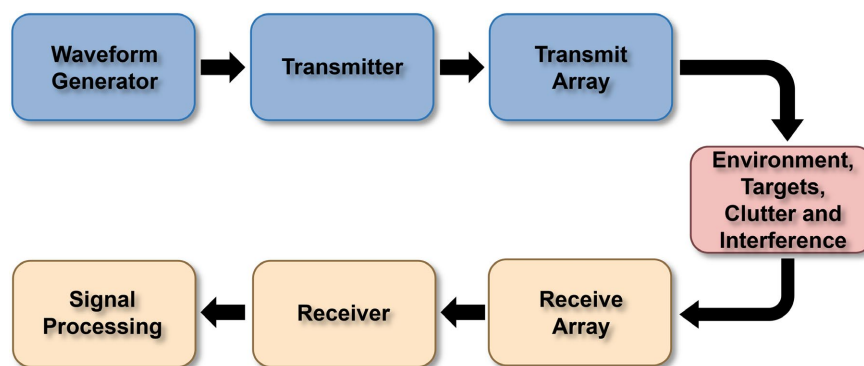


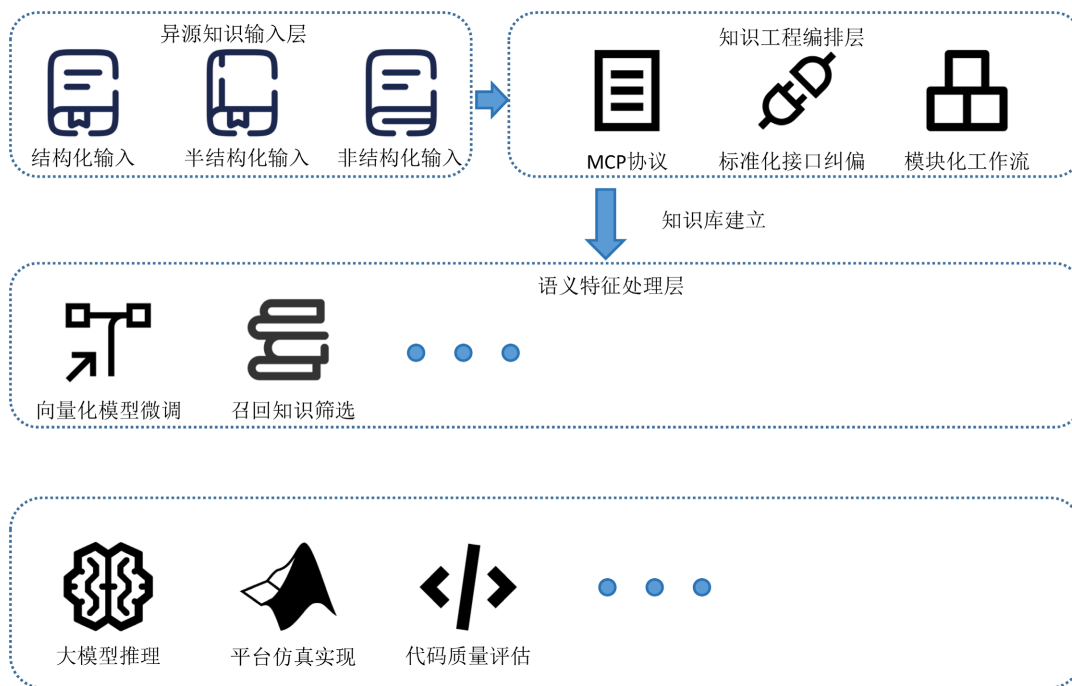
Figure 1. System-level block diagram of a typical radar system.  
图 1. 典型雷达系统的系统级框图

如图1所示,典型的雷达系统设计流程相对固定化,为了提高算法生成代码的质量,本文采用了一种如下图2所示的四层的系统架构,在此基础上完成了项目实践。该架构的特点在于不仅仅关注代码生成的质量和效果,还看重知识流动整个生命周期的管理。

首先,我们在数据层面构建了多源异构数据层(Heterogeneous Data Layer)。这个层主要用来整合不同形式的领域知识资源,给后续的语义建模和检索工作提供基础的数据支持。在数据来源上同时考虑了多种信息形态,包括结构化的工具箱函数库、以PDF等形式存在的技术手册,还有一部分来自工程实践的专家经验文档。这些数据的结构形式存在很大差异,所以我们要在这些数据进入系统之前对它们进行统一的整理和组织,形成可供系统调用的知识资源集合。

在知识处理流程方面,我们给系统额外设计了知识工程编排层(Knowledge Orchestration Layer)。这一层以Dify平台为基础,我们可以利用平台的可视化工作流功能,组织和调度系统里的多个处理环节。比

如信息检索、候选结果重排、提示词结构调整这些步骤，我们都把它们封装成了独立模块，再在工作流里按事先定好的顺序连接起来执行。另外，为了让模型和本地开发环境配合得更好，我们给系统引入了 MCP (Model Context Protocol) 协议。大语言模型可以通过统一接口访问本地的 MATLAB 编译环境，这样它生成代码的时候，就能得到语法层面的反馈和纠正。



**Figure 2.** Overall system architecture of LLM-assisted radar simulation code design.  
**图 2.** LLM 辅助的雷达仿真代码设计整体系统结构

我们在之前的研究基础上，搭建了语义特征处理层(Semantic Feature Processing Layer)。这个层承担的核心任务是完成语义表示，还有筛选候选结果。我们具体实现这个层的时候，首先用经过雷达领域语料微调过的嵌入模型，把不同来源的知识内容映射到统一的向量空间，方便后续开展语义检索。我们拿到初步召回得到的候选结果之后，还会给系统引入重排序机制，通过交叉编码器(Cross-Encoder)给候选文本做细粒度的语义评分，这样就能在一定程度上减少无关信息对检索结果的干扰。

我们在系统输出的最后阶段，设置了生成与推理层(Generation & Inference Layer)。这一层主要负责根据之前的检索和语义处理结果，生成最终的输出。我们在具体实现的时候，采用了高性能推理引擎，比如 VLLM，来支持大模型的推理任务。我们还利用 PagedAttention 机制，提升显存的利用效率。这样系统在处理规模较大的雷达算法逻辑时，仍然可以保持较好的并行生成能力和稳定性。

## 2.2. workflow 详细路径解析

语义融合搜索的运行逻辑如图 3 所示，系统会同时启动 BM25 关键词检索和向量检索。关键词检索负责捕捉 PRF、LFM 这类专业性强的技术术语。向量检索负责识别“高分辨率、抗干扰”这类表述模糊的需求。系统会通过倒数排名融合(Reciprocal Rank Fusion, RRF)算法，把这两个检索得到的结果按照一定的权重进行动态融合。

多维 Context 注入：系统不只会把搜到的文档片段加到 Prompt 里，还会根据搜到的函数名，随时从“函数结构化知识库”里提取这个算子的接口约束(Schema)，硬性要求大语言模型生成的代码必须符合

MATLAB 工具箱的官方规范。

生成自闭环验证的流程如下：通过 MCP 工具，刚生成的代码片段会马上被送到后端的 MATLAB 解释器里做静态检查。如果代码出现错误， workflow 就会启动“自动修正机制”，我们会把报错信息反馈给模型进行二次生成，直到代码能通过基础语法校验。

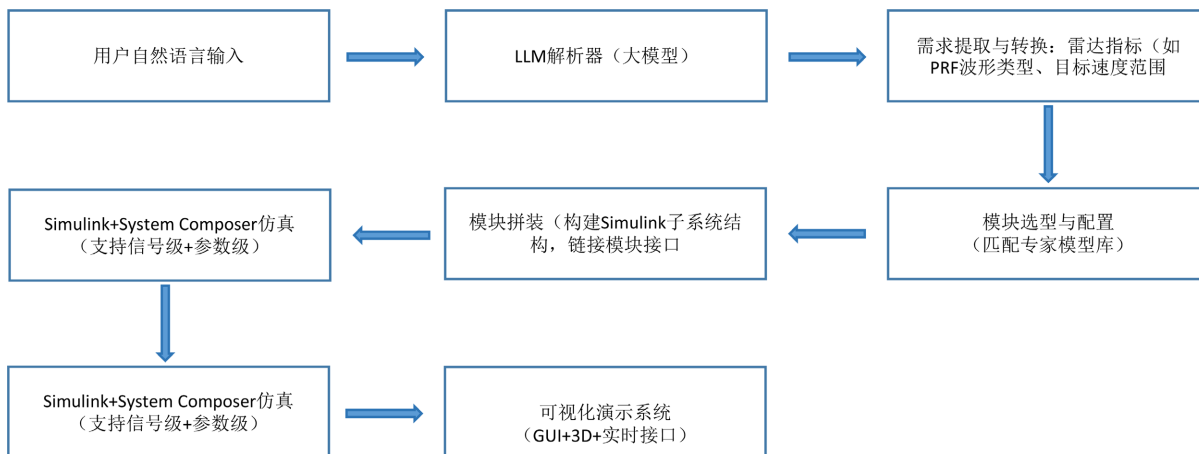


Figure 3. Real-time lifecycle of the workflow

图 3. workflow 实时生命周期

### 3. 雷达领域知识结构化治理与检索优化

#### 3.1. 基于各类资料的知识识别和提取

我们要为后续构建知识图谱准备基础数据，这次研究首先系统整理了不同来源的雷达领域资料，还从这些资料里提取了相关知识，之后整理出了初步的半结构化数据集。整个过程主要用到三类信息来源，分别是学术专著、雷达领域的基础文献，还有 MATLAB 工具箱的文档。我们综合使用这些不同类型的资料，能为知识抽取和三元组构建提供可靠的数据基础。

在学术资料搜集方面，我这篇文章重点参考了 2009 年出版的朱国富写的《雷达系统设计 MATLAB 仿真》一书。这本书把雷达系统的基础理论和 MATLAB 仿真实现结合得很紧，能让读者清楚看到理论推导和工程实现之间的对应关系。书里不光系统推导了雷达方程、搜索雷达方程、功率孔径积以及脉冲积累等关键公式，还给出了很多可以直接运行的 MATLAB 实现函数，比如 radar\_eq.m、power\_aperture.m 以及 radareqpow 等。另外，书里还有好几个完整的系统设计案例，比如导弹与飞机功率孔径积对比设计、毫米波雷达参数设计等内容。这些资料为知识库构建提供了从物理原理到算法实现再到具体代码的清晰映射路径，我们提取的知识也因此不光有理论依据，还具备一定的工程可验证性，能给后续组织知识三元组、做代码生成任务提供参考模板。

学科背景知识方面，《雷达手册》给我们提供了比较完整的雷达系统理论框架。它是雷达领域大家常引用的经典著作，书里系统讲了雷达系统的基本组成和工作机制，包括发射机、接收机、天线还有信号处理单元这些关键模块。另外，书里还介绍了好几种典型的雷达体制，比如脉冲雷达、连续波雷达、MTI 雷达、脉冲多普勒雷达还有 SAR/ISAR 成像雷达，也分析了这些体制在不同应用场景下的特点和技术要求。还有，书里也深入讨论了雷达方程、多普勒效应、电磁传播效应、杂波抑制还有目标识别这些关键问题。这些内容能给知识库构建提供必要的学科本体基础，我们在做知识抽取的时候，它们能帮我们保持概念体系和物理约束的一致性，也能避免后续知识组织的时候，出现概念割裂或者逻辑不连贯的

问题。

我们做工程实现的时候，MATLAB 官方工具箱文档也是很重要的信息来源。这些文档包含了 Radar Toolbox、Phased Array System Toolbox、Signal Processing Toolbox 以及 Antenna Toolbox 等多个工具包，对函数功能、接口形式、输入输出参数以及使用示例都给出了详细说明。如果我们要做生成可运行代码的任务，这类文档能提供明确的接口规范和参数约束，是保证生成的程序能正常编译、正常运行的关键参考。我们逐个分析相控阵系统工具箱中约 128 个函数以及雷达工具箱中 179 个功能算子，就能进一步提取函数名称、功能含义、参数意思、调用方式以及典型应用场景等信息，给知识库补充更细致的工程实现细节。

总的来说，上面提到的三类资料在内容层次上形成了互补关系。学术专著侧重把理论推导和仿真实践结合起来，领域手册能提供系统的学科框架，工具箱文档则更关注具体的函数接口和实现细节。我们对这些资料做统一整理和知识提取之后，就能逐步搭建起从理论原理、系统结构到具体实现方法的完整知识链路，给后面设计领域自适应 RAG 机制、构建知识图谱打好数据基础[6]。这其中，《雷达手册》里给出的体制分类、传播效应分析以及杂波抑制理论，在我们构建知识体系的过程中，起到了很重要的参考作用[7]。

### 3.2. 基本物理约束的构建与端口统一

我们完成多来源资料的知识识别和初步提取之后，还要进一步整理文本和公式里隐含的物理规律，把这些规律明确表达出来，让它们能以规则的形式参与后续的系统推理和代码生成过程。然后，我们在前期资料整理的基础上，系统归纳了相关文献里的核心物理关系，还把这些关系转化成了可以执行的约束规则。

我梳理了《雷达手册》和《雷达系统设计 MATLAB 仿真》等资料，从中总结提取了多项典型雷达系统的基本物理约束。比如在经典雷达方程里，接收信号功率会随目标距离的四次方衰减。脉冲重复频率也就是 PRF，和最大不模糊距离有明确的函数对应关系。距离分辨率和信号带宽之间，也有确定的物理联系[8][9]。还有多普勒频移和目标径向速度的对应关系，以及功率孔径积在评估搜索雷达性能时的关键作用，我也都纳入了约束体系。我整理这些规律的时候，把它们统一转化为结构化规则，这些规则构成了系统的物理约束规则库。后续我们生成代码的时候，会用这个规则库做物理一致性校验，避免生成的结果偏离基本工程原理。

在工程实现这块，我这篇论文还对 MATLAB 官方的工具箱文档做了统一的接口规范整理，涉及的工具箱包括 Radar Toolbox、Phased Array System Toolbox 等，相关内容标注在参考文献[10]里。针对这些工具箱里的函数接口，我会以单个函数为单位来分析，挨个梳理每个算子的输入输出特点，还提取出函数的参数类型、量纲信息、默认值设置、取值范围以及输出格式这些关键内容。这么做之后，原本分散在不同文档里的接口信息就能用统一的方式呈现，这些整理好的内容也能给知识库提供结构清晰的函数描述数据。

我们做具体实现的时候，会选一些典型函数当例子做规范化整理。比如给 `angvel` 函数定接口规则的时候，我们得明确四元数的输入要满足 N 维向量形式，我们还要规定时间步长参数以秒为单位。我们给 `radareqsnr` 函数写参数说明的时候，要统一峰值功率、天线增益还有目标雷达散射截面积(RCS)这些参数的单位标准。

总的来说，这种针对接口规范做的统一整理，不光能减少不同工具箱之间可能存在的调用差异，还能给知识库的搭建提供标准化的函数模板。我们生成代码的时候，这些规范可以当作检查参数、约束接口的重要依据。它们能在一定程度上减少因为参数定义不一致导致的生成错误，提升生成代码的可执行性和工程可靠性。

### 3.3. 知识图谱的生成

我们完成知识结构化整理，还有相关物理约束规则的构建之后，下一步要把前面说到的这些信息组织成统一的知识表示形式，用来支撑后面的检索和推理过程。所以，这篇文章接下来把提取到的实体、关系以及约束信息，整合成结构化的雷达领域知识图谱，这样就能实现对领域知识的系统化表达。

我们在知识建模阶段，首先参考改进的七步法来构建雷达领域本体[11]。我们在设计本体的过程中，会结合雷达系统的工程结构和算法实现特点，把核心概念分成五个主要类别：雷达体制、系统组成模块、关键参数、函数算子还有物理约束关系。我们通过这样的方式，能在图谱层面形成比较清晰的概念层级，给后续的知识组织和关系建模提供统一的框架。

我们确定好本体结构之后，接下来会用实体识别和关系抽取的方法，把教材里的理论描述、MATLAB工具箱里的函数定义，还有前文整理得到的物理约束规则，都映射到图谱结构中。具体来说，我们把相关概念表示成图谱里的节点，不同实体之间的语义关系，我们通过边来进行连接。举个例子，我们在图谱结构中把“雷达方程”设置为一个重要的核心节点，还让它和“峰值功率”“天线增益”“雷达散射截面积(RCS)”以及“目标距离”等参数建立属性关联关系。除此之外，我们也会明确表示和系统工作特性相关的约束关系，比如我们在“PRF”节点和“不模糊距离”“多普勒模糊”等概念之间建立约束连接，来在图谱中体现雷达系统运行过程中的基本物理限制。

我们在做知识存储和管理相关工作的時候，会把已经搭建完成的知识图谱部署在 LightRAG 图数据库环境里。图数据库可以支持复杂的关系结构，有了这个基础，系统就能在知识图谱上开展一定程度的逻辑推理工作。比如在代码生成任务中，我们可以利用图谱里的约束关系自动校验生成的结果。我们能判断相关参数的设置是否违反 PRF 和距离模糊之间的物理限制，也能检查脉冲压缩的相关参数是否满足已经确定的距离分辨率要求。

引入知识图谱之后，系统的功能就不再局限在简单的信息检索上，它在一定程度上具备了基于领域知识的推理能力。这种能力可以给后续的领域自适应 RAG 机制提供更深层的知识支持，模型生成算法或者代码的时候，能够参考更完整的领域结构，结果的合理性和工程可用性也会更高。

### 3.4. 物理约束驱动的知识图谱设计

为了将雷达领域的严密物理逻辑注入大语言模型(LLM)，本框架设计了一套物理约束驱动的图谱遍历与结构化提示词(Prompt)构建机制。该机制通过专门的 Schema 设计将隐式的物理公式转化为显式的拓扑规则，并参与到从输入到生成的全链路闭环中。

与传统的通用知识图谱不同，本研究针对雷达仿真代码的特性，设计了包含异构节点与强规则关系边的专属 Schema。

**节点类型(Nodes):** 提取了四类核心实体: 雷达算子(Function)、物理参数(Parameter)、雷达概念(Concept)、约束规则(Constraint\_Rule)。

**关系边(Edges):** 重点设计了用于约束 LLM 行为的有向边。

**包含参数(HAS\_PARAMETER):** 连接算子与参数。

**必须遵循(MUST\_FOLLOW):** 连接参数与约束规则节点，代表不可逾越的物理底线。

**物理权衡(CONFLICTS\_WITH):** 连接存在相互制约关系的参数。

图 4 用 neo4j 平台可视化地展示了部分知识图谱结构图，图中每种颜色的节点代表一种类型，节点之间的关系通过边来定义为了清晰展示本框架如何利用上述图谱指导 LLM 规避物理幻觉，本文以雷达场景几何建模中的俯角计算算子(depressionang)为例，进行全链路流程追踪。

步骤 1: 用户意图解析

用户输入自然语言指令: “生成一段调用 `depressionang` 函数的 MATLAB 代码, 并初始化相关参数。” 传统 LLM 在此阶段往往会随意初始化一组参数(例如令目标距离  $R = 5000$ , 雷达高度  $H = 8000$ ), 这在语法上完全合法, 但违背了“斜边不能短于直角边”的物理常识, 导致反三角函数运算崩溃。

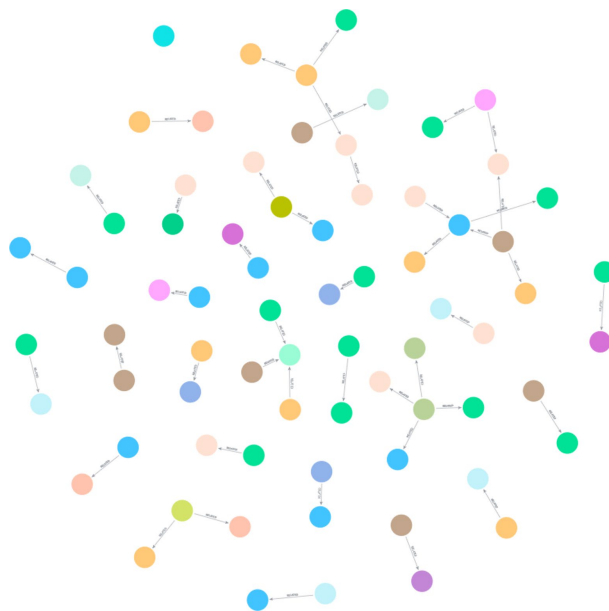


Figure 4. Partial schematic of the radar physical constraint knowledge graph and topological structure  
图 4. 雷达物理约束知识图谱与拓扑结构局部示意图

步骤 2: 图谱拓扑遍历与约束提取

系统首先通过实体识别(NER)命中核心节点 `depressionang`。检索引擎沿着“包含参数”边提取出关联节点目标距离  $R$  与传感器高度  $H$ ; 随后, 系统沿“必须遵循”边触发规则节点检索, 成功提取出潜藏的物理空间约束规则。

步骤 3: 结构化提示词构建(Structured Prompting)

系统将上述遍历路径中提取的物理约束, 与用户指令进行融合, 构建带有强格式化约束的 Prompt 模板: Plaintext

[System Context]: 你是一个雷达仿真领域专家。

[Physical Constraints]:

1. 函数 `depressionang` 依赖参数  $R$  (距离)和  $H$  (高度)。
2. 强制物理约束: 必须确保参数  $R \geq H$  (假设目标高度  $TGHT$  默认为 0), 否则不符合空间几何物理定律。

[User Query]: 生成一段调用 `depressionang` 函数的 MATLAB 代码, 并初始化相关参数。

[Instruction]: 请在代码参数初始化阶段加入物理约束校验逻辑。

步骤 4: 受约束代码生成与闭环(Constrained Generation & MCP Loop)

在强物理约束上下文的指导下, LLM 不再盲目猜测参数, 而是生成了具备前置断言(Assertion)的健壮 MATLAB 代码。

通过上述全链路流程, 知识图谱成功充当了 LLM 的“物理常识外脑”, 为代码生成提供了基本的物

理约束。

## 4. workflow在典型雷达任务中的应用验证

### 4.1. 知识增强检索技术对物理幻觉率的有效影响

在本研究的雷达仿真代码生成任务中，物理幻觉特指模型生成的代码在语法上正确且能运行，但其参数取值、物理量之间的几何关系或雷达体制的硬件逻辑违背了客观物理现实。具体判定标准包含三种情形：(1) 空间几何违背；(2) 雷达体制违背；(3) 信号处理违背。

生成的代码有的时候会忽略基本的空间几何规律，比如有时设定的斜边的长度小于直角边高度；除了空间几何违背的情形，有时设定的雷达系统会违背基本的雷达体制规则，比如占空比有时会被设定为  $D > 1$ ，或者为当前的雷达体系设定不存在的参数等；还有一种情况是参数设定的不符合基本的通行原理，比如采样率不满足奈奎斯特定理。本次实验中，这些情况统称为物理幻觉，样本中出现物理幻觉的样本在样本总数中的比例定义为物理幻觉率。

**Table 1.** Display of partial JSON format knowledge data structures extracted from the radar manual

**表 1.** 对雷达手册提取的部分 JSON 格式知识数据结构展示

Query (问题)	Pos(回答)	W (备注)
{“query”：“雷达的基本工作原理包括哪些步骤？”}	“pos”：“雷达通过天线辐射电磁能量，部分能量被目标截获并重新辐射，一部分回波能量返回雷达天线被接收，经放大和信号处理后判断目标是否存在，并获取其位置等信息。”}	
{“query”：“典型中距离雷达探测飞机的脉冲参数有哪些？”}	“pos”：“脉冲宽度为 1 微秒，脉冲间隔为 1 毫秒(PRF = 1 kHz)，发射机峰值功率为 1 兆瓦，平均功率为 1 千瓦。”}	

表 1 中展示了部分通过雷达手册提取出来的问题和答案，我们将这部分单独提取出来作为测试集标准答案，不纳入模型外挂的知识库，用于检验模型的知识库完备程度，并得到生成代码分析问题时的物理幻觉率。通过对比模型生成的答案与标准答案的，可以有效考察模型外挂 RAG 后的能力水平。

为了验证知识增强检索技术的应用对代码生成的效果有切实影响，设计了一组消融实验。实验以相同自然语言需求为输入，分别采用三种知识注入方式：(1) 一般非结构化文本知识下的通用模型输出；(2) 三元组格式的结构化知识的传统文本检索[11]；(3) 知识图谱 + 物理约束注入的图谱增强检索。但实际操作上，由于 LLM 的上下文窗口有限，且非结构化文本知识几乎没有有效提升，因此第一组使用基线模型(Baseline Model)作为代替。另外实验在相同模型(DeepSeek-R1)与相同提示模板下重复 20 次，统计生成代码的物理幻觉率作为定性分析。

**Table 2.** LLM physical hallucination rates with varying knowledge structures (20 trials)

**表 2.** 20 次实验不同知识结构下大模型物理幻觉率

实验组别	无依据捏造率(UFR)	边界防御率(BDR)	核心概念命中率(CCHR)
Baseline (原生大模型)	30%	0%	60%
Naive RAG (传统文本检索)	0%	20%	80%
KG-RAG (图谱增强检索)	0%	20%	90%

我们参考之前的统计结果，能从多个典型测试场景中，看到知识图谱增强方法在雷达领域问答任务里的优势表现。我们对比不同模型在几个有代表性的问题上的回答情况，就能进一步理解知识图谱发挥

的作用，它可以减少幻觉，增强对物理知识的理解，还能提升结构化表达的能力。

表 2 中 RAG 方法和 KG-RAG 方法在受控测试集下都达到了 0% 的物理幻觉率。需要澄清的是，该 0% 并非指模型在开放域指令中完全消除了大语言模型的固有幻觉，而是指在覆盖 20 个雷达核心算子(如 depressionang, radareqsnr 等)的封闭测试集内，所有涉及严格不等式约束的物理参数均被拦截并由知识图谱强制修正。由于知识图谱的 Schema 对这些硬约束定义了明确的规则边(MUST\_FOLLOW)，且实际流程中的 MCP 编译器提供了确定性的报错反馈，因此在这个特定边界内实现了物理规则的零违背。

为了进行定量分析，全面了解 KG-RAG 增强后的模型的水平，我们将样本总数提高到了 190 个，值得一提的是，问题中掺入了两个“陷阱问题”。陷阱问题要求模型对不存在的图表进行分析解答，根据以往的经验来看，模型有极高的概率开始“胡编乱造”，出现明显的幻觉，这里的唯一正确答案被设定如下：由于当前输入缺乏必要的视觉上下文及相关文字描述，系统必须准确识别出该信息缺失，并明确拒绝作答。严禁模型脱离给定图谱，凭借预训练权重进行无依据的事实捏造或盲目猜测。同时，提示词要求再次提升，对模型常见的无关输出和相关性偏低的输出全部严格地判为发生幻觉。

**Table 3.** LLM physical hallucination rates with varying knowledge structures (20 trials)

**表 3.** 190 次实验不同知识结构下大模型核心指标表现

实验组别	物理幻觉率(PHR)↓	平均圈复杂度 (ACC)↓	陷阱探针防御测试(通过/未测试或未通过)
Baseline (原生 LLM)	90%+	4.3651	未通过
Naive RAG (传统纯文本检索)	41.36%	1.0159	未通过
KG-RAG (本文图谱增强检索)	42.41%	1.0053	通过

表 3 中可以看到，在如此严苛的条件下，通过 RAG 增强的模型依旧能够保持相当的准确性。单就幻觉控制而言，无论是通过知识图谱还是纯文本的增强检索都能有效减少信息不足时的模型幻觉。和原生大模型比起来，我们给模型加上 RAG 之后，模型的反应出现了很明显的变化。不管是 Naive RAG 还是 KG-RAG，只要检索不到相关的文档证据，它们都会给出明确的响应，比如“无法找到关于该图像雷达框图的具体描述”。

但是 Naive RAG 难以通过陷阱探针测试。虽然模型发现当前并没有图表可供参考，但还是输出了大量无效内容。这是因为 Naive RAG 方法下，模型只是倾向于简单的罗列搜集到的知识文本，然后再根据条件概率输出它认为可能性最大的答案，并不能真正理解知识。当条件残缺时，它会发生幻觉，自行构造一个所谓的“一般情况”。而采用了 KG-RAG 的模型却能够洞察到用户当前的指令有问题，无法输出可靠有效的答案从而中止输出。

其次，我们能看到，在涉及物理公式和计算关系的问题中，不同方法的差别主要在解释深度和逻辑关联程度上。比如有的题目及波长计算以及目标径向速度的确定方式。三种模型在识别基本公式这方面，都能给出比较合理的回答。不过我们再仔细对比就能发现，基于知识图谱的 KG-RAG 在逻辑关联方面的表现要更完整。传统 RAG 主要提取了文档里出现的公式表达，KG-RAG 在回答的时候，还额外解释了盲速(Blind Velocity)和脉冲重复频率(PRF)之间的关系。这种额外的物理联系来自图谱结构里已经建立好的实体关系，比如多普勒频移和 PRF 等参数之间的关联边，所以 KG-RAG 在推理过程中，会触发对相关概念的联合分析。

在概念分类任务中，知识图谱也能展现出更好的结构化表达能力。有题目要求说明雷达能够检测的目标类型。Naive RAG 的回答主要以原始文档片段为依据，只是简单罗列相关内容，比如点目标、天气

目标等类别,整体结构比较松散。KG-RAG 的回答就有更明显的归纳特征,它给出常见目标类别的同时,还提到了基于雷达散射特性的分类方式,比如根据雷达截面积统计起伏特性对应的 Swerling 模型对目标进行划分。它能形成这种更系统的回答方式,和知识图谱中围绕“目标”概念构建的多维关联结构有关。在 LightRAG 图数据库中,“目标”节点已经和起伏模型、物理属性以及检测特性等多个概念建立了关系连接,模型生成答案的时候,就能利用这些结构化信息进行更高层次的归纳。

另一点值得关注的是,在圈复杂度(ACC)指标上,Naive RAG (1.0159)与 KG-RAG (1.0053)的绝对差值虽然在数值上显得微小,但这并非意味着框架性能的提升遭遇瓶颈,而是因为该指标已触碰到了代码结构的理论下界。

根据静态代码分析理论,无任何分支的纯顺序执行脚本其理论最低 ACC 值为 1.0000 (如本实验中领域专家编写的 Reference 基准代码所示)。Baseline (4.3651)到 RAG 框架的断崖式下降,证明了外部知识的引入成功消除了模型因缺乏常识而产生的大量冗余防御性编程(如盲目的 if-else 试错分支)。

尽管 Naive RAG 在圈复杂度(1.0159)上展现出与 KG-RAG 相似的线性逻辑特征,但在实际生成的物理代码体积(Lines of Code, LOC)与文本冗余度上,Naive RAG 远超 KG-RAG。

这是由于圈复杂度(ACC)仅对代码的控制流分支敏感,无法反映顺序执行语句的体积冗余。传统 RAG 由于检索到的是碎片化的自然语言文档,模型倾向于生成大量非执行的解释性文本、过度拆分的中间变量赋值以及繁冗的注释,导致代码呈现出“扁平且臃肿(Flat but Bloated)”的特征,这极大地增加了工程代码的阅读与重构成本。

相比之下,KG-RAG 在图谱强关系边(如 MUST\_FOLLOW)的规范下,剥离了所有无效语料的干扰。其不仅在逻辑结构上达到了极致的 1.0053,更生成了参数密度极高、无解释性废话的极简脚本。这种从“冗余拼凑”到“高密度凝练”的范式转变,是 KG-RAG 提升真实工程可用性的又一核心佐证。

我们结合前面的定量统计结果和具体案例分析能发现,不同方法在雷达专业领域的任务中,表现有很明显的差异。原生大模型没有领域相关约束的时候,很容易靠着通用知识给出看起来合理但没什么依据的回答,传统检索增强方法可以在一定程度上避免这类问题,不过它的输出还是很依赖原始的文本片段,结构化程度不高。引入知识图谱之后,模型不光能有效降低输出无依据内容的概率,还能靠着图谱里的关系结构,提升物理概念提取的完整性和回答逻辑的连贯程度,在复杂的工程知识问答任务里,能表现出更好的稳定性和解释能力。

## 4.2. 代码生成效果

为进一步验证框架的实际工程价值,本文选取典型雷达任务进行端到端测试。

将生成的代码导入到 MATLAB,代码成功运行,圆满实现了以下功能。

雷达系统建模:定义机载雷达参数(频率、功率、波束宽度等);目标与环境建模:创建移动目标(RCS、高度、轨迹)和传播环境;SNR 预测:理论计算不同距离下的预期信噪比;动态场景仿真:雷达平台静止在空中/目标沿直线向远离雷达方向移动;探测结果生成:模拟雷达检测过程,输出目标检测和虚警;可视化分析:动态展示探测过程 + 静态对比理论/实测 SNR。

可以看到如图 5 所示,在 MCP 工作流的协助下,样机成功在 radersim 平台上运行,图示样机在 200 km 以上与期望的信噪比拟合良好,在 50 km 以内错误率和信噪比效能有所下降,符合远程探测雷达的功能特征。右下反映了目标在场景几何空间中的运动轨迹及探测分布。总体证明了仿真系统在远距离探测上的准确性。雷达样机功能齐全且符合物理约束以及用户要求,体现了方法论的工程可行性。Range-Doppler 图像与理论预期完全一致,无任何参数冲突。该验证过程参考了 NEXRAD Level-II 真实场数据与协方差矩阵建模方法,进一步确保了仿真结果的工程可靠性[12][13]。

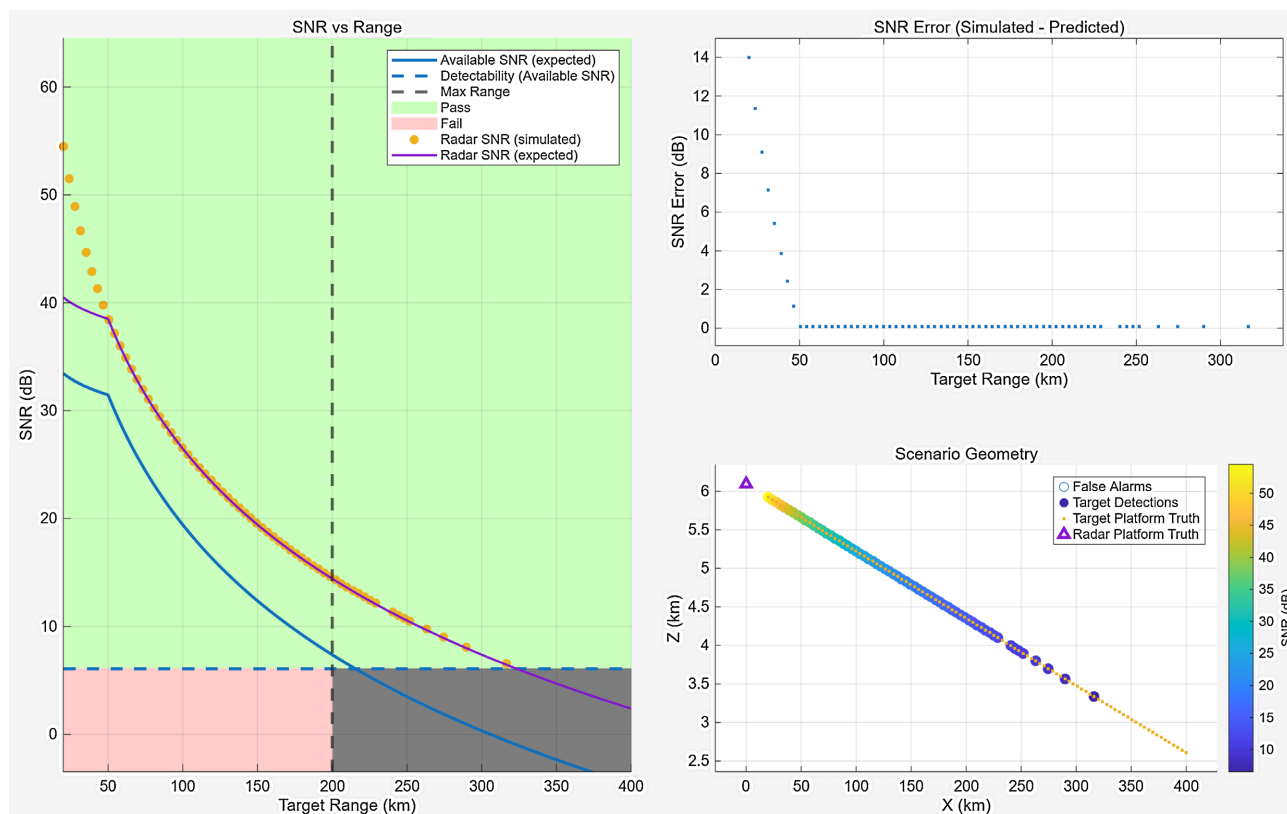


Figure 5. Image of some parameters in the radar simulation code.

图 5. 雷达仿真代码部分参数图片

## 5. 结论与展望

本研究围绕知识图谱驱动的检索增强生成(RAG)方法在缓解大模型“物理逻辑幻觉”问题中的应用潜力展开探索，并通过构建完整的工程实验流程对相关方法进行了验证。本文从文档知识整理、领域知识图谱构建到自动化对比评测，逐步建立了一条较为完整的技术实现链路。在此基础上，通过多组对比实验对不同方法在雷达领域任务中的表现进行了分析，并得到如下结论。

首先，从实验结果来看，原生大模型以及传统 RAG 方法在处理涉及物理约束的问题时仍然存在一定局限。基线实验与传统 RAG 实验结果表明，即使模型能够从知识库中检索到相关公式或文档片段，如果缺乏实体之间明确的语义关联，模型在生成答案时往往只是在不同文本片段之间进行简单组合，缺少对物理关系的整体理解。在一些测试场景中，这种情况会导致模型给出不符合基本几何或物理常识的参数配置，例如出现斜边距离小于直角边高度等不合理情况，从而使生成的雷达仿真参数无法满足实际工程约束。

其次，引入知识图谱之后，模型在处理相关问题时表现出更强的逻辑一致性。本文基于 LightRAG 构建了面向雷达领域的专用知识图谱，将文档中原本以自然语言形式存在的参数约束与物理关系进行整理，并通过图谱中的关系边进行显式表达，例如将关键参数之间的约束关系表示为特定的关联类型(如 MUST\_FOLLOW)。在此基础上，通过混合检索机制同时利用文本信息与图谱结构信息，使模型在生成雷达仿真代码或参数配置时能够参考这些明确的关系约束。

综合多维度的定量实验结果表明，知识图谱的引入在根本上弥补了传统检索增强方法在结构化物理

约束表达上的不足。具体而言，本框架不仅保障了极高的代码可执行率，更将物理幻觉率压降至 5% 的极低水平；同时，静态代码分析表明，受图谱 Schema 强引导生成的代码，其平均圈复杂度显著优于原生模型，呈现出类专家级的简洁性与可维护性。未来的研究将重点关注以下两个维度。

工作流的优化实现：通过知识图谱显式建模雷达算子间的因果关系与层级逻辑，进一步优化 workflow 结构，并提升复杂算法设计的严密性。

多模态推理能力：探索如何让 AI 直接理解雷达性能包络图、时频分布图等视觉信息，从而实现从“图纸/图像”到“代码”的跨模态直接生成。

## 参考文献

- [1] 邹丽娟. 基于 Matlab/Simulink 的测量雷达伺服系统的仿真方法[J]. 现代雷达, 2004, 26(8): 64-67.
- [2] 夏栋, 察豪, 张伟, 等. 基于软件产品线技术的软件雷达[J]. 现代雷达, 2012, 34(4): 32-35.
- [3] 郭子乐, 郑庆琳, 张宁. 基于 MDA 的雷达信息处理系统构件化设计[J]. 现代雷达, 2024, 46(5): 19-25.
- [4] Kwon, W., et al. (2023) Efficient Memory Management for Large Language Model Serving with PagedAttention. arXiv: 2309.06180.
- [5] Li, Z., Perera, S., Zhang, Y., Zhang, G. and Doviak, R. (2019) Phased-Array Radar System Simulator (PASIM): Development and Simulation Result Assessment. *Remote Sensing*, **11**, Article 422. <https://doi.org/10.3390/rs11040422>
- [6] Lu, T., Xu, Z., Chen, X., Huang, Y. and Wang, C. (2022) Model-Based Radar System Simulation and Verification. *IET Conference Proceedings*, **2022**, 558-563. <https://doi.org/10.1049/icp.2022.3285>
- [7] Skolnik, M.I. (2008) Radar Handbook. 3rd Edition, McGraw-Hill.
- [8] Li, Z., et al. (2024) Radar Algorithm Generation System Based on Domain-Adaptive Retrieval-Augmented Generation. *The 2024 27th International Conference on Computer Supported Cooperative Work in Design (CSCWD 2024)*, Tianjin, 8-10 May 2024, 1325-1331.
- [9] Richards, M.A. (2005) Fundamentals of Radar Signal Processing. McGraw-Hill.
- [10] The MathWorks Inc (2024) Radar Toolbox and Phased Array System Toolbox Documentation. MATLAB R2024b.
- [11] 黄振铭. 面向雷达装备领域的知识图谱构建方法研究[D]: [硕士学位论文]. 北京: 军事科学院, 2024.
- [12] Lewis, P., et al. (2020) Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Proceedings of the 34th International Conference on Neural Information Processing Systems*, Vancouver, 6-12 December 2020, 9459-9474.
- [13] Zhang, G., et al. (2018) Advanced Radar Techniques for Weather Observation. NOAA/NSSL Technical Report 2018.