

Human Digital Twin for Personal Conversation Automation Using Supervised Machine Learning Approaches

Aya Salama

Abstract—Digital Twin has emerged as a compelling research area, capturing the attention of scholars over the past decade. It finds applications across diverse fields, including smart manufacturing and healthcare, offering significant time and cost savings. Notably, it often intersects with other cutting-edge technologies such as Data Mining, Artificial Intelligence, and Machine Learning. However, the concept of a Human Digital Twin (HDT) is still in its infancy and requires further demonstration of its practicality. HDT takes the notion of Digital Twin a step further by extending it to living entities, notably humans, who are vastly different from inanimate physical objects. The primary objective of this research was to create an HDT capable of automating real-time human responses by simulating human behavior. To achieve this, the study delved into various areas, including clustering, supervised classification, topic extraction, and sentiment analysis. The paper successfully demonstrated the feasibility of HDT for generating personalized responses in social messaging applications. Notably, the proposed approach achieved an overall accuracy of 63%, a highly promising result that could pave the way for further exploration of the HDT concept. The methodology employed Random Forest for clustering the question database and matching new questions, while K-nearest neighbor was utilized for sentiment analysis

Keywords—Human Digital twin, sentiment analysis, topic extraction, supervised machine learning, unsupervised machine learning, classification and clustering.

I. INTRODUCTION

THE concept of the digital twin can be traced back to NASA's inception of the Apollo Program. In this pioneering endeavor, NASA employed a pair of digital twins, representing two identical spacecraft, to continuously monitor the conditions and performance in the vast expanse of space [1]. A Digital Twin is essentially a digital replica responsible for simulating physical entities within virtual environments, exerting control over decisions to enhance overall performance. In contrast, an Augmented Digital Twin actively interacts with the entire real-world environment surrounding the physical entity. It's from this perspective that the concept of a Human Digital Twin emerged, prompting considerations regarding its practical implementation [2].

Digital twin technology has predominantly found application in the realm of smart manufacturing. Within this context, various methodologies leveraged blockchain technologies, while others adopted event-driven approaches. Furthermore, digital twin technology has seen extensive utilization in the healthcare sector, with semi-digital twins

A. Salama is Assistant Professor at Faculty of Mathematics and Computational Sciences, University of Prince Edward Island, Cairo Campus, hosted by Universities of Canada (e-mail: aya.salama@uofacanada.edu.eg).

proving valuable in smart healthcare by facilitating disease severity detection and preventive measures [3]–[5]. To achieve these objectives, a range of technologies have been harnessed, including cloud computing, artificial intelligence, and machine learning, to identify patterns within samples. Additionally, deep learning methodologies have been employed through the utilization of multiple deep convolutional neural networks, enabling the extraction of distinct features from patients' samples [6]–[9].

Digital Twin technology, in general, offers valuable insights, predictions, and control capabilities to assist real-world entities, enabling individuals to make informed decisions, enhance performance, and extend the lifespan of physical objects. However, the concept of a Human Digital Twin involves creating a digital replica of a human being. As previously mentioned, a Human Digital Twin can encompass various aspects, including replicating human behavior in virtual environments and monitoring health conditions for healthcare applications. Consequently, the influence of the digital twin extends to the extent of potentially substituting humans in certain aspects of their lives.

Given the demands of our fast-paced lives, people often seek virtual substitutes to assist with their daily activities. The realm of social media platforms provides an ideal virtual environment for emulating human behavior, with applications like WhatsApp, Facebook Messenger, Instagram, and numerous similar social messaging platforms effectively replicating human interactions [10]. The number of people using social messaging applications is significantly increasing in billions around the world as illustrated in Fig. 1 [11] and WhatsApp is considered the mostly used application as shown in Fig. 2. 14.8% of the internet consumers are using Instagram and 15.7% are using WhatsApp, However, people on Instagram are usually surfing images and trending videos but not doing much of talking like on WhatsApp. People usually spend about 19 hours per month to reply to their WhatsApp messages. So, the goal of this paper is to automatically reply to these tremendous number of messages to save several hours of human's life.

However, previous approaches for automatic replies generation were generic chat bots or generic conversational applications, so they were not modelling certain individuals behaviour in specific [12]. For example, in [13], artificial neural network was used to work on corpus generated from social media such as tweets to generate generic answers just as "good luck" and "good day".

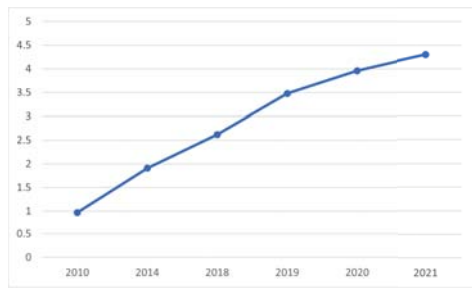


Fig. 1 Number of people using social messaging apps all over the world (y-axis in billions)

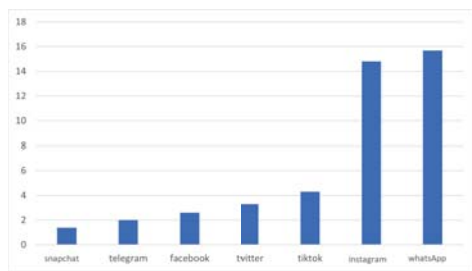


Fig. 2 Percentage of internet users on different social messaging apps

Considerable research has been dedicated to sentiment analysis in various languages, particularly in the context of social conversations. In alignment with the paper's objectives, sentiment analysis becomes essential when there are no analogous questions available for a newly posed query. This plays a crucial role in facilitating the use of emojis as responses to these questions.

Sentiment analysis is recognized as a text mining method employed for classification purposes. Consequently, it can be applied as a machine learning approach, encompassing both supervised and unsupervised methods, or as a lexicon-based approach that relies on a designated corpus [14]. The sentiment analysis can also be either document level or sentence level [15]. According to literature review, the sentence first must be determined if it is a fact or an opinion [16]. However, in this research, only personal responses were analyzed. As a result, sentiment analysis was considered in this study as pure machine learning approach to classify responses as positive, or negative.

Various classification methods were explored during the sentiment analysis phase. Additionally, a series of clustering and classification experiments were executed to group similar questions together, enhancing the precision of response generation.

Another crucial aspect of this research was topic extraction. This involved extracting topics based on key phrases. Numerous established and extensively studied key phrase extraction techniques from the field of machine learning were employed on the document corpus [17]. Key-phrase extraction from documents was based on defining significant key phrases by computing feature values for each key [18]. Other approaches also used statistical methods by giving scores for the key phrases that are extracted from documents depending on how much they are repeated and if they are

informative or not. Related to this work, some studies also used annotated data to evaluate and match key phrases [19], [20].

The paper's primary objective was to develop a Human Digital Twin capable of responding to messages in a manner identical to how the individual would reply. This digital twin would also model the user's interactions with both human contacts and digital twins of users on their friends list. Furthermore, the approach proposed in this paper involved the incorporation of sentiment analysis to ensure the use of emojis in responses for a more human-like interaction.

The paper's structure was as follows: Section II detailed the collection and processing of the examined dataset, while Section III provided an explanation of the proposed approach. In Section IV, the paper outlined the evaluation metrics and described the various experiments conducted. Finally, Section V presented the key findings and conclusions drawn from the research.

II. DATASET COLLECTION

A. Conversations Crawling

In this phase, we explored various online datasets to assess their suitability for evaluating our proposed approach. A challenge with the existing datasets is that they predominantly collect conversations between different individuals across various social media platforms. However, our paper's objective is to create a human digital twin capable of autonomously generating responses for a single individual across diverse topics. Consequently, in this study, we collected conversations exclusively from the author's WhatsApp application, all originating from one individual.

To gather this data, we opted to utilize the WhatsApp web browser interface instead of the mobile phone application. This choice was made because the WhatsApp web browser interface allowed us to collect conversations without encountering any restrictions or blocks on the extraction of individual conversations.

B. Extracting Questions and Answers

After crawling the data, questions were extracted from within these conversations. This was achieved by extracting the text that ends with question mark, exclamation mark, or having the question form (verb before subject) from all senders' sides. Then, attaching the subject's reply to that candidate question. Sample of the collected data after extracting pairs of questions and answers is shown in Table I.

TABLE I
 SNIPPET OF THE QUESTION/ANSWER KNOWLEDGE BASE

Question	Answer
How much are you paying taxes	about 25%
Do you prefer the IB education	yes
How old is selim now	3.5
Is he sick	no he is good
When is Maghrib	at 7 sharp
did you forget my birthday	you know I am so busy
do you know his email	let me check

III. THE PROPOSED APPROACH FOR GENERATING DIGITAL TWIN

The proposed approach is summarized in Figs. 3 and 4 respectively.

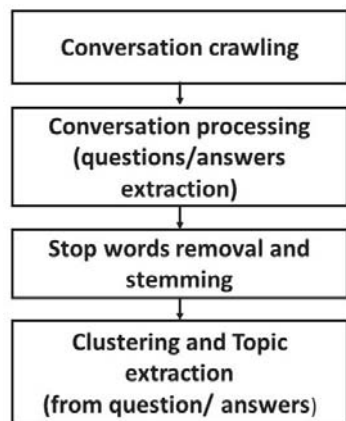


Fig. 3 Data set collection and processing

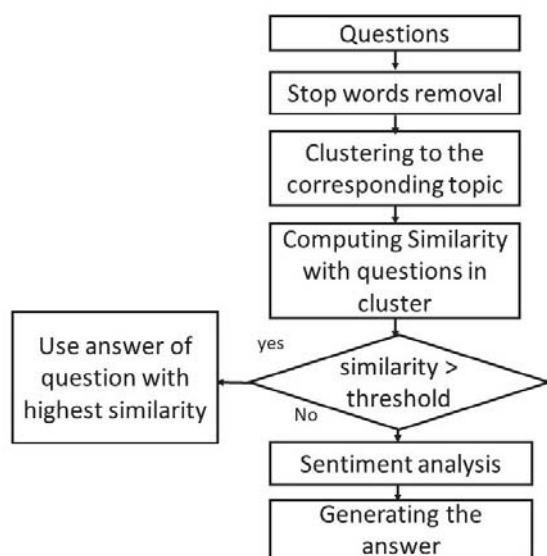


Fig. 4 Proposed approach human digital twin for answering new questions

A. Data Processing

The preprocessing was vital for removing noise so as not to affect the accuracy of finding similar answered questions.

In nearly all text mining methodologies, the initial preprocessing step involves eliminating stop words. In the context of our proposed approach, removing stop words was imperative since the reply generation relied on word frequency.

The corpora commonly employed for research with similar objectives contain numerous stop words that may play a crucial role in identifying the appropriate response. For instance, a widely used stop words removal tool, the Natural Language Toolkit (NLTK), is often applied in these scenarios [21] was

studied, but it did not include many noisy words in the collected data set such as the Franco Arabic words. Examples of the Franco Arabic words: "mashy" and "tayeb" mean okay. "Khalas" means finished and many other similar words are used frequently within all the questions in conversations. Therefore, to make sure that clustering would be efficient, list of stop words had to be built from scratch. Sample of stop words list to be removed is illustrated in Table II:

TABLE II
 SAMPLE OF THE STOP WORDS

to	yet	because
also	about	an
a	my	mine
your	from	in
tayeb	mashy	khalas

The unprocessed form of the question was retained in the database since it remained essential for the ultimate phase of associating the posed or test question with the dataset. Keywords like "who," "when," "why," and so on, were vital for pinpointing the correct response.

B. Clustering

Clustering played a pivotal role in the proposed methodology, as the effectiveness of clustering directly impacted the precision of response generation. Clustering, in this context, was responsible for determining the topic of each question, facilitating the grouping of related questions. Consequently, the question was exclusively compared to others within the same cluster.

Clustering can take on either a supervised or unsupervised approach. Unsupervised clustering involves the model working without any prior data knowledge. In contrast, supervised clustering functions as a classification model, necessitating an annotated dataset containing various types of questions [22]. In the proposed approach, supervised clustering was used to give more accurate results. In order to adopt the supervised learning approach, we constructed an annotated dataset comprising question-answer pairs organized into various topics.

For the purpose of clustering similar questions and creating a corpus of potential topics, we extracted frequent key phrases. This involved experimenting with different N-grams for this step. N-grams represent recurring features in text, and they can range from single words (unigrams), such as "son," to consecutive word pairs (bigrams) like "3 years," and even sequences of three consecutive words (trigrams), as in "three years old."

We tested unigrams, bigrams, and trigrams to determine which provided the most accurate predictions for generating responses. Following the extraction of the primary recurring words from the questions and answers, we readily assigned labels to these topics to establish a lexicon of key phrases corresponding to each of the defined topics. In Table III, an example is illustrated of a list of the most frequent words with their corresponding expected topics from the crawled conversations.

Based on the conducted experiments, it was evident that unigrams yielded the most favorable outcomes. Notably,

TABLE III
 SAMPLES KEY PHRASES FOR DIFFERENT TOPICS

Schools	education
Learn	
Teacher	
Graduate	
Bus	
Certificate	
Bag	
IB	
son	Motherhood
age	
Selim	
Children	
Pregnancy	
Vaccinate	
Trimester	

single words, or unigrams, such as "school," "teacher," and "curriculum," were particularly significant in the context of education-related topics. Subsequently, an annotated dataset was established to facilitate the training of supervised classification methods. In this phase, various approaches were put to the test. The initial algorithm under evaluation was the K-nearest neighbor, which, as per the literature review, was recognized as one of the most efficient mechanisms for research with similar objectives [23]. Nevertheless, based on the experiments conducted in this study, it was evident that Random Forests outperformed K-nearest neighbor in terms of accuracy. As a result, Random Forests were selected for implementation in the clustering process. These clusters served the purpose of organizing the processed questions while retaining their question keywords as features, a critical aspect for their use in the ultimate stage of matching questions with corresponding question-answer pairs.

C. Final Stage: Answering the Questions

Prior to addressing each question, the incoming message underwent a preprocessing step to confirm its interrogative nature. This verification was achieved by examining the presence of a question mark, related symbols, or the presence of a verb preceding the subject.

Once the question's category was determined for clustering purposes, further text processing was applied, involving the removal of stop words and stemming to refine the question. Subsequently, the raw state of the question was examined to extract the question keyword, as it played a pivotal role in the similarity measurement.

The final stage involved measuring the edit similarity between the new question and the list of previously asked questions within the cluster. A threshold was established at this juncture. If the similarity measure surpassed this threshold, the answer was employed. In scenarios where multiple questions exceeded the defined threshold, the response from the question with the highest similarity was utilized. Based on the results of our experiments, the most accurate threshold value was determined to be 0.49.

D. Sentiment Analysis

In cases where a matching question was not found for the new query, a sentiment analysis was employed. This sentiment analysis was carried out using machine learning techniques, which utilize feature vectors to effectively represent various emotional states. The objective of the sentiment analysis was to classify question-answer pairs as either positive or negative.

In this research, N-grams features were employed for sentiment analysis. Specifically, the study explored the impact of unigrams, bigrams, and trigrams on the accuracy of sentiment analysis. Furthermore, the analysis took into account negation and modification features. For instance, sentences like "I don't like parenting" and "I like parenting" both include the word "like," yet one conveys a negative sentiment, while the other conveys a positive sentiment. Therefore, the approach also encompassed the examination of negation features.

IV. RESULTS AND EVALUATION

A. Clustering Related Questions

As mentioned before in Section III, the quality of the overall system was dependent on the quality of the clustering phase. For this reason, different clustering approaches were tested.

Unsupervised learning experiment was conducted before deciding to refuge to the supervised learning approaches. K-means clustering was tested and the accuracy achieved was almost below the 30

Annotated data set was built in this work as discussed in this section to be able to apply Supervised learning approaches. Having prior information of the data set when training the model gave more efficient results than unsupervised learning.

The first tested algorithm was the K-nearest neighbor (KNN) since it was the most commonly used algorithm for similar research purposes, and it also proved its efficiency [23]. In KNN classification, the input is assigned to the cluster that got the maximum number of votes from its k nearest neighbors by measuring distance which is commonly the Euclidean distance. Random forest is also known to be powerful for purpose of clustering and classification [24]. Accordingly, random forest was also tested and compared to K-nearest neighbour. Random forest is dependent on several decision trees and each one of these tree creates its own prediction and the tree that gets the highest votes is used as the prediction model.

Naive Bayes was also tested in this experiment. Naive Bayes uses small training set to calculate the product of 2 Bayes probabilities, the prior probability and the posterior probability. Prior probability is usually conducted by previous experiment and both of them are conditional probabilities.

The evaluation of clustering was done by measuring the following metrics:

Accuracy: Is the probability that questions are assigned to the right cluster/topic [25].

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

TP, FP, FN and TN can here be summarized as in Table IV [26]:

TABLE IV
 TP, FP, TN AND FN DEFINITIONS

	Same cluster	Different cluster
Same category	TP	FN
Different category	FP	TN

As illustrated in Table IV, TP is the number of questions assigned to the correct topic. TN is the number of the questions not assigned to incorrect topic. FP is the number of questions assigned to incorrect topic. FN is the number of questions not assigned to the correct topic.

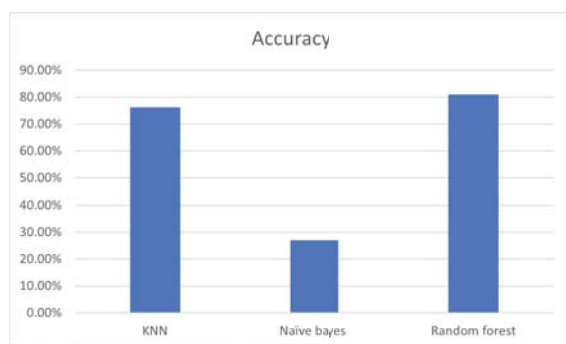


Fig. 5 Clustering results and evaluation

As shown in Fig. 5, K-nearest neighbour was supposed to result in the best accuracy. However, random forest clustering approach achieved the highest accuracy of 81% which is slightly higher than the k-nearest neighbour by almost 5%. While Naive Bayes gave the lowest accuracy of 27%. This comparison also included testing different N-grams. For the politics topic, using uni-grams was efficient, trying bi-grams did not make any difference. For the education topic, using uni-grams led to some incorrect classifications, for example, "school expensive" and "expensive", each one of them are keywords/key phrase. However, for education topic: bi-grams such as, "IB education", "American education" and "online learning" did not add to the accuracy, so using uni-grams was enough. The motherhood topic was the most challenging topic because it had too many common key phrases with the other topics. It was also the main theme topic of the crawled data set. The greetings topic such as, "happy birthday" and "happy new year" was the easiest topic because it was very similar to chat-bots. However, it gave better accuracy with using Bi-grams or even tri-grams. Furthermore, some Franco Arabic sentences were added to the corpus as key phrase such "kol sana wento tayebe" which means "wish you a happy new year".

To extract key phrases that contributed to the creation of the learning dataset for clustering, we experimented with three different approaches: the statistical approach, data annotation (key phrase extraction), and a hybrid approach combining the statistical and data annotation methods.

Data annotation involved assigning topics to each question-and-answer pair and was indispensable for the evaluation of the tested approaches. Notably, the accuracy of topic extraction using the annotated approach was twice as

high as the accuracy achieved through the statistical approach. This discrepancy emerged because, as mentioned earlier, the statistical approach was originally designed for mining extensive documents. When applied to the question-answer pairs dataset, it resulted in numerous sparse topics and a high rate of false negatives. However, combining the annotated approach with the statistical approach in the hybrid method showed a modest improvement in results of 7%.

As a final step, edit distance between the asked question and the related questions in the same cluster was computed. If similarity was higher than 0.49, then answer of the similar question was used.

B. Sentiment Analysis Experiment Using Machine Learning

According to the proposed approach, in case that the last step of finding similar answer did not work, then sentiment analysis was applied. The tested approaches were Naive Bayes classifier, K-nearest neighbour classifier, Random Forest and SVM. The input to these experiments was the question and the output was the sentiment towards these questions which was either positive or negative. To be able to represent the questions to be understood by the tested classifiers, feature vectors needed to be constructed for each question as follows: (term1: calculated frequency for term 1, term2: calculated frequency for term 2... , "positive or negative")

In this research, we adopted a supervised learning approach, which necessitated labeling all the test instances as either positive or negative. This labeling was essential for generating feature vectors that would be used to train the classifiers. We employed 10-fold cross-validation in all our experiments to ensure robust evaluation.

Before delving into sentiment analysis, the crawled questions underwent preprocessing, which entailed the removal of stop words and tokenization. The results of our experiments were meticulously analyzed to determine the optimal list of stop words, as some words exhibited an influence on sentiment analysis and could not be classified as stop words.

The tested approaches included Support Vector Machine (SVM), K-nearest neighbor, and Naive Bayes. The Support Vector Machine method involves mapping features to a higher-dimensional space, aiming to maximize the hyperplane between different classes by utilizing an optimization function to achieve the most efficient separation between the two categories: positive and negative.

To evaluate our sentiment analysis, we employed two main metrics: accuracy and root mean square error (RMSE). RMSE was used in conjunction with accuracy, as it is a commonly utilized metric in sentiment analysis, and it was calculated as in (2) [27]:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_{Pi} - Y_{Ai})^2} \quad (2)$$

where $Y_P = [y_{P1}, y_{P2}, \dots, y_{Pn}]$ is the predicted value of n samples and $Y_A = [y_{A1}, y_{A2}, \dots, y_{An}]$ is the actual value.



Fig. 6 Sentiment analysis results

10 fold cross validation was applied in all the tested approaches.

As illustrated in Fig. 6, KNN achieved the lowest RMSE and the highest accuracy of 77%. SVM and Naïve Bayes classifier achieved slightly similar accuracy of 73.13% and 74.62% respectively

C. Whole System Evaluation: Automatic Reply Generation

Based on the conducted experiments, Random forest was used for clustering and K-nearest neighbour was used for sentiment analysis as they achieved the highest accuracy results. After clustering, edit distance was measured and then if no similar question was found then sentiment analysis was applied. After testing a set of 77 new questions, the system was able to correctly answer 43 out of the testing set based on personal judgment. Sentiment analysis was also included in these results. Sample of the correct results:

Question: "Did you decide the school?"

Answer: "Yes GEMS"

Question: "Arent you tired of driving?"

Answer: "I hate driving"

While these responses were initially deemed accurate, it was imperative to subject them to further processing to ensure they aligned more seamlessly with the new questions.

As previously mentioned, the overarching objective of this research was to establish the practicality of implementing a reliable human digital twin capable of responding to everyday messages. While the results showed great promise, several challenges were encountered. One significant challenge was the limited diversity of topics within the dataset. The topics that were extracted closely resembled one another, which occasionally resulted in incorrect classification due to the prevalence of shared phrases. For example: "how schools are so expensive" in education cluster and "why is everything expensive" in politics cluster. However, such cases can be handled by scoring the different matches to clusters and then if both got the same score then both clusters had to be checked, and the answer of the most similar question in any of the matched clusters would be used. Moreover, some of the questions were very similar but required totally different answers such as: "what did you eat today?" was new question that went under the generic topic. Then it was matched with the question: "did you eat today?" so the answer was yes.

Another challenge was using Franco Arabic dialect since most of the people in the middle east uses Franco Arabic. However, in this research, Franco Arabic words was removed in preprocessing phase because it was considered as noise. Except for greetings topics as it was considered as one phrase such as "kol sana wento tayeben" as mentioned before in Section IV. Another reason that led to most of the incorrect replies was that there were no similar questions, and the sentiment analysis did not substitute well. As in the flowing example:

Question: "Did you hear about the professor who resigned today from AUC?"

Answer: emoji of a sad face.

This answer was because of the "resigned" feature which led to the negative decision by the classifier. Such errors would eventually be handled by using larger datasets. Therefore, the most important reason behind all these challenges was that the data set contained more of stories than questions.

V. CONCLUSION AND FUTURE WORK

The concept of the Digital Twin has garnered significant attention and interest among researchers due to its relevance across various domains. Furthermore, sentiment classification remains a key focus area for researchers in the fields of natural language processing and machine learning. The primary aim of this study was to demonstrate the feasibility of creating a Human Digital Twin capable of autonomously responding to social messaging applications, leveraging machine learning.

A series of experiments were conducted, encompassing both supervised and unsupervised machine learning. The study's findings led to the determination that the most effective approach involved a hybrid method. This approach combined supervised question classification using Random Forest with supervised sentiment analysis through K-nearest neighbor classification. This hybrid approach resulted in an overall system accuracy of approximately 63% which is a promising result that can be improved in future work. The primary challenge encountered revolved around the nature of the collected dataset, which primarily consisted of stories rather than questions. This highlighted the need for a more extensive knowledge base specific to questions. Nonetheless, this research paved the way for further exploration in this area, as previous work had largely focused on generic chatbots lacking the capacity for customization to individual human users. It also brought us closer to realizing the potential of Human Digital Twins, wherein these digital entities can interact autonomously, bridging the gap with human-like conversation without human intervention.

This study focused on the interactions within a single individual's WhatsApp conversations. Future research should involve crawling conversations from multiple individuals to simulate a wider range of behaviors, though obtaining consent for this can be challenging. Furthermore, the use of the Franco-Arabic dialect should be taken into account to enhance the model's adaptability. The algorithm should also be designed to learn new topics, possibly involving human interaction to label and update the model with fresh topics or questions.

Additionally, a valuable feature to add to the Human Digital Twin would be the ability to provide a summary of daily conversations, offering a concise overview of the day's interactions.

REFERENCES

- [1] R. Rosen, G. Wichert, G. Lo, R. Rosen, et al. (2015), About the importance of autonomy and digital twins for the future of manufacturing, IFAC-Papers OnLine 48-3 567–572.
- [2] hengli W.(2021), Is Human Digital Twin Possible? Computer Methods and Programs in Biomedicine Update, p. 100014
- [3] V. Damjanovic-Behrendt, W. Behrendt, (2019), An open source approach to the design and implementation of digital twins for smart manufacturing, Int. J.Comput. Integr. Manuf., doi:10.1080/0951192X.2019.1599436.
- [4] C. Mandolla, A. Messeni, Petruzzelli, al. (2019), Building a digital twin for additive manufacturing through the exploitation of blockchain: a case analysis of the aircraft industry, Comput. Ind. 109, 134–152.
- [5] D. Guivarch, E. Mermoz, Y. Marino, et al., (2019), Creation of helicopter dynamic systems digital twin using multibody simulations, doi:10.1016/j.cirp.2019.04.041.
- [6] G. Sannino, G. De Pietro,(2019), L. Verde, Healthcare systems: an overview of the most important aspects of current and future m-health applications, doi:10.1007/978-3-030-27844.
- [7] S.U. Amin, M. Alsulaiman, G. Muhammad, (2019), Deep learning for EEG motor imagerybased cognitive healthcare, doi:10.1007/978-3-030-27844-1.
- [8] A.E. Saddik, H. Badawi, R.A.M. Velazquez, et al., (2019), Dtwins: a digital twins ecosystem for health and well-being, IEEE COMSOC MMTC Commun.–Front. 14 (2).
- [9] N.K. Chakshu, J. Carson, I. Sazonov, et al., (2019), A semi-active human digital twin model for detecting severity of carotid stenoses from head vibration - acoupled computational mechanics and computer vision method, doi:10.1002/cnm.3180.
- [10] (online) Available: broadbandsearch.net/blog/most-popular-social-networking-sites.
- [11] (online) Available: <https://blog.hootsuite.com/whatsapp-stats/>
- [12] E. Adamopoulou, L. Moussiades, (2020), Chatbots: History, technology, and applications Machine Learning with Applications, 2, p. 100006,
- [13] W. Medhat, A. Hassan and H. Korashy, (2014), "Sentiment analysis algorithms and applications: A survey", Ain Shams engineering journal, vol. 5, no. 4, pp. 1093-1113.
- [14] Akhtar, M. S., Kumar, A., Ghosal, D., Ekbal, A., Bhattacharyya, P. (2017). A multilayer perceptron based ensemble technique for fine-grained financial sentiment analysis. In Proceedings of the Conference on Empirical Methods on Natural Language Processing.
- [15] dos Santos, C. N., Gatti, M. (2014). Deep convolutional neural networks for sentiment analysis for short texts. In Proceedings of the International Conference on Computational Linguistics.
- [16] J. serrano, F. Romero and J. olives, (2021), Ordered Weighted Averaging for Emotion-Driven Polarity Detection, Serrano-Guerrero J, Romero FP, Olivas JA. Ordered weighted averaging for emotion-driven polarity detection. Cognit Comput..
- [17] P. Turney., (2000), Learning Algorithms for Keyphrase Extraction. Information Retrieval, 2(4):303–336.
- [18] Tomokiyo, Takashi; Hurst, Matthew, (2003), A language model approach to keyphrase extraction, In Proceedings of the ACL 2003 workshop on Multiword expressions: analysis, acquisition and treatment, Vol. 18. Association for Computational Linguistics, Stroudsburg, PA, USA, 33-40.
- [19] Jain, S., Pareek, J., (2009), KeyPhrase Extraction Tool (KET) for Semantic Metadata Annotation of Learning Materials," 2009 International Conference on Signal Processing Systems, vol., no., pp.625-628, 15-17.
- [20] Wartena, C., Brussee, R., (2008), Topic Detection by Clustering Keywords," Database and Expert Systems Application, 19th International Workshop, pp.54-58, 1-5,
- [21] (online) Available: Edward Loper and Steven Bird. Nltk: The natural language toolkit, 2002.
- [22] Ranjan B, Schmidt F, Sun W, Park J, Honardoost MA, Tan J, et al., (2021), combining supervised and unsupervised clustering for cell type identification in single-cell RNA sequencing data. BMC Bioinform.; 22(1):186.
- [23] D. A. Adeniyi, Z. Wei and Y. Yongquan, (2016), Automated Web usage data mining and recommendation system using K-nearest neighbor (KNN) classification method, Appl. Comput. Inform., vol. 12, no. 1, pp. 90-108.
- [24] J. Dhaliwal, L. Erdman, E. Drysdale, F. Rinawi, J. Muir, T. D. Walters, I. Siddiqui, A. M. Griffiths, and P. C. Church, (2021), Accurate Classification of Pediatric Colonic Inflammatory Bowel Disease Subtype Using a Random Forest Machine Learning Classifier. Journal of Pediatric Gastroenterology Nutrition, vol. 72, no. 2, pp. 262–269.
- [25] Goadrich, M., Oliphant, L., Shavlik, J. (2004). Learning ensembles of first-order clauses for recall-precision curves: A case study in biomedical information extraction. Proceedings of the 14th International Conference on Inductive Logic Programming (ILP). Porto, Portugal.
- [26] Raghavan, V., Bollmann, P., Jung, G. S. (1989). A critical investigation of recall and precision as measures of retrieval system performance. ACM Trans. Inf. Syst., 7, 205–229.
- [27] Markus, B. H., Lancianese, S. L., Nagarajan, M. B., Ikpot, I. Z., Lerner, A. L., Wism, A. (2011) Prediction of Biomechanical Properties of Trabecular Bone in MR Images With Geometric Features and Support Vector Regression, IEEE Transactions on Biomedical Engineering, 58, 6, 1820-1826.



Aya Salama is Assistant Professor at Faculty of Mathematics and Computational Sciences, University of Prince Edward Island, Cairo Campus, hosted by universities of Canada, She has received her Ph.D degree from Faculty of Computer Science and Artificial Intelligence, Cairo University and her M.Sc from Faculty of Computer Science, The American University in Cairo and her B.Sc from Faculty of Computer Science, Ain Shams University. She has several publication in the research areas: Machine Learning, Deep Learning,

Image Processing and Artificial Intelligence.