الجمهورية اليمنية
جامعة العلوم والتكنولوجيا
كلية الحاسبات وتقنية المعلومات
قسم تقنية المعلومات

Republic of Yemen
University of Science and Technology
College of Computing
Department of Information Technology

# البيانات وتحليل الاصطناعي الذكاء على معتمد مؤسسي ذكي مساعد

Smart Institutional Assistant Based on Artificial Intelligence and Data Analytics

**إعداد الطلاب:**
وليد محسن يحيى العنسي
محمد رشيد أحمد الشريف
أحمد علي محمد المتوكل
عبدالكريم قاسم عبده العبدلي
محمد نجيب العواضي

**تحت إشراف:** د. نبيل المخلافي

أُنجز هذا المشروع استكمالاً لمتطلبات التخرج في كلية الحاسبات وتقنية المعلومات

**2027 / 2026**

# تصريح الطلاب

نُقِرّ نحن الطلاب الموقعين أدناه بأن هذا المشروع هو نتاج جهدنا الشخصي، وقد تم بإشراف وتوجيه من مشرفنا الأكاديمي. يعكس هذا العمل ما توصلنا إليه من خلال دراستنا وبحثنا. جميع النصوص والنتائج المأخوذة من مصادر أخرى قد تم توثيقها والإشارة إليها بشكل كامل.

| اسم الطالب | التوقيع |
|---|---|
| وليد محسن يحيى العنسي | .... |
| محمد رشيد أحمد الشريف | .... |
| أحمد علي محمد حسن المتوكل | .... |
| عبد الكريم قاسم عبده مسعود العبدلي | .... |
| محمد نجيب العواضي | .... |

**المشرف:** د. نبيل المخلافي

**التاريخ:** ..... / ..... / 2026 / 2027

# إقرار المشرف

نشهد بأن إعداد هذا المشروع بعنوان:

**«مساعد ذكي مؤسسي معتمد على الذكاء الاصطناعي وتحليل البيانات»**

من إعداد الطلاب المذكورة أسماؤهم أعلاه، قد تم تحت إشرافي، استكمالاً لمتطلبات الحصول على درجة البكالوريوس في تقنية المعلومات بقسم تقنية المعلومات.

**اسم المشرف:** د. نبيل المخلافي

**التوقيع:** .....

**التاريخ:** ..... / ..... / 2026 / 2027

# تأكيد

**عنوان المشروع:** مساعد ذكي مؤسسي معتمد على الذكاء الاصطناعي وتحليل البيانات

**أعد وقدّم هذا المشروع كل من:**

| الرقم الجامعي | اسم الطالب |
|---|---|
| 202310100941 | وليد محسن يحيى العنسي |
| 202310100869 | محمد رشيد أحمد الشريف |
| 202310101668 | أحمد علي محمد حسن المتوكل |
| 202310101670 | عبد الكريم قاسم عبده مسعود العبدلي |
| (يُملأ لاحقاً) | محمد نجيب العواضي |

تم اعتماد هذا المشروع من حيث النطاق والجودة والعرض التقديمي كمشروع تخرج.
قدم هذا التقرير استكمالاً لمتطلبات الحصول على درجة البكالوريوس في تقنية المعلومات.

كلية الحاسبات وتقنية المعلومات
جامعة العلوم والتكنولوجيا، اليمن

**المشرف:** د. نبيل المخلافي

**التوقيع:** .....

**التاريخ:** ..... / ..... / 2026 / 2027

# الإهداء

بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

الحمد لله رب العالمين، والصلاة والسلام على أشرف الأنبياء والمرسلين، نبينا محمد وعلى آله وصحبه أجمعين. نحمد الله سبحانه وتعالى الذي منَّ علينا بالقوة والقدرة والجهد لإنجاز هذا المشروع في وقته المحدد، فبدون فضله ورحمته وتوفيقه لما كان لنا أن نُتم هذا العمل.

إلى مَن بذل الله بهما نور الوجود، ومنحاني معنى الحياة .. والديّ الكريمين. إلى مَن آمنوا بي حين تعثرت، وشجعوني حين ترددت .. إخوتي وأسرتي الكريمة. إلى كل معلم علمنا حرفًا أضاء لنا دربًا في طريق العلم والمعرفة. إلى زملاء الدرب الذين شاركونا لحظات الجد والتعب والفرح بإنجاز هذا العمل.

نهدي هذا الجهد المتواضع، ثمرة سنوات من الدراسة والاجتهاد.

# الشكر والتقدير

نتقدم بجزيل الشكر والامتنان إلى مشرفنا الأكاديمي، **الدكتور نبيل المخلافي**، على أفكاره القيمة وتوجيهاته المستمرة طوال مسيرة إنجاز مشروع التخرج هذا، والتي كان لها الأثر الكبير في إخراجه بالشكل المطلوب.

كما نُقدِّر عالياً أعضاء لجنة المناقشة على ملاحظاتهم القيّمة ونقدهم البنّاء خلال العروض التقديمية، والتي ساعدتنا في تجاوز نقاط الضعف وتحسين الجوانب التي كانت بحاجة إلى تعديل.

وجزيل الشكر والعرفان إلى آبائنا وأمهاتنا الكرام، الذين كانوا خير سند وداعم لنا، وقدَّموا لنا التشجيع المعنوي الذي مكَّننا من إنجاز هذا المشروع بنجاح. كما نخص بالشكر زملاءنا في فريق العمل، الذين كانوا عوناً لنا وسنداً في مراحل التطوير المختلفة.

كل الجهود التي أبقتنا على الطريق الصحيح لن تُنسى، وخالص امتناننا موصول لجميع.

# الملخص

مع تزايد حجم البيانات والمستندات في المؤسسات الحديثة، أصبح من الضروري تطوير حلول ذكية تمكن المستخدمين من الوصول إلى المعلومات واتخاذ القرارات بسرعة وكفاءة. تعتمد كثير من المؤسسات اليوم على وسائل تقليدية مثل الاستعلامات المباشرة لقواعد البيانات من قبل متخصصين، والبحث اليدوي في المستندات غير المفهرسة، وإعداد التقارير بشكل يدوي، وهي وسائل تفتقر إلى السرعة والشمولية وسهولة الاستخدام. ولمعالجة هذه الإشكالات، يقدم هذا المشروع نظام «المساعد الذكي المؤسسي» (بيان)، وهو نظام محادثة ذكي متعدد الوكلاء (Multi-Agent System) يمكّن المستخدمين من التفاعل مع بيانات ومستندات المؤسسة باللغة العربية الطبيعية.

يعتمد النظام على معمارية متقدمة تضم خمسة وكلاء ذكاء اصطناعي متخصصين: وكيل المشرف لتوجيه الاستعلامات، ووكيل SQL لتحويل اللغة الطبيعية إلى استعلامات قواعد بيانات، ووكيل RAG للاسترجاع المعزز بالمستندات، ووكيل المحلل لاستخراج الرؤى والاتجاهات، ووكيل التقارير لتنسيق المخرجات. وتعد طبقة الـViews الذكية للبيانات من أبرز ابتكارات النظام، حيث ترفع دقة ترجمة اللغة الطبيعية إلى SQL من ~70% إلى ~90%+. يوفر النظام واجهة محادثة عربية (RTL) تدعم الفصحى والعامية، إلى جانب لوحة تحكم للإدارة تتيح تبديل النماذج اللغوية بين المحلية (Ollama) والسحابية، ومراقبة الأداء، وإدارة الصلاحيات والأمان المتدرج (Public، Masked، Restricted، Blacklist).

يستهدف النظام خمس فئات من المستخدمين: المديرين الذين يحتاجون إلى رؤى استراتيجية سريعة، والمحللين الذين يجرون استعلامات معقدة، والموظفين العاديين الباحثين عن معلومات محددة، ومسؤولي النظام المشرفين على التشغيل والأمان، ومشغلي التهيئة المسؤولين عن الإعداد الأولي. وتشمل الوظائف الرئيسية: المحادثة العربية الذكية، تحويل اللغة الطبيعية إلى SQL، الاسترجاع المعزز بالمستندات، التحليل الذكي للبيانات، توليد التقارير، الإعداد المؤتمت خلال 8 ساعات، والأمان المتدرج، وتتبع الوكلاء لعرض مسار التنفيذ.

ويركز المشروع على تقديم حل تقني عملي مصمم ليتوافق مع البيئة المؤسسية المحلية واحتياجات المؤسسات في اليمن والدول العربية، بهدف تمكين الوصول للبيانات للجميع، رفع كفاءة اتخاذ القرار، وتعزيز الشفافية والمساءلة، وتقليل الجهد اليدوي والتكلفة. ويشكل نظام «بيان» خطوة مهمة نحو التحول الرقمي في قطاع إدارة المؤسسات، بما يسهم في رفع مستوى رضا المستخدمين، وتحسين الكفاءة التشغيلية، وتعزيز القدرة على الابتكار من خلال توفير رؤى عميقة قائمة على البيانات.

**الكلمات المفتاحية:** الذكاء الاصطناعي، تحويل اللغة الطبيعية إلى SQL، الاسترجاع المعزز بالتوليد (RAG)، الأنظمة متعددة الوكلاء، معالجة اللغة العربية، LangGraph، البيانات المؤسسية.

# Abstract

With the increasing volume of data and documents in modern institutions, it has become essential to develop intelligent solutions that enable users to access information and make decisions quickly and efficiently. Many institutions still rely on traditional methods such as direct database queries by specialists, manual search through unindexed documents, and manual report preparation — methods that lack speed, comprehensiveness, and usability. To address these issues, this project presents the ``Smart Institutional Assistant'' (Bayan), an intelligent multi-agent conversational system that enables users to interact with the institution's data and documents in natural Arabic language.

The system relies on an advanced architecture comprising five specialized AI agents: a Supervisor agent for routing queries, an SQL agent for translating natural language into database queries, a RAG agent for retrieval-augmented generation over documents, an Analyst agent for extracting insights and trends, and a Reporter agent for formatting outputs. The intelligent `Views` layer is one of the system's key innovations, raising NL2SQL accuracy from $\sim$70% to $\sim$90%+. The system provides an Arabic RTL conversational interface supporting both Modern Standard Arabic and dialects, alongside an administrative dashboard that allows switching between local (Ollama) and cloud language models, monitoring performance, and managing tiered permissions (`Public`, `Masked`, `Restricted`, `Blacklist`).

The system targets five user roles: executives needing fast strategic insights, analysts running complex queries, regular employees searching for specific information, system administrators overseeing operations and security, and onboarding operators responsible for initial setup. Key features include: intelligent Arabic conversation, NL2SQL, retrieval-augmented generation, smart data analytics, report generation, automated onboarding within 8 hours, tiered security, and agent-trace transparency.

**Keywords:** Artificial Intelligence, NL2SQL, RAG, Multi-Agent Systems, Arabic NLP, LangGraph, Enterprise Data.

# قائمة الاختصارات

| الاختصار | التعريف |
|---|---|
| API | واجهة برمجة التطبيقات التي تتيح التواصل بين الأنظمة المختلفة. |
| BIRD | معيار قياسي لتقييم أنظمة NL2SQL على قواعد بيانات واقعية. |
| CSV | صيغة ملف نصي لتخزين البيانات الجدولية مفصولة بفواصل. |
| ERD | مخطط العلاقات بين الكيانات في قاعدة البيانات (Entity-Relationship Diagram). |
| GPU | معالج الرسوميات المستخدم لتسريع معالجة النماذج. |
| LangGraph | مكتبة لبناء وتنسيق وكلاء الذكاء الاصطناعي المتعددين. |
| LLM | النموذج اللغوي الكبير (Large Language Model). |
| NL2SQL | تحويل اللغة الطبيعية إلى استعلامات SQL (Natural Language to SQL). |
| Odoo | نظام ERP مفتوح المصدر، يُستخدم كبيئة محاكاة جامعية. |
| Onboarding | عملية الإعداد والتهيئة الأولية للنظام. |
| pgvector | امتداد PostgreSQL لدعم تخزين والبحث في المتجهات. |
| Qdrant | قاعدة بيانات شعاعية مفتوحة المصدر. |
| RAG | تقنية الاسترجاع المعزز بالتوليد (البحث في المستندات ثم توليد إجابة اعتمادًا على نتائج البحث). |
| RTL | الكتابة من اليمين إلى اليسار — دعم اللغة العربية. |
| SaaS | البرمجية كخدمة تُقدَّم عبر الإنترنت دون الحاجة لتثبيتها محليًا. |
| Schema Linking | تقنية تصفية الجداول والأعمدة ذات الصلة قبل توليد SQL. |
| Spider | معيار قياسي لتقييم أنظمة NL2SQL. |
| SQL | لغة الاستعلام البنيوية المستخدمة للتعامل مع قواعد البيانات. |
| UML | لغة النمذجة الموحدة (Unified Modeling Language). |
| Views | طبقة وسيطة مبسطة فوق قاعدة البيانات لرفع دقة NL2SQL. |

# المحتويات


تصريح الطلاب i

إقرار المشرف ii

تأكيد iii

الإهداء iv

الشكر والتقدير v

الملخص vi

Abstract vii

قائمة الاختصارات viii

1 المقدمة 1

1.1 الخلفية (Background) . . . . . 2

2.1 أهداف المشروع (Objectives) . . . . . 2

3.1 مشكلة المشروع (Problem Statement) . . . . . 3

4.1 تعريف المشروع (System Definition) . . . . . 4

5.1 معايير القبول (Acceptance Criteria) . . . . . 5

6.1 نطاق المشروع (Project Scope) . . . . . 5

1.6.1 النطاق الجغرافي (Geographical Scope) . . . . . 5

2.6.1 النطاق الوظيفي (Functional Scope) . . . . . 6

3.6.1 أصحاب المصلحة (Stakeholders) . . . . . 6

7.1 قيود المشروع (Limitations) . . . . . 6

8.1 المنهجية المستخدمة في مشروع بيان (منهجية أجايل) . . . . . 6

1.8.1 مرحلة التخطيط (Planning) . . . . . 7

2.8.1 مرحلة التصميم (Design) . . . . . 7

3.8.1 مرحلة التطوير (Development) . . . . . 7

4.8.1 مرحلة الاختبار (Testing) . . . . . 7

5.8.1 مرحلة الإطلاق (Deployment) . . . . . 8

6.8.1 مرحلة المراجعة (Review) . . . . . 8
9.1 الجدول الزمني للمشروع (Project Schedule) . . . . . 8
10.1 متطلبات الأدوات (Required Tools) . . . . . 8
1.10.1 الأدوات المادية (Hardware Tools) . . . . . 9
2.10.1 الأدوات البرمجية (Software Tools) . . . . . 9
11.1 هيكل فصول المشروع (Structure of Project) . . . . . 10

**2 الخلفية النظرية ومراجعة الأدبيات 13**
1.2 مقدمة الفصل . . . . . 14
2.2 الخلفية النظرية . . . . . 14
1.2.2 النماذج اللغوية الكبيرة (Large Language Models - LLMs) . . . . . 14
2.2.2 تقنية تحويل اللغة الطبيعية إلى SQL (NL2SQL) . . . . . 15
3.2.2 تقنية الاسترجاع المعزز بالتوليد (RAG) . . . . . 15
4.2.2 الأنظمة متعددة الوكلاء (Multi-Agent Systems) . . . . . 16
5.2.2 تحديات معالجة اللغة العربية (Arabic NLP) . . . . . 16
3.2 الأنظمة والدراسات السابقة . . . . . 17
1.3.2 الأنظمة في مجال تحويل اللغة الطبيعية إلى SQL . . . . . 17
2.3.2 الأنظمة في مجال الاسترجاع المعزز بالتوليد (RAG) . . . . . 18
3.3.2 الأنظمة في مجال الأنظمة متعددة الوكلاء . . . . . 18
4.3.2 المساعدات الذكية في التعليم العالي . . . . . 18
5.3.2 الدراسات العربية . . . . . 19
4.2 مقارنة الأنظمة السابقة . . . . . 19
5.2 الفجوة البحثية . . . . . 19
6.2 الخلاصة . . . . . 21

**المراجع 23**

# قائمة الجداول

2.1 الأدوات البرمجية المعتمدة في تطوير مشروع بيان. . . . . 10
1.1 الأدوات المادية المطلوبة للمشروع. . . . . 12
1.2 مقارنة شاملة بين الأنظمة السابقة ومشروع بيان. . . . . 20
2.2 ملخص الفجوة البحثية ومساهمة مشروع بيان في سدها. . . . . 21

# قائمة الأشكال

1.1 نموذج أجايل (Agile Model) المتبع في مشروع بيان. . . . . 7
2.1 مخطط جانت لمراحل المشروع – الجزء الأول. . . . . 8
3.1 مخطط جانت لمراحل المشروع – الجزء الثاني. . . . . 9
4.1 مخطط جانت لمراحل المشروع – الجزء الثالث. . . . . 9
5.1 مخطط جانت لمراحل المشروع – الجزء الرابع. . . . . 11
1.2 مخطط تدفق تقنية RAG في نظام بيان. . . . . 15
2.2 معمارية النظام متعدد الوكلاء القائم على LangGraph. . . . . 17

الفصل

# الأوّل

Chapter 1

## المقدمة

Introduction

## 1.1 الخلفية (Background)

شهد العالم خلال السنوات الأخيرة تحولًا كبيرًا في طريقة إدارة المؤسسات، حيث أصبح الاعتماد على الأنظمة الرقمية والذكاء الاصطناعي عاملًا رئيسيًا في تحسين جودة الخدمات المقدمة، وتعزيز كفاءة التشغيل، وتقليل الجهد اليدوي. ومع النمو المتسارع في حجم البيانات والمستندات المؤسسية وازدياد عدد المستخدمين والخدمات المرتبطة بها، أصبحت إدارة المعلومات أكثر تعقيدًا من أي وقت مضى، ولم تعد الأساليب التقليدية قادرة على مواكبة حجم العمل ومتطلبات المستخدمين.

في المؤسسات الحديثة، تتنوع المهام اليومية التي تتولى الإدارة الإشراف عليها، بدءًا من استعلامات البيانات وتحليلها، مرورًا بالبحث في المستندات واللوائح، ووصولاً إلى إعداد التقارير واتخاذ القرارات الاستراتيجية. هذه المهام تحتاج إلى نظام مركزي موثوق يضمن تنظيمها وتوثيقها، ويمنع ضياع المعلومات أو تداخلها، خاصة مع اعتماد المستخدمين المتزايد على الخدمات السريعة والموثوقة.

ورغم هذا التطور العالمي، ما زالت الكثير من المؤسسات في اليمن وفي العديد من الدول العربية تُدار بطرق تقليدية تُعرقل سير العمل وتؤثر على كفاءة اتخاذ القرار. ما يزال الاعتماد على الاستعلامات المباشرة لقواعد البيانات من قبل متخصصين شائعًا، وكذلك البحث اليدوي في المستندات غير المفهرسة، الأمر الذي يفتقر إلى السرعة والدقة ويؤدي إلى ضياع الوقت أو تكرار الأخطاء. كما أن إعداد التقارير والتحليلات تقتصر غالبًا على الجهد اليدوي دون وجود أي أدوات تحليلية تمكّن المستخدم من فهم اتجاهات البيانات أو اتخاذ قرارات مبنية على المعلومات.

ومع غياب نظام موحّد يجمع كل هذه الوظائف، تصبح عملية إدارة المعلومات مرهقة وغير فعّالة، ويعاني المستخدم من نقص الشفافية وضعف التواصل مع إدارة المؤسسة. كما تواجه الإدارة تحديات مستمرة في متابعة استعلامات المستخدمين، والتحقق من دقة المعلومات، وتحديد احتياجات التدريب، وإصدار تقارير تساعد في تحسين مستوى الخدمة.

من جهة أخرى، أدى التقدم في مجال الذكاء الاصطناعي وتعلم الآلة إلى فتح الباب أمام حلول رقمية جديدة في مجال إدارة المؤسسات، سواء عبر المساعدات الذكية أو أنظمة المحادثة المتقدمة، مما جعل عملية التحول الرقمي أكثر قابلية للتطبيق حتى في البيئات محدودة الموارد. هذه الحلول أصبحت اليوم ضرورة وليس خيارًا، خاصة في ظل احتياج المستخدمين إلى خدمات أسرع وأكثر دقة وأعلى جودة.

وبناءً على ذلك، نشأت الحاجة إلى تطوير نظام شامل لإدارة المؤسسات يربط بين المستخدمين والإدارة في إطار واحد، ويتيح تبادل المعلومات بشكل فوري، ويعزز الشفافية، ويوفّر سجلًا موحدًا لكل العمليات اليومية. ومن هنا جاء مشروع «المساعد الذكي المؤسسي (بيان)» ليقدّم نظامًا ذكيًا مصممًا خصيصًا لمعالجة التحديات التي تواجه المؤسسات في الواقع المحلي، مع مراعاة سهولة الاستخدام، وقابلية التوسع، والقدرة على تقديم تجربة استخدام احترافية تجمع بين البساطة والفعالية.

يقدم نظام «بيان» خطوة مهمة نحو التحول الرقمي في إدارة المؤسسات، حيث يجمع بين أحدث تقنيات الذكاء الاصطناعي وتطوير التطبيقات وبين فهم عميق لاحتياجات المستخدمين والإدارة، مما يجعله حلًا مبتكرًا وعمليًا يمكن تطبيقه مباشرة في المؤسسات، ويسهم في رفع مستوى الخدمات وتحسين كفاءة العمل واتخاذ القرارات.

## 2.1 أهداف المشروع (Objectives)

يهدف مشروع «بيان» إلى تطوير نظام مؤسسي متكامل مدعوم بتقنيات الذكاء الاصطناعي وتحليل البيانات، لتمكين المؤسسات من الوصول الذكي إلى بياناتها ومستنداتها باللغة العربية الطبيعية، وتحويل الأسئلة والمستندات إلى إجابات وتقارير دقيقة وفورية. ويتفرع من هذا الهدف الرئيسي مجموعة من الأهداف التفصيلية الآتية:

- **بناء مساعد محادثة ذكي باللغة العربية:** التفاعل مع المستخدمين باللغة العربية الطبيعية (الفصحى والعامية).
- **تمكين الاستعلام الذكي عن البيانات** (NL2SQL): تحويل الاستعلامات الصادرة باللغة الطبيعية إلى استعلامات قواعد بيانات (SQL) دون حاجة المستخدم لمعرفة تقنية مسبقة.
- **تطوير نظام استرجاع المستندات** (RAG): البحث في المستندات واللوائح المؤسسية وتوليد إجابات دقيقة وموثقة بمصادرها.
- **توفير أدوات التحليل الذكي:** تحليل البيانات واستخراج الاتجاهات والقيم الشاذة وتقديم رؤى إحصائية تدعم اتخاذ القرار.
- **توليد التقارير المنسقة:** إتاحة إمكانية إنتاج تقارير متعددة الأشكال (نصية، جدولية، ورسوم بيانية) قابلة للتصدير بصيغ مختلفة.
- **ضمان الشفافية والرقابة:** توفير سجل موحد (Audit Log) لتتبع الاستعلامات والإجابات وسجلات الوصول تعزيزًا للمساءلة.
- **تحسين التواصل المؤسسي:** تعزيز التفاعل بين المستخدمين وإدارة المؤسسة عبر نظام إشعارات وتنبيهات داخلية موحدة.

## 3.1 مشكلة المشروع (Problem Statement)

تكمن المشكلة الرئيسية في تعثر الوصول الذكي والسريع إلى البيانات والمستندات المؤسسية نتيجة الاعتماد على الوسائل التقليدية واليدوية، مما يترتب عليه بطء في اتخاذ القرار، وهدر الموارد التشغيلية، وانخفاض مستوى جودة الخدمات المقدمة.

**مظاهر المشكلة**

1. **صعوبة الوصول المباشر للبيانات:** اقتصار استعلام قواعد البيانات على المتخصصين التقنيين (SQL)، مما يحرم الإدارة والموظفين غير المختصين من الاستعلام المباشر ويؤخر الحصول على المعلومة.
2. **شتات المستندات المؤسسية:** غياب آلية موحدة للبحث في المستندات والسياسات واللوائح غير المفهرسة، مما يجعل الوصول إلى نصوص القرارات واللوائح أمراً مستغرقاً للوقت.
3. **بطء وتكلفة إعداد التقارير:** الاعتماد على تجميع البيانات وإعداد التقارير والتحليلات يدوياً، مما يزيد من احتمالية الأخطاء التشغيلية ويؤخر صُنّاع القرار.
4. **غياب الشفافية وسجلات التتبع:** تفتقر العديد من العمليات التشغيلية إلى سجل تتبع موحد للاستعلامات والإجابات، مما يعوق تقييم الأداء والرقابة.
5. **ضعف التواصل المؤسسي:** غياب القنوات الذكية المباشرة للتفاعل بين المستفيدين والإدارة، مما ينعكس سلباً على الشفافية ورضا المستخدمين.

**الحلول المقترحة لمعالجة المشكلة**

1. **تطوير محرك تحويل اللغة الطبيعية إلى استعلامات** (NL2SQL): لتمكين المستخدمين من الوصول المباشر للبيانات باستعلامات نصية بسيطة دون معرفة برمجة SQL.
2. **تطبيق تقنية الاسترجاع المعزز بالتوليد** (RAG): لفهرسة وتنظيم المستندات والبحث فيها فورياً واستخراج الإجابات الدقيقة المدعمة بالمصادر.
3. **تطوير وحدة آلية للتحليل وتوليد التقارير:** لأتمتة استخراج الاتجاهات والقيم الشاذة، وإنتاج تقارير منسقة بصيغ متعددة فور طلبها.
4. **بناء سجل تدقيق ومتابعة شامل** (Audit Log): لتسجيل جميع الاستعلامات والعمليات وضمان شفافية التفاعل والمساءلة.
5. **توفير واجهة محادثة ذكية وشبكة إشعارات:** لربط المستفيد بالإدارة عبر مساعد ذكي باللغة العربية يدعم الفصحى والعامية.

## 4.1 تعريف المشروع (System Definition)

يُعرَّف نظام «المساعد الذكي المؤسسي (بيان)» بأنه منصة محادثة ذكية متقدمة قائمة على معمارية الأنظمة متعددة الوكلاء (Multi-Agent System)، تمكّن المستخدمين من التفاعل التكاملي مع قواعد البيانات والمستندات المؤسسية باللغة العربية الطبيعية. وتتكون البنية الأساسية للنظام من المكونات التالية:

- **واجهة محادثة عربية** (RTL): واجهة تفاعلية تدعم اتجاه الكتابة من اليمين إلى اليسار لطرح الأسئلة واستقبال الردود الفورية.
- **لوحة تحكم ويب** (Web Admin Dashboard): منصة إدارية لمراقبة أداء النظام، وتتبع الوكلاء، وإدارة الصلاحيات والأمان.
- **خمسة وكلاء ذكاء اصطناعي متخصصين:** يعملون في إطار منسق عبر تقنية LangGraph وهم: الوكيل المشرف، وكيل SQL، وكيل RAG، وكيل التحليل، ووكيل التقارير.
- **قاعدة بيانات موحدة وآمنة:** لحفظ سجلات المستخدمين، والأنشطة، والمحادثات المخزنة.
- **طبقة العرض الذكية** (Smart Views Layer): طبقة وسيطة مخصصة تبسط بنية قاعدة البيانات لرفع دقة تحويل النص إلى استعلامات (NL2SQL).

ويعمل النظام على رقمنة عمليات الوصول للبيانات وأتمتتها من خلال الآليات التالية:

1. **المحادثة التفاعلية الذكية:** معالجة اللغة الطبيعية العربية وفهم سياق الأسئلة.
2. **تحويل النص إلى استعلام** (NL2SQL): ترجمة أسئلة المستخدمين مباشرة إلى استعلامات قواعد بيانات تنفيذية.
3. **الاسترجاع المعزز بالتوليد** (RAG): الفهرسة والبحث النصي الذكي في المستندات الرسمية وإسناد الإجابات للمصادر.

4. **التحليل الإحصائي الآلي:** استخراج المؤشرات، والاتجاهات العامة، والانحرافات الإحصائية من البيانات.
5. **توليد التقارير:** صياغة وتصدير تقارير الجداول والرسوم البيانية بصيغ متعددة.
6. **التهيئة الذكية** (Smart Onboarding): عملية مؤتمتة تربط النظام ببيانات المؤسسة وفهرسة وثائقها خلال زمن قدره 8 ساعات.
7. **الأمان المتدرج:** تصنيف أمن الجداول إلى أربعة مستويات أمنية (Public، Masked، Restricted، Blacklist).
8. **تتبع مسار التنفيذ** (Agent Trace): الشفافية الكاملة في إظهار الخطوات التحليلية التي اتبعها الوكلاء للوصول إلى النتيجة.

## 5.1 معايير القبول (Acceptance Criteria)

الميزات التي تميز نظامنا وتشجع المديرين على استخدام النظام هي:

- الحصول على رؤى وتقارير تنفيذية سريعة دون الحاجة لمعرفة تقنية.
- القدرة على الوصول إلى جميع البيانات العامة والمقنعة بسهولة.
- دعم اللغة العربية الفصحى والعامية في المحادثة.
- تتبع مسار تنفيذ الاستعلامات لضمان الشفافية.
- توليد تقارير قابلة للتصدير بصيغ متعددة.
- تكامل قوي مع قواعد البيانات المؤسسية القائمة.
- أمان متدرج يحمي البيانات الحساسة.
- واجهة مستخدم بسيطة وسهلة الاستخدام.
- إمكانية تبديل النماذج اللغوية بين المحلية والسحابية.
- سجل تدقيق شامل لكل العمليات.

## 6.1 نطاق المشروع (Project Scope)

يحدد هذا القسم النطاق الجغرافي والوظيفي للمشروع، بالإضافة إلى أصحاب المصلحة الرئيسيين المستفيدين من النظام.

### 1.6.1 النطاق الجغرافي (Geographical Scope)

يستهدف المشروع بشكل أساسي المؤسسات والجهات الأكاديمية في الجمهورية اليمنية، وخاصة جامعة العلوم والتكنولوجيا التي اتُخذت كدراسة حالة. ومع ذلك، يمكن تعميم النظام وتطبيقه في أي مؤسسة عربية تطلب بيئة محادثة ذكية باللغة العربية للوصول إلى بياناتها ومستنداتها.

### 2.6.1 النطاق الوظيفي (Functional Scope)

يشمل النطاق الوظيفي تطوير نظام محادثة ذكي يدمج بين أربع تقنيات أساسية: تحويل اللغة الطبيعية إلى SQL، والاسترجاع المعزز بالتوليد، والأنظمة متعددة الوكلاء، ومعالجة اللغة العربية. ويقدم النظام للأنظمة الجامعية حلاً شاملاً للاستعلام عن البيانات والبحث في المستندات وتوليد التقارير، مع ضمان الأمان والشفافية.

### 3.6.1 أصحاب المصلحة (Stakeholders)

تشمل قائمة أصحاب المصلحة الرئيسيين:

- **المديرون التنفيذيون:** يحتاجون رؤى استراتيجية سريعة وموثقة.
- **المحللون:** يجرون استعلامات معقدة على البيانات.
- **الموظفون:** يبحثون عن معلومات محددة بسرعة.
- **مسؤولو النظام:** يشرفون على التشغيل والأمان.
- **مشغلو التهيئة:** مسؤولون عن الإعداد الأولي للمؤسسة.
- **الإدارة الأكاديمية:** تتبني النظام كأداة تحول رقمي.

## 7.1 قيود المشروع (Limitations)

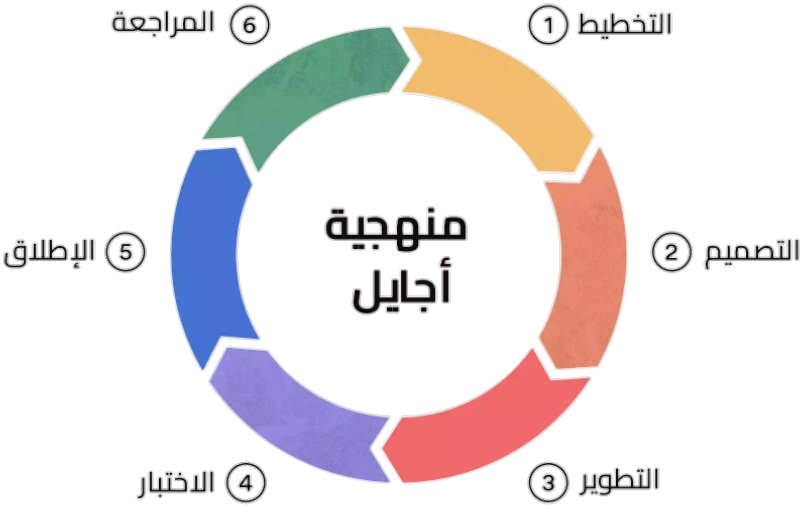
يواجه المشروع مجموعة من القيود التي يجب أخذها بعين الاعتبار:

- **قيود البنية التحتية:** محدودية الموارد الحاسوبية محلياً، مما يستوجب الاعتماد على نماذج لغوية متوسطة الحجم.
- **قيود اللغة:** ضعف دعم اللغة العربية في النماذج اللغوية التجارية، مما يستدعي حلولاً تكميلية.
- **قيود البيانات:** ندرة قواعد البيانات العربية الكاملة للاختبار.
- **قيود الوقت:** كون المشروع ضمن متطلبات التخرج يفرض جدولاً زمنياً محدوداً.
- **قيود التمويل:** عدم توفر ميزانية للاعتماد على نماذج سحابية مدفوعة، مما يُلزم بالاعتماد على حلول مفتوحة المصدر.
- **قيود الكوادر:** محدودية الفريق المعني بالتطوير والصيانة.

## 8.1 المنهجية المستخدمة في مشروع بيان (منهجية أجايل)

تم اعتماد منهجية أجايل (Agile) في تطوير مشروع بيان، نظراً لمرونتها وقابليتها للتكيّف مع المتطلبات المتغيرة. وتعتمد هذه المنهجية على تقسيم العمل إلى دورات تطوير قصيرة (Iterations)، مع التركيز على التسليم التدريجي والتغذية الراجعة

المستمرة.




**شكل 1.1:** نموذج أجايل (Agile Model) المتبع في مشروع بيان.

### 1.8.1 مرحلة التخطيط (Planning)

في هذه المرحلة يتم تحديد متطلبات المشروع بشكل عام، وجمع احتياجات أصحاب المصلحة، وتحديد أولويات الميزات. يتم إعداد قائمة الميزات المطلوبة وتقدير الجهد الزمني لكل منها، مع تحديد أهداف الدورة القادمة بشكل واضح وقابل للقياس.

### 2.8.1 مرحلة التصميم (Design)

يتم في هذه المرحلة تصميم بنية النظام والمكونات الرئيسية، بما في ذلك تصميم قاعدة البيانات، وواجهات المستخدم، ووكلاء الذكاء الاصطناعي. يعتمد التصميم على مخططات UML ومخططات العلاقات (ERD) لضمان وضوح البنية وسهولة التطوير.

### 3.8.1 مرحلة التطوير (Development)

يتم فيها كتابة الشفرة البرمجية بناءً على التصميم المعتمد. تشمل هذه المرحلة تطوير الواجهة الخلفية (FastAPI)، والواجهة الأمامية (Next.js)، وتدريب وتهيئة وكلاء الذكاء الاصطناعي عبر LangGraph.

### 4.8.1 مرحلة الاختبار (Testing)

يتم فيها اختبار كل مكون من مكونات النظام بشكل منفصل (اختبار الوحدة)، ثم اختبار النظام ككل (اختبار التكامل). تشمل الاختبارات قياس دقة NL2SQL، ودقة الاسترجاع في RAG، وأداء وكلاء الذكاء الاصطناعي.

### 5.8.1 مرحلة الإطلاق (Deployment)

بعد التأكد من جاهزية النظام واكتمال الاختبارات، يتم إطلاق النسخة الأولية لاستخدام من قبل المؤسسة التجريبية. تشمل هذه المرحلة تشغيل النظام في البيئة الحقيقية، ومتابعة أدائه، وتقييم مدى تحقيقه لأهداف المشروع. يتم النشر باستخدام Docker Compose لتسهيل تشغيل جميع الخدمات (الواجهة، الخلفية، قواعد البيانات، نموذج LLM) في بيئة موحدة. كما يتم جمع الملاحظات من المستخدمين التجريبيين لتحسين النسخ القادمة.

### 6.8.1 مرحلة المراجعة (Review)

في هذه المرحلة تتم مراجعة العمل المنجز خلال دورة التطوير، وتقييم ما تم تحقيقه، وتحليل المشكلات التي ظهرت أثناء العمل. تهدف هذه المرحلة إلى تحسين العملية في الدورات التالية وضمان تطوير النظام بشكل أفضل. تُعد المراجعة عنصرًا أساسيًا في منهجية أجايل لأنها تساعد على التعلم المستمر وتحسين الأداء، خاصة في مشروع يعتمد على تقنيات الذكاء الاصطناعي التي قد تتطلب تعديلات مستمرة بناءً على نتائج الاختبارات وتغذية المستخدمين الراجعة.

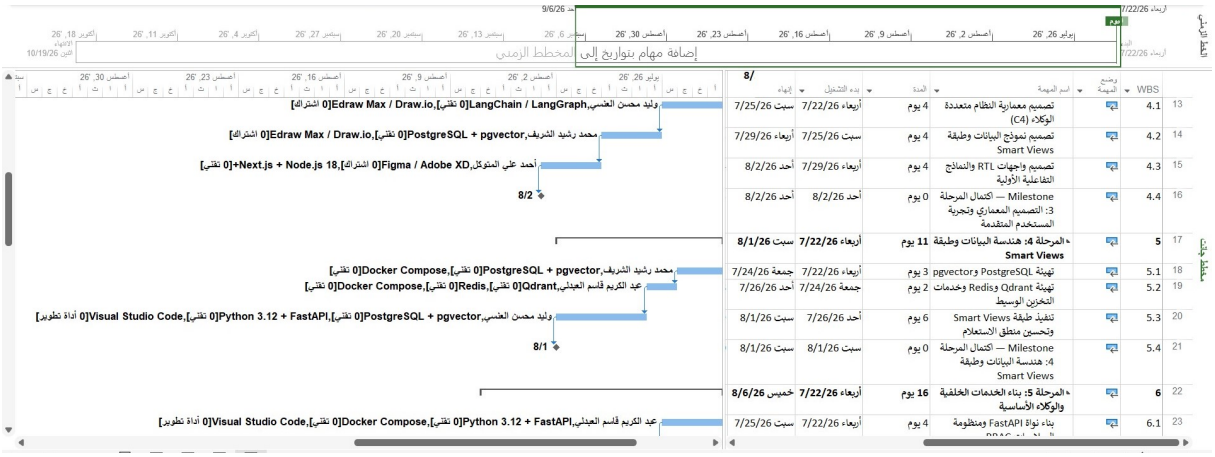
## 9.1 الجدول الزمني للمشروع (Project Schedule)

تم تصميم جدول المشروع ليوضح تسلسل الأنشطة المشاركة في تطوير هذا المشروع. وقد قُسّم المشروع إلى مراحل رئيسية مع توزيع المهام على أعضاء الفريق والمدد الزمنية المتوقعة لإنجاز كل مرحلة.

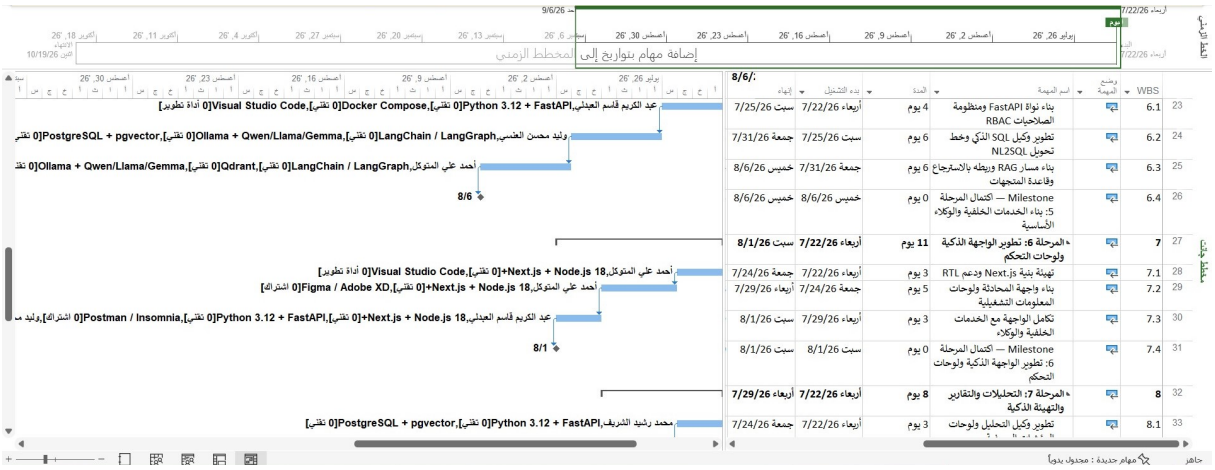
**شكل 2.1:** مخطط جانت لمراحل المشروع — الجزء الأول.

## 10.1 متطلبات الأدوات (Required Tools)

يستند تطوير مشروع بيان إلى مجموعة من الأدوات المادية والبرمجية التي تضمن جودة المنتج النهائي وكفاءة عملية التطوير. وفيما يلي عرض تفصيلي لهذه الأدوات.




**شكل 3.1:** مخطط جانت لمراحل المشروع — الجزء الثاني.




**شكل 4.1:** مخطط جانت لمراحل المشروع — الجزء الثالث.

### 1.10.1 الأدوات المادية (Hardware Tools)

يوضح الجدول 1.1 المواصفات المادية المطلوبة لكل عضو في فريق التطوير، إلى جانب الأجهزة اللازمة لتشغيل بيئة الخوادم.

### 2.10.1 الأدوات البرمجية (Software Tools)

يستعرض الجدول 2.1 الأدوات البرمجية الرئيسية المعتمدة في تطوير المشروع، مع ذكر الاستخدام المخصص لكل أداة.

| البرنامج (Software) | الاستخدام |
|---|---|
| Windows 10 | نظام تشغيل لتشغيل النظام على حواسيب التطوير، مع دعم WSL2 لتشغيل بيئات Linux على Windows. |
| Microsoft Word 2010 | لتوثيق المشروع. |
| Microsoft Project 2010 | لإعداد الجدول الزمني للمشروع. |
| Microsoft PowerPoint 2010 | لإعداد عرض تقديمي للمشروع. |

| البرنامج (Software) | الاستخدام |
|---|---|
| Edraw Max | لإنشاء المخططات والنماذج الخاصة بالمشروع. |
| Figma أو Adobe XD | لتصميم واجهات النظام. |
| Visual Studio Code | بيئة التطوير المتكاملة لكتابة كود المشروع (Python، JavaScript، TypeScript). |
| Python 3.12+ | لغة البرمجة الأساسية لتطوير الواجهة الخلفية (FastAPI) ووكلاء LangGraph. |
| Node.js 18+ | لغة البرمجة لتطوير الواجهة الأمامية (Next.js) وإدارة الحزم عبر npm. |
| FastAPI | إطار عمل لبناء الواجهة الخلفية (Backend API) وخدمات الوكلاء. |
| Next.js | إطار عمل لبناء الواجهة الأمامية (Frontend) مع دعم RTL للغة العربية. |
| LangGraph / LangChain | مكتبات لبناء وتنسيق الوكلاء المتعددين (Multi-Agent System). |
| PostgreSQL مع pgvector | قاعدة بيانات علائقية لتخزين البيانات المؤسسية والمتجهات (لـ RAG). |
| Qdrant | قاعدة بيانات شعاعية (Vector Database) لتخزين واسترجاع تضمينات المستندات. |
| Redis | تخزين مؤقت (Cache) وطابور رسائل لتنسيق الوكلاء وإدارة الجلسات. |
| Ollama | أداة لتشغيل النماذج اللغوية محلياً (مثل Qwen، Gemma، Llama). |
| Docker Desktop | منصة للحاويات لتشغيل جميع الخدمات (PostgreSQL، Qdrant، Redis، Ollama) بشكل موحد. |
| GitHub / Git | نظام التحكم في الإصدارات وإدارة المستودع البرمجي للمشروع. |
| Draw.io | إنشاء المخططات والنماذج (مخططات التدفق، ERD، UML، مخططات C4). |
| **Postman / Insomnia** | اختبار واجهات برمجة التطبيقات (APIs) للواجهة الخلفية. |

LT

**جدول 2.1:** الأدوات البرمجية المعتمدة في تطوير مشروع بيان.

## 11.1 هيكل فصول المشروع (Structure of Project)

يتكون هذا المشروع من ستة (6) فصول رئيسية، وقد تم تناول كل فصل بشكل منفصل كما يلي:

**شكل 5.1:** مخطط جانت لمراحل المشروع — الجزء الرابع.

**الفصل الأول: المقدمة.** يقدم هذا الفصل الخلفية العامة للمشروع، وبيان المشكلة، وأهداف المشروع، ونطاق العمل، والجدول الزمني المقترح، ومنهجية التطوير المتبعة، بالإضافة إلى الأدوات المستخدمة في بناء وتنفيذ نظام «بيان».

**الفصل الثاني: مراجعة الأدبيات والدراسات السابقة.** يشرح هذا الفصل الفكرة العامة للمشروع من خلال استعراض الدراسات السابقة والبحوث المرتبطة بالمساعدات الذكية المؤسسية، وأهم التقنيات والنظريات التي اعتمد عليها الباحثون الآخرون في تنفيذ أنظمة مشابهة. يساعد هذا الفصل في بناء فهم شامل وواضح للمبادئ التي يقوم عليها المشروع المقترح.

**الفصل الثالث: التحليل.** يتناول هذا الفصل تحليل المشروع بشكل تفصيلي عبر دراسة الجدوى، والمتطلبات الوظيفية وغير الوظيفية، وحالات الاستخدام (Use Cases)، بالإضافة إلى تقديم المخططات الأساسية للنظام مثل مخطط قاعدة البيانات ومخططات الفئات (Class Diagrams).

**الفصل الرابع: التصميم.** يتناول هذا الفصل جميع جوانب تصميم النظام، بما في ذلك تصميم قاعدة البيانات، وتصميم واجهات التطبيق ولوحة التحكم، بالإضافة إلى عرض الهياكل المعمارية والرسومية للنظام.

**الفصل الخامس: التنفيذ.** يعرض هذا الفصل الجزء العملي من المشروع، حيث يتم شرح كيفية تنفيذ قاعدة البيانات، وتطوير الواجهات القابلة للتشغيل لكل من موقع الإدارة والتطبيق المخصص للمستخدمين، إلى جانب عرض أهم الشفرات البرمجية المستخدمة.

**الفصل السادس: الخاتمة والتوصيات.** يتضمن هذا الفصل ملخصًا لما تم إنجازه في المشروع، والاستنتاجات الرئيسية، بالإضافة إلى التوصيات والمقترحات المستقبلية لتطوير النظام مثل إضافة مزايا جديدة وتحسين الأداء في الإصدارات القادمة.

**جدول 1.1:** الأدوات المادية المطلوبة للمشروع.

| الأداة | الكمية | الاستخدام | المواصفات |
|---|---|---|---|
| اتصال بالإنترنت | --- | للبحث، وتصحيح الأخطاء، وتحميل النماذج والأدوات، والتواصل مع الفريق | سرعة مناسبة للبحث والعمل الجماعي وتحميل ملفات بحجم عدة جيجابايتات (مثل نماذج Ollama). |
| لابتوب<br><br>**الرام** (RAM): 16 GB كحد أدنى، ويفضل 32 GB لتشغيل النماذج المحلية.<br>**مساحة التخزين:** 512 GB SSD أو أكثر (لأنظمة التشغيل، قواعد البيانات، ونماذج الذكاء الاصطناعي). | 4 | لكل عضو جهاز للعمل على تطوير المشروع | **المعالج:** Core i7 أو ما يعادله (لتشغيل بيئة التطوير). |
| خادم محلي أو حاسوب إضافي<br><br>**الرام** (RAM): 32 GB أو أكثر.<br>**مساحة التخزين:** 1 TB SSD أو أكثر.<br>**نظام التشغيل:** Linux (Ubuntu) أو Windows مع WSL2. | 1 | تشغيل خدمات الخلفية الأساسية مثل قواعد البيانات (PostgreSQL، Qdrant، Redis) ونماذج Ollama محلياً | **المعالج:** Core i7 أو ما يعادله مع دعم لتسريع CUDA (بطاقة رسوميات NVIDIA). |
| ذاكرة فلاش (Flash memory) | 2 | لعمل نسخ احتياطية للمشروع، وملفات التطوير، والبيانات الأولية | سعة 64 GB أو أكثر، ويفضل أن تدعم USB 3.0 للسرعة. |

الفصل

# الثاني

Chapter 2

## الخلفية النظرية ومراجعة الأدبيات

Theoretical Background and Literature Review

## 1.2 مقدمة الفصل

يستعرض هذا الفصل الأسس النظرية والمراجعة النقدية للبيئة البحثية والتقنية التي يستند إليها مشروع «المساعد الذكي المؤسسي (بيان)». تتوزع مراجعة الأدبيات في هذا الفصل على خمسة محاور رئيسية تشمل: النماذج اللغوية الكبيرة (LLMs)، وتقنيات تحويل اللغة الطبيعية إلى استعلامات قواعد البيانات (NL2SQL)، ونظم الاسترجاع المعزز بالتوليد (RAG)، والمعماريات القائمة على الأنظمة متعددة الوكلاء (Multi-Agent Systems)، بالإضافة إلى تحديات معالجة اللغة العربية (Arabic NLP). ويُختتم الفصل بتحليل مقارن يبرز الفجوة البحثية والتطبيقية التي يستهدف المشروع معالجتها.

تكتسب هذه المراجعة أهمية مضاعفة في مشروع بيان، نظراً لأنه يدمج أربع تقنيات متقدمة في إطار موحد. لذا، يتطلب المشروع فحصاً دقيقاً لأسس كل تقنية على حدة، ثم تحليل كيفية تكاملها في الأدبيات السابقة. كما يسهم هذا الفصل في تحديد الفجوة البحثية بدقة، وتبرير القرارات التصميمية المُتخذة في الفصل الأول، خاصة اختيار LangGraph لإطار التنسيق بين الوكلاء، واعتماد طبقة Views الذكية لتحسين دقة NL2SQL، واستخدام Odoo ERP كبيئة محاكاة واقعية.

يتضمن الفصل ستة أقسام رئيسية: يبدأ بالخلفية النظرية للتقنيات الأساسية، ثم يستعرض الأنظمة والدراسات السابقة، يلي ذلك قسم مقارن يقدم جدولاً تحليلياً، ثم تحليل الفجوة البحثية، وأخيراً الخلاصة. ويعتمد الفصل على نظام استشهاد IEEE مع استمرارية الترقيم من الفصل الأول.

## 2.2 الخلفية النظرية

تستند الخلفية النظرية لمشروع بيان إلى خمس ركائز علمية مترابطة، تشكل مجتمعةً الأساس النظري للنظام. وفيما يلي عرض تفصيلي لكل ركيزة مع التركيز على التطورات الحديثة والأدوات التجارية المتاحة.

### 1.2.2 النماذج اللغوية الكبيرة (Large Language Models - LLMs)

تُعرَّف النماذج اللغوية الكبيرة بأنها نماذج ذكاء اصطناعي ضخمة تُدرَّب على كميات هائلة من النصوص بهدف فهم اللغة البشرية وتوليدها. وقد شهدت هذه النماذج تطوراً مذهلاً منذ عام 2020، حيث انتقلت من نماذج متخصصة في مهمة واحدة إلى نماذج عامة قادرة على أداء آلاف المهام بدقة عالية.

تعدّ عائلة GPT من OpenAI الأبرز في هذا المجال، حيث أطلقت الشركة نموذج GPT-3.5 أواخر 2022 ثم GPT-4 في مارس 2023 [16]. يُقدَّر أن GPT-4 يضم نحو 7.1 تريليون معامل، ويتميز بقدرات استدلالية متقدمة وفهم سياقي عميق. وفي المنطقة العربية، أطلقت شركة Inception الإماراتية نموذج «Jais» عام 2023 وهو أول نموذج لغوي عربي كبير، كما أطلقت Alibaba نموذج «Qwen» الذي يدعم العربية بشكل جيد [6].

تعتمد هذه النماذج على معمارية المحول (Transformer) التي قدمتها ورقة Vaswani وآخرين عام 2017، والتي أحدثت ثورة من خلال آلية الانتباه الذاتي (Self-Attention). هذه الآلية تتيح للنموذج فهم العلاقات بين الكلمات بغض النظر عن مواقعها في الجملة، مما يحسّن الفهم السياقي بشكل كبير.

رغم قوتها الواضحة، تعاني النماذج اللغوية الكبيرة من قيود جوهرية: الهلوسة (Hallucination) أي توليد معلومات خاطئة بثقة عالية، وعدم القدرة على الوصول للمعرفة الخاصة بالمجال، وقيود حجم السياق. هذه القيود هي ما دفع الباحثين لتطوير تقنيات RAG التي تُعالج لاحقاً في هذا الفصل.

### 2.2.2 تقنية تحويل اللغة الطبيعية إلى SQL (NL2SQL)

تُعدّ تقنية NL2SQL إحدى أهم تطبيقات الذكاء الاصطناعي في مجال قواعد البيانات، حيث تُمكّن المستخدمين من الاستعلام عن البيانات بلغتهم الطبيعية دون الحاجة لمعرفة SQL [4]. وقد شهدت هذه التقنية قفزة نوعية منذ ظهور النماذج اللغوية الكبيرة، حيث ارتفعت الدقة من نحو 35% عام 2018 إلى أكثر من 85% عام 2024 على المعايير القياسية.

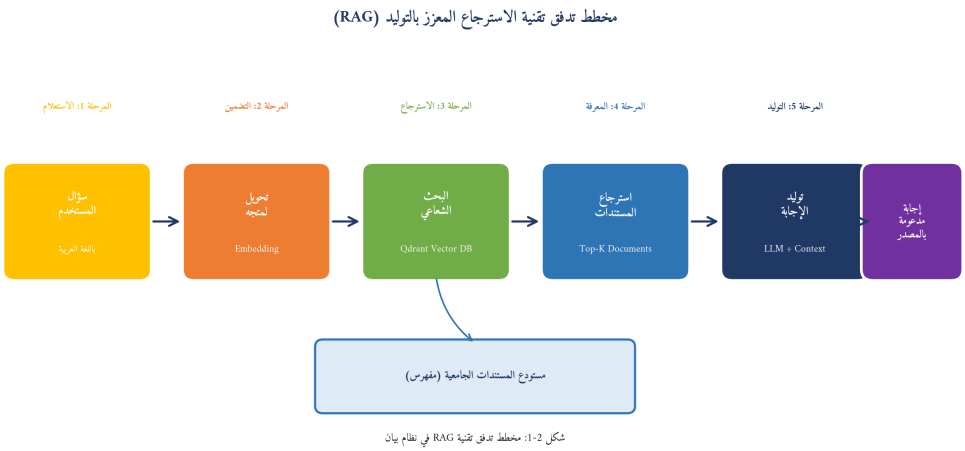
من أبرز الأنظمة الحالية في هذا المجال Vanna.ai، وهو إطار عمل مفتوح المصدر يدمج بين RAG و NL2SQL، ويُحقق دقة تتجاوز 80% عند توفير السياق المناسب [2]. يعتمد Vanna على تدريب نموذج خاص بكل قاعدة بيانات، ويُستخدم على نطاق واسع في التطبيقات التجارية. كما أطلقت Snowflake أوائل 2025 نموذج Arctic-Text2SQL-R1 الذي حقق 83.71% دقة على معيار BIRD الأكثر تحدياً، باستخدام تقنية التعلم المعزز في التدريب [3].

من أهم التحديات في NL2SQL ربط المخطط (Schema Linking)، حيث يجب على النظام تحديد الجداول والأعمدة ذات الصلة بالسؤال من بين مئات العناصر في قاعدة البيانات. ولهذا السبب، يلجأ كثير من المطورين إلى تبسيط المخطط عبر طبقات وسيطة، وهو ما يبرر ابتكار طبقة Views الذكية في مشروع بيان.

### 3.2.2 تقنية الاسترجاع المعزز بالتوليد (RAG)

تُمثل تقنية RAG إطاراً هجيناً يجمع بين قدرات النماذج اللغوية على التوليد، وقواعد المعرفة الخارجية على توفير المعلومات الدقيقة [5]. وقد أصبحت هذه التقنية المعيار الذهبي لتطبيقات LLM في البيئات المؤسسية منذ عام 2023، حيث تُعالج مشكلة الهلوسة وتُمكّن النموذج من الوصول للمعرفة الخاصة بالمجال.

يعمل نظام RAG وفق سير عمل من ثلاث مراحل: الاسترجاع (Retrieval) حيث يتم تحويل سؤال المستخدم إلى متجه وبحث عن المستندات الأكثر صلة، ثم التوليد المعزز (Augmented Generation) حيث يُمرَّر المستندات المسترجعة مع السؤال إلى النموذج لتوليد إجابة دقيقة، وأخيراً الاستشهاد (Citation) حيث يُربط كل جزء من الإجابة بالمصدر المستخرج منه.




**شكل 1.2:** مخطط تدفق تقنية RAG في نظام بيان.

من أبرز الأطر الحالية لبناء أنظمة RAG، يأتي LangChain [9] في المقدمة كإطار العمل الأكثر انتشاراً، ويقدم مكونات قابلة للتخصيص للبحث والتوليد. ويأتي LlamaIndex [10] كإطار متخصص في RAG، يقدم أدوات

متقدمة لفهرسة المستندات وإدارة السياق. كما تقدم Microsoft خدمة Azure AI Search المتكاملة مع RAG، وتقدم Google خدمة Vertex AI Search.

يعتمد أداء RAG بشكل جوهري على جودة نظام الاسترجاع، والذي يتطلب قاعدة بيانات شعاعية (Vector Database) متخصصة. وقد ظهرت عدة حلول مفتوحة المصدر وتجارية في السنوات الأخيرة، أبرزها Qdrant [8] الذي اعتمد عليه مشروع بيان، و Pinecone التجاري، و Milvus، إضافة إلى امتداد pgvector لـ PostgreSQL.

### 4.2.2 الأنظمة متعددة الوكلاء (Multi-Agent Systems)

تُعرَّف الأنظمة متعددة الوكلاء في سياق الذكاء الاصطناعي الحديث بأنها أنظمة تُكلف فيها عدة وكلاء مستقلين بإجراء مهام متخصصة، مع وجود آلية تنسيق تضمن تكامل مخرجاتهم. وقد ظهر هذا المفهوم بقوة منذ عام 2023 مع تطور قدرات النماذج اللغوية، حيث أصبح كل وكيل عبارة عن LLM مُوجَّه بمهمة وشخصية محددة.

يُعدّ إطار LangGraph [7] من LangChain الأبرز في هذا المجال، ويستخدم نظرية المخططات لتمثيل تدفق العمل بين الوكلاء كعقد (Nodes) وحواف (Edges). يتميز بتحكم دقيق في التدفق ودعم الحلقات التكرارية والتفرعات المشروطة، وهو ما اعتمد عليه مشروع بيان. ويأتي إطار AutoGen [11] من Microsoft Research كخيار مرن قائم على المحادثات بين الوكلاء، لكنه يعاني من زيادة في استهلاك الـ Tokens.

إطار CrewAI [12] هو إطار حديث أطلق عام 2024، يركز على محاكاة فرق العمل البشرية بأدوار محددة، ويتميز بسهولة الاستخدام. أما إطار MetaGPT [13] فيقدم نهجاً مختلفاً، حيث يحاكي فريق تطوير برمجيات تقليدي (Product Manager, Architect, Engineer) لإنتاج شفرات برمجية متكاملة، وقد حقق تفوقاً ملحوظاً في إنتاج الكود الصحيح.

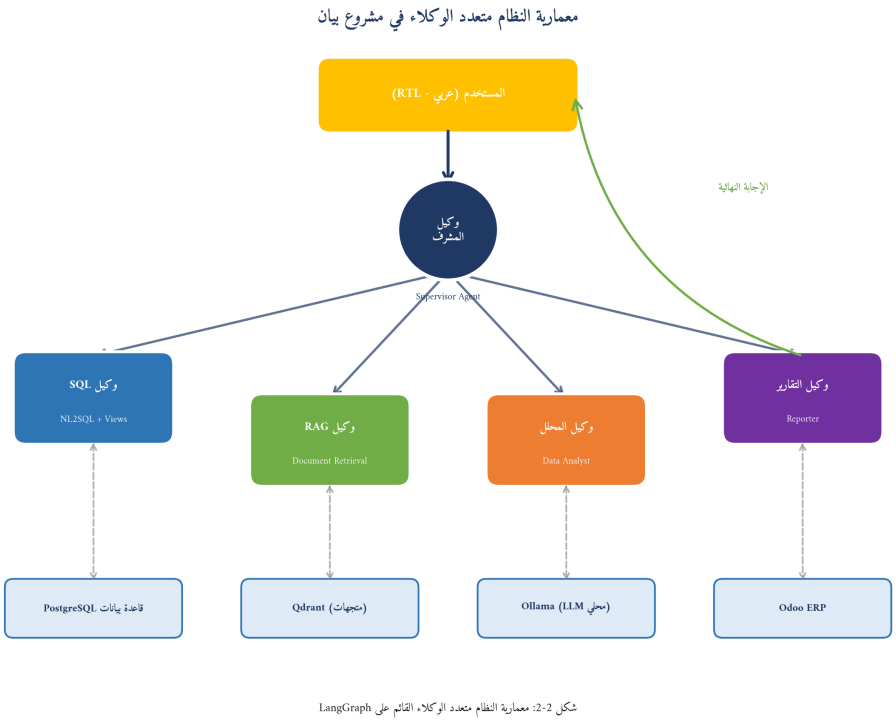
تُظهر هذه المراجعة أن LangGraph هو الإطار الأنسب لمشروع بيان، نظراً لكونه يدعم بناء أنظمة معقدة ذات وكلاء متخصصين، مع إمكانية تتبع مسار التنفيذ (Agent Trace) المطلوبة للشفافية في البيئة الجامعية.

### 5.2.2 تحديات معالجة اللغة العربية (Arabic NLP)

تواجه معالجة اللغة العربية طبيعياً تحديات جوهرية تختلف عن نظيراتها في اللغة الإنجليزية، وتعود هذه التحديات إلى عوامل لغوية وتقنية [6]. فمن الناحية اللغوية، تتميز العربية بالصرف الغني حيث يمكن اشتقاق عشرات الكلمات من جذر واحد ثلاثي، مما يزيد من حجم المفردات وصعوبة التطابق. كما تتميز بتركيب الجملة المرن الذي يصعّب التحليل النحوي الآلي، وظاهرة الإعراب التي تُغير الحركة النهائية للكلمات.

أما من الناحية التقنية، فتعاني العربية من فجوة الموارد الحادة؛ إذ تشير الإحصائيات إلى أن نسبة المحتوى العربي على الإنترنت لا تتجاوز 2%.1 من الإجمالي، في حين أن المتحدثين بالعربية يشكلون نحو 5% من سكان العالم [6]. وتعتمد معظم التطبيقات التجارية على نماذج إنجليزية تُظهر أداءً ضعيفاً على العربية، وتشمل النماذج المدعومة للعربية: Jais النموذج الإماراتي المتخصص في العربية، و AraBERT النموذج المبني على BERT والمتخصص في النصوص العربية، و Qwen 2.5 من Alibaba الذي يدعم العربية جزئياً ويتميز بكفاءة عالية في المهام البرمجية.

تتفاقم هذه التحديات في السياق اليمني تحديداً، حيث تُستخدم لهجة محلية تختلف عن الفصحى والعاميات الخليجية والمصرية، وغير مدعومة بأي موارد برمجية معروفة. ولهذا السبب، يعتمد مشروع بيان استراتيجية لغوية مرنة تقبل الفصحى والعامية، مع الاستفادة من قدرة النماذج اللغوية الحديثة على التعميم بدلاً من تدريب نموذج متخصص.




**شكل 2.2:** معمارية النظام متعدد الوكلاء القائم على LangGraph.

## 3.2 الأنظمة والدراسات السابقة

تم تصنيف الأنظمة والدراسات السابقة المراجعة في خمس فئات رئيسية، وفقاً لمجال التخصص وارتباطها بمشروع بيان. وقد رُوعي في الاختيار التركيز على الأنظمة الحديثة (2021-2025) والتجارية المتاحة فعلياً.

### 1.3.2 الأنظمة في مجال تحويل اللغة الطبيعية إلى SQL

يُعدّ معيار Spider الذي أُطلق عام 2018 المرجع القياسي لتقييم أنظمة NL2SQL، حيث يضم 10,181 سؤالاً عبر 200 قاعدة بيانات. وفي عام 2023، أُطلق معيار BIRD الأكثر تحدياً لتركيزه على قواعد البيانات الواقعية مع قيم شاذة وأخطاء. حققت أفضل الأنظمة 40% فقط على BIRD مقابل 60% على Spider، مما يُبرز فجوة الأداء في البيئات الواقعية [4].

من أبرز الأنظمة الحالية في هذا المجال، يُبرز نظام Vanna.ai [2] كأحد أكثر الأطر مفتوحة المصدر انتشاراً. يعتمد Vanna على دمج RAG مع NL2SQL، ويتدرب على قاعدة بيانات كل عميل لرفع الدقة. أما نظام Snowflake Arctic-Text2SQL-R1 [3] فهو نموذج متخصص أطلق أوائل 2025، يستخدم تقنية التعلم المعزز في التدريب ويحقق 83%.71 على معيار BIRD.

تجدر الإشارة إلى أن جميع هذه الأنظمة مصممة للبيئة الإنجليزية، ولا تقدم دعماً كافياً للغة العربية. كما أنها تعمل كأنظمة أحادية (Single-Agent)، ولا تستفيد من البنية متعددة الوكلاء في المهام المعقدة.

### 2.3.2 الأنظمة في مجال الاسترجاع المعزز بالتوليد (RAG)

منذ تقديم مفهوم RAG بشكلها الحديث عام 2020 [5]، شهدت هذه التقنية تطوراً متسارعاً وأصبحت المعيار الذهبي لتطبيقات LLM المؤسسية. أظهرت الدراسات أن RAG يقلل الهلوسة بنسبة تصل إلى 50% مقارنة بالنماذج اللغوية التقليدية، مما جعله خياراً مفضلاً للتطبيقات التي تتطلب دقة عالية.

من أبرز الأطر الحالية لبناء أنظمة RAG، يأتي LangChain [9] في المقدمة كإطار العمل الأكثر انتشاراً، ويقدم مكونات قابلة للتخصيص للبحث والتوليد. ويأتي LlamaIndex [10] كإطار متخصص في RAG، يقدم أدوات متقدمة لفهرسة المستندات وإدارة السياق. كما تقدم Microsoft خدمة Azure AI Search المتكاملة مع RAG، وتقدم Google خدمة Vertex AI Search.

على صعيد قواعد البيانات الشعاعية، يبرز Qdrant [8] كحل مفتوح المصدر يتميز بأداء عالٍ ودعم فلترة المتجهات بالبيانات الوصفية، وهو ما اعتمد عليه مشروع بيان. كما تقدم Pinecone حلاً تجارياً سحابياً، ويقدم Milvus حلاً مفتوحاً موزعاً. وفي سياق قواعد البيانات العلائقية، يقدم امتداد pgvector لـ PostgreSQL إمكانية البحث الشعاعي دون الحاجة لقاعدة بيانات منفصلة.

### 3.3.2 الأنظمة في مجال الأنظمة متعددة الوكلاء

شهد مجال الأنظمة متعددة الوكلاء تطوراً متسارعاً منذ عام 2023، مع ظهور عدة أطر تجارية ومفتوحة المصدر. يُعدّ إطار LangGraph [7] من LangChain الأبرز، ويستخدم نظرية المخططات لتمثيل تدفق العمل بين الوكلاء. يتميز بتحكم دقيق في التدفق ودعم الحلقات التكرارية والتفرعات المشروطة، وهو ما اعتمد عليه مشروع بيان.

يأتي إطار AutoGen [11] من Microsoft Research كخيار مرن قائم على المحادثات بين الوكلاء، ويتيح بناء تطبيقات معقدة بسرعة. ومع ذلك، يعاني AutoGen من زيادة في استهلاك الـ Tokens بسبب المحادثات الطويلة. أما إطار CrewAI [12] فهو إطار حديث أطلق عام 2024، يركز على محاكاة فرق العمل البشرية بأدوار محددة، ويتميز بسهولة الاستخدام.

إطار MetaGPT [13] يقدم نهجاً مختلفاً، حيث يحاكي فريق تطوير برمجيات تقليدي (Product Manager, Architect, Engineer) لإنتاج شفرات برمجية متكاملة. حقق MetaGPT تفوقاً ملحوظاً في إنتاج الكود الصحيح، لكنه متخصص في مجال تطوير البرمجيات وليس عاماً لكل التطبيقات المؤسسية.

تُظهر هذه المراجعة أن LangGraph هو الإطار الأنسب لمشروع بيان، نظراً لكونه يدعم بناء أنظمة معقدة ذات وكلاء متخصصين، مع إمكانية تتبع مسار التنفيذ (Agent Trace) المطلوبة للشفافية في البيئة الجامعية.

### 4.3.2 المساعدات الذكية في التعليم العالي

رغم الانتشار الواسع للمساعدات الذكية في مجال خدمة العملاء، تبقى التطبيقات المتخصصة في التعليم العالي محدودة. أطلقت OpenAI عام 2024 خدمة ChatGPT Edu المخصصة للجامعات، التي تتيح بناء مساعدات مخصصة لكل مؤسسة. ومع ذلك، تظل هذه الخدمة عامة ولا تدمج NL2SQL للاستعلام عن قواعد البيانات الجامعية.

تقدم Microsoft خدمة Copilot for Education مع تكاملات محدودة مع أنظمة إدارة التعلم (LMS) مثل Canvas و Blackboard، لكنها لا تدعم العربية بشكل كامل. كما تقدم Anthropic خدمة Claude for Education التي تركز على المساعدة الأكاديمية للطلاب.

على الصعيد العربي، أطلقت العديد من الجامعات مساعدات ذكية بسيطة للأسئلة الشائعة، لكنها تعتمد على قواعد معرفة محدودة ولا تستفيد من تقنيات RAG أو NL2SQL. وأكدت دراسة Al-Khalifa وآخرين [14] أن معظم الجامعات العربية تستخدم chatbots بسيطة دون الاستفادة من التقنيات المتقدمة.

تُبرز هذه المراجعة أن الميدان يفتقر إلى نظام متكامل يجمع بين NL2SQL و RAG والأنظمة متعددة الوكلاء في بيئة جامعية عربية، وهو ما يسعى مشروع بيان لسده.

### 5.3.2 الدراسات العربية

على الصعيد العربي، تظل الدراسات في مجال المساعدات الذكية المؤسسية محدودة. دراسة Magdy وآخرين [15] استعرضت واقع البحث في استرجاع المعلومات العربية، وأكدت أن الأنظمة التجارية تعاني من ضعف 30-40% في الدقة على العربية مقارنة بالإنجليزية. كما أكدت دراسة El-Haj [6] المرجعية في معالجة اللغة العربية طبيعياً، أن المشاريع العربية تواجه ثلاث تحديات رئيسية: ندرة corpora عربية عالية الجودة، عدم وجود معايير قياسية موحدة، وضعف التمويل البحثي.

في البيئة اليمنية تحديداً، تكاد تنعدم الدراسات المنشورة في مجال الذكاء الاصطناعي المؤسسي. وأكد تقرير الأمم المتحدة للتحول الرقمي [1] أن اليمن تحتل المرتبة 175 عالمياً في مؤشر تطور الحكومة الإلكترونية. ويُعزى ذلك إلى عوامل متعددة، تشمل: محدودية البنية التحتية الرقمية، وضعف الاستثمار في البحث والتطوير، وندرة الكوادر المتخصصة. هذا الواقع يُبرز أهمية مشروع بيان كمساهمة بحثية وتطبيقية في سد هذه الفجوة.

## 4.2 مقارنة الأنظمة السابقة

يقدم هذا القسم تحليلاً مقارناً للأنظمة والدراسات السابقة المراجعة، مع التركيز على ستة أبعاد رئيسية: التقنيات المستخدمة، اللغة المدعومة، البيئة التطبيقية، دقة NL2SQL، دقة RAG، والبنية المعمارية. ويُلخص الجدول 1.2 أبرز الأنظمة ومقارنتها بمشروع بيان.

تكشف المقارنة عن عدة ملاحظات جوهرية. أولاً، تركز معظم الأنظمة في NL2SQL على البيئة الإنجليزية، مع غياب شبه تام للأنظمة العربية المتخصصة. ثانياً، تعتمد الأنظمة المتقدمة مثل Vanna.ai على دمج NL2SQL و RAG، لكنها لا تستخدم البنية متعددة الوكلاء، مما يُحدها في المهام المعقدة. ثالثاً، تفتقر التطبيقات في التعليم العالي إلى التكامل بين الاستعلام عن البيانات الديناميكية والبحث في المستندات الثابتة.

كما يُلاحظ أن مشروع بيان هو الوحيد الذي يجمع بين التقنيات الأربع (NL2SQL، RAG، Multi-Agent، Arabic NLP) في إطار موحد، مع تخصيصه للبيئة الجامعية اليمنية. هذا التكامل يمثل نقطة قوة، لكنه في الوقت ذاته يفرض تحديات تصميمية معقدة، خاصة في تنسيق الوكلاء وضمان جودة الأداء عبر المكونات المختلفة.

## 5.2 الفجوة البحثية

يتميز مشروع «بيان» بكونه النظام الوحيد — في حدود ما اطلعت عليه الدراسة — الذي يجمع بين التقنيات الأربع الأساسية (NL2SQL، RAG، Multi-Agent Systems، Arabic NLP) في إطار موحد ومحلي الاستضافة، مع تخصيصه المباشر للبيئة المؤسسية والتعليمية. ورغم أن هذا التكامل يمثل نقطة قوة واضحة، إلا أنه يفرض في الوقت ذاته تحديات تصميمية وتنفيذية معقدة، خاصة في تنسيق عمل الوكلاء البرمجيين وضمان جودة الأداء والاعتمادية عبر المكونات المختلفة.

**جدول 1.2:** مقارنة شاملة بين الأنظمة السابقة ومشروع بيان.

| النظام/الإطار | التقنيات المستخدمة | اللغة | البيئة التطبيقية | دقة NL2SQL | دقة RAG | البنية |
|---|---|---|---|---|---|---|
| [2] Vanna.ai (2024) | NL2SQL + RAG | إنجليزية | تجارية | +80% | غير مذكورة | أحادية |
| Arctic-[3] Text2SQL (2025) | NL2SQL + RL | إنجليزية | تجارية | 83%.71 (BIRD) | --- | أحادية |
| [9] LangChain (2023) | RAG + LLM | متعددة | عامة | --- | عالية | أحادية |
| LlamaIndex (2023) [10] | RAG + LLM | متعددة | عامة | --- | عالية | أحادية |
| [11] AutoGen (2023) | Multi-Agent | إنجليزية | عامة | --- | --- | متعددة |
| [12] CrewAI (2024) | Multi-Agent | إنجليزية | عامة | --- | --- | متعددة |
| [13] MetaGPT (2024) | Multi-Agent | إنجليزية | تطوير برمجي | --- | --- | متعددة |
| ChatGPT Edu (2024) | LLM + RAG | متعددة | تعليم عالٍ | --- | جيدة | أحادية |
| [8] Qdrant (2024) | Vector DB | متعددة | عامة | --- | عالية | --- |
| **مشروع بيان (المقترح)** | NL2SQL + RAG + Multi-Agent | عربية (فصحى وعامية) | جامعي (يمني) | 85% ≤ (مستهدف) | 80% ≤ (مستهدف) | متعددة (LangGraph) |

بناءً على المراجعة الشاملة للأنظمة والدراسات السابقة والتحليل المقارن، يمكن تحديد الفجوة البحثية والتطبيقية التي يسعى مشروع «بيان» لسدها في **خمسة محاور رئيسية**:

**المحور الأول — الفجوة التقنية:** رغم وجود أنظمة NL2SQL متقدمة (Vanna.ai، Arctic-Text2SQL) وأنظمة RAG فعالة (LangChain، LlamaIndex)، إلا أنه لا يوجد نظام يجمع بين التقنيتين في إطار موحد مع بنية متعددة الوكلاء. تعمل هذه الأنظمة بشكل منفرد، مما يُفقدها التكامل المطلوب في البيئات المؤسسية المعقدة. مشروع بيان يسد هذه الفجوة عبر الدمج العضوي للتقنيات الأربع (NL2SQL، RAG، Multi-Agent، Vector DB).

**المحور الثاني — الفجوة اللغوية:** أكدت دراسات El-Haj [6] و Magdy [15] على الفجوة الواضحة في دعم اللغة العربية في أنظمة الذكاء الاصطناعي. تعاني الأنظمة التجارية من انخفاض الدقة 30-40% على العربية مقارنة بالإنجليزية، كما أنها غير مصممة لفهم اللهجات المحلية. مشروع بيان يعالج هذه الفجوة عبر دعم العربية الفصحى والعامية في واجهة محادثة RTL، مع الاستفادة من نماذج Qwen و Jais التي تدعم العربية.

**المحور الثالث — الفجوة التطبيقية** (Applied Gap): تفتقر الأنظمة التطبيقية في المؤسسات والقطاع الأكاديمي إلى التكامل بين الاستعلام عن البيانات الرقمية والمالية (NL2SQL) والبحث في اللوائح والسياسات والمستندات التنظيمية

(RAG). تركز الأنظمة الحالية (مثل ChatGPT Enterprise أو ChatGPT Edu) على جانب واحد فقط، مما يحد من قدرتها على تلبية الاحتياجات التشغيلية الكاملة للمستفيدين. يقدم نظام «بيان» حلاً موحداً يدمج الجانبين عبر وكلاء متخصصين.

**المحور الرابع — الفجوة السياقية:** الأنظمة التجارية مثل Vanna.ai و ChatGPT Edu مصممة لبيئات تجارية غربية، ولا تأخذ في الاعتبار خصوصية البيئة الجامعية اليمنية (اللوائح المحلية، اللهجة، البنية التحتية المحدودة). مشروع بيان مصمم خصيصاً لهذه البيئة، مع استخدام Odoo ERP كبيئة محاكاة واقعية تُحقن ببيانات تعادل 15 عاماً.

**المحور الخامس — فجوة الابتكار:** لا توجد دراسة سابقة قدمت طبقة Views ذكية كآلية لتبسيط مخطط قاعدة البيانات وتحسين دقة NL2SQL. هذا الابتكار يمثل مساهمة بحثية أصلية يمكن البناء عليها في دراسات مستقبلية، وقد يفتح آفاقاً جديدة في مجال Database Abstraction Layer.

يُلخص الجدول 2.2 الفجوة البحثية ومساهمة مشروع بيان في سدها عبر المحاور الخمسة.

**جدول 2.2:** ملخص الفجوة البحثية ومساهمة مشروع بيان في سدها.

| المحور | الفجوة المحددة | مساهمة مشروع بيان |
|---|---|---|
| تقنية | غياب الدمج بين NL2SQL و RAG و Multi-Agent | دمج التقنيات الأربع في إطار موحد |
| لغوية | ضعف دعم العربية واللهجات | دعم فصحى وعامية في واجهة RTL |
| تطبيقية | فصل الاستعلام عن البيانات عن البحث في المستندات | نظام موحد يجمع الجانبين |
| سياقية | غياب حلول مخصصة للبيئة المؤسسية والمحلية | تخصيص كامل للبيئة المؤسسية مع بيئة محاكاة لأكثر من نوع من قواعد البيانات |
| ابتكارية | غياب طبقة تجريدية مبسطة فوق قواعد البيانات | ابتكار طبقة Views ذكية |

## 6.2 الخلاصة

استعرض هذا الفصل الإطار النظري الشامل الذي يستند إليه «**مشروع المساعد الذكي المؤسسي (بيان)**»، بدءاً من المفاهيم الأساسية في النماذج اللغوية الكبيرة (LLMs)، مروراً بالتقنيات المتخصصة في NL2SQL و RAG والأنظمة متعددة الوكلاء (Multi-Agent Systems)، وصولاً إلى تحديات معالجة اللغة العربية طبيعياً (Arabic NLP). وقد كشف الاستعراض النظري أن المشروع يعتمد على ركائز علمية متينة ومثبتة، مع الاستفادة من أحدث الأطر والأدوات المتاحة (2021-2025) في كل مجال.

كما قدم الفصل مراجعة شاملة للأنظمة والدراسات السابقة ومقارنة مرجعية بينها. وأظهرت المراجعة أن مجال NL2SQL شهد تقدماً كبيراً في السنوات الثلاث الأخيرة مع ظهور أطر مثل Vanna.ai و Arctic-Text2SQL. وفي مجال RAG، أثبتت الأطر المتقدمة مثل LangChain و LlamaIndex كفاءتها العالية. وفي مجال الأنظمة متعددة الوكلاء، أظهر إطار **LangGraph** تفوقاً ملحوظاً في إدارة الحالات (State Management) والتحكم في مسارات التنفيذ.

كشف التحليل المقارن أن مشروع «بيان» يمثل خطوة متقدمة في الدمج الشامل للتقنيات الأربع في بيئة عمل مؤسسية، حيث اتُخذت جامعة العلوم والتكنولوجيا كدراسة حالة وتطبيق تجريبي. وخلص الفصل إلى تحديد خمسة محاور رئيسية للفجوة البحثية (تقنية، لغوية، تطبيقية، سياقية، وابتكارية) يسعى المشروع لسدها، مما يُبرر القرارات التصميمية والهندسية المتخذة، ويهيئ القارئ للفصل الثالث الذي يقدم تحليلاً تفصيلياً للمتطلبات والمنهجية المتبعة في تطوير النظام.

# المراجع

# المصادر


[1] United Nations, ``E-Government Survey 2024: Accelerating Digital Transformation for Sustainable Development," *United Nations Department of Economic and Social Affairs*, 2024. [Online]. Available: https://publicadministration.un.org/egovkb/

[2] Vanna.ai, ``Vanna: Open-Source Python RAG Framework for SQL Generation," 2024. [Online]. Available: https://vanna.ai

[3] Snowflake AI Research Team, ``Arctic-Text2SQL-R1: A Reinforcement-Learned Text-to-SQL Model," 2025. [Online]. Available: https://github.com/Snowflake-Labs/arctic

[4] T. Yu, R. Zhang, K. Yang, M. Yasunaga, D. Wang, Z. Li, J. Ma, I. Li, Q. Yao, S. Yu, T. Zhang, and D. Radev, ``Spider: A Large-Scale Human-Labeled Dataset for Complex and Cross-Domain Semantic Parsing and Text-to-SQL Task," in *Proc. EMNLP*, 2018, pp. 3911--3921.

[5] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, ``Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, 2020, pp. 9459--9474.

[6] M. El-Haj, ``Arabic Natural Language Processing: Challenges and Opportunities," *arXiv preprint* arXiv:2307.10543, 2023.

[7] LangChain, ``LangGraph: Building Stateful Multi-Actor Applications with LLMs," 2024. [Online]. Available: https://github.com/langchain-ai/langgraph

[8] Qdrant Team, ``Qdrant: Vector Similarity Search Engine," 2024. [Online]. Available: https://qdrant.tech

[9] LangChain, ``LangChain: Building Applications with LLMs through Composability," 2023. [Online]. Available: https://github.com/langchain-ai/langchain

[10] LlamaIndex, ``LlamaIndex: Data Framework for LLM Applications," 2023. [Online]. Available: https://www.llamaindex.ai

[11] Q. Wu, G. Bansal, J. Zhang, Y. Wu, B. Li, E. Zhu, L. Jiang, X. Zhang, S. Zhang, J. Liu, *et al.*, ``AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation," *arXiv preprint* arXiv:2308.08155, 2023.

[12] CrewAI Inc., ``CrewAI: Framework for Orchestrating Role-Playing Autonomous AI Agents," 2024. [Online]. Available: https://github.com/joaomdmoura/crewAI

[13] S. Hong, M. Zhuge, J. Chen, X. Zheng, Y. Cheng, C. Zhang, J. Wang, Z. Wang, S. K. S. Yau, Z. Lin, *et al.*, ``MetaGPT: Meta Programming for A Multi-Agent Collaborative Framework," *arXiv preprint* arXiv:2308.00352, 2023.

[14] H. S. Al-Khalifa and M. Al-Razgan, ``Arabic Educational Chatbots: A Review of Current Status and Future Directions," *Journal of Educational Technology & Society*, vol. 26, no. 2, 2023.

[15] W. Magdy, K. Darwish, *et al.*, ``Arabic Information Retrieval: A Survey of the State of the Art," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 23, no. 3, pp. 1--38, 2024.

[16] OpenAI, ``GPT-4 Technical Report," *arXiv preprint* arXiv:2303.08774, 2023.