كلية الهندسة
قسم الهندسة المعلوماتية
هندسة الذكاء الصنعي وعلوم البيانات

نظام بحث متعدد الوسائط للمحتوى التعليمي

Multimodal search for educational content

مشروع تخرج ١ - قدم استكمالا لمتطلبات الحصول على درجة البكالوريوس في هندسة المعلوماتية- هندسة الذكاء الصنعي وعلوم البيانات

اعداد
بيان الرحيل – مريم عبد السلام

باشراف الدكتور

انس عبد العزيز

2026

Faculty of Engineering
Department of Informatics Engineering
Artificial Intelligence and Data Science

Multimodal Search for Educational Content

Senior1 Project- Completed the requirements for obtaining a bachelor's degree in Informatics Engineering - Artificial Intelligence and Data Science Engineering

Prepared by
Bayan Alrahil Mariam abd ulslam

Supervised

By

Eng. Anas Abdulaziz

2026

Supervisor certification

I certify that the preparation of this project entitled Multimodal search for educational content, prepared by Bayan alrahil and Mariam abdalsalam was made under my supervision at Department of Artificial intelligence & data science as part of the Requirements for the fulfillment of the Bachelor Degree in Artificial intelligence and data science .

Name: Eng. Anas Abdulaziz Signature:……..…. Date: ……

الملخص

يركز هذا المشروع على تطوير نظام مبتكر للبحث واسترجاع المعلومات متعدد الوسائط مخصص للمحتوى التعليمي، يهدف إلى تجاوز محدودية أنظمة البحث التقليدية التي تعتمد حصرياً على المطابقة النصية للكلمات المفتاحية .يسعى النظام إلى تحسين تجربة التعلم من خلال تمكين الطلاب والمعلمين من استرجاع الموارد التعليمية (نصوص، صور، ومخططات علمية) بناءً على المعنى الدلالي والسياق البصري تعتمد المنهجية المقترحة على بنية تقنية متطورة تدمج بين إطار عمل LangChain وLangGraph لتنظيم سير العمليات المعقدة، مع استخدام قاعدة البيانات الشعاعية Milvus لفهرسة التمثيلات المتجهية عالية الأبعاد التي يتم توليدها بواسطة نموذج التضمين BGE ولضمان دقة الإجابات، تم توظيف تقنيات الاسترجاع المعزز بالتوليد (RAG) التي تربط المحتوى المسترجع من مجموعة بيانات ScienceQA بالنماذج اللغوية الكبيرة (LLMs) لتوليد ردود تعليمية دقيقة وموثوقة.

تساهم نتائج هذا البحث في تقديم حل ذكي يعزز من كفاءة الوصول للمصادر التعليمية، حيث أظهر التحليل الاستكشافي للبيانات (EDA) قدرة النظام على التعامل مع سياقات تعليمية متنوعة تشمل الفيزياء والأحياء والكيمياء للمراحل الدراسية المختلفة .يمثل هذا النظام خطوة نحو بيئات تعلم إلكتروني أكثر تفاعلية وذكاءً وشمولية

ABSTRACT

This project focuses on developing an innovative Multimodal Information Retrieval system tailored for educational content, aimed at overcoming the constraints of traditional search engines that rely solely on keyword matching. The system enhances the learning experience by enabling students and educators to retrieve educational resources—comprising text, images, and scientific diagrams—based on semantic meaning and visual context.

The proposed methodology utilizes an advanced technical architecture integrating the LangChain and LangGraph frameworks to orchestrate complex workflows. It employs the Milvus vector database to index high-dimensional embeddings generated by the BGE embedding model. To ensure response accuracy, Retrieval-Augmented Generation (RAG) techniques are implemented, linking retrieved content from the ScienceQA dataset with Large Language Models (LLMs) to generate precise and reliable educational answers.

The findings of this research contribute to providing an intelligent solution that improves access to educational materials. Exploratory Data Analysis (EDA) demonstrated the system's capability to handle diverse educational contexts across physics, biology, and chemistry for various academic levels. This system represents a significant step toward more interactive, intelligent, and inclusive e-learning environments.

فهرس المحتويات

Supervisor certification ٣

الملخص ٤

ABSTRACT ٥

فهرس الاشكال ٩

فهرس الجداول ١٠

قائمة الاختصارات(List of Abbreviations) ١١

الفصل الاول : المقدمة ١٢

١. المقدمة(Introduction) ١٣
٢. مشكلة البحث(Problem Statement) ١٣
٣. أهداف المشروع(Project Objectives) ١٣
٤. النظام المقترح(Proposed System) ١٤
٥. تنظيم التقرير(Report Organization) ١٤
٦. الملخص(Summary) ١٥

الفصل الثاني: الدراسات المرجعية والأعمال ذات الصلة ١٦

2.1مقدمة الفصل(Introduction) ١٧
2.2المفاهيم الأساسية(Fundamental Concepts) ١٧
2.2.1الذكاء الاصطناعي متعدد الوسائط(Multimodal Artificial Intelligence) ١٧
2.2.2البحث الدلالي(Semantic Search) ١٧
2.2.3التضمينات والتمثيلات المتجهية(Embeddings and Vector Representations) ١٧
2.2.4النماذج اللغوية الكبيرة(Large Language Models) ١٨
2.2.5الاسترجاع المعزز بالتوليد(Retrieval-Augmented Generation) ١٨
2.2.6 LangChain ١٨
2.2.7 LangGraph ١٩
2.3مراجعة الدراسات الخاصة بأنظمة البحث متعددة الوسائط ١٩
2.4مراجعة الدراسات المتعلقة باسترجاع المحتوى التعليمي ٢٠

الفصل الثالث: مجموعة البيانات وآليات معالجتها وتحضيرها ٢٣

3.1مقدمة الفصل ٢٤
3.2نظرة عامة على مجموعة بياناتScienceQA ٢٤
3.3بنية مجموعة البيانات ٢٥
3.3.1مكونات كل عينة بيانات ٢٥
3.4الإحصائيات العامة لمجموعة البيانات ٢٦
3.5أنواع البيانات(Multimodal Data) ٢٦
3.5.1البيانات النصية ٢٦
3.5.2البيانات الصورية ٢٦
3.6التحليل الاستكشافي للبيانات(Exploratory Data Analysis – EDA) ٢٨

3.6.1توزيع طول الأسئلة النصية ٢٨
3.6.2توزيع أنواع الصور ٢٨
3.6.3توزيع المجالات العلمية ٣٠
3.6.3الخاتمة ٣٦
3.7معالجة البيانات وتحضيرها(Data Preprocessing) ٣٦
3.8تمثيل البيانات والفهرسة ٣٧
3.9ملاءمة مجموعة بيانات ScienceQA للنظام المقترح ٣٧
3.10ملخص الفصل ٣٨
الفصل الرابع: المنهجية المقترحة ٣٩
4.1مقدمة الفصل ٤٠
4.2النظرة العامة للنظام ٤٠
4.3إطار تنسيق العمليات باستخدام LangChain وLangGraph ٤١
4.4معالجة مدخلات المستخدم(User Input Handling) ٤٤
4.4.1المدخل النصي(Text-Based Input) ٤٤
4.4.2إعادة الصياغة والتوسيع الدلالي(Rephrasing and Semantic Expansion) ٤٤
4.4.3تحويل النص إلى تمثيل متجهي(Text Embedding) ٤٤
4.5معالجة المدخلات الصورية(Image-Based Input) ٤٥
4.5.1تحليل الصورة واتخاذ قرارOCR ٤٥
4.5.2استخراج النص من الصورة(OCR Processing) ٤٥
4.5.3تمثيل المحتوى البصري(Visual Embedding) ٤٥
4.6الربط بين النص والصورة(Metadata Linking) ٤٥
4.7 الاسترجاع متعدد الوسائط(Multimodal Retrieval) ٤٦
4.8 التوليد المعزز بالاسترجاع(RAG-based Answer Generation) ٤٦
4.9دمج الوسائط والتحقق من الاتساق ٤٦
4.10 مميزات المنهجية المقترحة ٤٦
4.11ملخص الفصل ٤٨
الفصل الخامس: التنفيذ ٤٩
5.1مقدمة الفصل ٥٠
5.2بيئة التطوير والأدوات المستخدمة ٥٠
5.2.1لغة البرمجة وإطار العمل ٥٠
5.2.2النماذج الذكية المستخدمة ٥٠
5.2.3قاعدة البيانات المتجهية ٥٠
5.2.4واجهة المستخدم ٥١
5.3تنفيذ معالجة المدخلات النصية ٥١
5.3.1استقبال السؤال النصي ٥١
5.3.2تصحيح السؤال لغويًا ٥١
5.3.3إعادة كتابة وصياغة السؤال(Question Rewriting) ٥٢
5.3.4التوسيع الدلالي للسؤال(Semantic Expansion) ٥٢
5.3.5تقسيم السؤال المركّب(Question Decomposition) ٥٢
5.3.6تحويل السؤال النهائي إلى تمثيل متجهي(Embedding) ٥٣

5.4تنفيذ معالجة المدخلات الصورية ٥٣
5.4.1استقبال الصورة وتحليلها..... ٥٣
5.4.2قرار استخدامOCR ٥٣
5.4.3تحويل المحتوى إلىEmbeddings ٥٤
5.5تنفيذ الفهرسة وقاعدة البيانات المتجهية ٥٥
5.5.1تخزين البيانات..... ٥٥
5.5.2البحث والاسترجاع..... ٥٥
5.6تنفيذ نظام الاسترجاع المعزز بالتوليد(RAG) ٥٧
5.7تنفيذ واجهة المستخدم ٥٨
5.8التحديات التقنية والحلول ٥٩
5.9ملخص الفصل..... ٥٩
الفصل السادس: التقييم ٦٠
6.1 مقدمة الفصل..... ٦١
6.2 أهداف التقييم..... ٦١
6.3بيانات التقييم ٦١
6.4منهجية التقييم ٦١
6.4.1تقييم مرحلة الاسترجاع..... ٦١
6.4.2تقييم مرحلة التوليد..... ٦١
6.5مقاييس التقييم المستخدمة ٦٢
6.6إعدادات التجربة ٦٢
6.7نتائج التقييم وتحليلها..... ٦٢
6.7.1نتائج الاسترجاع ٦٢
6.7.2نتائج التوليد : ٦٢
6.8 مناقشة النتائج..... ٦٢
6.9حدود التقييم ٦٢
6.10خلاصة الفصل..... ٦٣
الفصل السابع: الخاتمة والآفاق المستقبلية..... ٦٤
7.1مقدمة الفصل ٦٥
7.2هيكلية التقرير وأهداف الفصول..... ٦٥
7.3آفاق العمل المستقبلي(Future Work) ٦٦
7.4ملخص الفصل ٦٧
المراجع(References) ٦٨

فهرس الاشكال

رسم توضيحي ١:مثال توضيحي لبنية السؤال في مجموعة بيانات ScienceQA ٢٤
رسم توضيحي ٢ تصنيف مجالات ومهارات الأسئلة ضمن مجموعة بيانات ScienceQA ٢٥
رسم توضيحي3 : مثال لصورة تعليمية مرفقة بسؤال من ScienceQA ٢٧
رسم توضيحي4 : توزيع أطوال الأسئلة النصية والصور في مجموعة ScienceQA ٢٨
رسم توضيحي5 : توازن البيانات ٢٩
رسم توضيحي6 : توافق البيانات بين النصوص والصور ٢٩
رسم توضيحي7 : تنفيذ EDAعلى مجموعة البيانات ٣٢
رسم توضيحي8 : توضيح مخطط سير العمل ٤١
رسم توضيحي ٩: بنية النظام باستخدام langchain & langgraph ٤٣
رسم توضيحي ١٠: مخطط سير عمل الRAG ٤٧
رسم توضيحي ١١: عمل cross modal ٤٧
رسم توضيحي ١٢:واجهة المستخدم للنظام المقترح القائم على الاسترجاع المعزز بالتوليد (RAG). ٥١
رسم توضيحي13 :مخطط يوضّح مسار معالجة الصورة من الإدخال حتى التخزين. ٥٤
رسم توضيحي14 : مخطط يظهر كيفية التخزين فيmilvus وآلية top k ٥٦
رسم توضيحي ١٥: مخطط يوضّح تكامل الاسترجاع مع التوليد (RAG Pipeline). ٥٧

فهرس الجداول

قائمة الاختصارات 1 ..... Error! Bookmark not defined.
جدول الدراسات المرجعية الاول 2 ..... ٢٠
جدول الدراسات المرجعية الثاني3 ..... ٢١
جدول محتوى كل عينة من البيانات 4 ..... ٢٥
جدول5 توزيع العينات فيScienceQA ٢٦
جدول6 أنواع الصور فيScienceQA ٢٨
جدول7 توزيع الأسئلة حسب المجال العلمي..... ٣٠

قائمة الاختصارات(List of Abbreviations)

الاختصار	التعريف
AI	Artificial Intelligence
NLP	Natural Language Processing
LLM	Large Language Model
RAG	Retrieval-Augmented Generation
OCR	Optical Character Recognition
EDA	Exploratory Data Analysis
API	Application Programming Interface

الفصل الاول : المقدمة

١. المقدمة(Introduction)

يقدّم هذا الفصل العناصر الأساسية لمشروع **نظام البحث متعدد الوسائط للمحتوى التعليمي**، حيث يوضّح المشكلة التي يسعى المشروع إلى معالجتها، ورؤية المشروع العامة، والمكوّنات الرئيسية للنظام، بالإضافة إلى بنية التقرير. يهدف هذا الفصل إلى إعطاء نظرة شاملة وعالية المستوى عن الحل الذكي المقترح للبحث، مع تسليط الضوء على أهمية تحسين إمكانية الوصول إلى المحتوى التعليمي من خلال توظيف تقنيات الذكاء الاصطناعي متعددة الوسائط.

٢. مشكلة البحث(Problem Statement)

تعتمد العديد من المنصات التعليمية الحالية على أنظمة بحث قائمة حصريًا على الاستعلامات النصية، مما يؤدي إلى محدودية في إمكانية الوصول إلى المحتوى، وانخفاض في دقة نتائج البحث، وعدم دعم المتعلمين الذين يعتمدون على المدخلات البصرية أو السمعية أو الهجينة. غالبًا ما يواجه الطلاب صعوبة في العثور على معلومات محددة ضمن مقاطع الفيديو، المستندات، أو الصور التعليمية، وذلك بسبب القيود المفروضة على البحث بالكلمات المفتاحية وغياب القدرة على فهم السياق الدلالي للاستعلامات.
تؤدي هذه القصور إلى عوائق في عملية التعلم الفعّال، مما يبرز الحاجة إلى تطوير محرك بحث ذكي متعدد الوسائط قادر على تفسير وربط المدخلات النصية، الصوتية، والصورية مع المحتوى التعليمي المناسب، وبما يحقق تجربة تعلم أكثر كفاءة وشمولية.

٣. أهداف المشروع(Project Objectives)

يهدف هذا المشروع بشكل رئيسي إلى تطوير نظام بحث ذكي متعدد الوسائط يساهم في تحسين تجربة التعلم من خلال تمكين المستخدمين من استرجاع المحتوى التعليمي باستخدام أنماط إدخال متعددة، مثل النصوص والصور. ويسعى المشروع إلى تحسين دقة نتائج البحث، وتقليل الزمن اللازم للوصول إلى المعلومات، ودعم أنماط التعلم المختلفة لدى المستخدمين.
كما يهدف النظام إلى دمج آليات فهرسة المحتوى، والبحث الدلالي، ومعالجة البيانات متعددة الوسائط باستخدام تقنيات الذكاء الاصطناعي، لضمان تقديم نتائج بحث عالية الصلة عبر وسائط تعليمية متنوعة.
يركّز نطاق المشروع على بناء منصة بحث موحّدة تدعم المدخلات متعددة الوسائط، والاسترجاع الذكي للمحتوى، والتفاعل السلس مع المستخدم، إضافة إلى قابلية التوسّع لمعالجة كميات كبيرة من البيانات التعليمية.

٤. النظام المقترح(Proposed System)

تم تصميم نظام البحث متعدد الوسائط للمحتوى التعليمي ليعمل كمنصة ذكية عالية المستوى تتيح الاسترجاع الفعّال للمواد التعليمية عبر صيغ وسائط متعددة. ومن خلال توظيف تقنيات الذكاء الاصطناعي متعددة الوسائط، يوفّر النظام قدرات متقدمة مثل البحث الدلالي والمطابقة عبر الوسائط المختلفة، مما يسمح للمستخدمين بالبحث باستخدام النصوص، الصوت، أو الصور، والوصول الفوري إلى المحتوى التعليمي الأكثر ارتباطًا.
علاوة على ذلك، يدمج النظام آليات ذكية لفهرسة المحتوى وتصنيفه، بهدف تحسين دقة البحث وتعزيز تجربة المستخدم. كما يقدّم واجهة موحّدة للطلاب، والمعلمين، والباحثين، مما يسهل عملية التنقل ضمن المستودعات التعليمية الكبيرة ويضمن الوصول السريع إلى الموارد المفيدة. وبفضل دعمه لأنماط إدخال متعددة وقدرته على الفهم الآلي للمحتوى، يساهم النظام في تحسين إمكانية الوصول، تقليل زمن البحث، ومواكبة متطلبات التعلم الحديثة.

٥. تنظيم التقرير(Report Organization)

تم تنظيم هذا التقرير على النحو التالي:

- **الفصل الأول: المقدمة**
  يقدّم هذا الفصل نظرة عامة عن المشروع، ويستعرض الخلفية العامة، مشكلة البحث، أهداف المشروع، والمساهمات الرئيسية.
- **الفصل الثاني: الدراسات المرجعية والأعمال ذات الصلة**
  يستعرض هذا الفصل المفاهيم الأساسية المتعلقة بالبحث متعدد الوسائط والذكاء الاصطناعي في المجال التعليمي، إضافة إلى مراجعة الدراسات والأبحاث السابقة ذات الصلة بالمشروع.
- **الفصل الثالث: مجموعة البيانات وتحضير البيانات**
  يركّز هذا الفصل على وصف مجموعات البيانات المستخدمة في المشروع، كما يوضّح خطوات معالجة البيانات وتحضيرها لاستخدامها في النظام المقترح.
- **الفصل الرابع: المنهجية المقترحة**
  يشرح هذا الفصل النهج المقترح لبناء نظام البحث متعدد الوسائط، ويعرض البنية العامة للنظام، وآلية معالجة المدخلات المختلفة، وأساليب البحث الدلالي والاسترجاع الذكي للمحتوى التعليمي.
- **الفصل الخامس: التنفيذ**
  يتناول هذا الفصل تفاصيل تنفيذ النظام عمليًا، بما في ذلك الأدوات والتقنيات المستخدمة، وآليات دمج نماذج الذكاء الاصطناعي، وقاعدة البيانات المتجهية، إضافة إلى واجهة المستخدم.
- **الفصل السادس: التقييم**
  يقدّم هذا الفصل تقييمًا لأداء النظام المقترح من خلال مجموعة من التجارب والاختبارات، مع تحليل نتائج البحث متعدد الوسائط ومدى فعاليته في تحسين دقة الاسترجاع وجودة النتائج.
- **الفصل السابع: الخلاصة والأعمال المستقبلية**
  يختتم هذا الفصل التقرير بعرض الخلاصة العامة للمشروع، وتلخيص الإنجازات، ومناقشة القيود الحالية، إضافة إلى اقتراح اتجاهات مستقبلية لتطوير النظام.

٦. الملخص(Summary)

قدّم هذا الفصل نظرة عامة على مشروع **نظام البحث متعدد الوسائط للمحتوى التعليمي**، موضحًا الحاجة إلى تطوير حلول ذكية تتجاوز قيود البحث التقليدي القائم على الكلمات المفتاحية. كما استعرض أهداف المشروع التي تركز على تحسين دقة البحث ودعم أنماط إدخال متعددة باستخدام تقنيات الذكاء الاصطناعي. تم تقديم عرض عام للنظام المقترح ودوره في تحسين إمكانية الوصول إلى المحتوى التعليمي وتوفير تجربة تعلم أكثر فعالية. واختُتم الفصل بتوضيح هيكلية التقرير، مما يمهّد للانتقال إلى الفصول التالية التي تتناول الجوانب النظرية والتقنية للمشروع.

الفصل الثاني: الدراسات المرجعية والأعمال ذات الصلة

2.1مقدمة الفصل(Introduction)

يهدف هذا الفصل إلى تقديم الإطار النظري والبحثي الذي يستند إليه مشروع **نظام البحث متعدد الوسائط للمحتوى التعليمي**، وذلك من خلال استعراض المفاهيم الأساسية المرتبطة بالذكاء الاصطناعي متعدد الوسائط، والبحث الدلالي، والنماذج اللغوية الكبيرة، بالإضافة إلى مراجعة الدراسات والأعمال السابقة ذات الصلة.
تبرز أهمية الدراسة المرجعية في تحديد موقع المشروع ضمن السياق البحثي الحالي، وفهم الحلول القائمة، وتحليل نقاط القوة والقصور فيها. كما تساهم هذه الدراسة في تبرير اختيار المنهجية المقترحة، وتوضيح الفجوة البحثية التي يسعى المشروع إلى معالجتها.
ونظرًا لأن المحتوى التعليمي الحديث يتضمن وسائط متعددة مثل النصوص، الصور، والمخططات، فإن البحث متعدد الوسائط أصبح ضرورة ملحّة لتحسين الوصول إلى المعرفة التعليمية. وعليه، يساعد هذا الفصل في بناء فهم شامل للأسس النظرية والتقنيات المستخدمة، مما يمهّد لتقديم المنهجية المقترحة وتنفيذها في الفصول اللاحقة.

2.2المفاهيم الأساسية(Fundamental Concepts)

2.2.1الذكاء الاصطناعي متعدد الوسائط(Multimodal Artificial Intelligence)

يشير الذكاء الاصطناعي متعدد الوسائط إلى الأنظمة القادرة على معالجة وتحليل أكثر من نوع واحد من البيانات في آنٍ واحد، مثل النصوص، الصور، الصوت، والفيديو. بخلاف الأنظمة أحادية الوسيط، تتيح الأنظمة متعددة الوسائط فهماً أعمق للسياق من خلال دمج المعلومات المستخلصة من مصادر مختلفة.
في المجال التعليمي، يلعب الذكاء الاصطناعي متعدد الوسائط دورًا مهمًا في تحليل المحتوى التعليمي المعقّد، مثل الأسئلة المصوّرة، الرسوم التوضيحية، والمخططات العلمية. فعلى سبيل المثال، قد يتطلب فهم مسألة علمية تحليل نص السؤال بالإضافة إلى تفسير صورة مرافقة له، وهو ما لا يمكن تحقيقه بكفاءة باستخدام أنظمة البحث التقليدية.

2.2.2البحث الدلالي(Semantic Search)

يعتمد البحث التقليدي على مطابقة الكلمات المفتاحية بين استعلام المستخدم والمحتوى المخزّن، مما يؤدي في كثير من الأحيان إلى نتائج غير دقيقة أو غير مرتبطة بالسياق المطلوب. في المقابل، يهدف البحث الدلالي إلى فهم المعنى الكامن وراء الاستعلام، بدلاً من الاكتفاء بمطابقة الكلمات حرفيًا.
يتميز البحث الدلالي بقدرته على التعامل مع المرادفات، المفاهيم المتقاربة، والصياغات المختلفة لنفس الفكرة، مما يجعله أكثر دقة وملاءمة للاستخدام في المجال التعليمي. وبذلك، يتيح هذا النوع من البحث استرجاع محتوى تعليمي مرتبط بالمعنى، حتى وإن اختلفت صياغة السؤال عن النص الأصلي للمحتوى.

2.2.3التضمينات والتمثيلات المتجهية(Embeddings and Vector Representations)

تُعد التضمينات (Embeddings) من المفاهيم الأساسية في أنظمة البحث الدلالي، حيث يتم من خلالها

تحويل البيانات، سواء كانت نصوصًا أو صورًا، إلى تمثيلات عددية على شكل متجهات في فضاء رياضي عالي الأبعاد. تعكس هذه المتجهات الخصائص الدلالية للمحتوى، مما يسمح بقياس درجة التشابه بين الاستعلامات والمستندات باستخدام مقاييس رياضية مثل تشابه جيب التمام.(Cosine Similarity)
في أنظمة البحث متعددة الوسائط، يتم استخدام نماذج متخصصة لتوليد تضمينات للنصوص والصور، ومن ثم مقارنة هذه التمثيلات داخل فضاء دلالي مشترك. يسهم هذا الأسلوب في تمكين البحث عبر الوسائط المختلفة، مثل البحث عن صورة باستخدام نص، أو العكس.

2.2.4 النماذج اللغوية الكبيرة(Large Language Models)

النماذج اللغوية الكبيرة هي نماذج ذكاء اصطناعي متقدمة تم تدريبها على كميات ضخمة من البيانات النصية، مما يمكنها من فهم اللغة الطبيعية، توليد النصوص، والإجابة على الأسئلة. تُستخدم هذه النماذج على نطاق واسع في تطبيقات متعددة مثل الترجمة الآلية، تلخيص النصوص، وأنظمة السؤال والجواب.
في السياق التعليمي، تلعب النماذج اللغوية الكبيرة دورًا محوريًا في تفسير استعلامات المستخدمين، إعادة صياغة الأسئلة، وتوليد إجابات تعليمية دقيقة بالاعتماد على المحتوى المسترجع. ويعزز دمج هذه النماذج مع أنظمة البحث الدلالي من قدرة النظام على تقديم إجابات ذات جودة عالية.

2.2.5 الاسترجاع المعزز بالتوليد(Retrieval-Augmented Generation)

يُعد الاسترجاع المعزز بالتوليد (RAG) نهجًا حديثًا يجمع بين أنظمة الاسترجاع وقوة النماذج اللغوية الكبيرة. في هذا النهج، يتم أولًا استرجاع المحتوى ذي الصلة من قاعدة المعرفة، ثم يتم استخدام هذا المحتوى كمدخل للنموذج اللغوي لتوليد الإجابة.
يساهم هذا الأسلوب في تقليل ظاهرة الهلوسة، حيث تعتمد الإجابات على مصادر معرفة محددة بدلًا من الاعتماد على المعرفة العامة للنموذج فقط. ويُعد RAG مناسبًا بشكل خاص للمجال التعليمي، إذ يضمن أن تكون الإجابات مبنية على محتوى تعليمي موثوق، وهو ما يتماشى مع أهداف المشروع المقترح.

2.2.6 LangChain

تُعد مكتبة LangChain إطارًا برمجيًا متقدمًا يهدف إلى تسهيل بناء التطبيقات الذكية المعتمدة على النماذج اللغوية الكبيرة .(Large Language Models) توفر LangChain مجموعة من الأدوات والمكونات الجاهزة التي تُمكّن المطورين من ربط النماذج اللغوية مع مصادر بيانات خارجية، مثل قواعد المعرفة، محركات البحث، وواجهات البرمجة المختلفة.
يعتمد LangChain على مفهوم سلاسل المعالجة(Chains) ، حيث يتم تنظيم خطوات معالجة المدخلات، الاسترجاع، والتوليد في تسلسل منطقي قابل للتخصيص. كما يدعم الإطار آليات متقدمة مثل **إدارة الذاكرة**، استخدام الأدوات(Tools) ، والاسترجاع الدلالي، مما يجعله مناسبًا لتطوير أنظمة البحث الذكي.
في هذا المشروع، تم استخدام LangChain لبناء سلاسل معالجة الأسئلة، وتنفيذ البحث الدلالي، ودمج نتائج الاسترجاع مع النماذج اللغوية لتوليد إجابات دقيقة ومدعومة بالمصادر التعليمية.

2.2.7 LangGraph

يُعد LangGraph امتدادًا متقدمًا لمكتبة LangChain ، ويُستخدم لبناء أنظمة ذكية تعتمد على مخططات الحالة (State Graphs) بدلًا من السلاسل الخطية التقليدية. يسمح هذا الإطار بتمثيل منطق النظام على شكل عقد (Nodes) وروابط (Edges) ، مما يوفّر تحكمًا أدق في تدفق البيانات واتخاذ القرار داخل النظام.
يمكّن LangGraph من تنفيذ سيناريوهات معقدة مثل:

- التفرّع الشرطي بناءً على نوع المدخل (نص أو صورة)
- الدمج بين عدة أدوات وخطوات معالجة
- إدارة الحالات المتعددة للنظام بطريقة منظمة وقابلة للتوسّع

في سياق هذا المشروع، تم استخدام LangGraph لتنظيم سير عمل نظام البحث متعدد الوسائط، حيث يتم توجيه المدخلات إلى مسارات مختلفة حسب نوعها، مثل مسار معالجة النصوص أو مسار معالجة الصور، ثم دمج النتائج في مرحلة الاسترجاع والتوليد النهائية. يسهم هذا النهج في تحسين مرونة النظام، وضمان قابلية التوسّع، وتحقيق فصل واضح بين مكونات المعالجة المختلفة.

2.3 مراجعة الدراسات الخاصة بأنظمة البحث متعددة الوسائط

(Literature Review for Multimodal Search Systems)

تناولت العديد من الدراسات البحثية تطوير أنظمة بحث متعددة الوسائط تهدف إلى دمج النص والصورة في عملية الاسترجاع. ركزت بعض الأبحاث على استرجاع الصور باستخدام النصوص، بينما اهتمت دراسات أخرى ببناء فضاءات دلالية مشتركة تتيح المقارنة بين أنواع مختلفة من البيانات.
أظهرت هذه الدراسات تحسنًا ملحوظًا في دقة الاسترجاع مقارنة بالأنظمة أحادية الوسيط، خاصة عند استخدام نماذج تضمين متقدمة. ومع ذلك، تعاني العديد من هذه الأنظمة من محدودية في التطبيق العملي، أو تركز على مجالات عامة مثل البحث التجاري أو الوسائط الاجتماعية، دون تخصيص كافٍ للمجال التعليمي.
كما أن بعض الأنظمة متعددة الوسائط لا تستفيد بشكل كامل من البحث الدلالي أو النماذج اللغوية الكبيرة، مما يحد من قدرتها على تقديم إجابات تعليمية متكاملة.

الرقم	الدراسة \ المرجع	نوع الدراسة	نقاط القوة	نقاط الضعف	المصدر
١	Multimodal Search (Concept Overview)	مفهوم البحث متعدد الوسائط	يعطي تعريفًا شاملًا لمفهوم Multimodal Search وكيفية دمج نص + صورة + صوت	ليس ورقة بحث، لكنه ممتاز لشرح المفاهيم الأساسية	https://en.wikipedia.org/wiki/Multimodal_search
٢	Using generative AI to do Multimodal Information Retrieval	Multimodal Retrieval + Gen AI	يقدم إطار حديث (GENIUS) للبحث متعدد الوسائط باستخدام الذكاء التوليدي	يركّز على الأبحاث الحديثة وقد لا تكون مطبّقة في التعليم	https://www.amazon.science/blog/using-generative-ai-to-do-multimodal-information-retrieval
٣	Multimodal Document Retrieval Challenge	Task Challenge	يناقش تحديات استرجاع مستندات متعددة الوسائط علىKaggle	يركّز على التحديات والمسابقات وليس تطبيق معيّن	https://www.kaggle.com/competitions/multimodal-document-retrieval-challenge/overview

جدول الدراسات المرجعية الاول 1

2.4مراجعة الدراسات المتعلقة باسترجاع المحتوى التعليمي

شهد مجال التعليم الذكي اهتمامًا متزايدًا بتطبيق تقنيات الذكاء الاصطناعي، حيث تم تطوير أنظمة تعليمية ذكية، وأنظمة دعم التعلم، ومنصات سؤال وجواب تعليمية. ركزت هذه الأنظمة على تحسين تجربة التعلم من خلال تخصيص المحتوى وتوفير إجابات فورية للطلاب.
كما ظهرت أبحاث متقدمة في مجال استرجاع المحتوى التعليمي المعتمد على الصور، مثل أنظمة Visual Question Answeringومجموعات البيانات التعليمية مثل ScienceQA، التي تجمع بين النص والصورة في سياق تعليمي واحد.
ورغم التقدم المحقق، إلا أن العديد من هذه الأنظمة لا تزال تعاني من نقص في التكامل بين الوسائط المختلفة، أو تفتقر إلى آلية واضحة للربط بين الاسترجاع الدلالي وتوليد الإجابة. ومن هنا، يبرز مشروع **نظام البحث متعدد الوسائط للمحتوى التعليمي** كحل متكامل يعتمد على الذكاء الاصطناعي، والاسترجاع المعزز بالتوليد، ويستهدف المجال التعليمي بشكل مباشر.

الرقم	الدراسة \المرجع	نوع الدراسة	نقاط القوة	نقاط الضعف	المصدر
١	Learn to Explain: Multimodal Reasoning via Thought Chains for Science QA	Dataset multimodal + Reasoning	أول dataset كبير للأسئلة العلمية متعددة الوسائط؛ يضم مقالات تعليمية وصور وسياقات تفسيرية	موجه لمهام QA في العلوم، وليس نظام بحث عام	https://scienceqa.github.io/
٢	*PDF-MVQA: Dataset for Multimodal Information Retrieval*	Retrieval & VQAعلى المستندات	يوفر بيانات تتضمن نص + جداول + صور متعددة الصفحات	تطبيقه يتطلب فهم الوثائق المعقدة؛ أقل تركيزًا على التعليم	https://arxiv.org/abs/2404.12720
٣	SPIQA: A Dataset for Multimodal QA on Scientific Papers	Multimodal QAعلى المقالات العلمية	dataset ضخم (270K أسئلة (لفهم الرسوم والجداول داخل الأوراق	معقد في التركيز على بيانات علمية بحثية أكثر من التعليمية العامة	https://openreview.net/forum?id=h3lddsY5nf
٤	CUE-M: Contextual Understanding and Enhanced Search with MLLM	Search & Context understanding	إطار لأفضل نتيجة بحث multimodal مع فهم السياق	دراسة حديثة لم تُنشر كـ dataset لكنها توفّر إطارًا قويًا	https://arxiv.org/abs/2411.12287

تعتبر *ScienceQA* من أهم مصادر البيانات متعدد الوسائط لفهم الأسئلة العلمية المعتمدة على نص + صورة، وقد استخدمت ربط Chain-of-Thought لتحسين الأداء على أسئلة متعددة الوسائط .
scienceqa.github.io
في سياقات علمية مشابهة، يقدم SPIQA dataset مجموعة ضخمة من أسئلة QA مع بيانات نص + صور من الأبحاث العلمية openreview.net .
كما تم اقتراح dataset مثل PDF-MVQA الذي يتعامل مع استرجاع معلومات في مستندات PDF متعددة الوسائط arxiv.org .
كذلك توفّر أطر بحث مثل CUE-M نموذجًا متقدمًا لتحسين فهم السياق والاسترجاع متعدد الوسائط .
arxiv.org

أظهرت الدراسات أن التحديات المرتبطة بتنفيذ نماذج الذكاء الاصطناعي تختلف باختلاف الخوارزميات المستخدمة في أنظمة البحث متعددة الوسائط. فقد تبيّن أن بعض النماذج تواجه تحديات تتعلق بمتطلبات

الموارد الحاسوبية وكثافة المعالجة، خاصة عند التعامل مع بيانات متعددة الوسائط مثل النصوص والصور. في المقابل، واجهت نماذج أخرى صعوبات مرتبطة باختيار المعاملات وضبطها بالشكل الأمثل، وهو ما يؤثر بشكل مباشر على دقة الاسترجاع وجودة النتائج.

كما أظهرت بعض الأساليب تحديات إضافية تمثلت في صعوبة تفسير أهمية الخصائص المستخرجة، والحاجة إلى مجموعات بيانات كبيرة لضمان أداء فعّال، إضافة إلى ضرورة إجراء عمليات ضبط دقيقة للنماذج بما يتناسب مع المهام التعليمية المحددة. وفي بعض الحالات، قد يؤدي استخدام نماذج معينة إلى أداء أقل عند التعامل مع بيانات تعليمية معقّدة تتطلب فهماً دلاليًا عميقًا وعلاقات متعددة الوسائط.

وفي الختام، تؤكد هذه الدراسة على أهمية اختيار المنهجية المناسبة في أنظمة البحث الذكية بناءً على طبيعة المجال التعليمي، وتعقيد البيانات المستخدمة، والقيود الحاسوبية المتاحة. فعلى الرغم من أن النماذج التقليدية في التعلم الآلي قد توفّر نتائج مستقرة وقابلة للتفسير، إلا أن النماذج العميقة والنماذج اللغوية الكبيرة تتيح أداءً متقدّمًا في فهم السياق متعدد الوسائط، وذلك على حساب زيادة المتطلبات الحاسوبية. ومع التطور المستمر لأنظمة البحث التعليمية المعتمدة على الذكاء الاصطناعي، تصبح قضايا التكامل بين المكونات، وقابلية التوسّع، والتحسين المستمر عناصر أساسية للحفاظ على جودة الأداء وتحقيق أهداف التعلم الحديثة.

الفصل الثالث: مجموعة البيانات وآليات معالجتها وتحضيرها

3.1مقدمة الفصل

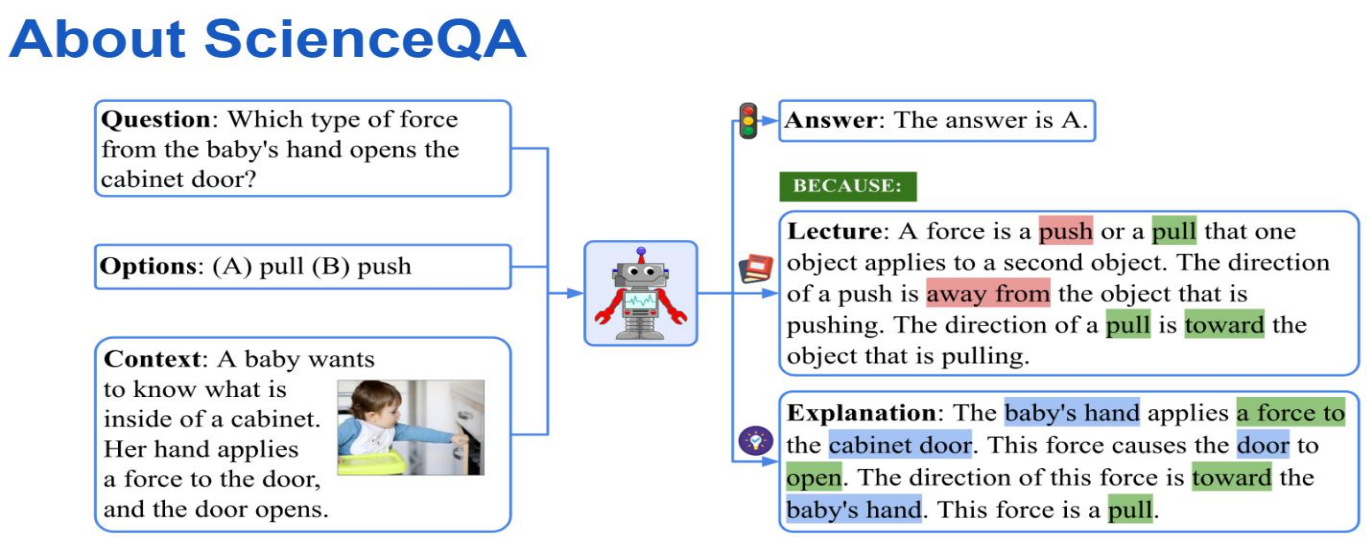
تُعد البيانات العنصر الأساسي في بناء أنظمة الذكاء الاصطناعي، لا سيما الأنظمة متعددة الوسائط التي تعتمد على تحليل ودمج أنواع مختلفة من البيانات مثل النصوص والصور. وتزداد أهمية البيانات في التطبيقات التعليمية، حيث يتنوّع المحتوى بين الشرح النصي، الرسوم التوضيحية، والمخططات العلمية، مما يتطلب نماذج قادرة على فهم السياق المتكامل للمعلومة.

في هذا المشروع، تم استخدام مجموعة بيانات **ScienceQA** بوصفها مجموعة بيانات تعليمية متعددة الوسائط، تم تصميمها خصيصًا لدعم مهام السؤال والجواب في المجالات العلمية. يهدف هذا الفصل إلى تقديم وصف تفصيلي لمجموعة البيانات المستخدمة، وتحليل بنيتها، وشرح خطوات تحضيرها ومعالجتها، إضافة إلى إجراء تحليل استكشافي للبيانات (EDA) يساهم في فهم خصائصها قبل استخدامها في النظام المقترح.

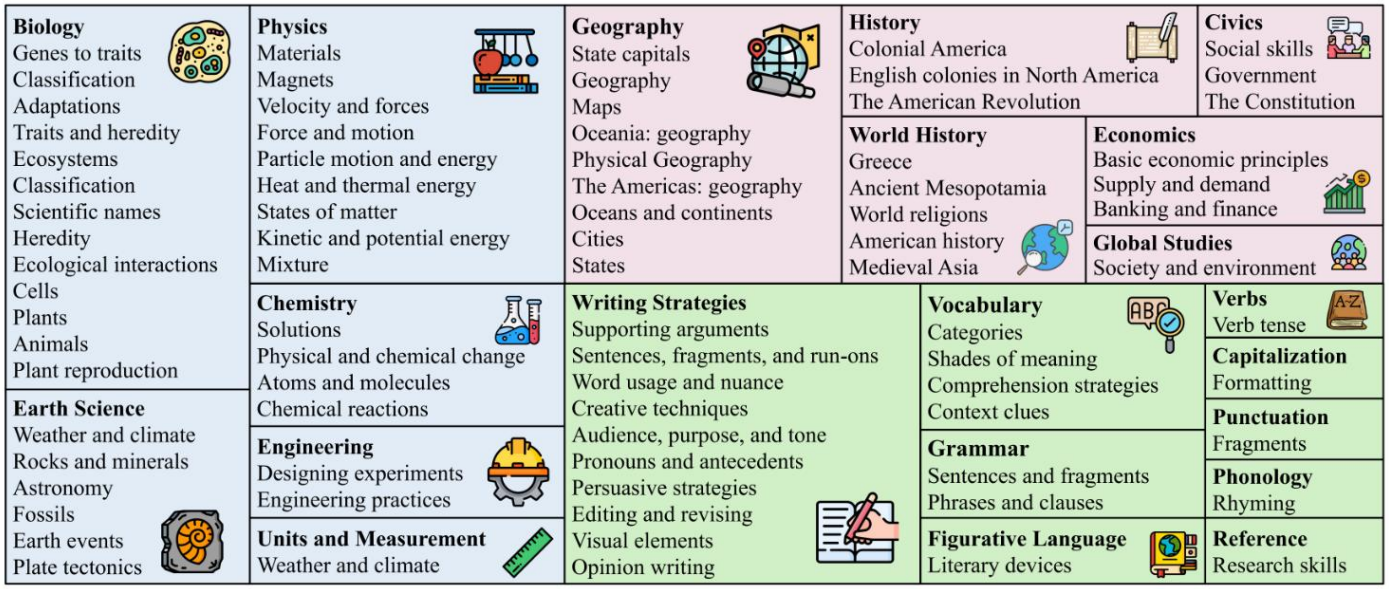
3.2نظرة عامة على مجموعة بياناتScienceQA

تُعد مجموعة بيانات **ScienceQA** من أهم مجموعات البيانات الحديثة في مجال الذكاء الاصطناعي التعليمي متعدد الوسائط. تم تطوير هذه المجموعة من قبل *Allen Institute for Artificial Intelligence (AI2)* بهدف دعم نماذج الذكاء الاصطناعي القادرة على الإجابة عن أسئلة علمية تتطلب دمج المعلومات النصية مع المعلومات البصرية.

تتضمن ScienceQA أسئلة تعليمية مأخوذة من سياقات تعليمية واقعية، حيث يتم تقديم السؤال مصحوبًا بصورة تعليمية، ومجموعة من الخيارات متعددة الاختيار، إلى جانب الإجابة الصحيحة وتفسير يوضّح سبب صحتها. يتيح هذا التصميم للنظام تقييم قدرته على الفهم والاستدلال، وليس فقط مطابقة الكلمات



رسم توضيحي ١:مثال توضيحي لبنية السؤال في مجموعة بيانات ScienceQA



رسم توضيحي ٢ تصنيف مجالات ومهارات الأسئلة ضمن مجموعة بيانات ScienceQA

3.3بنية مجموعة البيانات

تنقسم مجموعة بيانات ScienceQA إلى ثلاث مجموعات رئيسية لضمان تدريب وتقييم منضبط للنماذج:

- مجموعة التدريب(Training Set)
- مجموعة التحقق(Validation Set)
- مجموعة الاختبار(Test Set)

كما تحتوي البيانات على ملفات وصفية (JSON) تتضمن جميع المعلومات المرتبطة بكل سؤال.

3.3.1مكونات كل عينة بيانات

تتكون كل عينة بيانات من العناصر التالية:

3 جدول محتوى كل عينة من البيانات

الحقل	الوصف
Question Text	نص السؤال العلمي
Image Path	مسار الصورة المرتبطة بالسؤال
Choices	خيارات الإجابة(Multiple Choice)
Correct Answer	الإجابة الصحيحة
Explanation	تفسير الإجابة

3.4الإحصائيات العامة لمجموعة البيانات

يوضح الجدول التالي توزيع البيانات عبر مجموعات التدريب، التحقق، والاختبار:

جدول4 **توزيع العينات فيScienceQA**

مجموعة البيانات	عدد العينات	النسبة المئوية
Training	~70%	تقريبًا
Validation	~15%	تقريبًا
Test	~15%	تقريبًا
المجموع	100%	100%

3.5أنواع البيانات(Multimodal Data)

تعتمد مجموعة بيانات ScienceQA على نوعين رئيسيين من البيانات:

3.5.1البيانات النصية

- نص السؤال
- خيارات الإجابة
- التفسير النهائي

تمثل هذه البيانات المصدر الأساسي لفهم المشكلة العلمية المطروحة، وتتطلب معالجة لغوية دقيقة لفهم المعنى والسياق.

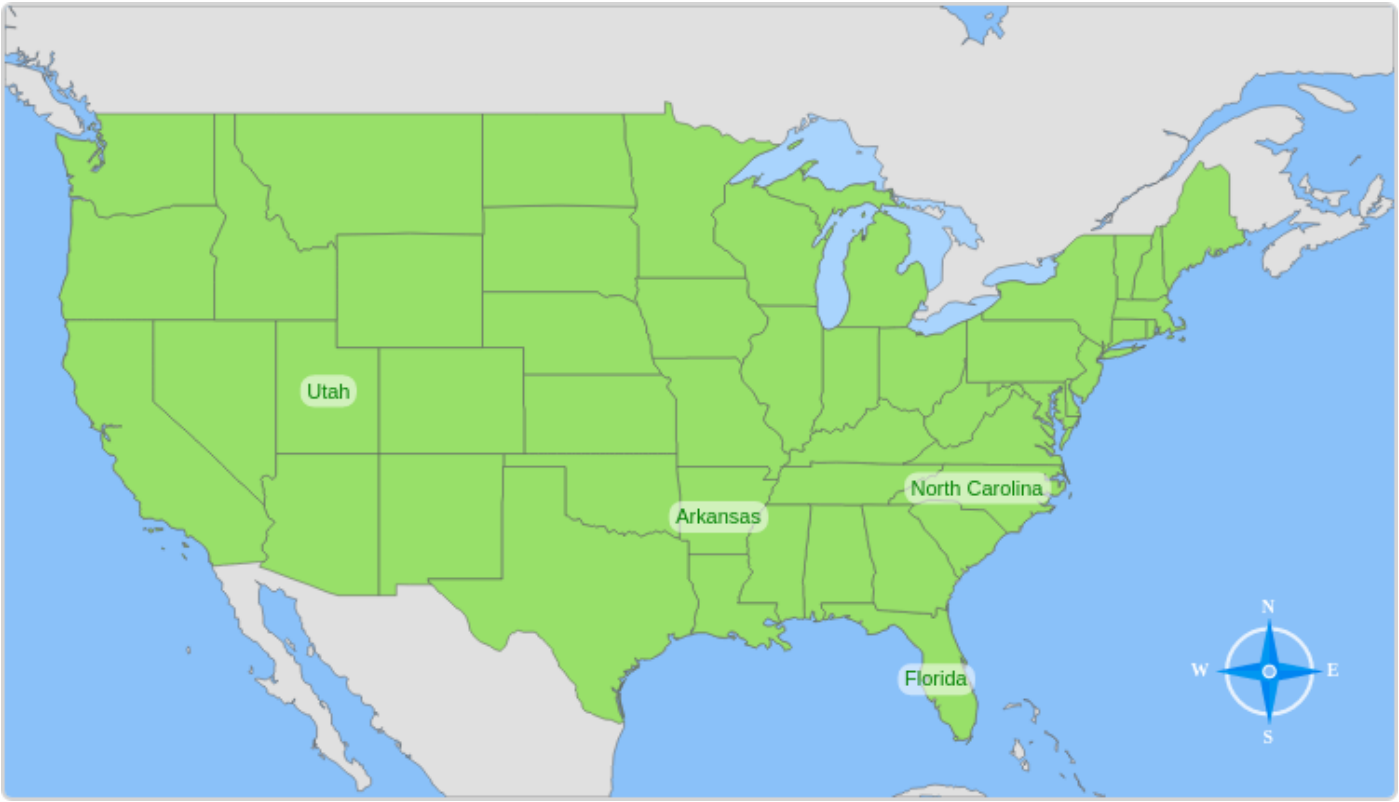
3.5.2البيانات الصورية

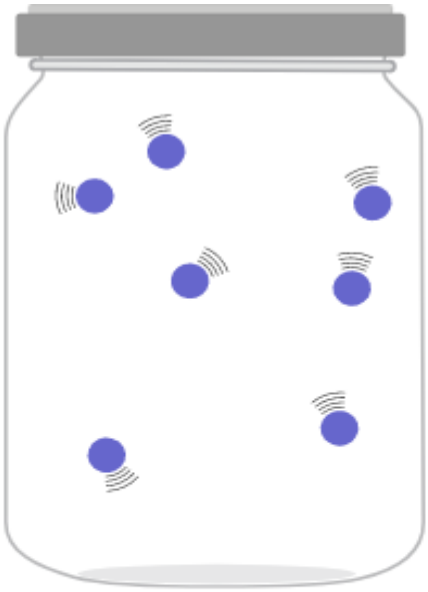
تشمل الصور التعليمية:

- الرسوم التوضيحية
- المخططات العلمية
- الصور التجريبية
-

غالبًا ما تحتوي هذه الصور على معلومات أساسية لا يمكن استنتاجها من النص فقط، مما يجعل تحليلها ضروريًا في أنظمة البحث متعددة الوسائط.

رسم توضيحي3 : مثال لصورة تعليمية مرفقة بسؤال من ScienceQA

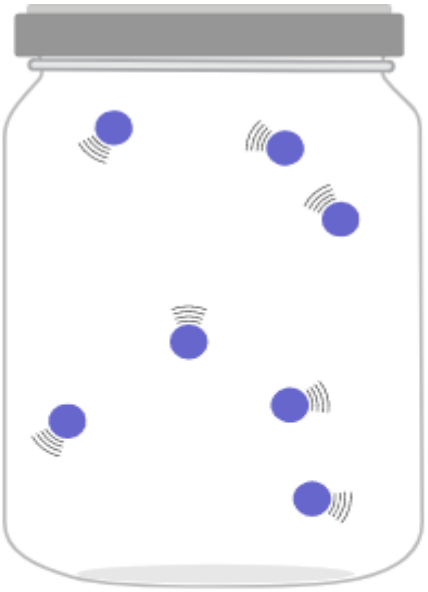




Sample A

Mass of each particle: 44 u

Average particle speed: 560 m/s



Sample B

Mass of each particle: 44 u

Average particle speed: 560 m/s

3.6التحليل الاستكشافي للبيانات(Exploratory Data Analysis – EDA)

يهدف التحليل الاستكشافي للبيانات إلى فهم الخصائص العامة لمجموعة البيانات قبل استخدامها في التدريب أو البحث الدلالي. يساعد هذا التحليل في اكتشاف الأنماط، التوزيعات، وأي مشكلات محتملة في البيانات.

3.6.1توزيع طول الأسئلة النصية

تم تحليل طول الأسئلة النصية من حيث عدد الكلمات، وتبيّن أن معظم الأسئلة تقع ضمن نطاق متوسط الطول، مما يجعلها مناسبة للمعالجة باستخدام النماذج اللغوية.

3.6.2توزيع أنواع الصور

يوضح التحليل أن الصور التعليمية تتنوع بين مخططات ورسوم توضيحية وصور علمية، كما هو موضح في الجدول التالي:

جدول5 **أنواع الصور فيScienceQA**

النسبة التقريبية	نوع الصورة
مرتفعة	**مخططات ورسوم**
متوسطة	**صور علمية**
منخفضة	**صور توضيحية أخرى**

رسم توضيحي4 : توزيع أطوال الأسئلة النصية والصور في مجموعة ScienceQA

Main Statistics

Statistic	Number
Total questions	21,208
Questions with text context	10,220 (48.2%)
Questions with image context	10,332 (48.7%)
* Image of natural format	≈2,960 (14.0%)
* Image of diagram format	≈7,372 (34.8%)
Questions with both contexts	6,532 (30.8%)
Questions without any context	7,188 (33.9%)
Questions with a lecture	17,798 (83.9%)
Questions with a explanation	19,202 (90.5%)
Different questions	9,122
Different lectures	261
Topic classes	26
Category classes	127
Skill classes	379
Average question length	12.11
Average choice length	4.40
Average lecture length	125.06
Average explanation length	47.66

Table 1: Main statistics in SCIENCEQA.

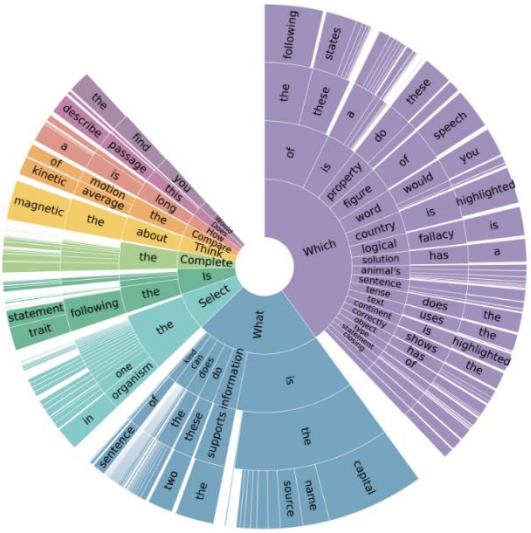
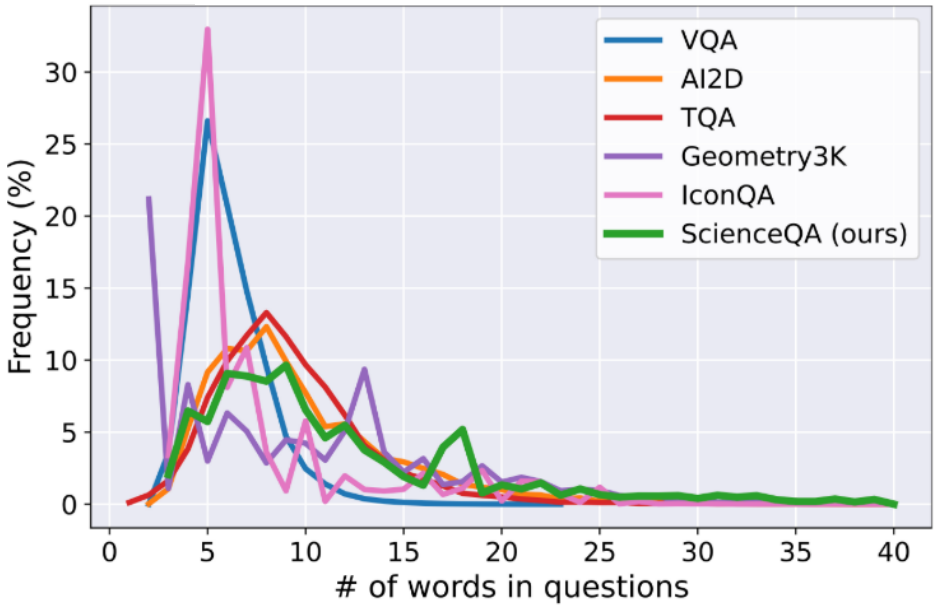


Figure 2: Question distribution in SCIENCEQA.

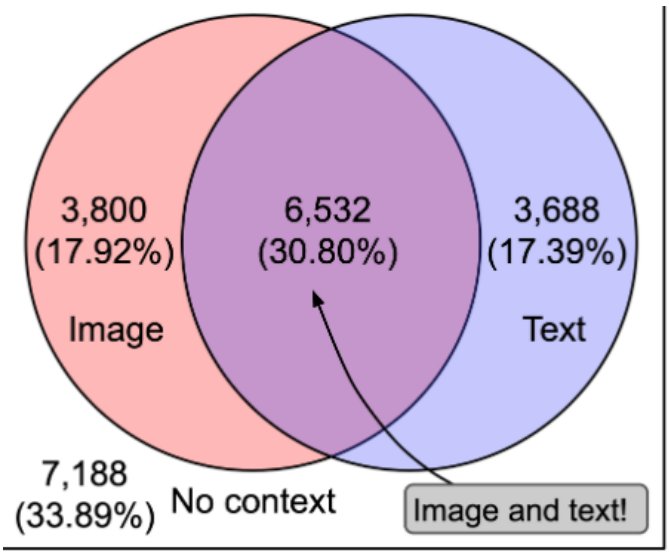
يبيّن التحليل التالي أن مجموعة بيانات ScienceQA لا تتركّز أسئلتها في أطوال قصيرة أو طويلة فقط، بل تمتد بشكل متوازن عبر أطوال مختلفة، مما يعكس تنوّعًا في مستوى تعقيد الأسئلة التعليمية وملاءمتها لتقييم النماذج متعددة الوسائط.

رسم توضيحي5 : توازن البيانات



تُظهر نتائج التحليل أن الغالبية العظمى من أسئلة ScienceQA تعتمد على سياق داعم، حيث تحتوي ٦٦,١١٪ من الأسئلة على سياق نصي أو بصري، بينما تجمع ٣٠,٨٠٪ منها بين النص والصورة، مما يؤكد الطبيعة متعددة الوسائط لهذه المجموعة وملاءمتها لتقييم أنظمة البحث متعددة الوسائط.

رسم توضيحي6 : توافق البيانات بين النصوص والصور



3.6.3 توزيع المجالات العلمية

تشمل الأسئلة مجالات علمية متعددة كما هو موضح في الجدول التالي:

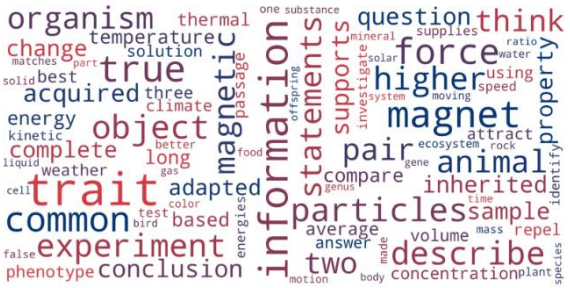
جدول 6 ***توزيع الأسئلة حسب المجال العلمي***

نسبة الأسئلة	المجال العلمي
مرتفعة	**الفيزياء**
متوسطة	**الأحياء**
متوسطة	**الكيمياء**
منخفضة	**علوم عامة**

يساهم هذا التنوع في تقييم قدرة النظام على التعامل مع محتوى علمي متنوع

تُظهر سُحب الكلمات الخاصة بالأسئلة أن مجموعة بيانات ScienceQA تغطي نطاقًا واسعًا من الموضوعات العلمية المتنوعة، مما يعكس شمولية المحتوى التعليمي الذي تتضمنه.

(a) Questions of all subjects.

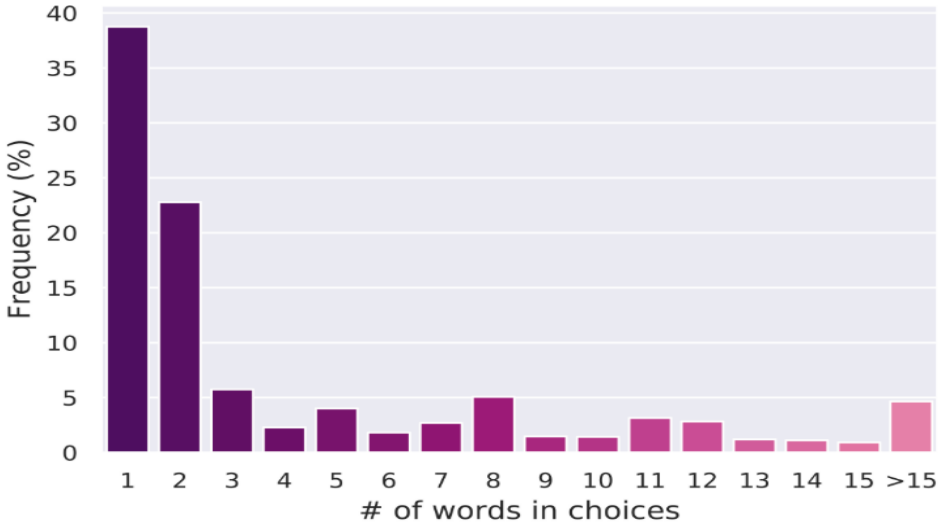


(b) Questions in natural science.

(c) Questions in social science.

(d) Questions in language science.

كما يتبيّن أن معظم خيارات الإجابة قصيرة، إذ تحتوي في الغالب على ما يصل إلى خمس كلمات، إلا أن توزيع أطوال الخيارات يمتد بذيل طويل، حيث تحتوي نحو ٥٪ من الخيارات على أكثر من ١٥ كلمة، وهو ما يشير إلى وجود أسئلة تتطلب إجابات أكثر تفصيلًا وتعقيدًا.



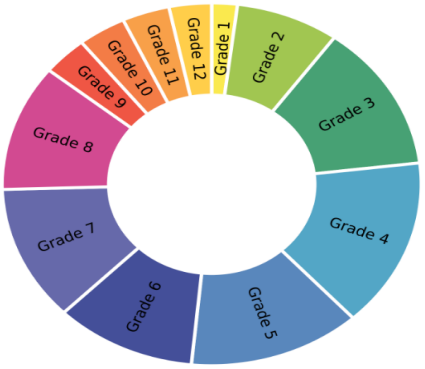
يُظهر توزيع الأسئلة حسب المرحلة التعليمية أن مجموعة بيانات ScienceQA تركز بشكل أساسي على المناهج الدراسية للمرحلة المتوسطة، مع تمثيل محدود نسبيًا لمستوى المرحلة الثانوية، مما يجعلها مناسبة لتقييم أنظمة البحث التعليمية الموجّهة لطلاب المراحل الأساسية والمتوسطة.

Grade Distributions

The majority of questions come from the middle level curriculum (i.e., from grade 3 to grade 8) while around 10% are taken from the high school curriculum (i.e., from grade 9 to grade 12).

Grades	Number	Percent
Grade 1	95	0.45%
Grade 2	1,678	7.91%
Grade 3	3,032	14.3%
Grade 4	3,544	16.71%
Grade 5	3,086	14.55%
Grade 6	2,450	11.55%
Grade 7	2,749	12.96%
Grade 8	2,546	12.0%
Grade 9	491	2.32%
Grade 10	558	2.63%
Grade 11	539	2.54%
Grade 12	440	2.07%

(a) Grade distribution statistics.



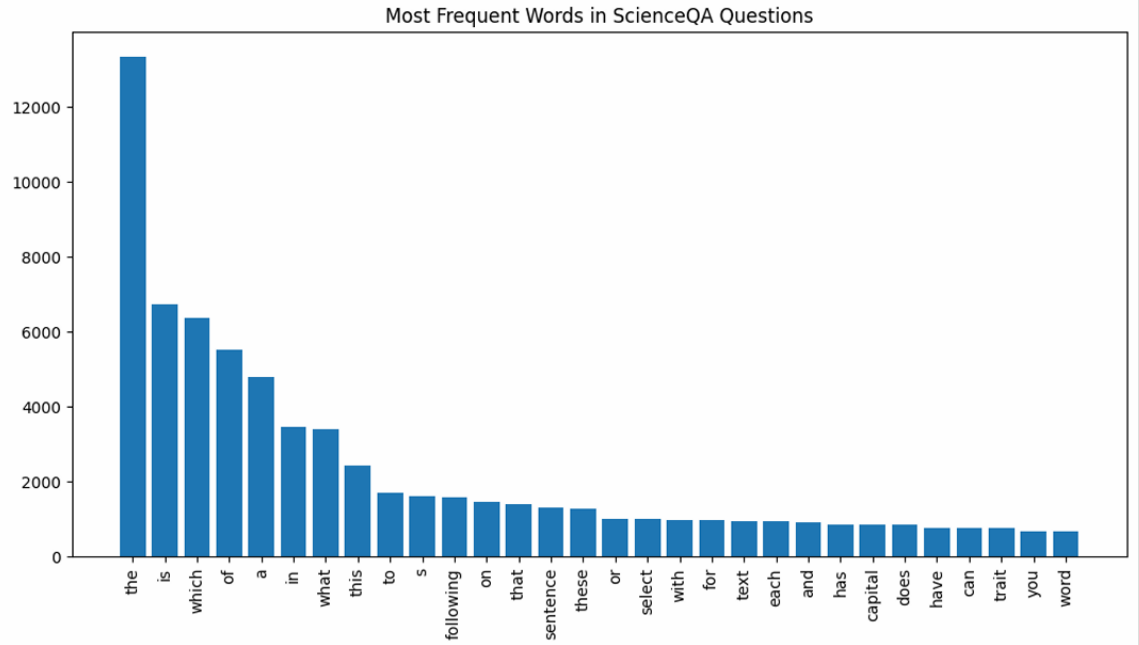
(b) Grade distribution visualization.

تم تنفيذ التحليل الاستكشافي للبيانات (EDA) برمجيًا بهدف دراسة خصائص مجموعة بيانات ScienceQAوفهم أنماطها الأساسية قبل استخدامها في النظام المقترح.

رسم توضيحي7 : تنفيذ EDAعلى مجموعة البيانات

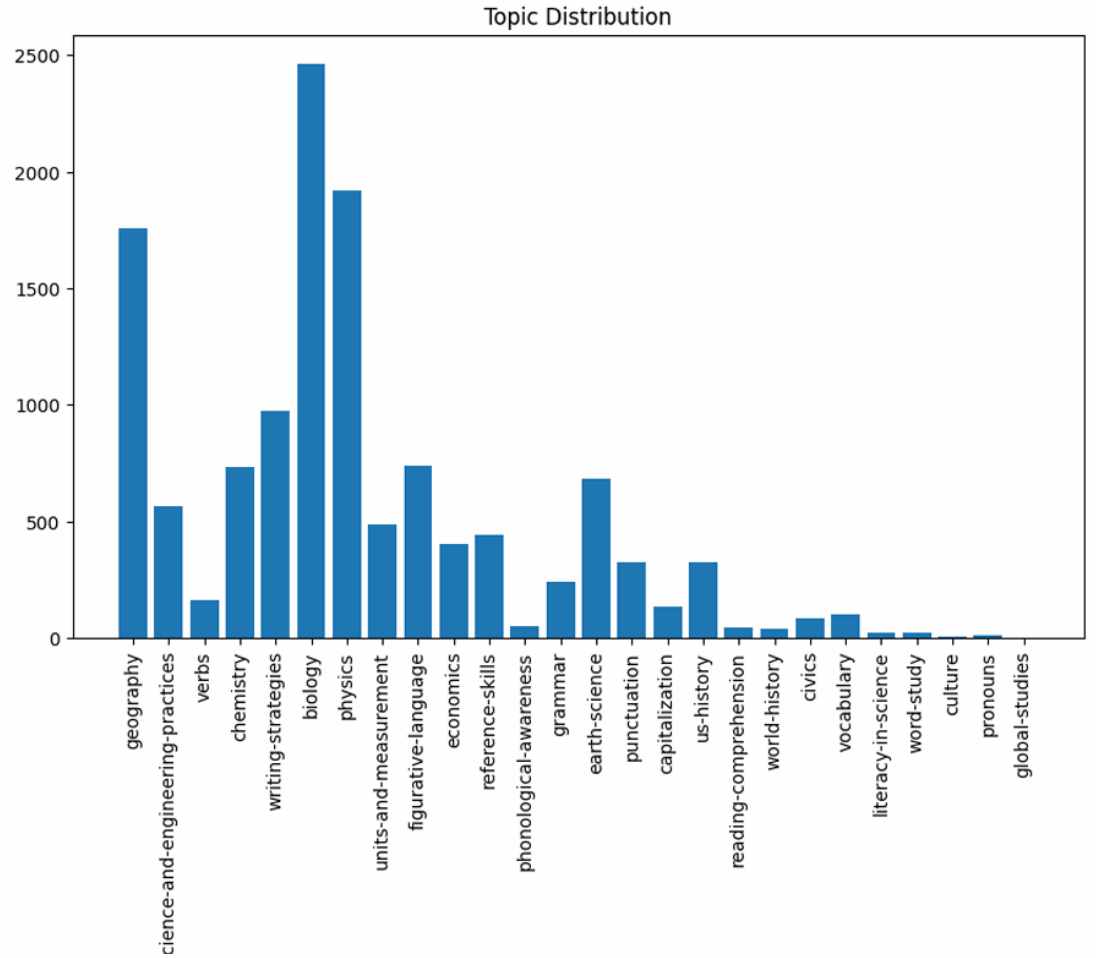
```
import re
from collections import Counter
import matplotlib.pyplot as plt
train = ds['train']
def clean_text(text):
    text = text.lower()
    text = re.sub(r"[^a-zA-Z]+", " ", text)
    return text

# دمج كل الأسئلة في نص واحد
all_q = " ".join([clean_text(x["question"]) for x in train])
words = all_q.split()
word_freq = Counter(words).most_common(30)   # أعلى 30 كلمة

# عرض النتائج
print(word_freq)

# رسم شريطي
keys = [w for w, c in word_freq]
values = [c for w, c in word_freq]

plt.figure(figsize=(12,6))
plt.bar(keys, values)
plt.xticks(rotation=90)
plt.title("Most Frequent Words in ScienceQA Questions")
plt.show()
```

```
[('the', 13324), ('is', 6733), ('which', 6360), ('of', 5506), ('a', 4779), ('in', 3444), ('what', 3403), ('this', 2422), ('t
```



```
# رسم
plt.figure(figsize=(10,6))
plt.bar(topic_counts.keys(), topic_counts.values())
plt.xticks(rotation=90)
plt.title("Topic Distribution")
plt.show()
```

```
Counter({'biology': 2464, 'physics': 1923, 'geography': 1756, 'writing-strategies': 974, 'figurative-language': 739, 'chemis
```

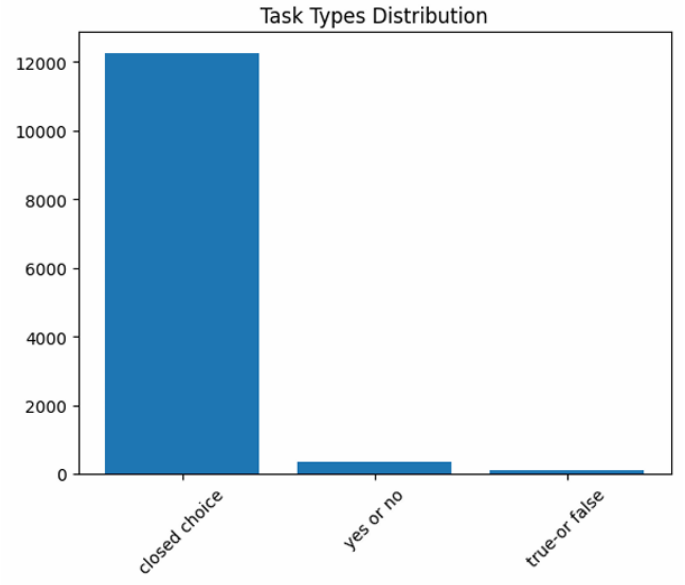


```
types = Counter([x["task"] for x in train])
print(types)

plt.bar(types.keys(), types.values())
plt.xticks(rotation=45)
```



```
plt.title("Task Types Distribution")
plt.show()
```

```
Counter({'closed choice': 12254, 'yes or no': 357, 'true-or false': 115})
```




```
from datasets import load_dataset
ds = load_dataset("derek-thomas/ScienceQA")
print(len(ds['train']), len(ds['validation']), len(ds['test']))
```

```
12726 4241 4241
```

```
train = ds['train']
img_count = sum(1 for x in train if x["image"] is not None)
print(img_count, "من أصل", len(train))
```

```
6218 12726 من أصل
```

```
import collections
subjects = [x["subject"] for x in train]
print(collections.Counter(subjects))
```

```
Counter({'natural science': 6873, 'language science': 3243, 'social science': 2610})
```

```
options_len = [len(x["choices"]) for x in train]
import statistics
print(set(options_len))  # أنواع عدد الخيارات
```

```
from datasets import load_dataset
ds = load_dataset("derek-thomas/ScienceQA")
print(len(ds['train']), len(ds['validation']), len(ds['test']))
```

```
12726 4241 4241
```

```
train = ds['train']
img_count = sum(1 for x in train if x["image"] is not None)
print(img_count, "من أصل", len(train))
```

```
6218 من أصل 12726
```

```
import collections
subjects = [x["subject"] for x in train]
```

```
طول السؤال:
أصغر: 3
أكبر: 141
المتوسط: 12.203206034889202
الوسيط: 9.0

طول الـ hint:
أصغر: 4
أكبر: 304
المتوسط: 35.29774633985853
الوسيط: 24.0

طول الحل solution:
أصغر: 4
أكبر: 437
المتوسط: 48.12540165002171
الوسيط: 37.0
```

```
import numpy as np

# حساب طول الكلمات لكل lecture
l_len = [len(x["lecture"].split()) for x in train if x["lecture"]]

# ملخص المحاضرة lecture
print("طول المحاضرة lecture:")
print("أصغر:", np.min(l_len))
print("أكبر:", np.max(l_len))
print("المتوسط:", np.mean(l_len))
print("الوسيط:", np.median(l_len))
```

```
طول المحاضرة lecture:
أصغر: 8
أكبر: 490
المتوسط: 125.08305274971941
الوسيط: 104.0
```

```
from datasets import load_dataset
ds = load_dataset("derek-thomas/ScienceQA")

print(ds['train'].column_names)
```

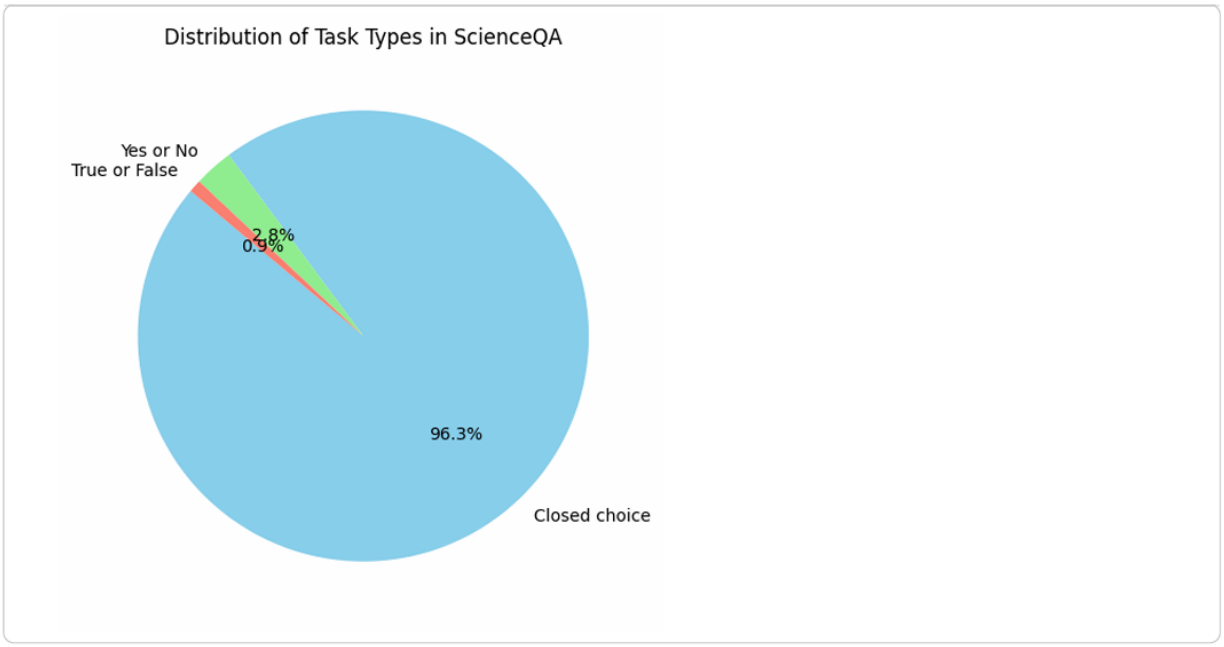
```
['image', 'question', 'choices', 'answer', 'hint', 'task', 'grade', 'subject', 'topic', 'category', 'skill', 'lecture', 'sol
```

```
grades = [x['grade'] for x in train]
print("توزيع الصفوف:", collections.Counter(grades))
```

```
توزيع الصفوف: Counter({'grade4': 2134, 'grade5': 1825, 'grade3': 1786, 'grade7': 1642, 'grade8': 1520, 'grade6': 1491, 'grade2
```

```
skills = [x['skill'] for x in train]
print("توزيع المهارات:", collections.Counter(skills))
```

```
توزيع المهارات: Counter({'Use guide words': 419, 'Inherited and acquired traits: use evidence to support a statement': 355, 'Com
```




```
lec_len = [len(x["lecture"].split()) for x in train if x["lecture"]]
print("متوسط طول lecture:", np.mean(lec_len))
```

3.6.3 الخاتمة

أظهر التحليل الاستكشافي للبيانات أن مجموعة بيانات ScienceQA تتميز بتنوعها من حيث أطوال الأسئلة، وأنواع السياقات النصية والبصرية، وتوزيع المستويات التعليمية والموضوعات العلمية. كما أكّد التحليل الطبيعة متعددة الوسائط لهذه المجموعة، حيث تعتمد نسبة كبيرة من الأسئلة على دمج المعلومات النصية والبصرية معًا. وتدل هذه الخصائص على ملاءمة مجموعة البيانات لتطوير وتقييم نظام البحث متعدد الوسائط المقترح، إذ توفر بيئة غنية لاختبار قدرة النظام على فهم السياق التعليمي المتكامل واسترجاع المحتوى ذي الصلة بدقة.

3.7 معالجة البيانات وتحضيرها (Data Preprocessing)

في هذا المشروع، لم يتم إجراء أي عمليات معالجة أو تعديل إضافي على مجموعة بيانات ScienceQA ، حيث تم استخدام البيانات بصيغتها الأصلية كما هي متوفرة. ويعود هذا القرار إلى كون مجموعة البيانات مُعدّة مسبقًا لأغراض تعليمية وبحثية، وتتمتع ببنية منظّمة وجودة عالية، مما يجعلها مناسبة للاستخدام المباشر في نظام البحث متعدد الوسائط المقترح. وقد أتاح هذا النهج تقييم أداء النظام في التعامل مع بيانات واقعية دون الاعتماد على عمليات تنقية أو تحسين قد تؤثر على طبيعة المحتوى الأصلي.

3.8تمثيل البيانات والفهرسة

بعد تجهيز مجموعة البيانات، تم تمثيل البيانات النصية باستخدام نموذج التضمين BGE (BAAI General Embedding)، الذي يُعد من النماذج الحديثة والفعّالة في تمثيل المعنى الدلالي للنصوص ضمن فضاء متجهي عالي الأبعاد. يتيح هذا النموذج تحويل الأسئلة التعليمية والمحتوى النصي إلى تمثيلات عددية دقيقة تسهّل قياس التشابه الدلالي بين الاستعلامات والمحتوى المخزّن.
أما البيانات الصورية، فقد تم تمثيلها باستخدام نماذج تضمين مناسبة تتيح تحويل الصور التعليمية إلى متجهات عددية تعكس خصائصها البصرية والدلالية، مما يدعم تنفيذ البحث متعدد الوسائط داخل النظام المقترح.
تم تخزين التمثيلات المتجهية الناتجة لكل من النصوص والصور داخل قاعدة بيانات متجهية تعتمد على Milvus، وهي قاعدة بيانات مصمّمة خصيصًا للتعامل مع عمليات الفهرسة والاسترجاع السريع للمتجهات عالية الأبعاد. يوفّر استخدام Milvus كفاءة عالية في عمليات البحث الدلالي وقابلية للتوسّع، مما يجعلها مناسبة للتعامل مع مجموعات بيانات تعليمية كبيرة ومتعددة الوسائط.
تشكل مرحلة التمثيل والفهرسة الأساس التقني لنظام Retrieval-Augmented Generation (RAG)المستخدم في هذا المشروع، حيث يتم توظيف نتائج الاسترجاع المتجهية من قاعدة بيانات Milvusلتزويد النموذج اللغوي بالمحتوى الأكثر صلة بالسؤال المطروح. وسيتم تناول التفاصيل التقنية المتعلقة بنموذجBGE ، وبنية قاعدة بياناتMilvus ، وآلية دمجها ضمن النظام بشكل موسّع في الفصول اللاحقة، ولا سيما في الفصل الرابع (المنهجية المقترحة) والفصل الخامس (التنفيذ).

3.9ملاءمة مجموعة بيانات ScienceQA للنظام المقترح

تُعد مجموعة بيانات ScienceQAمناسبة بشكل كبير للنظام المقترح للبحث متعدد الوسائط عن المحتوى التعليمي، وذلك لعدة أسباب جوهرية تتماشى مع أهداف المشروع ومنهجيته. أولًا، تتميز ScienceQA بدعمها الصريح لتعدد الوسائط، حيث تجمع بين البيانات النصية والصورية ضمن كل عينة بيانات، مما يتيح تقييم قدرة النظام على فهم ودمج مصادر معلومات مختلفة عند معالجة الاستعلامات التعليمية.

ثانيًا، تتسم مجموعة البيانات بطبيعتها التعليمية الواضحة، إذ تم تصميم الأسئلة لتغطي مفاهيم علمية مأخوذة من مناهج دراسية حقيقية، وبمستويات تعليمية متفاوتة. ويجعل ذلك ScienceQA مناسبة لتطوير أنظمة بحث تعليمية موجّهة للمتعلمين، حيث تتطلب الأسئلة فهمًا عميقًا للمحتوى بدلًا من الاعتماد على مطابقة الكلمات المفتاحية فقط.

علاوة على ذلك، تحتوي مجموعة بيانات ScienceQA على تفسيرات مرافقة للإجابات الصحيحة، مما يوفّر سياقًا إضافيًا يمكن الاستفادة منه في تقييم جودة نتائج الاسترجاع والإجابات التي يولدها النظام. كما تساهم هذه التفسيرات في دعم الاستخدام المستقبلي للنظام في تطبيقات تعليمية تفاعلية تهدف إلى تعزيز الفهم وليس مجرد تقديم الإجابة النهائية.

وأخيرًا، فإن البنية المنظّمة لمجموعة البيانات وقابليتها للتحويل إلى تمثيلات متجهية تجعلها ملائمة للاستخدام في أنظمة البحث الدلالي. ويسمح ذلك بدمجها بسهولة مع قواعد بيانات متجهية ونماذج تضمين متقدمة، الأمر الذي يدعم تنفيذ منهجية Retrieval-Augmented Generation (RAG)بكفاءة.

وبناءً على هذه الخصائص، تم اختيار ScienceQA لتكون الأساس التجريبي للنظام المقترح في هذا المشروع.

3.10 ملخص الفصل

استعرض هذا الفصل مجموعة البيانات المستخدمة في المشروع، مع تقديم وصف تفصيلي لبنيتها ومكوّناتها وأنواع البيانات التي تحتويها. كما تم عرض الخصائص الإحصائية لمجموعة بيانات ScienceQA وإجراء تحليل استكشافي للبيانات بهدف فهم توزيع الأسئلة، وتنوع السياقات النصية والبصرية، والمجالات التعليمية التي تغطيها.
بالإضافة إلى ذلك، تناول الفصل مراحل تحضير البيانات وتمثيلها وفهرستها باستخدام نماذج تضمين وقواعد بيانات متجهية، مبيّنًا دور هذه المراحل في دعم البحث الدلالي متعدد الوسائط. وفي الختام، تم توضيح أسباب اختيار مجموعة بيانات ScienceQA وملاءمتها للنظام المقترح. ويمهّد هذا الفصل للانتقال إلى الفصل الرابع، الذي يتناول المنهجية المقترحة وبنية النظام متعدد الوسائط المستخدمة لتحقيق أهداف المشروع.

الفصل الرابع: المنهجية المقترحة

4.1مقدمة الفصل

يستعرض هذا الفصل المنهجية المقترحة لتصميم وتنفيذ نظام بحث متعدد الوسائط مخصص للمحتوى التعليمي. يهدف النظام إلى تمكين المستخدم من استرجاع المعرفة التعليمية اعتمادًا على أكثر من نمط إدخال، بما في ذلك النصوص والصور، مع الاستفادة من تقنيات الذكاء الاصطناعي الحديثة مثل التضمينات الدلالية(Embeddings) ، النماذج اللغوية الكبيرة(LLMs) ، ونظام الاسترجاع المعزز بالتوليد (Retrieval-Augmented Generation – RAG).

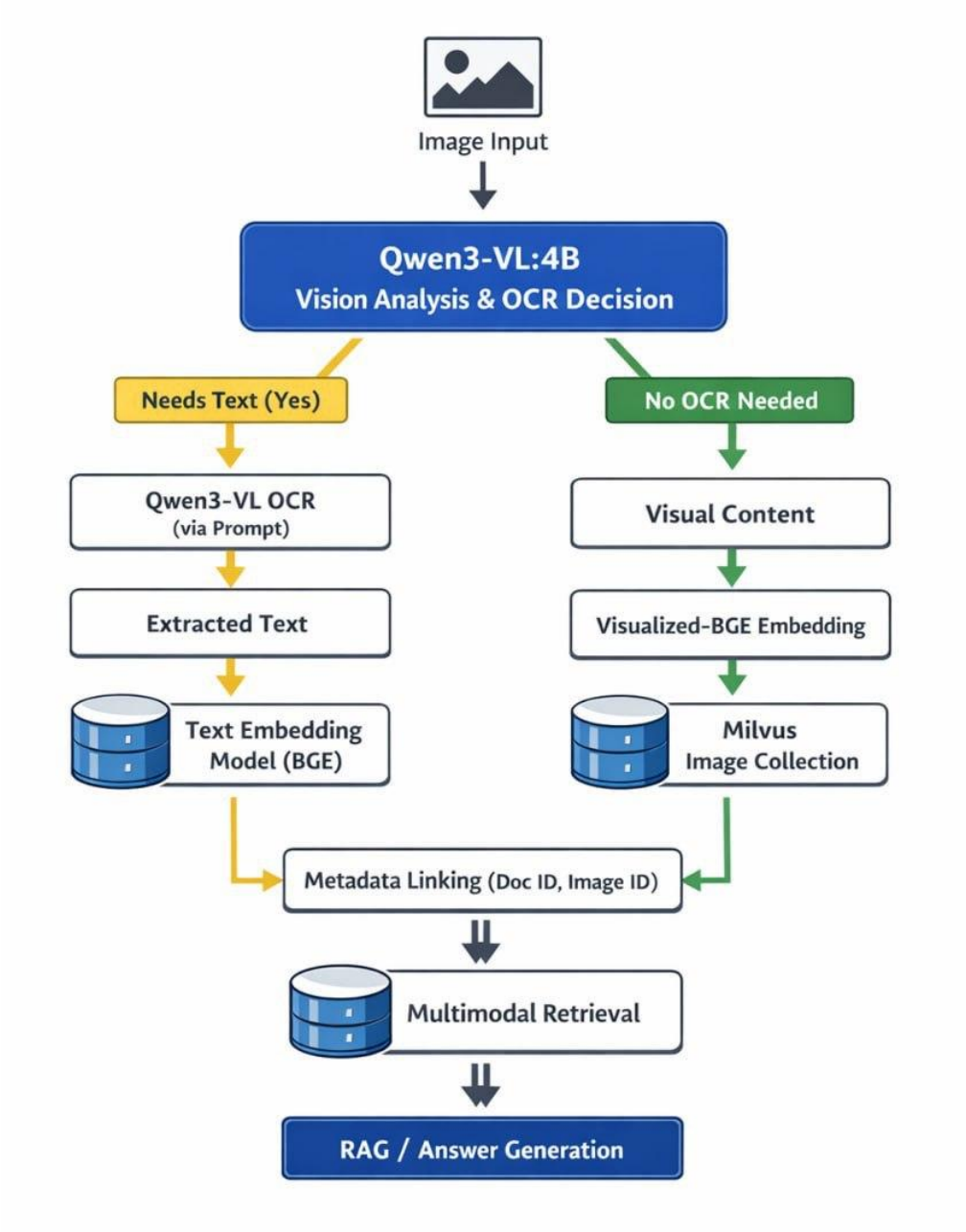
تعتمد المنهجية المقترحة على دمج المعالجة النصية والرؤية الحاسوبية ضمن إطار موحد، بحيث يمكن للنظام فهم المدخلات متعددة الوسائط، ومعالجتها دلاليًا، ثم استرجاع المحتوى التعليمي الأكثر صلة وتوليد إجابة دقيقة ومدعومة بالأدلة.

4.2النظرة العامة للنظام

يعتمد النظام المقترح على بنية متعددة المكونات تهدف إلى تحقيق تكامل فعّال بين معالجة البيانات متعددة الوسائط، وآلية الاسترجاع الدلالي، وقدرات التوليد اللغوي.
يتكوّن النظام بشكل عام من مراحل رئيسية:

- استقبال مدخل المستخدم (نص أو صورة).
- تحديد نوع المدخل ومعالجته وفق المسار المناسب.
- تحويل المدخل إلى تمثيل متجهي.(Embedding)
- تنفيذ عمليات تحسين السؤال (تصحيح، إعادة صياغة، وتوسيع دلالي عند الحاجة).
- استرجاع المحتوى ذي الصلة من قاعدة البيانات المتجهية.
- دمج النتائج النصية والبصرية.
- توليد الإجابة النهائية باستخدام.RAG

يبدأ النظام باستقبال سؤال المستخدم، والذي قد يكون نصيًا أو مرتبطًا بصورة تعليمية. بعد ذلك، يتم تحويل هذا الإدخال إلى تمثيل متجهي مناسب، ثم مقارنته مع البيانات المخزّنة في قاعدة البيانات المتجهية لاستخراج المحتوى الأكثر صلة. وأخيرًا، يتم تمرير النتائج المسترجعة إلى نموذج لغوي كبير لتوليد إجابة نهائية مدعومة بالسياق.



رسم توضيحي8 : توضيح مخطط سير العمل

4.3إطار تنسيق العمليات باستخدام LangChain وLangGraph

يعتمد النظام المقترح في هذا المشروع على إطار برمجي متقدم لتنسيق وتنفيذ مكونات المعالجة المختلفة، وذلك باستخدام كل من **LangChain**و **LangGraph**، لما يوفرانه من مرونة عالية في تصميم أنظمة قائمة على النماذج اللغوية الكبيرة وإدارة تدفق العمليات المعقدة داخل أنظمة الاسترجاع المعزز بالتوليد.(RAG)

تُستخدم مكتبة **LangChain**كطبقة ربط أساسية بين وحدات النظام المختلفة، حيث تتيح بناء سلاسل معالجة (Chains) منظمة تربط بين مراحل تصحيح السؤال، إعادة صياغته، تحويله إلى تمثيل متجهي،

تنفيذ عمليات الاسترجاع، وأخيرًا توليد الإجابة. يساهم هذا الإطار في توحيد التعامل مع النماذج اللغوية الكبيرة وقواعد البيانات المتجهية وأدوات المعالجة المختلفة ضمن بنية برمجية واحدة، مما يسهل عملية التطوير والتوسعة المستقبلية للنظام.

أما **LangGraph**، فيُستخدم لتمثيل سير العمل الكامل للنظام على شكل **مخطط حالات (State Graph)**، حيث يتم تعريف كل مرحلة من مراحل المعالجة كعقدة مستقلة داخل الرسم البياني، مع تحديد الانتقالات المنطقية بين هذه العقد. يتيح هذا النهج إدارة تدفق البيانات بشكل دقيق، خاصة في الأنظمة متعددة المسارات مثل النظام المقترح، الذي يدعم مسارًا نصيًا ومسارًا بصريًا لمعالجة مدخلات المستخدم.
يسمح استخدام LangGraph بالتحكم في التفرعات الشرطية داخل النظام، مثل التمييز بين المدخل النصي والمدخل الصوري، واتخاذ القرار المناسب بشأن تفعيل عمليات إضافية مثل التعرف الضوئي على الحروف (OCR)أو تنفيذ إعادة صياغة دلالية للسؤال. كما يساهم في منع التعارض بين تحديثات الحالة (State Updates)من خلال تنظيم التفاعل بين العقد المختلفة وضمان أن كل مرحلة تتعامل مع نسخة متسقة من بيانات الإدخال.

بالإضافة إلى ذلك، يدعم LangGraph تنفيذ عمليات متوازية عند الحاجة، مثل معالجة النصوص والصور بشكل منفصل ثم دمج نتائجها لاحقًا، مما يعزز كفاءة النظام وقابليته للتوسع. ويُعد هذا النهج مناسبًا بشكل خاص للأنظمة متعددة الوسائط، حيث تتطلب إدارة دقيقة لتدفق البيانات بين وحدات المعالجة المختلفة.

من خلال الدمج بين LangChain وLangGraph، يحقق النظام المقترح بنية معيارية وقابلة لإعادة الاستخدام، حيث يمكن تعديل أو استبدال أي مكوّن دون التأثير على بقية النظام. كما يضمن هذا التكامل تنفيذًا منظمًا لمنهجية الاسترجاع المعزز بالتوليد، ويعزز موثوقية النتائج من خلال التحكم الكامل في مراحل المعالجة المختلفة.

يوضح **الشكل (4.1)** البنية العامة لسير العمل داخل النظام باستخدام LangChain وLangGraph، حيث يُبيّن كيفية انتقال البيانات بين العقد المختلفة منذ لحظة إدخال السؤال وحتى توليد الإجابة النهائية المدعومة بالأدلة.

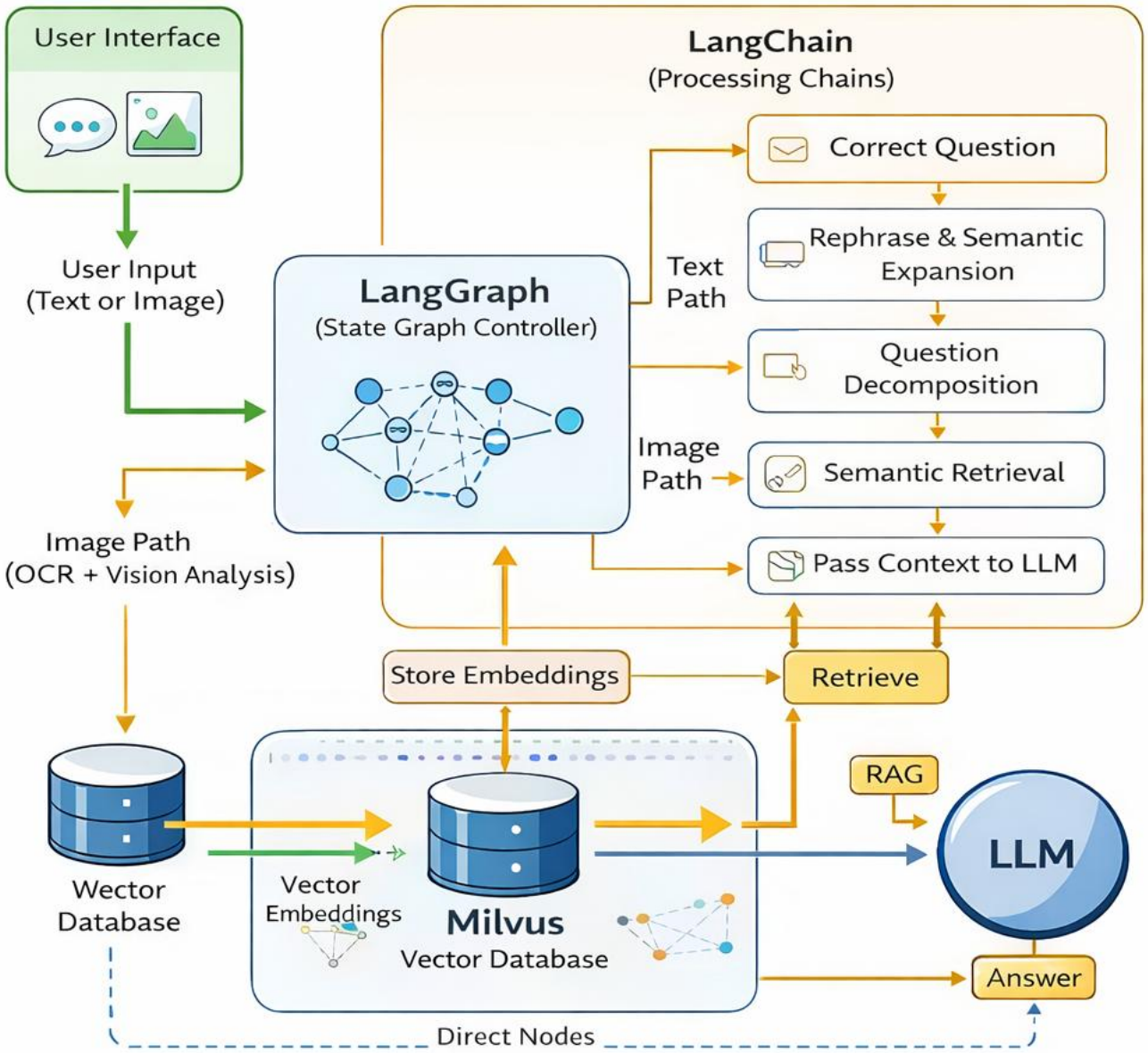


Figure 4.1: System Architecture and Workflow Using LangChain and LangGraph

رسم توضيحي ٩: بنية النظام باستخدام langchain & langgraph

4.4معالجة مدخلات المستخدم(User Input Handling)

4.4.1المدخل النصي(Text-Based Input)

عند إدخال المستخدم لسؤال نصي، يقوم النظام أولًا باستقبال النص الخام كما هو مكتوب من قبل المستخدم. بعد ذلك، يتم تمرير السؤال عبر وحدة تحليل لغوي تعتمد على نموذج لغوي كبير، بهدف:

- تصحيح الأخطاء اللغوية أو الإملائية إن وجدت.
- تحسين صياغة السؤال لغويًا ودلاليًا.
- إزالة الغموض أو الصياغات غير الدقيقة.

تهدف هذه الخطوة إلى ضمان أن يكون السؤال واضحًا وقابلًا للفهم من قبل مكونات النظام اللاحقة، مما يسهم في تحسين دقة نتائج البحث.

4.4.2إعادة الصياغة والتوسيع الدلالي(Rephrasing and Semantic Expansion)

بعد تصحيح السؤال، قد يقوم النظام بإعادة صياغته أو توليد عدة صيغ دلالية مكافئة له. يتم ذلك من أجل:

- التقاط المعنى الحقيقي للسؤال.
- زيادة فرص مطابقة السؤال مع محتوى قاعدة المعرفة.
- دعم البحث الدلالي بدلاً من الاعتماد على التطابق الحرفي للكلمات.
-

في بعض الحالات، يتم تقسيم السؤال المعقد إلى أسئلة فرعية أبسط، مما يسمح باسترجاع معلومات أكثر دقة وشمولية.

4.4.3تحويل النص إلى تمثيل متجهي(Text Embedding)

يتم تحويل السؤال النصي النهائي إلى تمثيل متجهي باستخدام نموذج تضمين نصي متقدم مثل **BGE (Bidirectional General Embedding)**. يتيح هذا التمثيل:

- التعبير عن المعنى الدلالي للنص رقميًا.
- إجراء عمليات مقارنة فعالة بين السؤال والمحتوى المخزن.
- دعم البحث الدلالي في قاعدة البيانات المتجهية.
-

يتم بعد ذلك استخدام هذا التمثيل للبحث عن أقرب المتجهات داخل قاعدة البيانات.

4.5معالجة المدخلات الصورية(Image-Based Input)

4.5.1تحليل الصورة واتخاذ قرارOCR

في حال إدخال المستخدم لصورة تعليمية، يتم تمرير الصورة إلى نموذج رؤية متعدد الوسائط مثل **Qwen3-VL**. يقوم النموذج بتحليل الصورة وتحديد ما إذا كانت تحتوي على نصوص ذات أهمية تعليمية أم لا.

- إذا كانت الصورة تحتوي على نص (مثل مخططات، أسئلة، أو شروحات)، يتم تفعيل مرحلة التعرف الضوئي على الحروف.(OCR)
- إذا كانت الصورة تعتمد على المحتوى البصري فقط، يتم التعامل معها مباشرة كمدخل بصري.

4.5.2استخراج النص من الصورة(OCR Processing)

في حال الحاجة إلىOCR ، يتم استخراج النص من الصورة باستخدام نموذج الرؤية نفسه أو عبر توجيه مخصص .(Prompt-based OCR) يتم بعد ذلك:

- تنظيف النص المستخرج.
- معالجته لغويًا.
- تحويله إلى تمثيل متجهي باستخدام نموذج BGE النصي.

4.5.3تمثيل المحتوى البصري(Visual Embedding)

بالنسبة للمحتوى البصري غير النصي، يتم تحويل الصورة مباشرة إلى تمثيل متجهي باستخدام نموذج تضمين بصري متوافق مع البحث متعدد الوسائط، مثل **Visual-BGE**. يتيح ذلك:

- تمثيل الخصائص البصرية للصورة بشكل رقمي.
- مقارنة الصور مع محتوى بصري آخر داخل قاعدة البيانات.
- دعم الاسترجاع العابر للوسائط.(Cross-Modal Retrieval)

4.6الربط بين النص والصورة(Metadata Linking)

بعد تحويل النصوص والصور إلى تمثيلات متجهية، يتم ربطها باستخدام بيانات وصفية (Metadata) مثل:

- معرف الوثيقة.(Document ID)
- معرف الصورة.(Image ID)
- نوع الوسيط (نص / صورة).

يسمح هذا الربط باسترجاع محتوى متكامل يجمع بين الشرح النصي والمعلومة البصرية المرتبطة بالسؤال.

4.7 الاسترجاع متعدد الوسائط(Multimodal Retrieval)

تعتمد مرحلة الاسترجاع على قاعدة بيانات متجهية مثل **Milvus**، حيث يتم:

- تنفيذ بحث Top-k باستخدام التمثيلات المتجهية.
- استرجاع العناصر الأكثر تشابهًا دلاليًا مع سؤال المستخدم.
- الجمع بين النتائج النصية والبصرية عند الحاجة.
-

تتميز هذه المرحلة بالكفاءة العالية والقدرة على التعامل مع كميات كبيرة من البيانات متعددة الوسائط.

4.8 التوليد المعزز بالاسترجاع(RAG-based Answer Generation)

بعد استرجاع المحتوى ذي الصلة، يتم تمريره إلى نموذج لغوي كبير ضمن إطار **Retrieval-Augmented Generation**. تقوم هذه المرحلة بـ:

- تحليل الأدلة المسترجعة.
- تنظيم المعلومات ذات الصلة.
- توليد إجابة نهائية دقيقة ومتماسكة.
-

يساعد هذا النهج على تقليل الهلوسة اللغوية، حيث تعتمد الإجابة على محتوى فعلي مسترجع من قاعدة المعرفة.

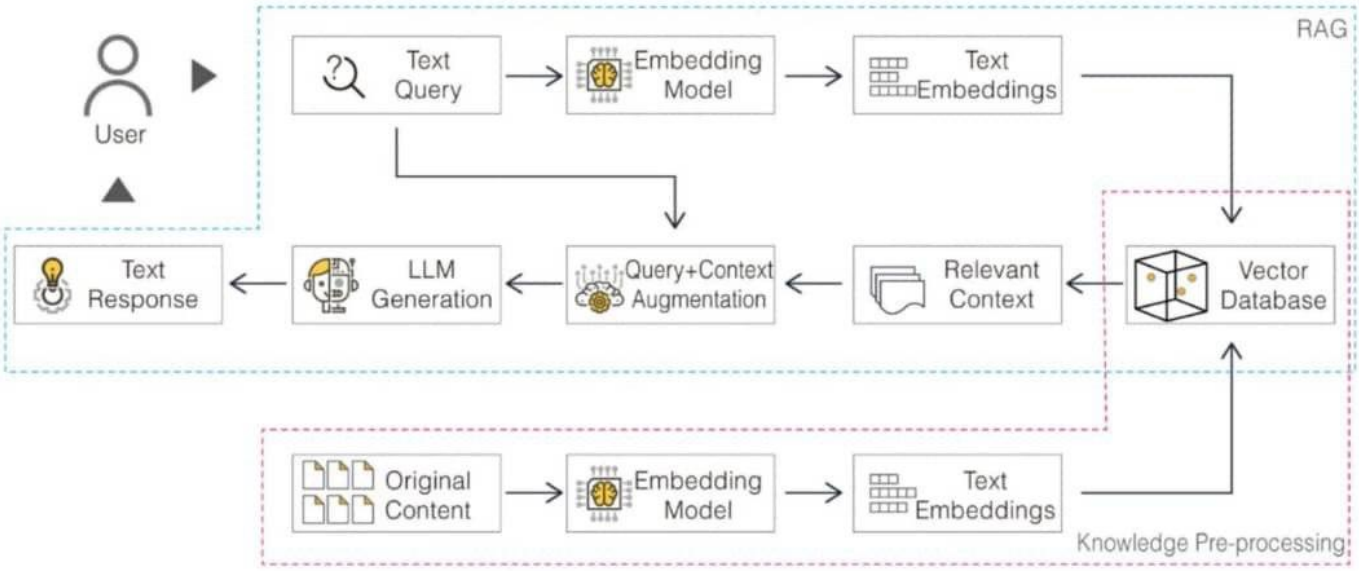
4.9 دمج الوسائط والتحقق من الاتساق

في حال وجود نتائج من أكثر من وسيط (نص + صورة)، يقوم النظام بدمج المعلومات والتحقق من اتساقها المنطقي والدلالي قبل عرض الإجابة النهائية للمستخدم.

4.10 مميزات المنهجية المقترحة

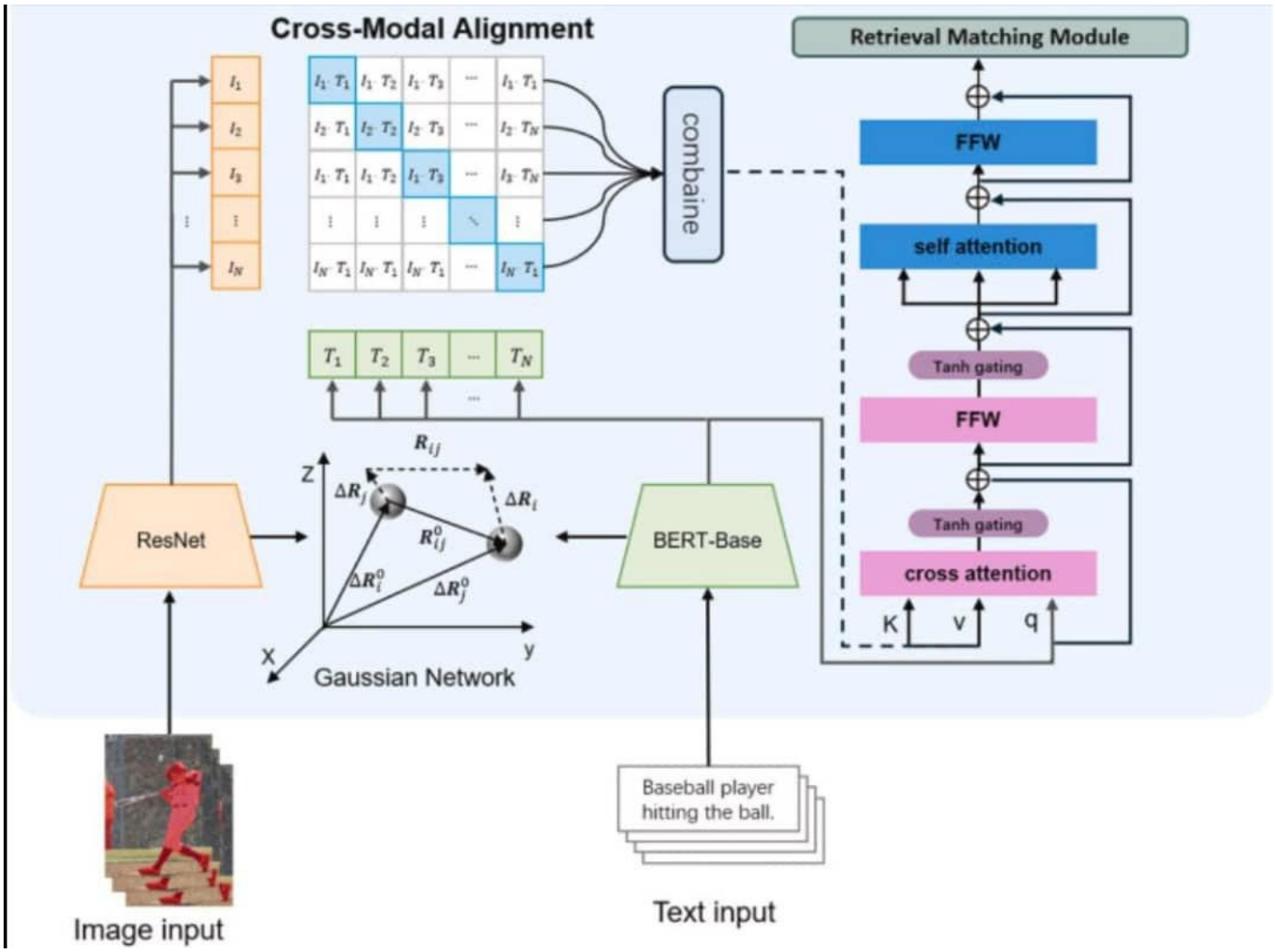
تتميز المنهجية المقترحة بعدة خصائص رئيسية، من أهمها:

- دعم حقيقي لتعدد الوسائط.
- اعتماد البحث الدلالي بدلًا من البحث التقليدي.
- دمج فعال بين الرؤية الحاسوبية والنماذج اللغوية.
- تحسين جودة الإجابات باستخدام RAG.
- ملاءمتها للمجال التعليمي.



Schematic workflow of a traditional text-based RAG system

رسم توضيحي ١٠: مخطط سير عمل الRAG



رسم توضيحي ١١: عمل cross modal

4.11 ملخص الفصل

تناول هذا الفصل عرضًا تفصيليًا للمنهجية المقترحة لتصميم وتنفيذ نظام بحث متعدد الوسائط مخصص للمحتوى التعليمي، حيث تم التركيز على دمج تقنيات الذكاء الاصطناعي الحديثة لمعالجة المدخلات النصية والبصرية ضمن إطار موحد وذكي. استعرض الفصل البنية العامة للنظام وآلية تدفق البيانات منذ لحظة إدخال المستخدم للسؤال وحتى توليد الإجابة النهائية، مع توضيح دور كل مكوّن من مكونات النظام.

تم شرح كيفية تعامل النظام مع المدخلات النصية من خلال مراحل تصحيح السؤال لغويًا، وإعادة صياغته دلاليًا، وتوسيعه عند الحاجة، ثم تحويله إلى تمثيل متجهي باستخدام نماذج تضمين متقدمة لدعم البحث الدلالي. كما تناول الفصل معالجة المدخلات الصورية، موضحًا كيفية تحليل الصور التعليمية باستخدام نماذج رؤية متعددة الوسائط، واتخاذ القرار المناسب بشأن الحاجة إلى استخراج النصوص منها عبر تقنيات التعرف الضوئي على الحروف(OCR) ، أو التعامل معها مباشرة كمحتوى بصري.

كذلك، ناقش الفصل آلية تمثيل النصوص والصور في فضاء متجهي موحد، وربطها باستخدام البيانات الوصفية، مما يتيح تنفيذ عمليات استرجاع متعددة الوسائط بكفاءة عالية. وتم توضيح دور قاعدة البيانات المتجهية في تسريع عملية البحث واسترجاع المحتوى الأكثر صلة، إضافة إلى دمج النتائج النصية والبصرية ضمن مرحلة واحدة.

واختُتم الفصل بشرح دور نظام الاسترجاع المعزز بالتوليد (RAG) في تحسين جودة الإجابات النهائية، حيث يعتمد على المحتوى المسترجع فعليًا من قاعدة المعرفة لتوليد إجابات دقيقة ومتسقة، مما يقلل من احتمالية الهلوسة اللغوية ويزيد من موثوقية النظام. وبذلك، يشكّل هذا الفصل الأساس المنهجي الذي تُبنى عليه مرحلة التنفيذ العملي، والتي سيتم تناولها بالتفصيل في الفصل الخامس.

الفصل الخامس: التنفيذ

5.1مقدمة الفصل

يركّز هذا الفصل على عرض الجوانب التطبيقية والتنفيذية للنظام المقترح للبحث متعدد الوسائط في المحتوى التعليمي. فبعد أن تم في الفصل السابق استعراض المنهجية المقترحة وبنية النظام من الناحية النظرية، يهدف هذا الفصل إلى توضيح كيفية تحويل ذلك التصميم إلى نظام عملي متكامل يعمل فعليًا على معالجة مدخلات المستخدم المتعددة (نصية وصورية) واسترجاع المحتوى التعليمي المناسب باستخدام تقنيات الاسترجاع المعزز بالتوليد.
يتناول هذا الفصل بيئة التطوير، الأدوات والتقنيات المستخدمة، وآلية تنفيذ كل مرحلة من مراحل النظام، بدءًا من استقبال المدخلات، مرورًا بمعالجتها وتحويلها إلى تمثيلات متجهية، وانتهاءً بتوليد الإجابة النهائية باستخدام نظام الاسترجاع المعزز بالتوليد .(RAG) كما يسلّط الضوء على التحديات التقنية التي واجهت عملية التنفيذ والحلول المعتمدة لمعالجتها.

5.2بيئة التطوير والأدوات المستخدمة

تم تنفيذ النظام باستخدام مجموعة من الأدوات والتقنيات الحديثة التي تتيح بناء نظام بحث ذكي وقابل للتوسّع. وقد تم اختيار هذه الأدوات بعناية بما يتناسب مع طبيعة المشروع ومتطلباته متعددة الوسائط.

5.2.1لغة البرمجة وإطار العمل

اعتمد تنفيذ النظام على لغة **Python** نظرًا لدعمها الواسع لمكتبات الذكاء الاصطناعي وتعلم الآلة، إضافة إلى سهولة التكامل مع نماذج التعلم العميق وقواعد البيانات المتجهية.

كما تم استخدام إطار **LangChain** لبناء سلاسل المعالجة اللغوية، وإطار **LangGraph** لإدارة تدفق الحالات (State Management) والتحكم في مسار المعالجة داخل النظام.

5.2.2النماذج الذكية المستخدمة

تم توظيف عدة نماذج ذكاء اصطناعي، من أبرزها:

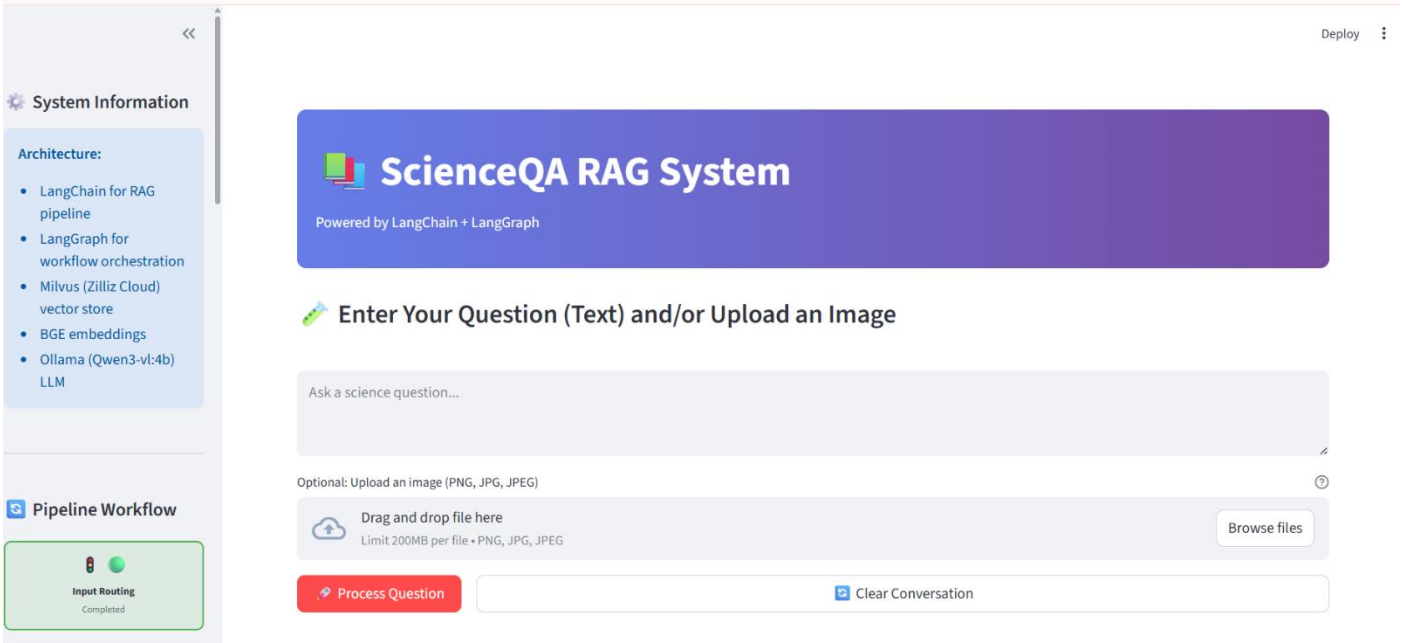
- نموذج **BGE Embedding** لتحويل النصوص والصور إلى تمثيلات متجهية.
- نموذج **Qwen3-VL** لمعالجة المدخلات الصورية وتحليل المحتوى البصري واتخاذ قرار الحاجة إلى استخدام OCR.
- نموذج لغوي كبير (LLM) لتوليد الإجابات النهائية بالاعتماد على السياق المسترجع.
-

5.2.3قاعدة البيانات المتجهية

تم استخدام قاعدة البيانات المتجهية **Milvus** لتخزين التمثيلات المتجهية (Embeddings) للنصوص والصور، وذلك لما توفره من كفاءة عالية في عمليات البحث الدلالي واسترجاع أقرب المتجهات.

5.2.4 واجهة المستخدم

تم تطوير واجهة المستخدم باستخدام **Streamlit**، مما أتاح إنشاء واجهة تفاعلية بسيطة وسهلة الاستخدام تُمكّن المستخدم من إدخال النصوص أو الصور وعرض النتائج بشكل مباشر.



رسم توضيحي ١٢: واجهة المستخدم للنظام المقترح القائم على الاسترجاع المعزز بالتوليد (RAG).

5.3 تنفيذ معالجة المدخلات النصية

عند قيام المستخدم بإدخال سؤال نصي، يمر هذا السؤال بعدة مراحل معالجة تهدف إلى تحسين جودة البحث ودقة النتائج.

5.3.1 استقبال السؤال النصي

يتم استقبال السؤال النصي من خلال واجهة المستخدم، حيث يُخزّن مؤقتًا ضمن حالة النظام ليتم تمريره إلى مراحل المعالجة اللاحقة.

5.3.2 تصحيح السؤال لغويًا

في هذه المرحلة، يتم فحص السؤال لغويًا باستخدام النموذج اللغوي الكبير، بهدف:

- تصحيح الأخطاء الإملائية والنحوية.
- تحسين صياغة السؤال دون تغيير معناه.

تُعد هذه الخطوة ضرورية لضمان أن يكون تمثيل السؤال دقيقًا عند تحويله إلى تمثيل متجهي.

📌

مخطط يوضّح عملية تصحيح السؤال لغويًا

5.3.3 إعادة كتابة وصياغة السؤال (Question Rewriting)

بعد إتمام عملية التصحيح اللغوي، يخضع السؤال لمرحلة **إعادة كتابة وصياغة** تهدف إلى تحسين وضوحه ودقته الدلالية دون التأثير على المعنى الأصلي. تعتمد هذه المرحلة على النموذج اللغوي الكبير (LLM) الذي يقوم بإعادة صياغة السؤال بأسلوب أكثر تنظيمًا واتساقًا لغويًا.

تكمن أهمية هذه الخطوة في معالجة الأسئلة التي قد تكون:

- غير واضحة الصياغة،
- تحتوي على جمل عامية أو غير أكاديمية،
- أو مصاغة بطريقة قد تؤثر سلبًا على جودة البحث الدلالي.
-

تساهم إعادة الصياغة في إنتاج سؤال قياسي ومفهوم، مما ينعكس إيجابًا على عملية تحويل السؤال إلى تمثيل متجهي (Embedding) أكثر دقة، وبالتالي تحسين نتائج الاسترجاع.

📌

مثال يوضّح الفرق بين السؤال قبل إعادة الصياغة وبعدها .

5.3.4 التوسيع الدلالي للسؤال (Semantic Expansion)

في هذه المرحلة، يتم توسيع السؤال دلاليًا عبر إضافة مفاهيم مرادفة أو مصطلحات ذات صلة بالسؤال الأصلي. يعتمد النظام على القدرات الدلالية للنموذج اللغوي الكبير لاستخلاص المعاني الضمنية المرتبطة بالسؤال.

يهدف التوسيع الدلالي إلى:

- تجاوز محدودية البحث القائم على الكلمات المفتاحية.
- تحسين فرص استرجاع محتوى تعليمي متنوع ومتكامل.
- التقاط السياق العلمي للسؤال بشكل أعمق.
-

تُعد هذه الخطوة ضرورية خصوصًا في المجال التعليمي، حيث قد تختلف المصطلحات المستخدمة للإشارة إلى نفس المفهوم العلمي.

5.3.5 تقسيم السؤال المركّب (Question Decomposition)

في حال كان السؤال المدخل مركّبًا أو يتضمن أكثر من مطلب، يقوم النظام بتنفيذ عملية **تقسيم السؤال** إلى مجموعة من الأسئلة الفرعية الأبسط. تهدف هذه الخطوة إلى معالجة كل جزء من السؤال بشكل مستقل لضمان دقة الإجابة وشموليتها.

على سبيل المثال، إذا احتوى السؤال على:

- أكثر من مفهوم علمي،
- أو طلب مقارنة بين ظاهرتين،
- أو سؤال تفسيري متعدد الأجزاء،

فإن النظام يقوم بتحليل بنية السؤال وتقسيمه إلى وحدات دلالية أصغر، يتم التعامل معها بشكل منفصل خلال مرحلة الاسترجاع.

يساعد هذا الأسلوب على:

- تحسين جودة نتائج البحث.
- منع إغفال أي جزء من السؤال.
- توليد إجابات أكثر اكتمالًا وتنظيمًا.

📌

مثال يوضّح تقسيم سؤال مركّب إلى أسئلة فرعية.

5.3.6 تحويل السؤال النهائي إلى تمثيل متجهي (Embedding)

بعد الانتهاء من جميع مراحل التحسين (التصحيح، إعادة الصياغة، التوسيع، والتقسيم)، يتم تحويل السؤال أو الأسئلة الناتجة إلى تمثيلات متجهية باستخدام نموذج **BGE Embedding**. تُستخدم هذه التمثيلات لاحقًا في عملية البحث الدلالي داخل قاعدة البيانات المتجهية Milvus.

(ممكن نحط صورة للembrdding)

5.4 تنفيذ معالجة المدخلات الصورية

يدعم النظام استقبال الصور التعليمية كمدخلات بحث، وهو ما يشكّل جوهر الطابع متعدد الوسائط للمشروع.

5.4.1 استقبال الصورة وتحليلها

عند رفع صورة، يتم تمريرها إلى نموذج **Qwen3-VL** الذي يقوم بتحليل المحتوى البصري وتحديد ما إذا كانت الصورة تحتوي على نص يحتاج إلى استخراج.

5.4.2 قرار استخدام OCR

بناءً على تحليل الصورة:

- إذا كانت الصورة تحتوي على نص، يتم تفعيل عملية OCR لاستخراج النص.
- إذا لم يكن هناك نص، يتم التعامل مع الصورة كمحتوى بصري فقط.
-

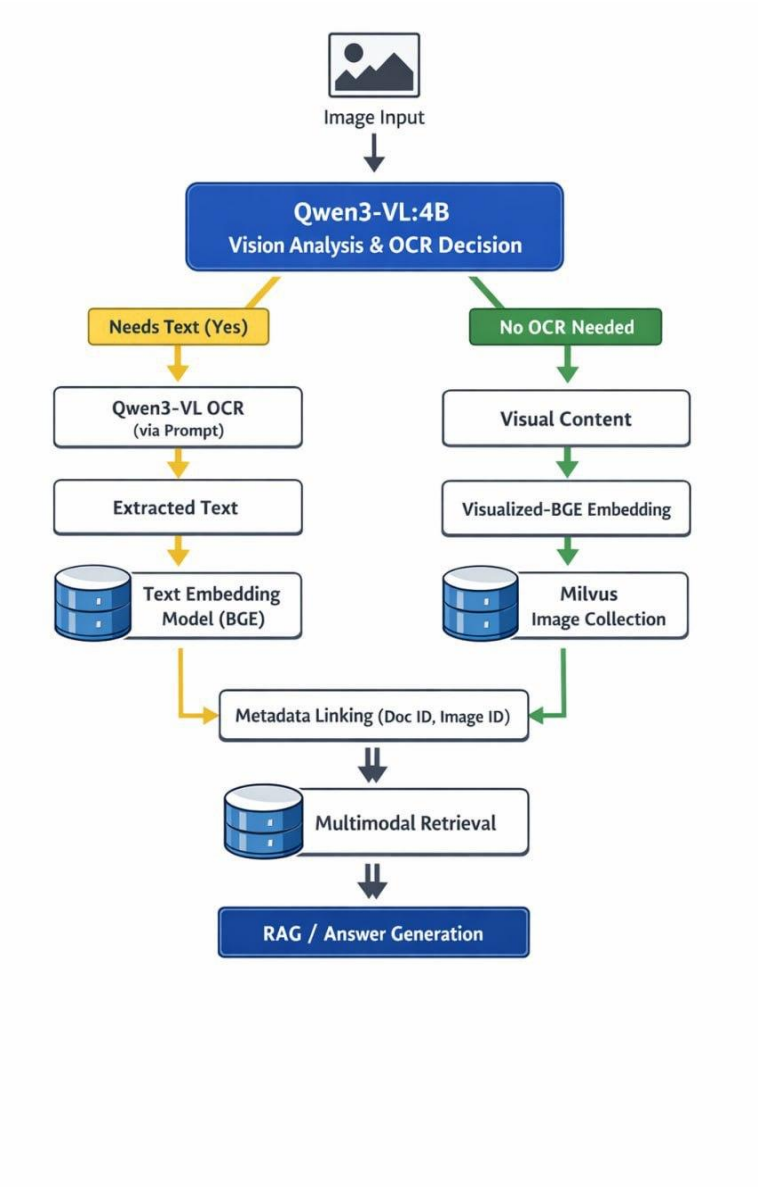
5.4.3 تحويل المحتوى إلى Embeddings

يتم تحويل:

- النص المستخرج (إن وجد) إلى Embeddings نصية.
- المحتوى البصري إلى Embeddings صورية.

ثم تُخزّن هذه التمثيلات في قاعدة البيانات المتجهية مع ربطها ببيانات وصفية مناسبة.

رسم توضيحي 13: مخطط يوضّح مسار معالجة الصورة من الإدخال حتى التخزين.



5.5تنفيذ الفهرسة وقاعدة البيانات المتجهية

تعتمد عملية الاسترجاع في النظام على البحث الدلالي باستخدام قاعدة بيانات.Milvus

5.5.1تخزين البيانات

يتم تخزين Embeddings النصوص والصور داخل Milvus مع ربطها ببيانات وصفية مثل:

- نوع المحتوى (نصي / صوري).
- معرف السؤال أو الصورة.
- مصدر البيانات.
-

5.5.2البحث والاسترجاع

عند تنفيذ عملية بحث، تتم مقارنة Embedding المدخل مع Embeddings المخزنة باستخدام مقاييس التشابه(Cosine Similarity) ، ويتم استرجاع أفضل النتائج.(Top-k)

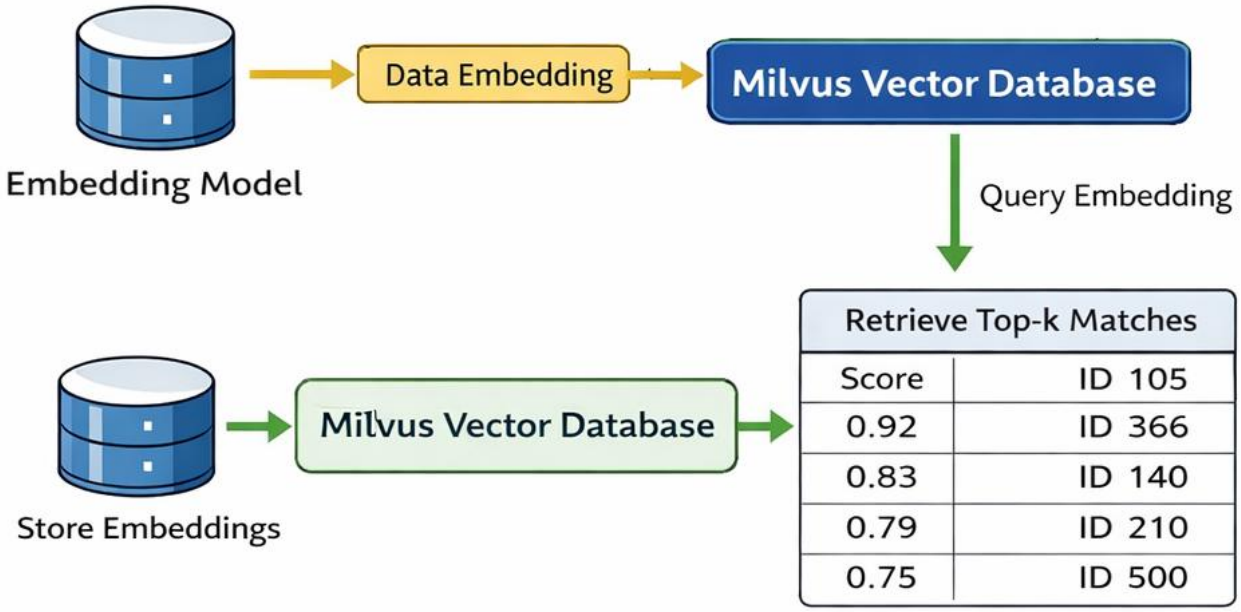


Figure 5.1: Diagram showing the process of storing embeddings in Milvus

$$similiarity = \frac{\cos\theta) = \vec{A} \cdot \vec{B}}{\sqrt{\sum_i^A B_i}\sqrt{\sum_i^i Bi}} = \cos\theta$$

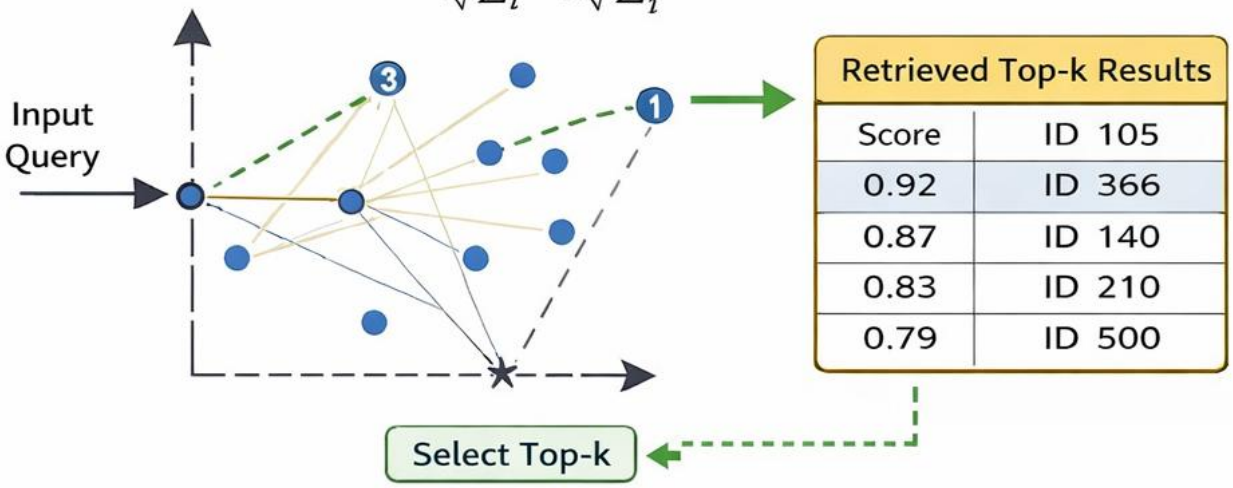


Figure 5.2: Diagram explaining cosine similarity and top-k selection.

5.6تنفيذ نظام الاسترجاع المعزز بالتوليد(RAG)

بعد استرجاع المحتوى الأكثر صلة، يتم تمريره إلى النموذج اللغوي الكبير ضمن إطار **RAG**. تقوم هذه المرحلة بـ:

- دمج السياق المسترجع مع السؤال.
- توليد إجابة دقيقة ومبنية على البيانات الفعلية.
- تقليل احتمالية الهلوسة في الإجابات.

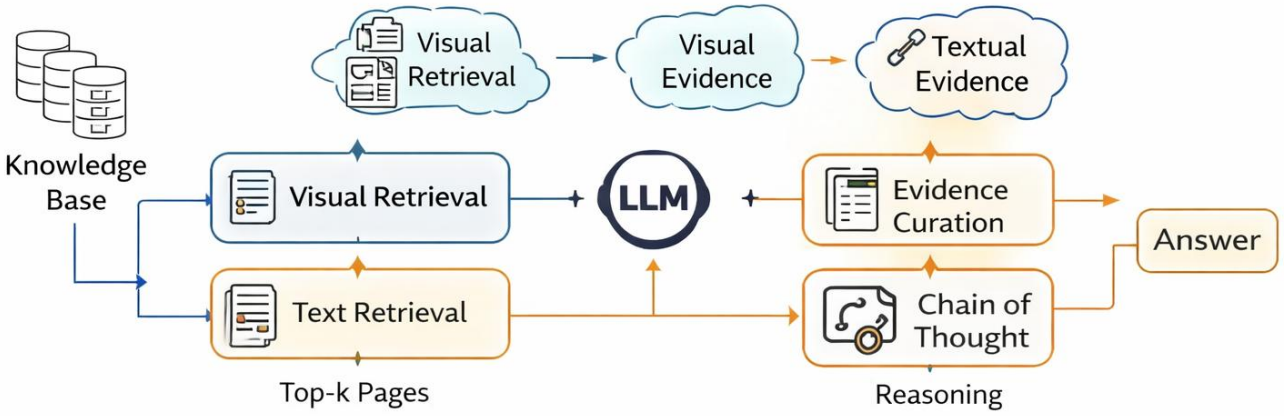


Figure 5.3: Diagram showing the integration of retrieval and generation (RAG Pipeline).

رسم توضيحي ١٥: مخطط يوضّح تكامل الاسترجاع مع التوليد (RAG Pipeline).

5.7تنفيذ واجهة المستخدم

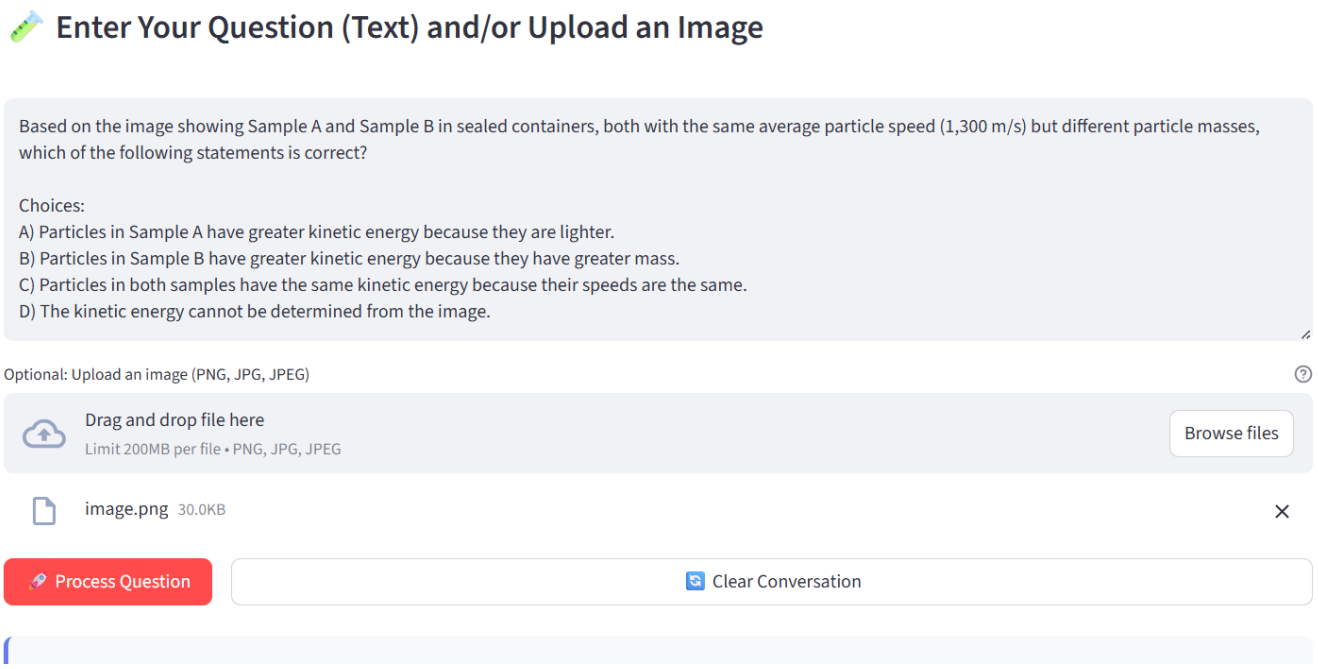
تم تصميم واجهة المستخدم لتكون:

- سهلة الاستخدام.
- داعمة للمدخلات النصية والصورية.
- واضحة في عرض النتائج.

تشمل الواجهة:

- حقل إدخال السؤال.
- زر تنفيذ البحث.
- عرض الإجابة النهائية.
- خيار البحث عبر الويب في حال عدم توفر إجابة واثقة.
- Workflow
- Debug
- JSON statment
- عرض مخطط حالة النظام.(State Graph Visualization)
-

Deplo



5.8التحديات التقنية والحلول

واجه تنفيذ النظام عدة تحديات، من أبرزها:

- إدارة الحالة في LangGraph ومنع التعارض بين الحالات.
- تحقيق توازن بين دقة الاسترجاع وسرعة الأداء.
- التعامل مع اختلاف طبيعة البيانات النصية والصورية.

وقد تم التغلب على هذه التحديات من خلال:

- تنظيم تدفق الحالات داخل الرسم البياني.
- تحسين منطق الاسترجاع.
- الفصل الواضح بين مسارات المعالجة النصية والصورية.

5.9ملخص الفصل

قدّم هذا الفصل شرحًا تفصيليًا لعملية تنفيذ نظام البحث متعدد الوسائط للمحتوى التعليمي، موضحًا بيئة التطوير، الأدوات المستخدمة، وآلية تنفيذ كل مرحلة من مراحل النظام. كما تم استعراض كيفية معالجة المدخلات النصية والصورية، وتنفيذ الفهرسة والاسترجاع الدلالي، ودمج النتائج ضمن إطار RAG لتوليد إجابات دقيقة وموثوقة. يمهّد هذا الفصل للانتقال إلى الفصل السادس، الذي يركّز على تقييم أداء النظام وقياس فعاليته.

الفصل السادس: التقييم

6.1 مقدمة الفصل

يهدف هذا الفصل إلى تقييم أداء النظام المقترح المبني على معمارية Retrieval-Augmented Generation (RAG)، وذلك من خلال تحليل كفاءة مرحلتي الاسترجاع (Retrieval) والتوليد (Generation) بشكل مستقل.
تم اعتماد مقاييس تقييم معيارية مستخدمة على نطاق واسع في الأبحاث الحديثة، لضمان موضوعية النتائج وإمكانية مقارنتها مع أنظمة مشابهة.

6.2 أهداف التقييم

يركز التقييم في هذا المشروع على تحقيق الأهداف التالية:

- قياس دقة نظام الاسترجاع في جلب المقاطع النصية ذات الصلة بالسؤال.
- تقييم جودة إجابات التوليد مقارنة بالإجابات المرجعية.
- تحليل استقرار النظام عند العمل على عدد كبير من الأسئلة.
- التأكد من أن النظام يعمل بشكل موثوق في بيئة واقعية.

6.3 بيانات التقييم

تم استخدام **1000سؤال** من مجموعة بيانات **ScienceQA**، حيث تم اختيار **500 سؤال** للتقييم النهائي لمرحلة التوليد.
تعتمد البيانات على أسئلة اختيار من متعدد، مع وجود إجابة مرجعية لكل سؤال.

6.4 منهجية التقييم

6.4.1 تقييم مرحلة الاسترجاع

تم تقييم مرحلة الاسترجاع بشكل مستقل عن مرحلة التوليد، وذلك باستخدام مقاييس **Recall@5** و**Precision@1**.
تم إنشاء Ground Truth للاسترجاع **بشكل آلي** عبر إجراء بحث متجهي (Vector Search) لكل سؤال وتخزين أعلى المقاطع تشابهًا كمرجع.
هذا النوع من التقييم يُستخدم لقياس **استقرار وكفاءة نظام الاسترجاع**، وليس الفهم الدلالي العميق للمحتوى.

6.4.2 تقييم مرحلة التوليد

تم تقييم مرحلة التوليد باستخدام أسلوب LLM-as-a-Judge، حيث يقوم نموذج لغوي كبير بدور الحكم على صحة الإجابة المولدة مقارنة بالإجابة المرجعية.تم إلزام النموذج المولد بإنتاج إجابة مقتصرة على أحد الخيارات المتاحة فقط، لتفادي الإجابات الخارجة عن نطاق السؤال
النموذج المستخدم للتوليد هو: gpt-oss:120b-cloud عبر منصة Ollama.

6.5 مقاييس التقييم المستخدمة

6.5.1 Recall@5 : يقيس هذا المقياس قدرة النظام على استرجاع المقاطع الصحيحة ضمن أول ٥ نتائج مسترجعة.
6.5.2 Precision@1: يقيس دقة أعلى نتيجة مسترجعة فقط، وهو مؤشر مهم لجودة الترتيب.
6.5.3 Generation Accuracy : تم حساب دقة التوليد كنسبة عدد الإجابات الصحيحة إلى العدد الكلي للأسئلة التي تم تقييمها.

6.6 إعدادات التجربة

- عدد الأسئلة المستخدمة في تقييم التوليد: **500 سؤال**
- النموذج المستخدم : **gpt-oss:120b-cloud**
- طريقة التقييم: **LLM-as-a-Judge**

6.7 نتائج التقييم وتحليلها

6.7.1 نتائج الاسترجاع

- **Average Recall@5: 0.6008**
- **Average Precision@1: 0.6660**

6.7.2 نتائج التوليد :

تم تقييم مرحلة التوليد على **500سؤال**، وكانت النتائج:

- عدد الإجابات الصحيحة : ٣٨١
- نسبة النجاح : ٧٦,٢٪

تعكس هذه النتيجة قدرة النموذج على توليد إجابات صحيحة اعتمادًا على السياق المسترجع.

6.8 مناقشة النتائج

تُظهر النتائج أن فصل مرحلتي الاسترجاع والتوليد ساهم في تحسين موثوقية التقييم رغم أن Ground Truth للاسترجاع تم إنشاؤه آليًا، إلا أن هذا الأسلوب شائع في أنظمة RAG الحديثة ويُستخدم لقياس استقرار الاسترجاع وليس الفهم الحقيقي للمعلومة.

6.9حدود التقييم

يعتمد هذا العمل على منهجية تقييم آلية شائعة في أنظمة الاسترجاع المعزَّز بالتوليد(RAG) ، حيث يتم إنشاء Ground Truth للاسترجاع باستخدام البحث المتجهي بدلًا من التوسيم البشري المباشر. يُستخدم هذا النهج على نطاق واسع في الأدبيات البحثية لكونه **قابلًا للتوسع وقابلًا لإعادة الإنتاج**، إلا أنه يركّز بالدرجة الأولى على قياس التشابه الدلالي التمثيلي، وليس على التحقق اليدوي من الصلة المعرفية الدقيقة بين السؤال والمعلومة المسترجعة.
كما تم تقييم مرحلة التوليد باستخدام نموذج لغوي كبير كحكم(LLM-as-a-Judge) ، وهو أسلوب معتمد حديثًا في تقييم نماذج اللغة عندما

يتعذر إجراء تقييم بشري شامل. وعلى الرغم من فعاليته العملية، فإن هذا الأسلوب قد يُظهر حساسية لصياغة الإجابة أو للقيود الشكلية المفروضة على مخرجات النموذج، خاصة في سياق الأسئلة متعددة الخيارات.
تتوافق هذه القيود مع ما هو مذكور في الأدبيات الحديثة، حيث تشير عدة دراسات إلى أن التقييم الآلي لأنظمة RAG يقيس **الأداء النسبي والاستقرار الوظيفي للنظام** أكثر مما يقيس الفهم الدلالي العميق للمعلومة، مما يجعل هذه النتائج مناسبة للمقارنة المنهجية ضمن نفس الإعداد التجريبي.

6.10خلاصة الفصل

في هذا الفصل، تم تقييم النظام المقترح باستخدام مقاييس معيارية وبمنهجية واضحة أظهرت النتائج أن النظام يحقق أداءً جيدًا في مرحلتي الاسترجاع والتوليد، مما يؤكد فعاليته كنظام RAG عملي وقابل للتطبيق.

الفصل السابع: الخاتمة والآفاق المستقبلية

7.1مقدمة الفصل

يقدّم هذا الفصل نظرة شاملة على هيكلية التقرير وتنظيم فصوله، مع توضيح الهدف من كل فصل ودوره في توثيق مراحل تطوير نظام البحث متعدد الوسائط للمحتوى التعليمي. يهدف هذا الفصل إلى إرشاد القارئ خلال التسلسل المنطقي للتقرير، وبيان كيفية ترابط الفصول المختلفة لتكوين دراسة متكاملة تبدأ من تحديد المشكلة وتنتهي بتقييم النظام المقترح واستخلاص النتائج.
يساعد هذا العرض العام على تسهيل فهم محتوى التقرير، ويوضح كيف تساهم كل مرحلة في بناء النظام الذكي القائم على تقنيات الذكاء الاصطناعي والاسترجاع المعزز بالتوليد.(RAG)

7.2هيكلية التقرير وأهداف الفصول

يتكون هذا التقرير من سبعة فصول رئيسية، تم تنظيمها بعناية لضمان عرض منهجي ومتدرّج لمختلف جوانب المشروع، كما يلي:

الفصل الأول: المقدمة(Introduction)

يركّز هذا الفصل على تقديم فكرة المشروع بشكل عام، مع تسليط الضوء على أهمية البحث متعدد الوسائط في المجال التعليمي. كما يتناول مشكلة البحث التقليدي القائم على النصوص فقط، ويعرض أهداف المشروع ومساهماته العلمية، إضافة إلى توضيح بنية التقرير وتنظيمه العام.

الفصل الثاني: الدراسة المرجعية والأعمال ذات الصلة

يقدّم هذا الفصل الخلفية النظرية اللازمة لفهم المشروع، حيث يستعرض المفاهيم الأساسية المرتبطة بالذكاء الاصطناعي متعدد الوسائط، والبحث الدلالي، والتضمينات المتجهية، والنماذج اللغوية الكبيرة، والاسترجاع المعزز بالتوليد .(RAG) كما يتضمن مراجعة للأبحاث والدراسات السابقة المتعلقة بأنظمة البحث متعددة الوسائط، واستخدام الذكاء الاصطناعي في استرجاع المحتوى التعليمي، مع تحليل نقاط القوة والقصور في هذه الأعمال وربطها بالمشروع المقترح.

الفصل الثالث: مجموعة البيانات وتحضيرها

يستعرض هذا الفصل مجموعة البيانات المستخدمة في المشروع(ScienceQA) ، من حيث بنيتها، ومكوناتها، وخصائصها الإحصائية. كما يتناول تحليل البيانات الاستكشافي (EDA) الذي تم إجراؤه لفهم توزيع الأسئلة، وأنواع الوسائط، والمستويات التعليمية. إضافة إلى ذلك، يوضّح الفصل خطوات تمثيل البيانات وتحضيرها للاستخدام في نظام البحث متعدد الوسائط، مع تبرير ملاءمة مجموعة البيانات للنظام المقترح.

الفصل الرابع: المنهجية المقترحة

يركّز هذا الفصل على شرح المنهجية المعتمدة في تصميم النظام متعدد الوسائط، موضحًا البنية العامة للنظام وآلية تدفق البيانات. يتناول الفصل كيفية معالجة المدخلات النصية والصورية، وتحويلها إلى تمثيلات متجهية، وتنفيذ البحث الدلالي متعدد الوسائط باستخدام قاعدة بيانات متجهية. كما يشرح دور إطار الاسترجاع المعزز بالتوليد(RAG) ، وآلية دمج الاسترجاع مع التوليد، إضافة إلى توضيح استخدام LangChain و LangGraphفي تنظيم تدفق العمليات واتخاذ القرار داخل النظام.

الفصل الخامس: التنفيذ العملي

يستعرض هذا الفصل الجوانب العملية لتطوير النظام، بما في ذلك الأدوات والتقنيات المستخدمة، وتنفيذ وحدات معالجة المدخلات، والتضمين المتجهي، والاسترجاع باستخدامMilvus ، وتوليد الإجابات باستخدام النماذج اللغوية الكبيرة. كما يتضمن شرحًا لتكامل مكونات النظام، مع توضيح الواجهات البرمجية ومسار تنفيذ العمليات، وإدراج أمثلة توضيحية من الكود وواجهات النظام.

الفصل السادس: التقييم

يتناول هذا الفصل منهجية تقييم النظام المقترح، حيث تم استخدام إطار RAGAS لتقييم جودة أنظمة الاسترجاع المعزز بالتوليد، إضافة إلى أسلوب التقييم القائم على النموذج اللغوي (LLM-as-a-Judge) باستخدام نموذج Mistral. يعرض الفصل مقاييس التقييم المعتمدة، مثل مدى الصلة، والدقة، والاتساق، وجودة الإجابة، إلى جانب مناقشة نتائج التقييم وتحليل أداء النظام.

الفصل السابع: نظرة عامة على التقرير

يقدّم هذا الفصل عرضًا تجميعيًا لبنية التقرير، موضحًا دور كل فصل في توثيق مراحل المشروع المختلفة. ويساعد القارئ على تكوين صورة شاملة عن تسلسل العمل، بدءًا من الخلفية النظرية وتحليل البيانات، مرورًا بالتصميم والتنفيذ، وصولًا إلى التقييم النهائي للنظام.

7.3آفاق العمل المستقبلي(Future Work)

على الرغم من النتائج الإيجابية التي حققها النظام المقترح، إلا أن هناك العديد من الجوانب التي يمكن تطويرها وتحسينها في الأعمال المستقبلية، سواء على مستوى توسيع قدرات النظام أو تحسين أدائه ودقته.
من حيث دعم الوسائط، يمكن توسيع النظام ليشمل أنماط إدخال إضافية مثل الصوت والفيديو، مما يتيح التعامل مع محتوى المحاضرات المسجلة، والتجارب التعليمية المصورة، وشرح المفاهيم العلمية المعقدة باستخدام وسائط غنية ومتنوعة. كما يمكن دمج تقنيات التعرف على الكلام (Speech Recognition) وتحليل الفيديو لدعم البحث داخل المحتوى السمعي والبصري.

على مستوى النماذج المستخدمة، يمكن استبدال أو تحسين نماذج التضمين والنماذج اللغوية بنماذج أحدث وأكثر كفاءة، أو نماذج مدربة خصيصًا على المحتوى التعليمي، مما قد يسهم في تحسين دقة الاسترجاع وجودة التوليد. كما يمكن دراسة تأثير استخدام نماذج متعددة (Ensemble Models) لتحسين أداء النظام في الحالات المعقدة.

فيما يتعلق بعملية التقييم، يمكن توسيعها لتشمل تقييمًا بشريًا (Human Evaluation) من قبل طلاب ومعلمين، بهدف قياس مدى فائدة النظام في بيئة تعليمية حقيقية، وتقييم جودة الإجابات من حيث الوضوح، والدقة، وسهولة الفهم. كما يمكن إجراء تجارب مقارنة بين النظام المقترح وأنظمة بحث تعليمية أخرى لقياس مستوى التحسن الذي يقدمه.

إضافة إلى ذلك، يمكن تطوير النظام ليكون أكثر تفاعلية وتخصيصًا، من خلال مراعاة مستوى الطالب، واهتماماته، والمجال الدراسي، مما يتيح تقديم إجابات وشروحات تتناسب مع احتياجات المستخدم التعليمية. كما يمكن دعم لغات متعددة لتوسيع نطاق استخدام النظام ليشمل شرائح أكبر من المتعلمين.

أخيرًا، يمكن دمج النظام ضمن منصات تعليم إلكتروني فعلية مثل أنظمة إدارة التعلم (Learning Management Systems – LMS)، ودراسة أثره على تحسين نواتج التعلم، وزيادة تفاعل الطلاب، وتعزيز الفهم العميق للمفاهيم التعليمية. كما يمكن أن يشكل هذا المشروع أساسًا لأبحاث مستقبلية تهدف إلى تحسين أنظمة الاسترجاع المعزز بالتوليد

في السياق التعليمي، وتقليل التحيز والهلوسة في النماذج اللغوية، وبناء أنظمة تعليمية ذكية أكثر موثوقية واستدامة.

7.4 ملخص الفصل

يُعد هذا الفصل دليلًا إرشاديًا لفهم تنظيم التقرير وكيفية ترابط فصوله المختلفة في عرض مشروع البحث متعدد الوسائط للمحتوى التعليمي. يضمن هذا التنظيم تقديم دراسة منهجية ومتكاملة، تبدأ بتحديد المشكلة وأسسها النظرية، ثم تنتقل إلى تصميم النظام وتنفيذه وتقييمه، مما يوفّر توثيقًا علميًا واضحًا وشاملًا للمشروع.

(References)المراجع

• Lu, P. et al. (2023). *ScienceQA: A Multimodal Question Answering Dataset*. Available at: https://arxiv.org/abs/2209.09513 .

• Lewis, P. et al. (2020). *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*. Available at: https://arxiv.org/abs/2005.11401.

• Reimers, N. and Gurevych, I. (2019). *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks*. Available at: https://arxiv.org/abs/1908.10084 .

• Milvus Documentation. *Vector Database for AI Applications*. Available at: https://milvus.io/docs .

• LangChain Documentation. *Building LLM-powered applications*. Available at: https://python.langchain.com/docs/ .

• LangGraph Documentation. *Stateful LLM workflows*. Available at: https://langgraph.readthedocs.io/ .

• Es, S. et al. (2023). *RAGAS: Automated Evaluation of Retrieval-Augmented Generation*. Available at: https://arxiv.org/abs/2309.15217 .