

yamam play

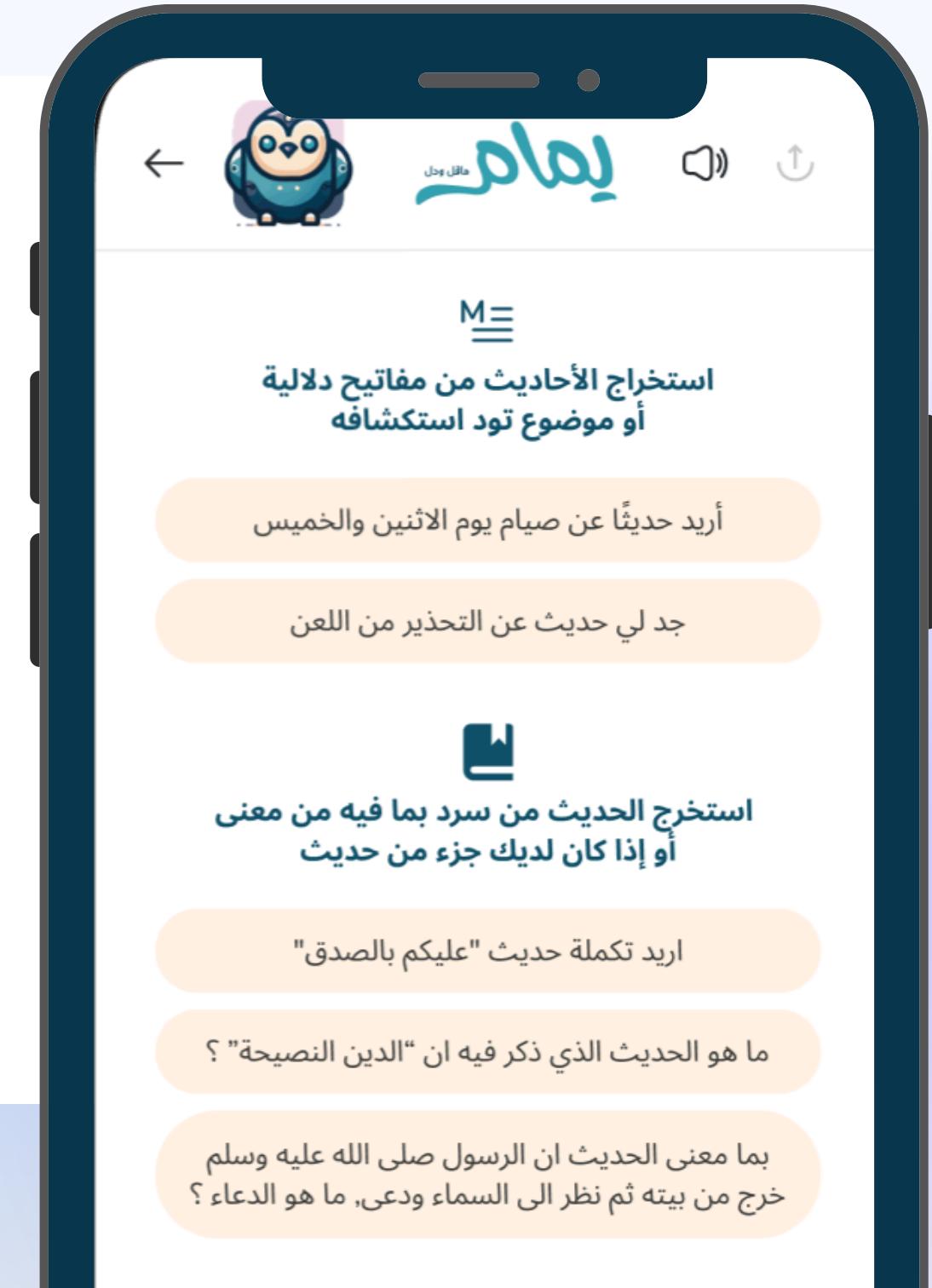
هائل ودل



Problem Description

- 1 Traditional research methods lack the ability to understand the deep dimensions of religious texts
- 2 The noticeable deficiency in comprehending linguistic contexts and deep meanings.
- 3 Traditional search engines generate results of limited relevance, requiring additional efforts and extensive time.

We present to you:



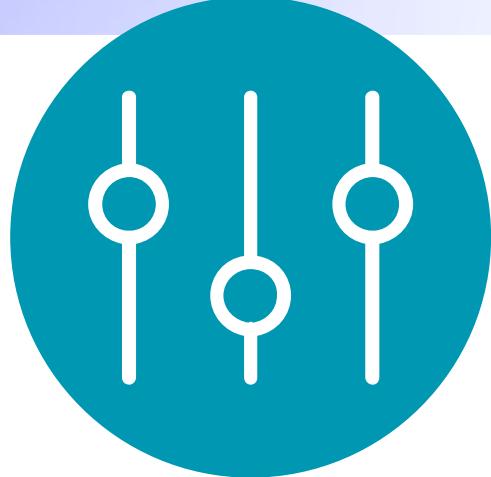
The smartphone screen shows the Yamam app's user interface. At the top, there is a navigation bar with a back arrow, a search icon (owl), and other icons. The main area has a light blue background with white text. It displays a section titled "استخراج الأحاديث من مفاتيح دلالية أو موضوع تود استكشافه" (Extracting Hadiths from Semantic Keys or Topics You Want to Explore). Below this, there are several orange speech bubbles containing Arabic text, which are examples of user queries the app can handle. The text in the bubbles includes:

- أريد حديثاً عن صيام يوم الاثنين والخميس
- جد لي حديث عن التحذير من اللعن
- استخرج الحديث من سرد بما فيه من معنى أو إذا كان لديك جزء من الحديث
- أريد تكميلة حديث "عليكم بالصدق"
- ما هو الحديث الذي ذكر فيه أن "الدين النصيحة"؟
- بما معنى الحديث أن الرسول صلى الله عليه وسلم خرج من بيته ثم نظر إلى السماء ودعى، ما هو الدعاء؟

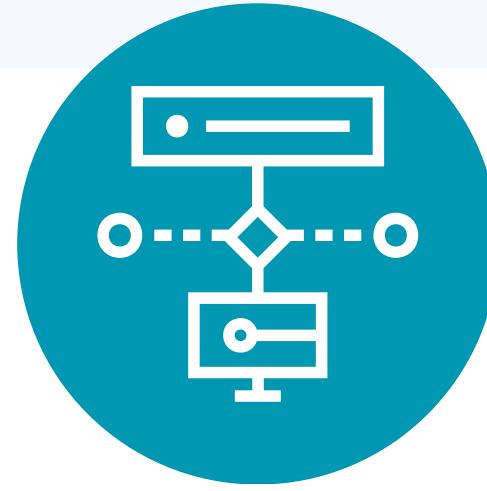
 Developing an advanced semantic search engine tailored for Prophetic Hadiths that relies on artificial intelligence technologies to understand the linguistic and semantic context of texts.

 Enabling users to search for Hadiths based on meanings and topics rather than just keywords, resulting in more accurate and in-depth results.

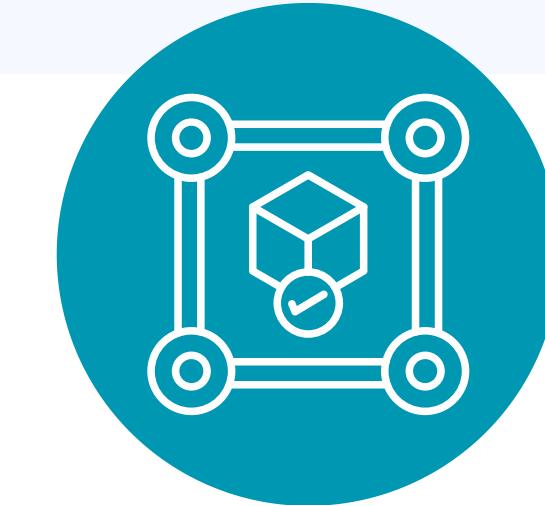
At the core of *yamam* مام



Environment Setup



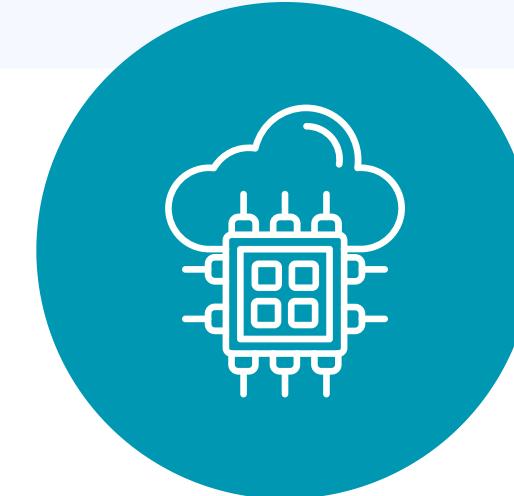
Dataset Preparation



Model Preparation



Data Processing



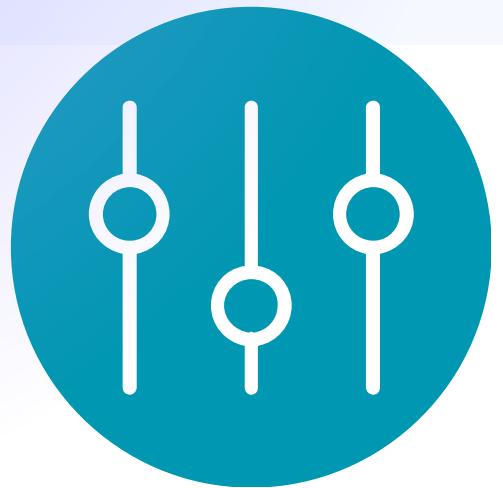
Model Training



Evaluation

At the core of *yamam* يمام

مأقل ودل



Environment Setup



The use of the following libraries :

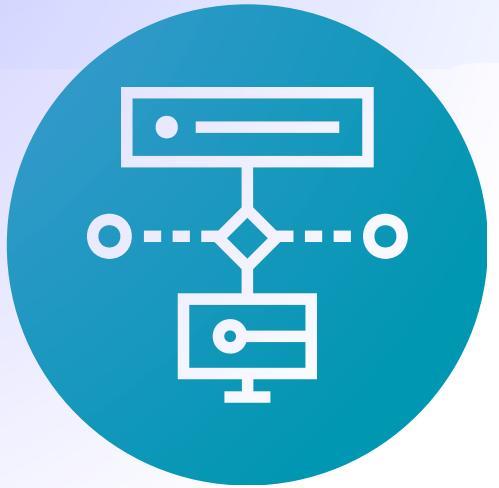
pandas: for data cleaning

transformers: for model

sklearn: for data separation

pytorch: for embedding

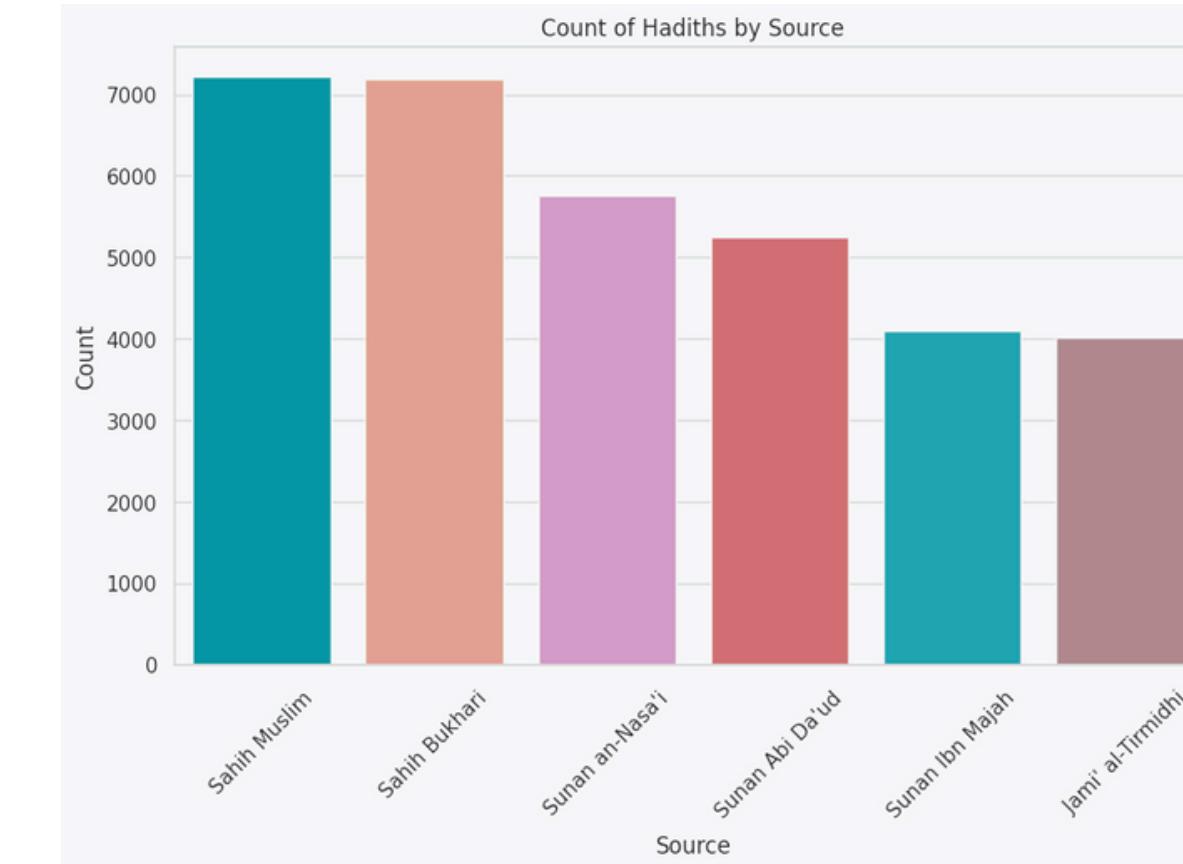
At the core of *yamam* يَمَامٌ مَّا قُلْ وَدَلْ



Dataset Preparation

- Sahih Muslim
- Sahih Bukhari
- Sunan an-Nasa'i
- Sunan Abi Da'ud
- Sunan Ibn Majah
- Jami' al-Tirmidhi

The data was extracted from the six renowned books of Hadith, which are:



The dataset originally contained 9 columns and 34,500 rows. After removing redundancy, it was reduced to 33,560 rows.

At the core of *yamam* يمام



Data Processing

1

Data Cleaning:

- Normalize text to lowercase and remove special characters using regex.
- Drop rows with missing values

2

Language-Specific Processing:

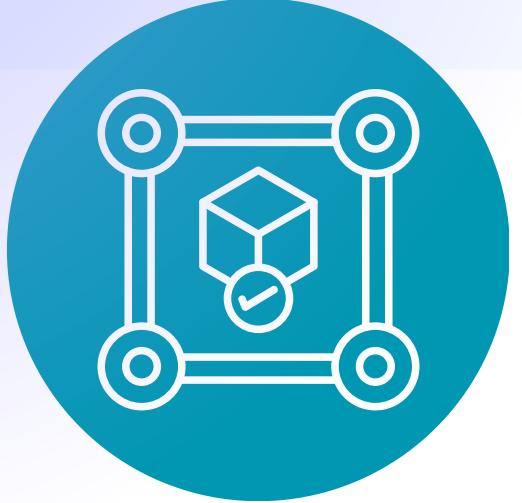
- The `remove_english_words` function filters chapter names to retain only Arabic content.

3

Data Splitting:

- Split the dataset into training, validation, and test subsets with `train_test_split`.

At the core of *yamam* يمام



Model Preparation

1

We utilized the [AraGPT](#) model, a LLM pre-trained on an extensive Arabic dataset

2

The initial step in preparing the model involved [tokenization](#) using an auto tokenizer.



At the core of *yamam* مام



Model Training

Approach 1: Language Model Fine-Tuning

- 1** Optimizer Configuration: AdamW optimizer
- 2** Training Process: the model learned from a vast corpus of Hadith texts. This process involves feeding the model examples and gradually reducing errors in prediction
- 3** Validation: to ensure that the model was not just memorizing the text but genuinely understanding it.

At the core of *yamam* يمام



Evaluation

Perplexity in Fine-Tuning:

- Used for evaluating coherence and accuracy in generated Arabic text.
- Provided insights into text quality and linguistic fluency.

At the core of *yamam* مام



Model Training

Approach 2: Embedding Generation

1

Training and validation datasets are loaded into PyTorch

- Encode Hadith texts into token embeddings for the AraGPT model.
- Train the model over several epochs with optimizers like AdamW to minimize loss and improve accuracy.

2

Regularization techniques such as dropout and layer normalization help ensure generalization and prevent overfitting.

At the core of *yamam* يمام



Evaluation

Cosine Similarity in Embedding:

- Evaluates the semantic relevance of search results to user queries.
- Measures how similar different Hadith texts are to the user's query.
- Semantic search engine computes cosine similarity scores between query and pre-computed embeddings of Hadith texts.
- Results range from 0.52 to 0.41, indicating contextual similarity and relevance.

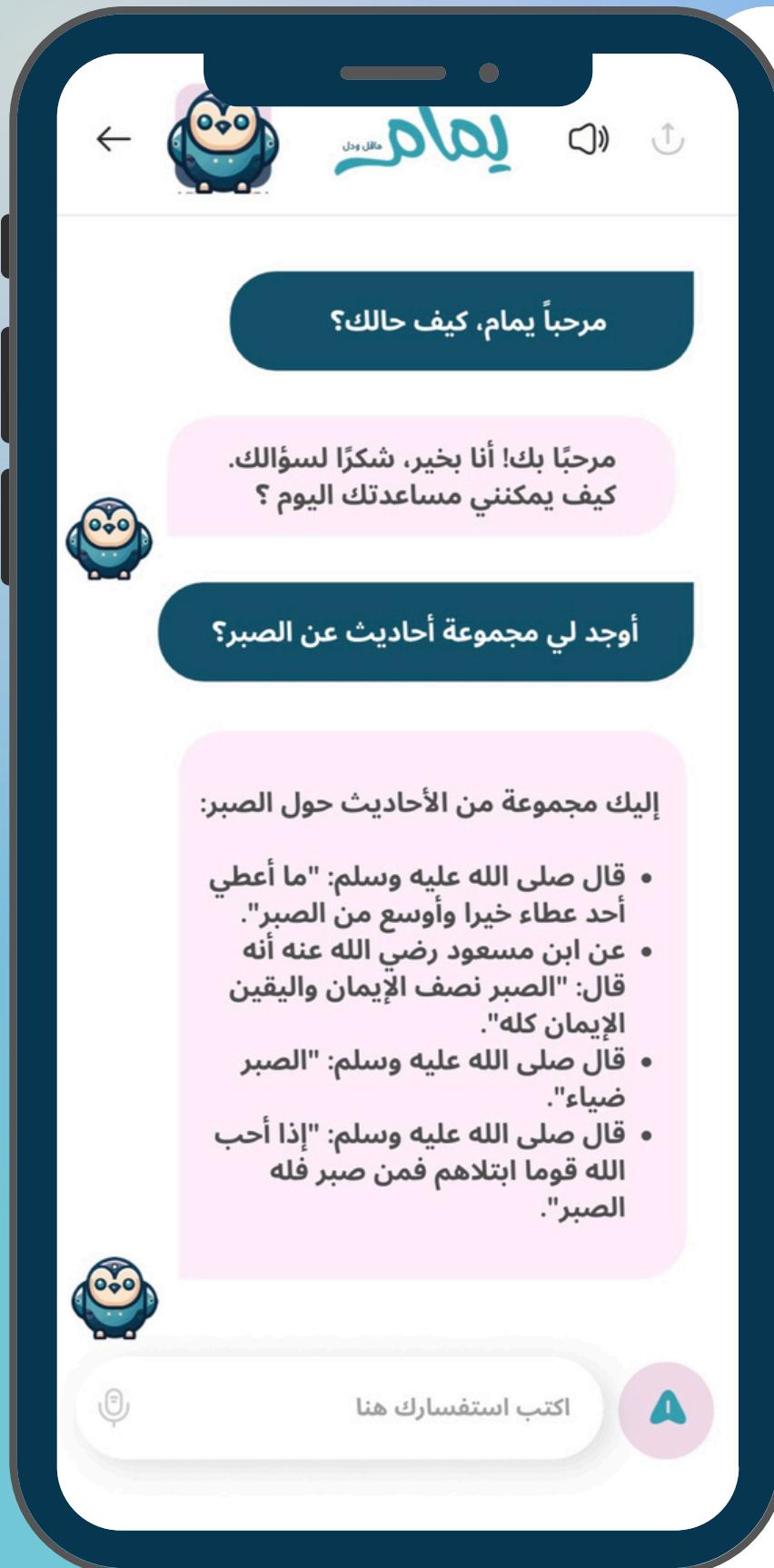
Results *yamam*

المدخل

”أوجد لي حديث عن أهمية الصلاة“

المخرج

Score: 0.5251553654670715 : حدثني حرملة بن يحيى، أخبرني ابن وهب، أخبرني يونس، ح وحدثنى أبو الطاهر، ومحمد بن سلمة المرادي قالا حدثنا ابن وهب، عن يونس، عن ابن شهاب، قال أخبرني عروة بن الزبير، عن عائشة، زوج النبي صلى الله عليه وسلم قالت خسفت الشمس في حياة رسول الله صلى الله عليه وسلم فخرج رسول الله صلى الله عليه وسلم إلى المسجد فقام وكبر وصف الناس وراءه فاقترا رسول الله صلى الله عليه وسلم قراءة طويلة ثم كبر فركع ركوعا طويلا ثم رفع رأسه فقال " سمع الله لمن حمده ربنا ولك الحمد " . ثم قام فاقترا قراءة طويلة هي أدنى من القراءة الأولى ثم كبر فركع ركوعا طويلا هو أدنى من الركوع الأول ثم قال " سمع الله لمن حمده ربنا ولك الحمد " . ثم سجد - ولم يذكر أبو الطاهر ثم سجد - ثم فعل في الركعة الأخرى مثل ذلك حتى استكمل أربع ركعات وأربع سجادات وانجلت الشمس قبل أن ينصرف ثم قام فخطب الناس قائني على الله بما هو أهلها ثم قال " إن الشمس والقمر آيتان من آيات الله لا يخسفان لموت أحد ولا لحياته فإذا رأيتموها فافزعوا للصلوة " . وقال أيضا " فصلوا حتى يفرج الله عنكم " .



Properties of *yamam* يمام



- 1 The deep understanding of textual context**
- 2 Searching based on meanings and topics**
- 3 Reducing the effort and time required to verify the results."**
- 4 The use of AI technologies**

To discover more about *yamam* يمام

هائل ودل



Presenters



Ryouf Bandar Alghamdi



Ryouf.119@gmail.com



0550771060



Rama Mohammed Alyoubi



Rama_mohammed_alyoubi@gmail.com



0500226517



Rimas Nawi Alshehri



reemasnawi@gmail.com



0547856515

Thank You