

AUS | الجامعة الأميركية في الشارقة
American University of Sharjah

**Hosted by the Department of
Arabic and Translation Studies**

March 6-7, 2024

THE PRAXIS OF DIGITAL HUMANITIES:

**Expanding Horizons and Transforming
Scholarship in the Arab/Islamic World**



**Symposium Program
and Abstracts**

Venue: LIB 011-AUS Library

Day (1): 6th March 2024	
9:00 - 9:30	Registration + morning coffee
9:45 - 10:00	Welcoming notes
10:00 - 11:00	Keynote speech: Prof. Umar Ryad Challenges and Opportunities in Digital Humanities for Arabic and Islamic Studies
11:00 - 11:30	Coffee break
11:30 - 1:30	<p>Session (1) - Chair: Mai Zaki</p> <p>David Joseph Wrisley & Suphan Kirmizialtin Exploring Gulf Manuscript Documents with Word Embeddings</p> <p>Lisa Mischer Discovering Cultural Geography of the Classical Islamic World: GISting Historiographical Sources</p> <p>David Wilmsen Digital methodologies for assessing the age of Arabic dialect features</p> <p>Saqer Almarri Topic Modeling as Finding Aid to the OpenITI Corpus</p>
1:30 - 2:30	Lunch
2:45 - 4:45	<p>Session (2) - Chair: Lahouari Taleb</p> <p>Wafaa kamel Fayed Developing a Computational Linguistic Software for Accurate Analysis of Arabic Trilateral Abstract Geminate Verbs Morphology Using Corpus-based Methods</p> <p>Tharwat El- Sakran Integrating ChatGPT into Academic Research: Avoiding Student Stigmatization</p> <p>Mahmoud Al-Ashiri & Magdalena Lewicka Islamic Multiword Expressions and Designing Curriculums for Arabic as a Second Language Learners: Towards a Corpus-based Linguistic List</p> <p>Nawar Golley Mapping the Production of Knowledge on Women and Gender in the Arab Region (MPK)</p>
7:00 pm	Symposium Dinner

Day (2): 7th March 2024	
10:00 - 10:30	Morning Coffee
10:30 - 12:30	<p>Session (3) - Chair: Imed Nsiri</p> <p>Amani Elshimi Journey through Digital Tools: An Analysis of Abū Ḥāmid al-Ghazālī's <i>The Criterion of Action</i></p> <p>Eid Mohamed & Talaat Mohamed Racio-religious Imaginary and Discursive Formation of Arabo-Islamic Identity in Al-Manar and Al-Risala: An LDA Topic Modeling Study</p> <p>Meis Al-Kaisi Digitizing the Past, Enriching the Future: Sharjah's House of Islamic Manuscripts</p> <p>Maxim Romanov Computational typology of texts as a means of modeling the development of the Arabic written tradition</p>
12:45 - 1:45	Lunch
2:00 - 4:00	<p>Session (4) - Chair: Meis Al-Kaisi</p> <p>Mohammad Ayish Digital Humanities for Arab Media Studies: Opportunities and Challenges</p> <p>Mohammed Al-Batineh & Abdul Gabbar Al-Sharafi Machine algorithms vs human creativity: The reception dynamics of AI and human literary Translation</p> <p>Samar Magdy, Karima Kadaoui & Muhammad Abdul-Mageed Beyond the Hype: An examination of ChatGPT translation of Arabic texts</p> <p>Mohamed Abduljawad, Ahmed Alghamdi, Khalid Almubarak & Fatimah Alzamzami A High-Style Arabic Machine Translation between English and Classical Arabic</p>
4:00 - 4:30	Closing remarks

DAY 1 SESSION 1

DH Symposium Abstracts

1. Exploring Gulf Manumission Documents with Word Embeddings

Suphan Kirmizialtin, New York University Abu Dhabi, UAE
David Joseph Wrisley, New York University Abu Dhabi, UAE

Amidst the complex and largely inaccessible archival landscape of the Arabian peninsula, the digitization of thousands of documents from the India Office Records (file IOR 1/15) by the Qatar Digital Library has opened the door to unprecedented modes of computational analysis. Our talk proposes an exploration of about 14000 pages (roughly 800K words) of manumission requests—short narratives recorded by British officials in the Gulf on behalf of enslaved or indentured people in the region, and other related documents spanning the period 1880-1940. Scholars have provided general overviews into the contents of these manumission documents and have urged for such documents to be analyzed critically. They have not been examined fully, however, with an eye for an understanding of the forms and spaces of violence that they portray (Zdanowski, 2011).

Our research is predicated on two distinct phases of computational intervention. The first is text creation adapted to the multilingual and multimodal nature of the corpus; the corpus includes handwritten English documents, typed English texts, and Arabic handwritten sources. Leveraging Handwritten Text Recognition (HTR), we curate a corpus from the QDL, thereby turning this collection of manumission documents into viable searchable and computable datasets.

The second phase of computation in our research involves the application of a computational text analysis technique known as word embeddings. This variety of machine learning consists in a method for translating words in a corpus into numerical representations in order to measure the similarity (or the collocative quality) of words or groups of words—nouns, verbs, qualifiers, proper names. Exploring patterns in the Gulf manumission documents surfaces clusters and undercurrents in the corpus, data for narrating and visualizing comprehensive, distant views of the sixty-year phenomenon, and offering new avenues for interpretation. By interweaving data science methodologies and historical and linguistic investigation, we aim to shed new light on the socio-cultural dynamics of the Gulf region during the historical period just before and after the collapse of the pearling industry in the 1920s. Furthermore, as part of our analysis, we reflect on the methodological challenges of working with such a corpus—both emerging from a complex, changing colonial scenario, and remediated using transformer AI models (whose training data most likely does not contain historical materials from the region). Finally, we reflect on ethical questions of working on cultural heritage materials found in digitized colonial archives that document peoples who were enslaved only some four generations ago in the region.

2. Discovering Cultural Geography of the Classical Islamic World: GISTing Historiographical Sources

Lisa Mischer, University of Hamburg, Germany

Biographical dictionaries are a keystone of Arabic and Islamic historiography. Recent advances in Arabic NLP allow us to harvest this information at a large scale. In our project, we are currently working with 100 texts - aiming eventually to include some 400 texts - covering the period of c. 600-1600 CE.

Applying our CAMEL-BERT-based fine-tuned model for toponym extraction and using our historical gazetteer to visualize the results, we can identify the geographical coverage of each and every biographical and historical text. These analytical maps not only give us insights into each source but also help us understand the development of the biographical and historiographical tradition in Arabic.

For example, we can easily distinguish sources that have a highly localized focus (maps of Andalusian sources have a strong focus on al-Andalus and al-Maghrib) and those that have a more general coverage (for example, vast collections from the Mamluk period covering the entire Islamic world by that period more evenly). We can see the correlation between the period, subject, and geography (for example, all dictionaries focusing on the Companions of the Prophet have similar geographies, revolving around Arabia, Lower al-Iraq, and al-Sham; while dictionaries of legal schools show their distinct geographies at specific periods). We can even get a sense of the expansion of the Islamic world over time by combining the data from these sources into a chronological representation.

In the paper, we will present our preliminary results, but already expanded onto some 400 texts, and will discuss what other insights this data can provide us with. Last but not least, we will discuss the technicalities and challenges of our approach as well as how the complexity of this data can be increased to provide us with more qualitative insights into the cultural geography of the classical Islamic world.

DAY 1 SESSION 1

DH Symposium Abstracts

3. Digital methodologies for assessing the age of Arabic dialect features

David Wilmsen

One of the more vexing questions in Arabic linguistics is that of the origins of the modern dialects. Are they descendants of fuṣḥā or are they independent branches descended from the same parent as fuṣḥā? Answering the question has proven elusive for lack of abundant attestations of Arabic before the Islamic era. Attestations from the earliest studies of the language, dating to the eighth century AD (1st/2nd centuries AH), indicate that some salient dialect features have been present since that time and probably before. Digital analysis of early Arabic texts provides reason for concluding that some dialect features of the present day were probably present before the codification of fuṣḥā in the eighth century into the transnational standard that it was to become.

Dialect phonologies, such as the palatalization of the voiceless velar stop [k], the so-called kashkasha, that is known to have been present in the 8th century, are difficult to assess in texts, because the nature of the writing does not provide the means for assessing some phonological variation. On the other hand, morphosyntactic variants are easier to assess, because they can be reflected in writing. Digitally assisted analysis of large texts can uncover variants in the texts that even close reading may fail to detect. The means for doing this are either with large electronically searchable corpora, incorporating many texts, or in searching individual electronically readable texts that are becoming increasingly available. Reliable results also require consulting variant manuscript readings, themselves becoming increasingly available in electronically readable formats. A few such variants, among others, are the use of ascriptive negator laysa in uninflected form and the use of pronominal object marker iyyāhu as a demonstrative. Techniques for discovering such rarities in writings and the implications of the discoveries themselves will be discussed.

4. Topic Modeling as Finding Aid to the OpenITI Corpus

Saqer Almarri, New York University Abu Dhabi, UAE

The OpenITI corpus comprises of 6,785 unique texts, and 2.2 billion words for the study of Arabic and Islamic(ate) intellectual traditions (Nigst et al. 2023). The OpenITI has a structure that allows for distant reading, and large-scale data analysis of the written tradition. Yet, workflows for how to use OpenITI effectively for the close reading of texts remain open and underdeveloped. When faced with specific literary or conceptual usages, one faces a daunting task of finding examples of such material among the corpus's 2.2 billion words. To address this, I propose a data-gathering workflow that uses the OpenITI corpus as a first stage to close reading analysis through the use of topic modeling methods.

I combine the use of AntConc and R-based LDA topic modeling to find variation in the use of concepts and statements (Anthony 2023, Blei 2012a and b). This data gathering method allows the scholar to conduct analyses on these concepts and statements that trace their development through re-use or citations over time and genre. My examples to demonstrate this workflow involves examining a topic model of all attestations of the conceptual term 'خنتى' across a variety of genres in the OpenITI, and similarly the development of the prophetic tradition 'علمان العلم' as authors make use of it within non-hadith literature across time.

DAY 1 SESSION 2

DH Symposium Abstracts

5. Developing a Computational Linguistic Software for Accurate Analysis of Arabic Trilateral Abstract Geminate Verbs Morphology Using Corpus-based Methods

Wafaa Kamel Fayed, Cairo University, Egypt

This research paper addresses the complex challenge faced by both learners and users of the Arabic language: the precise identification of morphological patterns and categories associated with the imperfect verb, particularly focusing on the trilateral abstract geminate verb. The ultimate objective of this project is to create a software application that streamlines this identification and determination process, thereby enhancing the learning journey and elevating proficiency in the Arabic language. In order to realize this objective, an exhaustive investigation was conducted into the relevant verbs, drawing from diverse Arabic dictionaries including *Al-fayrūzabādī's al-Qāmūs al-muḤIT Dictionary*, as well as entries from *Ibn MaḌūr's Lisān al-ḡarab* and *az-Zabidī's tāj al-ḡarūs min jawāhir al-qāmūs*.

This process yielded a comprehensive database of 562 entries, meticulously tabulated to include detailed morphological conjugations. Notably, considerations were extended to encompass the points and range of articulation, as these factors exert influence on the verb patterns. Subsequently, the developed software application underwent rigorous testing involving diverse groups of language users and learners, totaling 119 participants. Each participant was presented with a distinct test comprising a list of 10 verbs, tasked with determining the appropriate imperfect form. The accuracy of responses ranged from 40.67% to 50.80%, underscoring the significance and necessity of the software's development. Moreover, the study unveiled four distinct categories to which the verbs under investigation belong, presenting a pivotal revelation in the realm of Arabic verb analysis.

6. Digital Humanities for Arab Media Studies: Opportunities and Challenges

Mohammad Ayish, American University of Sharjah, UAE

Kirschenbaum defines Digital Humanities (DH), as "a field of study, research, teaching, and invention concerned with the intersection of computing and the disciplines of the humanities." As a by-product of the digital information and communications revolution of the late 20th Century, Digital Humanities is widely viewed as a disruptive interdisciplinary academic movement that involves collaborative, transdisciplinary, and computationally engaged research, teaching, and publishing in a wide range of areas, including media and communication (Burdick, et. al. The digital transitions in the media field have brought in new visions and methodologies for addressing human problems in virtual and digital communication settings. Yet, there seems to be no consensus among scholars about the value of Digital Humanities for media studies. DH proponents argue that the new field helps in capturing what it means to be "human" and "humane" in an increasingly complex cyberspace. On the other hand, DH opponents think Digital Humanities does not necessarily seem to fully promote the critical spirit of media studies as an area of knowledge that seeks to help understand and analyze how the media sphere is playing out in our human living experience. Across the Arab media studies landscape, Digital Humanities seems to have slowly established some foothold as the Web, social media networks, and artificial intelligence continue to define the Arab World's communications terrain. More Arab media researchers and educators are joining the Digital Humanities bandwagon by focusing on online and digital media topics and embracing Big Data analytics (both quantitative and qualitative), online data collection, and instructional methods. This paper focuses on the Digital Humanities features of Arab media studies (media scholarship and education) in the past decade. It will identify key DH trends by analyzing topics and methodologies across media education programs and scholarly works.

The writer argues that while Digital Humanities provides a new pedagogical and scholarly momentum to Arab media studies, we need always not to lose sight of the human essence embedded in Arabs' engagement with digital and online media channels. The analysis of Big Data sets can never be a substitute for qualitative, face-to-face interaction with human subjects (audiences or consumers). AI-supported algorithms are not the best tools to help us understand genuine human insights and emotions. There is always more to investigating human communication behavior (offline or online) than sentiment data analysis. The proliferation of fake news and deep-fake practices clearly undermines the quality of our engagement with online knowledge about audiences' attitudes and behaviors. If Digital Humanities is to be successfully integrated into Arab media studies, it should be always coupled with well-informed human interventions in addressing media issues and events.

DAY 1 SESSION 2

DH Symposium Abstracts

7. Islamic Multiword Expressions and Designing Curriculums for Arabic as a Second Language Learners: Towards a Corpus-based Linguistic List

Mahmoud Al-Ashiri, Qatar University, Qatar

Magdalena Lewicka, Foreign Language Academy, Torun, Poland

The aim of this study is to compile a list of religious Collocations (i.e. referring to Islam) and to present the proposed methodology for their classification. Classification not based on syntax pattern, linguistic structure, or lexical core, but based primarily on didactic goals. This is to help in the creation of curricula for teaching Arabic as a foreign language, considering the importance of knowing these collocations for the learner and the legitimacy of presenting it to him at a given level of language learning. Through this study, the authors attempt to compile the broadest possible list by means of manual assembling and semi-automatic computer work.

The manual method is understood as extracting collocations from dictionaries, specialist religious books and Arabic language textbooks, while semi-automatic computer work is understood as extracting them from a corpus containing eighty-one million words. In this way, a list of collocations will be created, and the basis for their classification will be significant linguistic and functional patterns.

8. Mapping the Production of Knowledge on Women and Gender in the Arab Region (MPK)

Nawar Al-Hassan Golley, American University of Sharjah, UAE

My presentation will introduce the audience to the international, collaborative, and interdisciplinary research project titled "Mapping the Production of Knowledge on Women and Gender in the Arab Region (MPK)". It will explain how the project came about, the objectives of the project, the strategies that the researchers adopted in their work, and the perceived benefits of the project. The MPK is a project of the University of California Davis Arab Region Consortium (UCDAR) in partnership with the Arab Council for the Social Sciences. It aims to evaluate the key theoretical and methodological frameworks guiding research in the field, the role of funding agencies in informing research, the role of local states in shaping knowledge production on women and gender in the Arab region, and the patterns of findings.

The MPK project aims to build a database of, and critically analyzes, the research on women and gender in the Arab region from 1970 to the present. The project assembles and critically evaluates over 50 years of research on the thematic of socialities: class, race, minorities, family and kinship; media and popular culture; religion; war, violence, refugees; space. The MPK database and critical reviews, which will go public as soon as possible, aim to provide early career scholars with resources to help them push the frontiers of research towards innovative work supporting social justice and equality. This is going to be the first publicly available comprehensive database and critical analyses of half a century of scholarship on women and gender in the Arab region.

DAY 2 SESSION 1

DH Symposium Abstracts

1. Journey through Digital Tools: An Analysis of Abū Ḥāmid al-Ghazālī's The Criterion of Action

Amani Elshimi, American University in Cairo, Egypt

Exploring how the Islamic worldview anchors the Muslim perception and conceptualization of the notions of education, knowledge, science and learning, this presentation selects as a case study Abu Hamid al-Ghazali's "The Criterion of Action" - a seminal Arabic work written by the renowned Islamic thinker in the late eleventh century - closely examining the linguistic and rhetorical devices employed. While various studies in the literature have had Islamic knowledge as their main concern, very few, if any, have looked at the concept through the lens of linguistic and rhetorical devices that reflect worldview. Both qualitative and quantitative methods are used for the study. The qualitative analysis is supported by Norman Fairclough's framework for Critical Discourse Analysis, and the quantitative analysis is conducted, using Lawrence Anthony's corpus concordancer, AntConc (version 4.0.7, released on April 10th, 2022).

The mixed methods together shed light on the linguistic choices made repeatedly, revealing the schemata that govern the perception. Conditional structures, speech acts, conceptual metaphors and imperatives are example structures analyzed to reveal aspects of the socio-cultural context, assumptions, aspirations, attitudes, power dynamics, and other factors that are shaped by the Islamic worldview and that, in turn, influence knowledge and education. The implications of the study are significant in their potential to understand and impact current philosophies of Islamic education. Alongside the research conclusions, this presentation also traces the research process, where multiple digital tools were examined, cast away, and/or replaced, eventually leading to a modification of the research question to overcome digital challenges in working with Arabic text. Sharing this experience is significant, since researchers rarely make public the challenges, confusions, faults and inconsistencies they go through. Young or novice scholars often read published literature, seeing only the end-product of research, and thinking the process is linear, neat or flawless. Revealing the obstacles along the journey may dispel such research myths.

2. Racio-religious Imaginary and Discursive Formation of Arabo-Islamic Identity in Al-Manar and Al-Risala: An LDA Topic Modeling Study

Eid Mohamed, Qatar University, Qatar

Talaat Mohamed, Jouf University, KSA and Al-Azhar University, Egypt

This paper focuses on the dynamics of culture, language, and race as integral to the discourse on Pan-Islamist/Pan-Arab national identity in Egypt during the premodern era, as understood through computational analysis. It engages the discursive formations of nationalism and racism and concomitant discursive negotiations of identity. This critical intervention aims at identifying racially marked nationalist discourse or language through zooming in on conceptual developments of certain racio-ethnic, racio-religious and racio-political clusters or cooccurrences.

This analysis is meant to trace the intellectual development in Rida and Al-Zayyat's writings through tracing conceptual shifts in his view of the 'other' on different levels, ethnic, social, cultural, religious, and whether that other was a minority at home or a foreign other. The discursive formations of the other in Al-Manar and Al-Risala relatively reflect Rida and Al-Zayyat's somewhat ambiguous approach to reform, being torn between reform pioneers' (like Al-Afghani and Abduh's) rational liberalism and conservative Salafism (puritan Wahhabism) and traditionalism. The identity discourse in both magazines - as manifested in literary and political writings - involves notions of race and ethnicity, selfness and otherness, as integral to the identity formation discourse, with the literary, the political, the social and the religious, entwining in the service of their largely hybridist Arabist ideology.

DAY 2 SESSION 1

DH Symposium Abstracts

3. Digitizing the Past, Enriching the Future: Sharjah's House of Islamic Manuscripts

Meis Al-Kaisi, American University of Sharjah, UAE

In the ever-evolving landscape of knowledge preservation, the digitization of Arabic / Islamic literary heritage stands as a transformative endeavor. This paper traverses the comprehensive journey of digitization at Sharjah's House of Islamic Manuscripts, from the ongoing projects and their challenges to the democratization of knowledge through technology, interdisciplinary collaborations, preservation, cultural heritage conservation, and the promising future prospects.

In its digitization endeavors, the House currently finds itself harnessing what might be considered primitive technology. The tools at its disposal, while functional, lack the sophistication and capabilities of advanced digitization methods. It's important to acknowledge that it is essentially in its infancy when it comes to digitization.

To underscore the House's need to overcome these challenges, we aim to bring them to the forefront of discussions. These challenges are not insurmountable obstacles but rather stepping stones toward progress. This presentation will share the achievements, however modest they may seem, as they signify significant milestones in the quest for comprehensive digitization. During our discussions, we will focus on the work that the House has accomplished thus far. We believe it is essential to shed light on the remarkable outcomes it has achieved despite the current limitations. These achievements serve as testaments to the dedication and resilience, encouraging the House to strive for greater success in the future. Moreover, our conversations will extend to explore potential solutions and strategies that can help the House navigate the complex landscape of digitization more effectively. We recognize that the path ahead may be challenging for the House, but these are opportunities for growth and innovation. With a clear vision and determination, the House can look forward to a future where its digitization efforts will be more sophisticated and transformative. Our current project and efforts serve as the foundation upon which we hope to help the House build more advanced and comprehensive digitization programs in the future. These early stages allow us to explore the possibilities and potential for development in the realm of digitization for Sharjah House of Islamic Manuscripts. By acknowledging and understanding the current limitations, we anticipate the growth of capabilities, the acquisition of advanced tools and techniques, and the expansion of expertise.

4. Computational typology of texts as a means of modeling the development of the Arabic written tradition

Maxim Romanov, University of Hamburg, Germany

Based on machine learning, computational modeling may allow us to address questions that have been difficult to imagine until quite recently. What are the main types of texts within the Arabic written tradition? (historical, biographical, legal, exegetical, grammatical, etc.) How did they evolve over the course of many centuries of Islamic history? How did they intertwine and interconnect? In this paper, I will showcase how different types of texts from the Arabic written tradition can be modeled and evaluated. For example, a model can be trained to recognize texts that resemble preselected exemplars of a specific type. Such a model can then be used to assess large quantities of texts—in our case, the OpenITI corpus with over 8,000 unique texts covering the entire span of Islamic history—in order to identify all texts that exhibit characteristics similar to the type we have modeled. An iterative withholding approach can be used to examine what secondary types are “mixed” into the main type that we are analyzing. For instance, after identifying all exegetical texts (tafsir), we can apply another model trained to recognize various types, with the exegetical type explicitly excluded.

This approach enables us to identify subtypes of exegetical texts that lean towards other types of texts. Thus, we discover that some are strongly inclined toward law (fiqh), others toward Tradition (Hadith or Prophetic Sira), and still others toward grammar (nahw) or scholastic theology (kalam). Furthermore, these models can be applied for the “rolling” evaluation of texts, in which texts are divided into consecutive overlapping sections, and each section is evaluated separately. This approach allows us to create a detailed typological map for each text. This approach is particularly useful for the analysis of extensive compilatory texts, which are often composed of sections of different types. The paper will present the overall approach, discuss some early findings, and speculate on the broader implications of such research.

DAY 2 SESSION 2

DH Symposium Abstracts

5. Integrating ChatGPT into Academic Research: Avoiding Student Stigmatization

Tharwat M. El-Sakran, American University of Sharjah, UAE

The advent of Chat GPT, a potent and conveniently accessible technological tool, has recently raised apprehensions regarding its potential to instigate plagiarism within educational environments. In response, academic institutions have commenced the dissemination of cautionary advisories to educators, urging the incorporation of explicit notifications within course outlines. These notifications serve to forewarn students that resorting to the utilization of this application may jeopardize their academic integrity, consequently resulting in a forfeiture of credit for their submitted work. While these concerns possess a degree of justifiability, it was anticipated that higher education institutions would facilitate preparatory instructional sessions for instructors, equipping them with adeptness in navigating Chat GPT and harnessing its pertinent attributes within pedagogical contexts.

This study is dedicated to acquainting the educational community with the functionality of Chat GPT, whilst concurrently delineating methodologies by which tertiary-level scholars can be galvanized to leverage its potential for the composition and refinement of their scholastic research papers. This intentional integration is envisioned to seamlessly align with the prescribed learning objectives as stipulated within syllabi and the didactic frameworks of academic writing curricula. The contention put forth by the author resides in the notion that the propensity for plagiarism through Chat GPT is contingent upon the inherent structure of the academic course in which it is implemented. As such, pedagogical facilitators are enjoined to embrace bespoke instructional paradigms and evaluative approaches, the exegesis of which constitutes a pivotal focal point of this investigation.

6. Machine algorithms vs human creativity: The reception dynamics of AI and human literary Translation

**Mohammed Al-Batineh, United Arab Emirates University, UAE
Abdul Gabbar Al-Sharafi, Sultan Qaboos University, Oman**

Machine Translation (MT) has witnessed significant breakthroughs with the advent of deep learning techniques and transformative language models like the Generative Pre-trained Transformer (GPT). These models, trained on massive datasets, employ advanced neural networks that 'understand' the nuances of language. This enables the models to 'learn' the structure, grammar, and vocabulary of a language, which helps them produce more fluent and accurate responses and translations. Despite the advanced research in the areas of MT, the application of GPTs in literary translation remains underexplored. To fill this gap, this study examines the evolution of GPTs and their potential use in literary translation and investigates the reception of GPT-based literary translation from Arabic into English in comparison to human translation, with a focus on target text quality parameters, such as clarity, fluency, and naturalness. Furthermore, this study investigates whether readers can distinguish between human-translated and AI translated literary excerpts to reveal the extent to which AI can mimic human-like translations.

Accordingly, this study aims to assess readers' willingness to read and purchase AI-translated novels under different pricing conditions. The data were collected through an online survey targeting English readers with varying reading habits and educational backgrounds. The findings revealed that the AI translation was perceived as idiomatic, fluent, and emotionally resonant, with participants unable to distinguish it from human-translated text. Despite the high quality of the AI-translated text, the participants showed reluctance to purchase AI-translated novels unless they were priced cheaper than human translations. While many participants were willing to read high-quality AI-translated literature, a significant proportion remained skeptical about AI matching human expertise. This study underscores the urgent need for further research on AI-translated literature and seeks to provide insights for publishers, authors, translation professionals, and educators to understand the evolving landscape of AI-based literary translations.

DAY 2 SESSION 2

DH Symposium Abstracts

7. Beyond the hype: An examination of ChatGPT translation of Arabic texts

Samar M. Magdy, Mohamed bin Zayed University of Artificial Intelligence, UAE
Karima Kadaoui, Mohamed bin Zayed University of Artificial Intelligence, UAE
Muhammad Abdul-Mageed, Mohamed bin Zayed University of Artificial Intelligence, UAE

ChatGPT (OpenAI, 2022) has recently emerged as a powerful platform for natural language processing. Nevertheless, it is not immediately clear how this generative AI model can handle intricacies of the host of languages and varieties we collectively refer to as Arabic in the context of machine translation (Almahairi et al., 2016; Habash & Hu, 2009; Junczys-Dowmunt et al., 2016; Mohamed et al., 2012; Salloum & Habash, 2011). This is a particularly important question given that differences between these varieties cut across almost all linguistic levels (lexical, semantic, syntactic, etc.) (Abdul-Mageed et al., 2020; Nagoudi et al., 2021; 2022). In this paper, we explore this question with a focus on Classical Arabic (CA), the variety used in the religious discourse as well as early literature during the first and second AH centuries. In particular, we carry out an empirical investigation of ChatGPT's ability to translate CA into English. We compare this ability to the model's capacity to translate Modern Standard Arabic (MSA), the modern variety of Arabic used in educational and governmental settings, and various Arabic dialects (i.e., Algerian, Egyptian, Emirati, Jordanian, Mauritanian, Moroccan, Palestinian, and Yemeni).

For each variety, we randomly sample 200 sentences from respective sources. These sources include classical books in the case of CA, news outlets for MSA (i.e., Aljazeera and BBC Arabic), and transcriptions of YouTube videos for the dialects. We manually translate these sentences into English. Translation is carried out by professionals to guarantee high quality. We then design efficient templates to prompt ChatGPT to acquire translations in each of these varieties. For evaluation, we employ both automated methods such as BLEU score (Papineni et al., 2002) and COMET (Rei et al., 2020). We also carry a human error analysis to identify areas where the model has issues. We also present our findings using descriptive statistics. Our main findings suggest that ChatGPT can carry out machine translation of CA at a sensible level, with some occasional errors such as in cases of polysemy and homonymy, but still lags behind the level of a professional translator. In addition, our careful error analysis covering all varieties studied reveals various areas where the model struggles especially on dialects. These include translation of named entities, idiomatic expressions, and certain cases of tense. Based on our study, we discuss a number of use cases of state-of-the-art machine translation technologies in Arabic and Islamic digital humanities, such as CA translation in museums, multilingual text analysis, and digitized manuscript collections.

8. A High-Style Arabic Machine Translation between English and Classical Arabic

Mohamed Abduljawad, United Arab Emirates University, UAE
Fatimah Alzamzami, University of Ottawa, Canada
Khalid Almubarak, Prince Sattam bin Abdulaziz University, KSA
Ahmed Alghamdi, The University of Edinburgh, Scotland

The concept modernization that took place during the 20th century has significantly influenced the development of Modern Standard Arabic (MSA), resulting in numerous Arabic lexical, phrasal, and stylistic features that began to resemble loan translations mainly from English; and hence existing machine translation systems from English to Arabic still predominantly deploy MSA-based translation methods, resulting in translations that sound foreign and disconnected from the Fasih or Classical Arabic (CA) expressions.

Moreover, due to the reliance of relevant machine translation datasets mainly consisting of MSA texts and/or translations, the gap between Modern Arabic translations and the elegant, idiomatic aspects of Classical Arabic, has been increased. Hence, this work is proposed to address this issue and introduce a revolutionary AI-powered solution aimed at providing a high-class Classical Arabic translations, presenting them in the very original form of Arabic "Fasih Arabic style" In this work, we present "Rasa'aef"; a new English-Classical Arabic parallel corpus constructed using high-quality historic Arabic books of well-known Arabic eloquent writers such as al-Shafi'i, al-Jahiz, and Ibn Hazm. These books containing 80,000 pairs of English and Classical Arabic sentences, have been translated into English by native to near native translators who are highly skilled in both English and Classical Arabic languages. We adopt different neural network architectures to train multiple neural machine translation (NMT) models using our proposed dataset "Rasa'aef". To further enhance the translation performance, we utilize transfer learning using in-domain datasets (i.e. relevant Classical Arabic datasets) such as Qur'an and Hadith to avoid model learning confusion that might occur when using non-Classical Arabic forms like MSA or informal dialects. Finetuning BERT and araGPT2 pre-trained models using the Rasa'aef dataset have yielded a translation performance of 8 BLEU scores whereas a significant translation performance have been achieved (BLUE score of 18.51) when finetuning Facebook's NLLB pre-trained language model using our proposed dataset. We believe that this work offers a significant contribution as it presents a new parallel English-Classical Arabic corpus which fills a significant research and translation-system gap. Not only this, our experiments have also illustrated a reliable machine translation model which in turn proves the validity of our proposed high-quality corpus and modeling approach.



Scan the QR code for
more information

THE PRAXIS OF DIGITAL HUMANITIES:

**Expanding Horizons and Transforming
Scholarship in the Arab/Islamic World**

