

Recommendation **ITU-T F.748.52 (03/2025)**

SERIES F: Non-telephone telecommunication services

Multimedia services

Requirements and evaluation methods for retrieval augmented generation of large scale pre-trained models

ITU-T F-SERIES RECOMMENDATIONS
Non-telephone telecommunication services

TELEGRAPH SERVICE	F.1-F.109
Operating methods for the international public telegram service	F.1-F.19
The gentex network	F.20-F.29
Message switching	F.30-F.39
The international telemesssage service	F.40-F.58
The international telex service	F.59-F.89
Statistics and publications on international telegraph services	F.90-F.99
Scheduled and leased communication services	F.100-F.104
Phototelegraph service	F.105-F.109
MOBILE SERVICE	F.110-F.159
Mobile services and multideestination satellite services	F.110-F.159
TELEMATIC SERVICES	F.160-F.399
Public facsimile service	F.160-F.199
Teletex service	F.200-F.299
Videotex service	F.300-F.349
General provisions for telematic services	F.350-F.399
MESSAGE HANDLING SERVICES	F.400-F.499
DIRECTORY SERVICES	F.500-F.549
DOCUMENT COMMUNICATION	F.550-F.599
Document communication	F.550-F.579
Programming communication interfaces	F.580-F.599
DATA TRANSMISSION SERVICES	F.600-F.699
MULTIMEDIA SERVICES	F.700-F.799
ISDN SERVICES	F.800-F.849
UNIVERSAL PERSONAL TELECOMMUNICATION	F.850-F.899
ACCESSIBILITY AND HUMAN FACTORS	F.900-F.999

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T F.748.52

Requirements and evaluation methods for retrieval augmented generation of large scale pre-trained models

Summary

Recommendation ITU-T F.748.52 establishes a framework of technical requirements and evaluation methods for retrieval augmented generation based on large scale pre-trained models. This Recommendation offers users a better understanding of retrieval augmented generation technology and helps them set up retrieval augmented generation applications more quickly and suitably. This Recommendation also provides guidelines for users to examine the level of retrieval augmented generation applications' capabilities and promotes the effectiveness of applications.

History *

Edition	Recommendation	Approval	Study Group	Unique ID
1.0	ITU-T F.748.52	2025-03-16	21	11.1002/1000/16242

Keywords

Retrieval augmented generation.

* To access the Recommendation, type the URL <https://handle.itu.int/> in the address field of your web browser, followed by the Recommendation's unique ID.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, and information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents/software copyrights, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the appropriate ITU-T databases available via the ITU-T website at <https://www.itu.int/ITU-T/ipr/>.

© ITU 2025

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

Table of Contents

	Page
1 Scope	1
2 References.....	1
3 Definitions	1
3.1 Terms defined elsewhere	1
3.2 Terms defined in this Recommendation.....	1
4 Abbreviations and acronyms	1
5 Conventions	2
6 Framework of requirements and evaluation methods for retrieval augmented generation of large scale pre-trained models	2
7 Requirements and evaluation methods of dataset construction	2
7.1 Data access	2
7.2 Data segmentation	3
7.3 Data indexing.....	3
7.4 Data updates	4
7.5 Data storage	4
8 Requirements and evaluation methods of technical capabilities	4
8.1 Retrieval capability.....	4
8.2 Generative capability	6
8.3 Optimization capability	7
9 Requirements and evaluation methods of application capabilities.....	9
9.1 Application maturity.....	9
9.2 Application stability	11
Annex A – Grading requirements for generation capability	15
A.1 Answers relevance grading requirements.....	15
A.2 Answers faithfulness grading requirements	15
A.3 Answers truthfulness grading requirements	16
A.4 Answers relevance grading requirements.....	16

Introduction

Retrieval augmented generation technology utilizes external knowledge sources to improve the accuracy and relevance of large scale pre-trained models and provides an efficient and safe solution for enterprises to apply large scale pre-trained models. Therefore, it is very important to evaluate the retrieval augmented generation technology capabilities. This proposal provides general requirements for retrieval augmented generation technology and products that can provide a specification and basis for both retrieval augmented generation technology parties and application parties.

Recommendation ITU-T F.748.52

Requirements and evaluation methods for retrieval augmented generation of large scale pre-trained models

1 Scope

This Recommendation focuses on retrieval augmented generation technology and applications. The scope of this Recommendation includes four parts:

- Framework of requirements and evaluation methods for retrieval augmented generation of large scale pre-trained models.
- The requirement and evaluation methods for data construction capabilities.
- The requirement and evaluation methods for technical capabilities.
- The requirement and evaluation methods for application capabilities.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

None.

3 Definitions

3.1 Terms defined elsewhere

None.

3.2 Terms defined in this Recommendation

This Recommendation defines the following term:

3.2.1 retrieval augmented generation: A technique that combines retrieval models and generative models, which utilizes information from large-scale databases or knowledge bases to improve the accuracy and reliability of generative artificial intelligence models.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

API	Application Programming Interface
HTML	Hyper Text Markup Language
HyDE	Hypothetical Document Embeddings
MAP	Mean Average Precision
RAG	Retrieval Augmented Generation
ROUGE	Recall-Oriented Understudy for Gisting Evaluation

SDK	Software Development Kit
TPS	Transactions Per Second
XML	Extensible Markup Language

5 Conventions

In this Recommendation:

- The keywords "**is required to**" indicate a requirement that must be strictly followed and from which no deviation is permitted if conformance to this document is to be claimed.
- The keywords "**is recommended to**" indicate a requirement that is recommended but which is not absolutely required. Thus, this requirement need not be present to claim conformance.

6 Framework of requirements and evaluation methods for retrieval augmented generation of large scale pre-trained models

The development route of retrieval augmented generation technology and products consists of large-scale databases or knowledge bases construction, information retrieval, answer generation and application ecology stages. This Recommendation broadly forms an evaluation method to evaluate the capability of retrieval augmented generation techniques for dataset construction, retrieval and generation, as well as application. Figure 1 shows the framework of requirements and evaluation methods for retrieval augmented generation of a large scale pre-trained model.

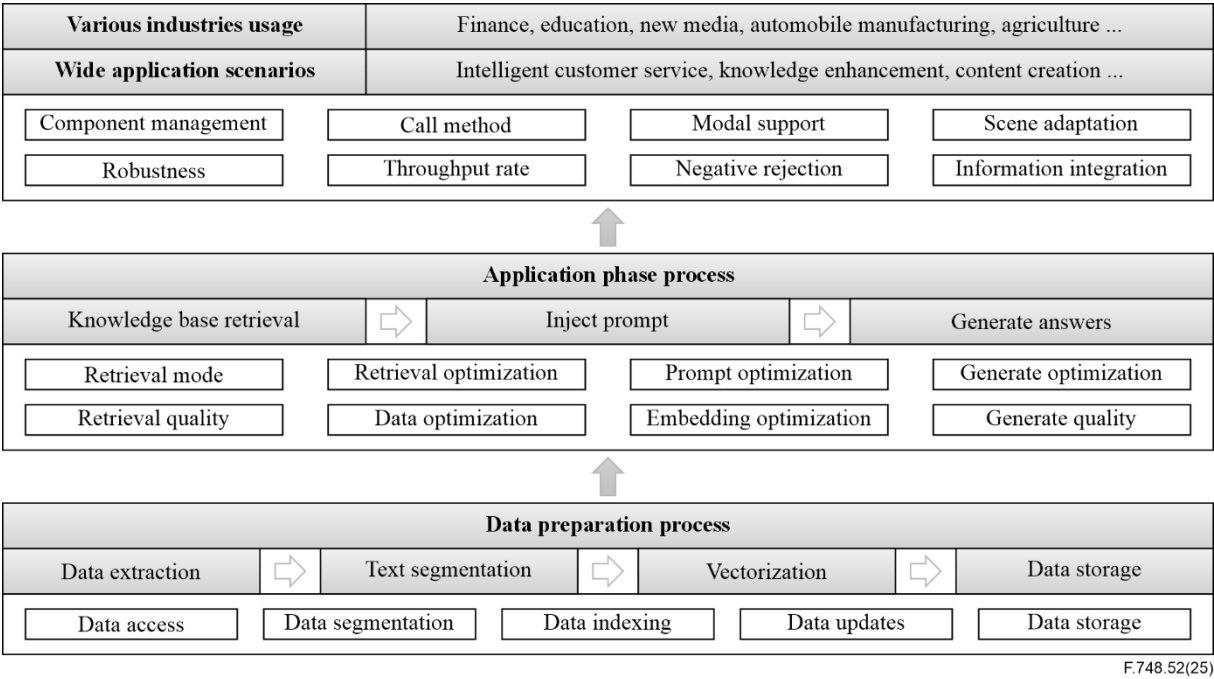


Figure 1 – Framework of requirements and evaluation methods for retrieval augmented generation of large scale pre-trained model

7 Requirements and evaluation methods of dataset construction

7.1 Data access

Technical requirements: The retrieval augmented generation technology is required to have the capability to handle multiple data types and support large amounts of data during the dataset construction process.