

An Evaluation of the Automatic Detection of Hate Speech in Social Media Networks

The Case of Arabic Posts on Facebook Regarding France's Muhammad Cartoon Controversy

Abdulfattah Omar¹

Department of English, College of Science & Humanities
Prince Sattam Bin Abdulaziz University, KSA
Faculty of Arts, Port Said University, Egypt

Mohamed Elarabawy Hashem²

Department of English, College of Science and Arts in
Tabarjal, Jouf University, KSA
Al-Azhar University, Cairo, Egypt

Abstract—Numerous approaches have been developed over recent years to detect hate speech on social media networks. Nevertheless, a great deal of what is generally recognized as hate speech cannot yet be detected. There remain many challenges to assuring the effectiveness and reliability of automatic detection systems in different languages, including Arabic. Social media platforms and networks such as Facebook continue to encounter difficulties regarding the automatic detection of hate speech in Arabic content. Given the importance of developing reliable artificial intelligence and automatic detection systems that can reduce the problems and crimes associated with the spread of hate speech on social media platforms, this study is concerned with evaluating the performance of the automatic detection and tracking of hate speech in Arabic content on Facebook. As an example, the study evaluates the period in October 2020 that came to be known as France's cartoon controversy. Two different corpora were designed. The first corpus comprised 347 posts deleted by Facebook, now known as Meta. The second corpus was composed of 1,856 posts that were randomly selected using the hashtag *الإسلام لله* (except the Prophet of Allah). The results indicate that there is a considerable amount of hate speech taken from or influenced by the Islamic religious discourse, but that automatic detection systems are unable to address the peculiar linguistic features of Arabic. There is also a lack of clarity in defining what constitutes "hate speech". The study suggests that social media networks, including Facebook, need to adopt more reliable automatic detection systems that consider the linguistic properties of Arabic. Political thinkers and religious scholars should be involved in defining what constitutes hate speech in Arabic.

Keywords—Artificial intelligence; automatic detection; Facebook; hate speech; Islamic discourse; social media networks

I. INTRODUCTION

In recent years, the spread of social media networks and platforms has resulted in the emergence of different forms of hate speech, which have negative impacts on the stability of societies [1]. Millions of users around the world today use these social media networks and platforms to spread hate against specific groups and individuals [2, 3]. It is clear that hate speech has a central role in various discussions, including those on immigration, politics, sports, religion, and even diseases [4-6]. Hate speech has also been associated with crime, racial hatred, and violence [7, 8]. In the face of the

increasing threats posed by hate speech to the lives of individuals and societies, social media networks have adopted a range of automatic detection systems with capabilities in different languages, especially Indo-European languages [9]. For his part, Mark Zuckerberg, the Chief Executive of Facebook, expressed his commitment to addressing the issue of hate speech on the platform. In a speech made at the ceremony for the newly established Axel Springer Award in Berlin on 25 February, 2016, Zuckerberg stressed that "hate speech has no place on Facebook and in our community". In a recent report, Facebook announced that the company removed 22.3 million pieces of content containing hate speech, down from 31.5 million in the second quarter of 2021, as shown in Fig. 1.

However, a report by the Wall Street Journal in 2021 highlighted that Facebook removed posts that generated just 2% of the hate speech viewed on the platform and that violated its rules [10]. In the face of these contradictory statistics, many users, groups, and organizations have questioned Facebook's figures and thus the reliability of automatic detection and the artificial intelligence systems adopted by Facebook for detecting and tracking hate speech in its content. Many users have criticized the lack of effectiveness of the company's procedures for curbing hate speech on the platform, for instance allowing ISIS members and supporters to use it. In contrast, others have described the company as taking a Big Brother approach in dictating what can and cannot be said [11].

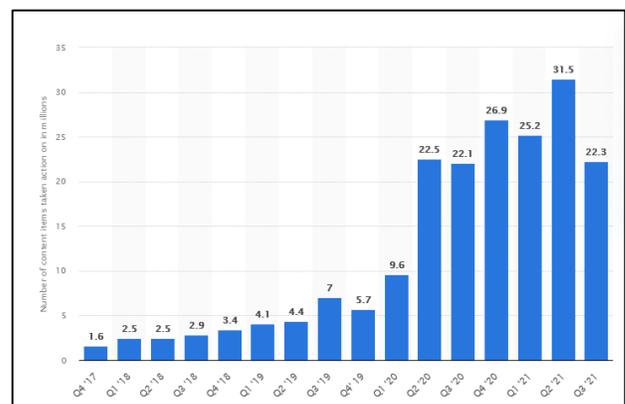


Fig. 1. Global Number of Hate Speech-containing Content removed by Facebook from 4th Quarter 2017 to 3rd Quarter 2021.

To illustrate the issue, this study evaluates the automatic detection of hate speech on Facebook in October 2020 during what came to be known as France's cartoon controversy. In October 2020, statements made by the French President Emmanuel Macron concerning Islam and the Prophet Muhammad led to many protests in the Arab and Muslim world. In these statements, Macron declared that his country would not stop publication of offensive cartoons of the Prophet, referring to them as freedom of expression. Macron's statements were warmly received by many activists, who described them as an assertion of France's "freedom to speak, to write, to think, to draw". Millions of Facebook users supported Macron's case, depicting Muslims as terrorists, especially after the brutal murder of a French teacher beheaded for showing his students cartoons of the Prophet Mohammed [12]. In turn, many commentators depicted Macron's statements as hate speech and a call for violence [13]. Furthermore, several hashtags trended in different Arab and Muslim countries through which activists described the statements of the French President as an insult to the Prophet of Islam and Muslims around the world. These hashtags included "except the Prophet of Allah", "boycott French products", "our prophet is a red line", "Macron offends the Prophet", and "stop insulting our Prophet". For its part, Facebook removed thousands of posts that were defined by the company as hate speech. In light of the above, this study seeks to evaluate the performance of artificial intelligence and automatic detection systems adopted by Facebook to understand how well they work and the extent to which they achieve their goals.

The remainder of this article is organized as follows. Section II provides a brief survey of automatic detection systems and approaches. Section III describes the methods and procedures. Section IV reports the results of the study. Section V is an interpretation of the results. Section VI concludes.

II. RELATED WORK

Recent years have seen increasing interest in "hate speech" in research studies. The phenomenon has been extensively studied in various disciplines, including discourse studies, social media research, sociology, and recently artificial intelligence, data mining, and information studies. This can be attributed to the increasing rates of crimes associated with hate speech on social media networks and platforms. Although the concept of "hate speech" was evident in different societies before the emergence of social media networks and platforms, the concept has recently been linked to social media [14]. Despite the usefulness and reliability of these networks and platforms for bringing people closer to each other, they have unfortunately also helped to disseminate user-generated content that gives rise to hate speech on heated political and religious topics [15, 16].

In the face of this issue, researchers have sought to develop automatic detection systems and algorithms with the capability of identifying hate speech in content so that such posts can be removed [1, 17]. Studies in this tradition are usually multidisciplinary. That is, they are based on different disciplines, including artificial intelligence, data mining, natural language processing, and computational linguistics [18,

19]. The underlying principle is that algorithms should be trained to identify linguistic content and detect forms of hate speech through artificial intelligence and data mining tools [20, 21]. In this regard, linguistics research has always been central to the development of automatic detection systems. Capozzi et al. [22] argue that hate speech can be deployed through various morphological structures and lexical choices with a myriad of nuances geared to the context of situation. In some languages, dictionaries of terms used in hate speech have been compiled.

As noted by Cobbe [23], artificial intelligence systems can usefully be employed to control and monitor hate speech on social platforms. Fortuna and Nunes [24] similarly argue that automatic detection methods are effective mapping tools for tracking the diffusion of hate speech on a large scale across regions. Nonetheless, the detection of hate speech can be challenging for machines, let alone humans, due to the complexity of determining lexical referentiality [25]. Natural language processing designers have developed operational frameworks focusing on representative features and based on semantic classifications [26], but these always have to be linked to the context for the meaning of the lexis to be effectively attributed to the notion of hate speech [27].

The literature indicates that much automatic detection research has focused on social media networks and platforms, including Facebook and Twitter. Since these networks exhibit different forms of hate speech, they provide good opportunities for researchers to test their models in different languages, including English, Spanish, Italian, and Chinese [28]. For instance, Poletto et al. [29] used the Twitter platform for data collection to detect hate speech communicated by Italian users on social media with regard to immigrants. Similarly, Vigna et al. [30] examined the hateful content of speech presented on Facebook.

Although there is extensive literature on the automatic detection of hate speech in different languages, including English and Chinese, very little has been done in Arabic due to the linguistic differences between Arabic and Western languages. However, the considerable spread of hate speech and abusive language on social media in recent years has led to pressure on the industry and researchers to find workable and reliable solutions for hate speech problems in the Arab world.

According to Bahaa-eddin [31], the rise in hate speech on social media in Arab countries can be described as a "tsunami" that has grave consequences for the stability of Arab societies. He suggests that the unprecedented growth in hate speech in recent years can be ascribed to the intermittent, but ongoing turmoil in the region, such as the Iraqi invasion of Kuwait, the 9/11 attacks that left Arabs with diverse views, the war on Iraq, the Israeli-Palestinian conflict, the clashes between Shias and Sunnis, and very recently the Arab Spring with all its repercussions. All these events and more have had a significant effect on the temper of the Arab public. Within this environment, social media platforms allow domains in which people can comment and use insulting and offensive language in their interactions.

In this regard, there have been various attempts in recent years to develop automatic detection systems to address hate speech in Arabic. Al-Hassan and Al-Dossari [32], for instance,

used deep learning within artificial neural networks to build a model that mimics layers of neurons to identify patterns in the text. Likewise, Watanabe et al. [33] proposed the use of n-gram features for detecting hate speech on Twitter. In addition to these efforts, the study of hate speech in Arabic content on social media platforms still accelerates in many respects.

III. METHODS, DATA AND PROCEDURES

This study is based on two different corpora built from Facebook posts covering France’s cartoon controversy in October 2020. The first corpus is composed of 1,347 posts deleted by Facebook, now known as Meta. The second corpus comprises 1,856 posts that were randomly selected using the hashtag *إلا رسول الله* (except the Prophet of Allah). Data were collected from October 18 through November 5, 2020. The study is limited to posts in Arabic.

The deleted posts from Facebook included terms that were described as of a threatening nature, as shown in Table I.

In the second corpus (based on the hashtag *إلا رسول الله* [except the Prophet of Allah]), posts were clustered using vector space clustering methods. The posts were classified into four main groups (clusters). The most distinctive lexical features of Cluster 1 included words such as coexistence, tolerance, understanding, values, peace, and mercy. The second cluster included words such as “terrorists”, “murderers”, “bloody”, and “beasts”. The third cluster included words such as “pigs”, “Jews”, “Christians”, and “enemies”. Finally, the last cluster included almost all the words in the third cluster and encompassing different writing styles.

TABLE I. LIST OF HATE SPEECH DELETED BY FACEBOOK

Arabic Terms	English Translation
انزل غضبك ومقتك	may you pour out your wrath and hatred
يا حثالة	O scumbags!
يا كلاب	O dogs!
الكلب ماكرون	Macron, the dog
انزل غضبك و سخطك	May you pour out your wrath and hatred
انتقم منهم اشد انتقام	take revenge on them
المتصهينين	The Zionists
طبع الله على قلوبهم	May Allah close off their hearts
لن يتغير اليهود	the Jews will never change
الخاسرين	the losers
قطع الله السننكم	may Allah silence you
ابن الكلب	son of a dog
مواجهة التوسع الصهيوني و الإيديولوجية المسيحية	confronting Zionist expansion and Christian ideology
أزل دولتهم	Remove their country
سلط داء ليس له دواء	may you grant them a disease which has no cure
احفظ لنا ماكرون وابن اليهودية في تلاجة الموتى	keep Macron, the son of Judaism, in the mortuary
يا عبيد البقر	O worshippers of cows
شل الإله لك اليمين	may Allah paralyze you
شلت أيديهم	may their hands be paralyzed
اللهم عليك بالظالمين	may you annihilate the wrongdoers
اللهم عليك بالكفر وأهله	may you wipe the unbelievers out
النازية	Nazism
شئت شملهم	dissolve their unity
فرق جمعهم	divide their gatherings

For the purposes of the study, Facebook’s Policy Rationale developed for the definition of hate speech is adopted, as shown in Fig. 2.

We define hate speech as a direct attack against people — rather than concepts or institutions— on the basis of what we call protected characteristics: race, ethnicity, national origin, disability, religious affiliation, caste, sexual orientation, sex, gender identity and serious disease. We define attacks as violent or dehumanizing speech, harmful stereotypes, statements of inferiority, expressions of contempt, disgust or dismissal, cursing and calls for exclusion or segregation. We also prohibit the use of harmful stereotypes, which we define as dehumanizing comparisons that have historically been used to attack, intimidate, or exclude specific groups, and that are often linked with offline violence.

Fig. 2. Facebook’s Policy of Hate Speech.

IV. RESULTS

As mentioned above, the posts in the second corpus were clustered into four distinct classes. To identify the thematic features of each group, a centroid-based lexical analysis was carried out. Based on Facebook’s policies and definition of hate speech, Clusters 2, 3, and 4 are classified as hate speech and harmful content. Posts in these clusters constitute around 67% of the overall posts in the corpus, as shown in Table II.

It was clear that many users employed undefined writing systems to deceive Facebook’s artificial intelligence algorithms. Arabic has a unique writing system, which is completely different from Western languages. In the Arabic orthographic system, dotting is a special characteristic that is used to address the problem of ambiguities in Arabic consonants [34]. According to Maroun [35], thirteen of the 28 Arabic letters include dots, which can be placed above or below letters. Some of these letters have one dot (e.g., ب /b/), while others have two (e.g., ي /j/) or three (e.g., ش /š/). Sometimes, just one dot can distinguish between two or more words (e.g., جديد حديد /hadi:d/ /dʒadi:d/ iron, new). Interestingly, Classical Arabic was used without dotting. According to Al-Azami [36], only context was used to identify the consonants, as shown in Fig. 3.

TABLE II. CLASSIFICATION OF THE FOUR LEXICAL CLUSTERS

Cluster	Number of posts	Percentage
Cluster 1	618	33%
Cluster 2	137	7%
Cluster 3	837	46%
Cluster 4	264	14%

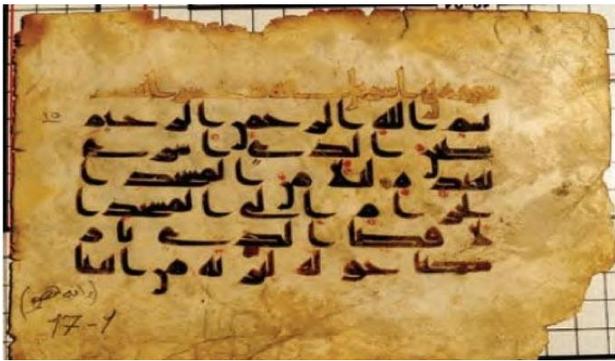


Fig. 3. An Example of Quranic Text.

Historically, with the expansion of the Arab and Muslim empire and the use of Arabic as a global language, it was difficult for many speakers of other languages to distinguish consonants. Thus, the dotting system was introduced in the 12th century [37, 38]. From that time on, Arabic has typically used dots for differentiation. Today, both standard Arabic and colloquial dialects are written using the standard dotting system, as shown in Fig. 4.

However, in the Facebook posts, contrary to usual practice in the standard writing system of Arabic, many users resorted to writing without dotting to circumvent Facebook's algorithms, which are trained to identify, track, and delete content that are classified as offensive and incite hatred in violation of its rules, as shown in Fig. 5.

ا	ا	ا	ا	alif	·	·
ب	ب	ب	ب	bā'	b	baby
ت	ت	ت	ت	tā'	t	tiet
ث	ث	ث	ث	thā'	th	thin
ج	ج	ج	ج	jīm	j	job
ح	ح	ح	ح	hā'	h	‡
خ	خ	خ	خ	khā'	kh	Ger. Buch\$
د	د	د	د	dāl	d	didt
ذ	ذ	ذ	ذ	dhāl	dh	then
ر	ر	ر	ر	rā'	r	error (trilled)
ز	ز	ز	ز	zā'	z	zone
س	س	س	س	sīn	s	sand
ش	ش	ش	ش	shīn	sh	sh, s shy
ص	ص	ص	ص	sād	\$	}
ض	ض	ض	ض	ḍād	ḍ	
ط	ط	ط	ط	ṭā'	ṭ	
ظ	ظ	ظ	ظ	ẓā'	ẓ	}
ع	ع	ع	ع	'ayn	·	
غ	غ	غ	غ	ghayn	gh	Fr. rien
ف	ف	ف	ف	fā'	f	fifty
ق	ق	ق	ق	qāf	q	♀
ك	ك	ك	ك	kāf	k	kin
ل	ل	ل	ل	lām	l	lilyt
م	م	م	م	mīm	m	maim
ن	ن	ن	ن	nūn	n	not
ه	ه	ه	ه	hā'	h	hat
و	و	و	و	wāw	w	watchō
ي	ي	ي	ي	yā'	y	yet □

Fig. 4. The Arabic Alphabet (Source Britannica.com).



Fig. 5. An Example of the use of Arabic without Dots on Facebook.

Among users, to help with this form of writing, different algorithms have been developed to help convert written forms and differentiate them (without using dots) so that their posts are not deleted by Facebook. This has also been used as a way of enabling users to keep their accounts active, rather than being blocked or deleted by Facebook. It was clear that the artificial intelligence algorithms developed by Facebook were not effective in dealing with these non-standard linguistic features of Arabic, which can still be understood by many users even without the dotting system.

V. ANALYSIS AND DISCUSSION

Based on the findings of the study, it seems that the artificial intelligence algorithms developed by Facebook for the automatic detection and tracking of hate speech tend not to be effective for content in Arabic. This can be attributed mainly to the design of standard automatic detection systems not being appropriate for Arabic content. Arabic, as a Semitic language, has a unique linguistic system that is completely different from Indo-European languages [39]. Today, Arabic is the fifth most widely spoken language globally. It is also ranked fourth in languages used on the Internet [40]. Thus, the linguistic features of Arabic should be considered in the development of artificial intelligence algorithms and automatic detection systems.

The findings of the study agree with the bulk of the related literature in that so far there is no consensus regarding the definition of hate speech. MacAvaney et al. [41] assert that there are disagreements concerning how hate speech should be defined. In our case, it was clear that much of the hate speech in the content identified by Facebook is related to the influence of the religion of Islam. Indeed, many, if not most, hate terms and phrases are taken from or influenced by religious Islamic discourse. For instance, the results showed that posts including the phrases لعنة الله عليهم (May Allah's curse be upon them) and القرود والخنازير (pigs and apes) were tracked and deleted. These phrases were classified by Facebook as inciting hatred against specific groups, namely Christians and Jews. Thus, millions of Facebook users sought to undermine the platform's recognition of these phrases as hate speech by finding ways of deceiving the artificial intelligence algorithms.

In certain interpretations of the Qur'ān, which is believed by Muslims to be the word of God revealed to His prophet Muhammad, the phrase لعنة الله عليهم (May Allah's curse be upon them) is a form of prayer or invocation used to ask Allah

to harm and curse others. According to Ibn manzūr, those who are thus cursed are rejected by Allah, shunned from his mercy, and hence damned. The verb *la'ana* means to curse, namely to call upon divine or supernatural power to inflict injury upon somebody. The word *la'ana* and its derivatives are mentioned 41 times in the Qur'ān, where it is invoked for specific rejected groups of people. For instance, the curse of Allah is invoked upon all those who reject faith in Allah, hypocrites, polytheists, and pagans.

Likewise, the two terms “apes” and “pigs” are used figuratively in the sense of “Carry on behaving like apes and pigs if you want to”, rather than literally [42]. This term of address is given to polytheists. Apes alone are mentioned in the Qur'ān in Chapter/Surat Al-Araf (The Heights) to refer to a specific group of Jews who are blamed by God for their disobedience and breaking the Sabbath by fishing. When the Qur'ān casts blame on Jews, Christians, or the followers of any other religion, it does so specifically on certain people for aberrant behavior, not on the adherents of the religion as a whole [43].

However, contrary to moderate interpretations of the Qur'ān, many phrases have been taken out of context and used to incite hatred against specific groups. Thus, there is a need for religious authorities to point out that such terms and phrases related to particular contexts and specific groups of people, based solely on their lack of belief, transgressions, disobedience, hypocrisy, or aggression, and that it is unacceptable to exploit religious texts, taking such terms and phrases out of context and using them as hate speech on social media.

VI. CONCLUSION

In recent years, hate speech on social media networks has become a serious challenge for both individuals and institutions. This study aimed to evaluate the performance of artificial intelligence algorithms developed by social media networks for the automatic detection of hate speech. The study was based on evaluating the automatic detection of hate speech in Arabic on Facebook during the 2020 cartoon controversy in France. It can be concluded that automatic detection in Arabic poses a major challenge both for research and social media platforms. This can be attributed to the peculiar linguistic features of Arabic, which are different from those of Western languages. Finally, hate speech in Arabic is greatly influenced by the Muslim religious discourse. Social media posts reproduce verses of Qur'anic text taken out of context and misinterpreting them. Religious organizations and leaders should emphasize that such words and expressions should not be used to disseminate hate or justify hatred and violence.

REFERENCES

- [1] Gagliardone, D. Gal, T. Alves, and G. Martinez, Countering online hate speech. United Nations Educational, Scientific and Cultural Organization, 2015.
- [2] C. R. Carlson, Hate Speech. Cambridge, MA: MIT Press, 2021.
- [3] F. Del Vigna, A. Cimino, F. Dell'Orletta, M. Petrocchi, and M. Tesconi, "Hate me, hate me not: Hate speech detection on facebook," in Proceedings of the First Italian Conference on Cybersecurity (ITASEC17), 2017, pp. 86-95.
- [4] S. Assimakopoulos et al., Online Hate Speech in the European Union: A Discourse-Analytic Perspective. Springer International Publishing, 2017.
- [5] J. Golbeck, Online Harassment. Springer International Publishing, 2018.
- [6] M. Mondal, L. A. Silva, and F. Benevenuto, "A measurement study of hate speech in social media," in Proceedings of the 28th ACM conference on hypertext and social media, 2017, pp. 85-94.
- [7] A. C. Nakaya, Social Media Hate Speech. ReferencePoint Press, 2020.
- [8] J. C. Pereira-Kohatsu, L. Quijano-Sánchez, F. Liberatore, and M. Camacho-Collados, "Detecting and monitoring hate speech in Twitter," Sensors, vol. 19, no. 21, p. 4654, 2019.
- [9] S. Modha, T. Mandl, P. Majumder, and D. Patel, "Tracking Hate in Social Media: Evaluation, Challenges and Approaches," SN Computer Science, vol. 1, no. 2, p. 105, 2020/03/28 2020.
- [10] E. Walsh. (2021, Oct 17, 2021) Facebook claims it uses AI to identify and remove posts containing hate speech and violence, but the technology doesn't really work. Insider. Available: <https://www.businessinsider.com/facebook-ai-doesnt-work-to-remove-hate-speech-and-violence-2021-10>.
- [11] J. Naughton. (2021, 12 Jun 2021) Big Brother is still watching you and he goes by the name Facebook. The Guardian. Available: <https://www.theguardian.com/commentisfree/2021/jun/12/big-brother-is-still-watching-you-and-he-goes-by-the-name-facebook-john-naughton>.
- [12] N. I. Ariffin and F. Hussain, "The 2020 France Attacks: A Framing Analysis of UK and US Newspapers," International Journal of Modern Trends in Social Sciences, vol. 4, no. 15, pp. 133-146, 2021.
- [13] K. Willsher, "Anger spreads in Islamic world after Macron's backing for Muhammad cartoons," in The Guardian, ed, 2020.
- [14] A. Omar and B. Deraan, "Towards a Linguistic Stylometric Model for the Authorship Detection in Cybercrime Investigations," International Journal of English Linguistics, vol. 9, no. 5, pp. 182-192, 2019.
- [15] R. Moon, Putting Faith in Hate: When Religion Is the Source or Target of Hate Speech. Cambridge: Cambridge University Press, 2018.
- [16] A. Guiora and E. A. Park, "Hate speech on social media," Philosophia, vol. 45, no. 3, pp. 957-971, 2017.
- [17] A. Omar and B. Deraan, "Cybercrime and authorship detection in very short texts," Opción, vol. 34, pp. 1765-1785, 2019.
- [18] A. Nalamothu, O. E. Theses, and D. Center, Abusive and Hate Speech Tweets Detection with Text Generation. Wright State University, 2019.
- [19] G. Rehm and T. Declerck, Language Technologies for the Challenges of the Digital Age: 27th International Conference, GSCL 2017, Berlin, Germany, September 13-14, 2017, Proceedings. Springer International Publishing, 2018.
- [20] G. Ignatov and R. Mihalcea, An Introduction to Text Mining: Research Design, Data Collection, and Analysis. SAGE Publications, 2017.
- [21] A. Choudhary, A. P. Agrawal, R. Logeswaran, and B. Unhelkar, Applications of Artificial Intelligence and Machine Learning: Select Proceedings of ICAAAIML 2020. Springer Singapore, 2021.
- [22] A. T. E. Capozzi et al., "Computational Linguistics Against Hate: Hate Speech Detection and Visualization on Social Media in the "Contro L'Odio" Project," in CLiC-it 2019 Italian Conference on Computational Linguistics, Bari, Italy, 2019: Proceedings of the Sixth Italian Conference on Computational Linguistics.
- [23] J. Cobbe, "Algorithmic censorship by social platforms: power and resistance," Philosophy & Technology, pp. 1-28, 2020.
- [24] P. Fortuna and S. Nunes, "A survey on automatic detection of hate speech in text," ACM Computing Surveys (CSUR), vol. 51, no. 4, pp. 1-30, 2018.
- [25] A. Schmidt and M. Wiegand, "A Survey on Hate Speech Detection using Natural Language Processing," in Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media, Valencia, Spain, 2017 pp. 1–10: Association for Computational Linguistics.
- [26] S. Sun, C. Luo, and J. Chen, "A review of natural language processing techniques for opinion mining systems," Information fusion, vol. 36, pp. 10-25, 2017.
- [27] E. M. Bender and D. T. Langendoen, "Computational Linguistics in Support of Linguistic Theory," Linguistic Issues in Language Technology, vol. 3, no. 2, pp. 1-31, 2010.

- [28] M. Sanguinetti, F. Poletto, C. Bosco, V. Patti, and M. Stranisci, "An Italian Twitter Corpus of Hate Speech against Immigrants," in Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan, 2018: European Language Resources Association (ELRA).
- [29] F. Poletto, M. Stranisci, M. Sanguinetti, V. Patti, and C. Bosco, "Hate speech annotation: Analysis of an Italian Twitter corpus," in Proceedings of the Fourth Italian Conference on Computational Linguistics (CLiC-it 2017), Rome, Italy, 2017, vol. 2006.
- [30] F. D. Vigna, A. Cimino, F. Dell'Orletta, M. Petrocchi, and M. Tesconi, "Hate me, hate me not: Hate speech detection on Facebook," in Proceedings of the First Italian Conference on Cybersecurity (ITASEC17), Venice, Italy, 2017.
- [31] M. M. Bahaa-eddin, *HateSpeak in Contemporary Arabic Discourse*. . Newcastle upon Tyne, UK: Cambridge Scholars Publishing, 2012.
- [32] A. Al-Hassan and H. Al-Dossari, "Detection of hate speech in social networks: a survey on multilingual corpus," *Computer Science and Information Technology*, vol. 9, no. 2, pp. 83–100, 2019.
- [33] H. Watanabe, M. Bouazizi, and T. Ohtsuki, "Hate speech on Twitter: a pragmatic approach to collect hateful and offensive expressions and perform hate speech detection," *IEEE Access*, vol. 6, pp. 13825–13835, 2018.
- [34] N. E. Gayar and C. Y. Suen, *Computational Linguistics, Speech And Image Processing For Arabic Language*. Singapore: World Scientific Publishing Company, 2018.
- [35] M. Maroun, "Diacritics and the Resolution of Ambiguity in Reading Arabic," PhD, Department of Psychology, University of Essex, Brighton, 2017.
- [36] M. M. Al-Azami, *The History of the Quranic Text*. Turath Publishing, 2020.
- [37] K. Versteegh, *Arabic Language*. Edinburgh University Press, 2014.
- [38] J. F. Healey and G. R. Smith, *A Brief Introduction to The Arabic Alphabet*. Saqi, 2012.
- [39] A. Omar, B. I. Elghayesh, and M. A. M. Kassem, "Authorship Attribution Revisited: The Problem of Flash Fiction A morphological-based Linguistic Stylometry Approach," *Arab World English Journal (AWEJ)*, vol. 10, no. 3, pp. 318-329, 2019.
- [40] G. Badaro et al., "A survey of opinion mining in Arabic: A comprehensive system perspective covering challenges and advances in tools, resources, models, applications, and visualizations," *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, vol. 18, no. 3, pp. 1-52, 2019.
- [41] S. MacAvaney, H.-R. Yao, E. Yang, K. Russell, N. Goharian, and O. Frieder, "Hate speech detection: Challenges and solutions," (in eng), *PLoS one*, vol. 14, no. 8, pp. e0221152-e0221152, 2019.
- [42] E. M. Badawi and M. Abdel Haleem, *Arabic-English Dictionary of Qur'anic Usage*. Leiden: Brill, 2008.
- [43] M. Abdel Haleem, *Exploring the Qur'an: Context and Impact*. London/New York: I.B. Tauris, 2017.

AUTHORS' PROFILE

Abdulfattah Omar is an Associate Professor of English Language and Linguistics in the Department of English, College of Science & Humanities, Prince Sattam Bin Abdulaziz University (KSA). Also, he is a standing lecturer of English Language and Linguistics in the Department of English, Faculty of Arts, Port Said University, Egypt. Dr Omar received his PhD degree in computational linguistics in 2010 from Newcastle University, UK. His research interests include computational linguistics, literary computing, digital humanities, discourse analysis, and translation studies.

[ORCID: 0000-0002-3618-1750](https://orcid.org/0000-0002-3618-1750)

Mohamed Elarabawy Hashem is an Assistant Professor of English language and Translation Studies at the College of Science and Arts in Tabarjal, Jouf University, KSA & a standing lecturer of English Islamic Studies at the Faculty of Languages and Translation, Al-Azhar University, Cairo. His research interests include Islamic Translation Studies, Semantics, Islamic Discourse Analysis and Hermeneutics.

[ORCID: 0000-0001-6818-799X](https://orcid.org/0000-0001-6818-799X)