Analyzing Digital Polarization on Hijab: A Dataset of Annotated YouTube Comments

Heba Al Heraki[1], Wajdi Zaghouani[2]

[1]Hamad Bin Khalifa University

[2]Northwestern University in Qatar

hebaheraki@gmail.com, wajdi.zaghouani@northwestern.edu

Abstract

This study presents an analysis of digital polarization on the topic of the Hijab by examining YouTube comments in Arabic. Employing a novel dataset of around 10K annotated comments, this research investigates the digital discourse using seven labels: Stance, Use of Sarcasm, Argumentation, Cordiality, Offensiveness, Hopefulness, and Apparent Gender of Commenters. The findings reveal significant insights into gender dynamics and the prevalence of specific rhetorical strategies within the debate. This study contributes to the broader field of polarization and argument mining, offering a unique lens on the intersection of digital culture and societal issues in the Arab context.

Datasets — https://zenodo.org/records/14652088.
DOI: 10.5281/zenodo.14650735

Introduction

Digital polarization on social media platforms has become a prevalent issue, reflecting and amplifying societal divisions. In the Arabic-speaking world, one such divisive topic is the Hijab, a headscarf symbolizing modesty and faith in Islam. The emergence of social media platforms like YouTube has transformed the discourse around the Hijab, allowing diverse perspectives to be shared and debated publicly. This study is motivated by the need to understand these digital conversations. The research examines YouTube comments from the Arab world to analyze sentiment trends, argumentative strategies, and gender representation within the discourse on the Hijab.

Despite the significant cultural and social importance of the Hijab in the Arab world, there is a noticeable gap in comprehensive studies and datasets exploring the digital discourse surrounding this topic. Existing literature lacks extensive datasets with a high number of labels that cover the nuances of online polarization. This research strives to fill this void by providing a detailed and large annotated dataset of YouTube comments, capturing the diverse perspectives and rhetorical strategies employed. It presents a novel dataset of approximately 10,000 annotated YouTube comments focused on the Hijab discourse in the Arab world, providing detailed annotations using seven labels: Stance, Use of Sarcasm, Argumentation, Cordiality, Offensiveness, Hopefulness, and Apparent Gender of Commenters. Relative to existing literature, this high number of labels and the large dataset size -in an underrepresented language and a niche topic that gives voice to nuanced perspectives- marks a clear territory for this study.

The primary objectives of this research are to analyze sentiment trends in YouTube comments regarding the Hijab, study the argumentative strategies employed in these comments, and explore gender representation within this context. Additionally, the research aims to contribute to developing online polarization guidelines and dataset creation in Arabic while providing a resource for other researchers to gain deeper insights into online polarization generally and on the issue of the Hijab specifically.

Literature Review

Digital Polarization Studies

Digital polarization studies with proximity to our topic focus on X (formerly known as Twitter) and YouTube. For instance, Weber et al. (2013) examined the ideological divide between secular and Islamist groups on Twitter, analyzing user interactions such as retweets and follows to infer ideological leanings. They found that users' retweet patterns and followings could indicate their political affiliations, drawing parallels to political polarization in the United States between Democrats and Republicans. Röchert et al. (2020) explored Hijab rights on YouTube, using sentiment and network analysis to reveal that users frequently encounter and express opposing views, contrary

to previous studies suggesting ideological homogeneity. Similarly, Siersdorfer et al. (2010) analyzed six million comments on 67,000 YouTube videos, finding that comment ratings depend on the language and sentiment expressed and that polarizing videos tend to have a wider variance in comment ratings.

Related Works in Arabic

As for relevant Arabic NLP studies, there are significant studies in both guidelines and dataset creation. In terms of guidelines creation, Zaghouani and colleagues worked in 2014 and 2015 on setting standards for non-native Arabic texts and error annotation. In 2016, they outlined guidelines for annotating Arabic punctuation and creating a discretized Arabic corpus. By 2018, they introduced frameworks for author profiling and text analytics on social media data. Elfardy and Diab (2016) focused on annotating social media posts about Egyptian ideological divisions, achieving a 75.7%-92% Inter Annotator Agreement (IAA). Almanea and Poesio (2022) annotated 2,000 Arabic tweets for misogyny, recommending the use of disaggregated data and "soft loss training," which utilizes the full range of annotator judgments rather than relying on a consensus or "gold standard."

As for annotated corpora creation, notable contributions include the development of dialect-specific corpora, such as the MADAR Arabic dialect corpus by Bouamor et al. (2018) and the multi-dialectal Arabic corpus by Charfi et al. (2019). Efforts in sentiment analysis have also been significant, exemplified by Zaghouani's (2018) corpus for youth depression detection and the dataset by Laabar and Zaghouani (2024) focusing on political discourse on Facebook. Other advancements include domain-specific resources like the DAICT irony corpus by Abbes et al. (2020) and Munazarat 1.0, a corpus of Arabic competitive debates by Khader et al. (2024). Additionally, linguistic resources such as the revised Arabic PropBank by Zaghouani et al. (2010) and treebank annotations by Maamouri et al. (2010) are crucial for syntactic analysis.

Argument Detection

Lycos et al. (2019) describe the evolution of argument mining from traditional models to opinion and argument mining of social media data. Cabrio and Villata (2018) survey argument mining with data-driven approaches, categorizing studies by their contributions to sentence or argument classification and the detection of relations or boundaries within argumentative content. Studies highlight the challenges in argument mining, such as the need to combine multiple sources of information and the lack of interoperability due to differing definitions of argumentation spans and relations (Lytos et al., 2019; Cabrio & Villata, 2018; Vecchi et al., 2021).

Arabic argument detection is sparse. Notable studies include Muhsen et al. (2021), who focused on argument detection from e-health data using lexical tags. Jasim et al. (2019) contributed by focusing on argument mining in legal documents. Hamdi (2022) utilized topic modeling to mine ideological discourse on extremism, while Azmi and Alzanin (2014) created the Ara' system for opinion polarity mining in Saudi public opinion.

Two particularly relevant studies in the field inspired our guidelines and annotation scheme. Bosc et al. (2016) annotated 4,000 tweets for argument detection, defining arguments based on opinions, factual information, or claims. Schaefer and Stede (2022) developed comprehensive argument-mining categories, including claims and evidence, while distinguishing between verifiable and unverifiable claims and various types of evidence.

Methodology

This section outlines our data collection and annotation approach for examining digital polarization on the topic of Hijab. We developed a systematic methodology to ensure reliable and comprehensive data gathering, processing, and annotation of YouTube comments.

Dataset Size

The dataset consists of two fully annotated corpora totaling 9,745 YouTube comments. The first corpus (4,921 comments) was annotated using four labels: stance, sarcasm, argumentation, and offensiveness—the last three of which were in binary form. In the second corpus (4,824 comments), the annotation scheme was expanded both in scope and granularity. Three additional labels were introduced: politeness, hopefulness, and apparent gender. Additionally, sarcasm, argumentation, and offensiveness were annotated using more detailed, multi-category frameworks. While the granularity of the labels varied between the two parts, both corpora were fully annotated and analyzed, with corresponding results and inter-annotator agreement (IAA) scores reported separately.

Data Sources

YouTube videos discussing the Hijab within the context of secularism, enforcement, and cultural debates were selected. Keywords used for searching videos included "Hijab," "polarization," "secularism," "Iran," "Turkey," "France," and "forcing Hijab." These keywords ensured a diverse set of videos from various contexts and viewpoints. Videos were sourced from both Western-based channels like BBC Arabic, France 24 Arabic, and DW Arabic, and originally Arabic channels like Al Jazeera, Orient News, and Al Arabiya. This approach aimed to capture a comprehensive and diverse range of perspectives on the Hijab debate.

Sampling

Comments were sampled based on recency and popularity. Initially, the most recent 100 to 300 comments from each video were collected. However, to improve the representativeness of the dataset, the most popular comments (those with the highest number of likes) were prioritized in subsequent sampling rounds. This approach aimed to capture the most relevant, impactful, and engaging comments in the dataset. The final dataset consisted of approximately 10K comments from 73 YouTube videos, ensuring a substantial and diverse collection of user-generated content.

Pre-Processing

Comments were pre-processed to remove duplicates, irrelevant content, and non-Arabic text. Comments that were incomprehensible, spam, or not directly related to the Hijab discussion were excluded. This step ensured that the final dataset was clean, relevant, and ready for annotation.

For the first corpus, the comments were already scraped starting from the most recent, and no additional processing for sampling was required. For the second corpus, the "votes" column, which represented the number of likes on the comment, was sorted from highest to lowest to determine popularity, with comments having higher numbers of likes ranked first.

Labels

The annotation labels for the first corpus were 1) stance, 2) sarcasm, 3) argumentation, and 4) offensiveness. For the second corpus, additional 5) politeness, 6) hope, and 7) apparent gender labels were added. All labels were created using Excel's Data Validation Rules feature. Each label had specific criteria and examples as a guideline for the annotators as follows.

Stance

This label had three options: 1. With Hijab, 2. Against Hijab, or 3. Neutral/Mixed, respective examples of which are as follows:

1. " تماما لهذا نطالبهم بالسماح بالحجاب بما أن قوانينهم تدعم الحرية الشخصية فلما النفاق يعني عندك حرية كما تتعرى بس ممنوع تتحجب فتوقف عن الدفاع عنهم لو كنت أرويا يامنبطح"
   Exactly, this is why we demand that they allow the Hijab since their laws support personal freedom. So why the hypocrisy, does it mean that you have the freedom to be naked, but you are forbidden to veil, so stop defending them as if you were a European, you lowlife.

2. " ايها المسلمون! قبل ان تبينوا نفاقكم فئ تعليقاتكم. فما رأيكم في الدول الاسلامية التي تفرض الحجاب ليس فقط في مكان العمل بل في كل مكان عام؟ "
   Oh Muslims! Before you show your hypocrisy in your comments. What do you think of Islamic countries that impose the Hijab not only in the workplace but in every public place?

3. " على كل حال هذه بلدانهم وهم احرار كما لا نريد نحن في بلداننا ما نريد فحقهم هم ايضا المشكل في المسلمين عليهم المغادرة من اجل انقاد دينهم وهنا يكون الابتلاء حقا والصادق من المنافق ماذا تنتظر من بلاد الكفر والفجور "
   In any case, these are their countries, and they are free, just as we do not want in our countries what we do not want it is their right too. The problem is with the Muslims; they should leave at the appropriate time in order to save their religion and here is the real showing the truthful from the hypocrite. What is expected from the lands of disbelief and injustice?

Sarcasm

This label had five options: 1. Meaning the Opposite, 2. Mockery/Contempt, 3. Subtle Mockery/Contempt, 4. Mocking/Sarcastic, and 4. No Sarcasm. Examples are respectively as follows:

1. "جماعة حقوق المرأة وينكم"
   Where are you the people of women's rights?
   "هي دي الحريه"
   This is the freedom

2. " مسلمين منبعصين من حظر الحجاب في الشركة وناسيين انهم حاظرين على الناس ياكلو بحرية في شهر رمضان لان مشاعرهم الرقيقة بتنجرح"
   Muslims are bothered with the ban on the Hijab in the company and forget that they prohibit people from eating freely during Ramadan because their tender feelings are hurt.
   " الاسلام امبراطورية عظمى حاربه امبراطوريات وفشلوا..كيف تاأتي حشرات وتقول سنوقفه "
   Islam is a great empire. Empires have fought it and failed. How can insects come and say we will stop it?

3. " الربط لديك مخزي اليس هناك اوروبيين مسلمين ايضا ؟؟؟؟؟ اليست تلك بلاد الحرية و الديمقراطية؟؟؟؟ لم ارى اي قانون ضد المسيحية في البلاد المسلمة ؟؟؟؟"
   Your linking logic is shameful. Aren't there Muslim Europeans too????? Isn't that a place of freedom and democracy???? I have not seen any law against Christianity in Muslim countries????

4. "و مذا ان ذهبت إلى العمل بطربوش على رأسها و كمامة على وجهها و وشاح حول عنقها.....مسموح حسب القانون....اذن فليكن" And if she goes to work with a fez on her head, a mask on her face, and a scarf around her neck...that is allowed according to the law...then so be it....

5. "اما احترام قانون البلد الذي انت فيه او اتركه ولا تفرض معتقداتك عليهم"
Either respect the law of the country you are in or leave it and do not impose your beliefs on them.

Politeness

This label had three options: 1. Very Polite, 2. Polite, or 3. Not Polite/Neutral, respective examples of which are as follows:

1. "حسبنا الله و نعم الوكيل ،أنصح بناتنا بعدم نزع الحجاب ، قال الله تعالى: ومن يتق الله يجعل له مخرجًا ويرزقه من حيث لا يحتسب..."
God is sufficient for us, and He is the best disposer of affairs. I advise our daughters not to remove the Hijab. God Almighty said: And whoever fears God, He will make for him a way out and provide for him from where he does not expect…

2. "ردي عليك بسبب قولك يجب ان يهاجر المسلمين قلت لك بالصين هناك مسلمين وفي الهند هناك مسلمين اين يذهب المسلم وهو في بلده الذي عاش فيه اجداده"
My response to you because of your saying that Muslims must immigrate. I told you that in China, there are Muslims, and in India, there are Muslims. Where does a Muslim go when he is in his country where his ancestors lived?

3. "لماذا هذا القانون و لماذا لم تجاوب على اسألتي ؟؟؟؟؟؟؟؟؟؟؟؟؟؟؟؟"
Why this law and why did you not answer my questions???????????????

Offensiveness

This label had three options: 1. Very Offensive, 2. Offensive, or 3. Not Offensive/Neutral, respective examples of which are as follows:

1. "وماذا عن فرنسيي الجنسيه المسلمين؟! إلى أين يذهبون... يا عزيزي جهلك وتخلفك واستحمارك من الغرب قد اعماك.. الاسلام لم ولن يكون للعرب فقط وإنما للعالم اجمع يوجد مسلمين في كل بقاع العالم... لم يصبح الاسلام ثاني اكبر طائفه دينيه بالعالم من فراغ! فاين يذهب المواطنين من سكان هذه البلاد في ظل هذه الأحكام والسياسات المخالفه لحقوق الإنسان وحرياته!"
And what about French Muslim nationals?! Where are they going... Oh, my dear, your ignorance, backwardness, and brainwashing by the West have blinded you... Islam is not and will not be for the Arabs only but for the entire world. There are Muslims in all parts of the world... Islam has not become the second-largest religious sect in the world out of thin air! Where will the citizens of this country go in light of these rulings and policies that violate human rights and freedoms!

2. "كل الجاليات تندمج في اوروبا إلا الجالية الاسلامية الغبية. الاسلام والاندماج ضدان لا يتلاقيان"
All communities are integrating into Europe except the stupid Islamic community. Islam and integration are two opposites that do not meet!

2. "كم النفاق بين صياحهم على حرية ملابس المرأة إن كانت سوف تخلع ملابسها و بين إجبارهم للسيدات في اوروبا على خلع الحجاب"
How much hypocrisy is there between their whining cries about the freedom of a woman's clothing if she will take off her clothes and their forcing women in Europe to take off their veils?

3. "مشكلة الغرب يدعون التحضر و الحرية وهم عكس ذلك تماما منافقون بدرجة إمتياز"
The problem is that the West claims civility and freedom, but they are exactly the opposite. They are hypocrites par excellence.

3. "فرنسا أصبحت دكتاتورية و لا عدالة فيها منذ ترسي ماكرون الحكم، و القادم أعظم"
France has become a dictatorship and there is no justice in it since Macron took power, and what is coming is greater.

Hopefulness

This label had three options: 1. Inspiring Hope, 2. Not Inspiring Hope, or 3.Mixed. Example for each are respectively as follows:

1. "لاشك أن هنالك تدبيراً إلهياً يحصل حين يشتد التضييق على الاسلام ومظاهره والمتمسكين به ، ولا يحصل الفرج والنصر من بعده إلا حينما يكون التمحيص..فالله الله اخواني وأخواتي المسلمين في كل مكان اثبتوا على الدين واصبروا .. إن الأرض لله يورثها من يشاء من عباده والعاقبة للمتقين"
There is no doubt that there is a divine planning that takes place when the restrictions on Islam, its practices, and those who adhere to it become severe, and relief and victory after it will not come except when there is testing... By God, my Muslim brothers and sisters everywhere, remain steadfast in the religion and be patient... The earth belongs

to God. He leaves it as an inheritance to whomever He wishes from among His servants, and the outcome is for the righteous.
"الله يفتح بصيرتك لترى الحق من الباطل"
May God open your heart to see and discern the truth from falsehood.

2. "محاكم التفتيش تعود من جديد"
The Inquisition courts returns again.
"روح تتصرف تشتغل بالسعوديه اذا مش عاجبها"
Let her go work in Saudi Arabia if she doesn't like it.

3. " إن تنصروا الله ينصركم لا ترد عليه هو شخص يضيع وقتك فقط سيستمر بتعليقات السخيفة هذا هو عمله"
If you stand by God, He will stand by you. Do not respond to him. He is someone who is just wasting your time. He will continue with ridiculous comments. This is his job.

Argumentation

This label had five options which were mainly based on the framework introduced by Schaefer and Stede in 2022. 1. Verifiable Claim - Internal Evidence (VC IE) which means claims that can be fact-checked or verified through evidence drawn from external sources such as statistics, studies, or authoritative sources. 2. Verifiable Claim - Internal Evidence (VC IE) which means claims that can be fact-checked or verified through evidence derived from personal experiences, anecdotes, or observations.
3.Unverifiable Claim (UC) which have none of the above types of evidence yet still use reasoning. 3. Statement of Opinion which present neither evidence nor reasoning. And lastly 4. Non-Arguments such as merely praying against the opposite end. Example for each are respectively as follows:

1. "الدول العربية تحتوي على نسبه كبيره من المسيح و الايزيد و الصابئة و الارميين والاذريين و السريان و الكاكائيين"
Arab countries contain a large percentage of Christians, Yazidis, Sabians, Armenians, Dharites, Syriacs, and Kakaites.
"يحاربون فقط المسلمين ، لايبالون لمن لهم وشم ... ذو معاني دينية أو قلادات للصليب"
They only fight Muslims, they do not care about those who have tattoos with religious meanings or pendants of the cross.

2. "و من قال لك بان ليسو المسلمين عنصريين؟ اكثر عنصريين هم المسلمين عندي صديق في السعوديه مسيحي دائما بهان و يشتم من قبل المسلمين و ما يحترمون ديانته و اعتقاداته بس تقولون ليش الغرب ما يحترم المسلمين و منع الحجاب"
And who told you that Muslims are not racist? The most racist are Muslims. I have a Christian friend in Saudi Arabia. He is always sworn at and insulted by Muslims, and they do not respect his religion and beliefs. But you say, why does the West not respect Muslims and ban the Hijab?

3. "الديمقراطية و الحريات بمختلف انواعها في الغرب هي مجرد فقاعة هواء لا يوجد احترام لأي ثقافة او اي عرق او اي دين او معتقد لحريات تطبق على اهوائهم الافكار الايجابية المزيفة التي وضعها او زرعها الإعلام في ادمغتنا عن الحرية و الديمقراطية و المساواة في الغرب كذبة اعلامية ليس إلا عندما تتحدث عن الحرية و الديمقراطية و المساواة يجب أن تعم الجميع و لا تتوقف الحرية عند شخص او عرق او لون او دين... يجب أن يحصل الجميع في الغرب او الشرق على حرية التعبير و الاختيار في اللباس و غيره و حرية اختيار الدين و حرية الفكر"
Democracy and freedoms of all kinds in the West are just an air bubble. There is no respect for any culture, any race, any religion or belief. Freedoms are applied according to their whims. The false positive ideas placed or implanted by the media in our brains about freedom, democracy, and equality in the West are a media lie and nothing more. When you talk about freedom, democracy, and equality, you must include everyone, and freedom does not stop at a person, race, color, or religion... Everyone, whether in the West or the East, must have freedom of expression and choice in dress and other matters, the freedom to choose one's religion, and freedom of thought.

4. "احقر دولة اوروبية هي فرنسا"
The most despicable European country is France.
"عفكره الاسلام اصغر بكتيررر مما تعتقدون"
By the way, Islam is much less significant than you think.

5. "ألا بعدا لفرنسا ومن على شاكلتها اللهم أنصر الإسلام والمسلمين وأعز المسلمين وأعل بفضلك كلمة الحق والدين اللهم أنصر الإسلام والمسلمين وأذل الشرك والمشركين ودمر أعداء الدين يا رب العالمين اللهم من أراد بنا كيد فاجعل كيده في نحره يا رب العالمين"
Away from France and those like it. O God, support Islam and the Muslims, cherish the Muslims, and raise, thanks to You, the word of truth and religion. O God, support Islam and the Muslims, humiliate polytheism and polytheists, and destroy the enemies of religion, O Lord of the Worlds. O God, whoever intends a plot against us, turn his plot against him, O Lord of the Worlds.

Apparent Gender

This label included three sub labels: 1. Male which assigned when the commenter's gender is identifiable as male based on their username or an explicit mention in the comment(s). Female which was assigned when the commenter's gender is identifiable as female based on their username or an explicit mention, or gender-specific linguistic clues in the comment. Unclear was assigned when the commenter's gender cannot be definitively determined based on their username or comment content. This includes ambiguous usernames or unisex names.

Annotator Pipeline

The annotation process involved training annotators using detailed guidelines with definitions and examples for each label. The full dataset was primarily annotated by the first author (24 years old) during the second year of her master's program.

To measure inter-annotator agreement (IAA), three additional annotators each annotated a subset of 100 comments: a 20-year-old medical student, a 22-year-old high school graduate, and a 29-year-old master's graduate. These 300 comments were used solely for consistency checks and not included in the final dataset analysis. The IAA process served to calibrate the labeling scheme and validate guideline clarity.

All annotators are female, native Arabic speakers. The implications of this demographic homogeneity are addressed in the Limitations section.

Analysis Results

The data were analyzed using PivotTables in Excel to calculate frequencies and percentages of different labels. Chi-square tests of independence were performed to examine the relationship between stance and other variables such as sarcasm, argumentation, offensive language, and gender. The analysis focused on identifying significant patterns and correlations within the dataset, providing insights into the dynamics of online discourse on the Hijab.

Results for Corpus (1)

From a total of 4921 YouTube comments, the majority (around 66%) were proponents of the Hijab, followed by mixed views (around 18%) and opponents (around 14%). Sarcasm was more prevalent among opponents (21%) and those with mixed views (16%), while supporters used sarcasm the least (9%). Logic was used most frequently by those with mixed views (78%), followed by supporters and opponents, both around 70%. Offensive language was most common among opponents (59%), with supporters and mixed views using offensive language to a lesser extent (around 27% for both). The Chi-square tests indicated a highly significant relationship between stance towards the Hijab and the use of sarcasm ($X2 = 83.39$, $p < .001$), logic ($X2 = 17.63$, $p < .001$), and offensive language ($X2 = 293.72$, $p < .001$).

Results for Corpus (2)

From a total of 4824 YouTube comments, the majority (78%) were proponents of the Hijab, followed by neutral/mixed views (14%) and opponents (8%). The breakdown of sarcasm showed that "mocking/funny arguments" were more prevalent among neutral/mixed and against Hijab comments. Statements of opinion were the most common form of argumentation across all stances, suggesting a prevalence of subjective viewpoints rather than evidence-based arguments. Gender distribution revealed an almost equal representation of males and females among proponents and neutral/mixed categories, while opponents were predominantly male.

Politeness analysis indicated that comments supporting the Hijab were generally more polite, with a significant portion labeled as "very polite" or "polite." In contrast, comments against the Hijab were less polite. The hopefulness analysis showed that comments supporting the Hijab were more likely to inspire hope, while those against the Hijab were predominantly neutral or pessimistic. The offensiveness analysis confirmed that comments against the Hijab were significantly more offensive, with a higher proportion of "very offensive" and "offensive" labels.

Sarcasm and Argumentation Analysis

The analysis of sarcasm types revealed distinct patterns across different stances. Comments against the Hijab frequently used "mockery/contempt" and "subtle mockery/contempt," indicating a higher tendency to ridicule opposing views. For example, comments such as "Finally, Europeans began to understand the stench of Islam" exemplified the use of direct mockery. In contrast, comments supporting the Hijab rarely used sarcasm, reflecting a more serious and respectful tone in their discourse. The analysis of argumentation showed that "statement of opinion" was the dominant form of argument across all stances, suggesting a prevalence of subjective viewpoints rather than evidence-based arguments. For instance, comments like "The Hijab is a personal choice and a sign of modesty" were common among supporters. This indicates that the debate on the Hijab is highly subjective and emotionally charged.

Gender Analysis

The gender analysis revealed interesting insights into the dynamics of online discourse on the Hijab. Male commenters were more prevalent in the "against Hijab" category, while female commenters were almost equally distributed across all stances. This suggests that men are

more likely to express opposition to the Hijab in online discussions.

Hopefulness Analysis

The analysis of hopefulness indicated that comments supporting the Hijab were more likely to inspire hope, with many comments expressing optimism and encouragement. For example, comments such as "Stay strong, sisters, and continue to wear your Hijab with pride" were common among supporters. In contrast, comments against the Hijab were predominantly neutral or pessimistic, with very few comments labeled as "inspiring hope."

Politeness and Offensiveness Analysis

The politeness analysis showed that comments supporting the Hijab were generally more polite, with a significant portion labeled as "very polite" or "polite." For example, comments such as "May God guide you to see the truth" were common among supporters. In contrast, comments against the Hijab were less polite and more likely to include offensive language. This difference in tone reflects the polarized nature of the debate and the varying rhetorical strategies used by different groups.

The offensiveness analysis confirmed that comments against the Hijab were significantly more offensive, with a higher proportion of "very offensive" and "offensive" labels. For example, comments such as "The Hijab is a symbol of oppression" were common among opponents. This suggests that opposition to the Hijab is often expressed in harsher and more aggressive terms, reflecting the contentious nature of the debate.

IAA Results

IAA Results for Corpus (1)

The IAA for Corpus (1) showed high agreement percentages among all three annotators, with Cohen's Kappa scores indicating substantial agreement for most labels. For example, stance had an agreement of 84% with a Kappa score of 0.68, sarcasm had an agreement of 84% with a Kappa score of 0.46, argumentation had an agreement of 76% with a Kappa score of 0.45, and offensiveness had an agreement of 86% with a Kappa score of 0.55. These high agreement scores indicate that the annotation guidelines were clear and consistently applied by the annotators.

IAA Results for Corpus (2)

For Corpus (2), the agreement percentages were relatively lower due to the increased complexity of the task. The lowest percentage and our only negative Cohen's Kappa score was with the "hope" label, which was due to a misunderstanding of the label's criteria by one of the annotators. Feedback from annotators highlighted areas where guidelines could be improved for better clarity and consistency. For example, the "hope" label had an agreement of 87% but a Kappa score of 0.20, indicating a need for clearer definitions and examples. Despite these challenges, the overall IAA results for Corpus (2) were satisfactory, with stance having an agreement of 62% with a Kappa score of 0.28, sarcasm having an agreement of 46% with a Kappa score of 0.03, argumentation having an agreement of 84% with a Kappa score of 0.61, offensiveness having an agreement of 84% with a Kappa score of 0.67, politeness having an agreement of 72% with a Kappa score of 0.56, and gender having an agreement of 86% with a Kappa score of 0.80.

Value of the Data and Possible Applications

This annotated dataset, with its extensive 10K comments in Arabic and multiple labels, is invaluable for exploring connections between stance, sarcasm, argumentation, offensive language, and more. Researchers can investigate gender dynamics and how different genders employ sarcasm, argumentation, and offensive language within this dataset. Globally, our study stands out as it is the first to annotate comments for argumentation, sarcasm, and toxic language in Arabic. Unlike Schaefer and Stede's (2022) study on tweets, ours encompasses more categories and is based on YouTube comments. The stability and straightforward nature of YouTube data make it advantageous for NLP studies, as noted by Uryupina et al. (2019). Moreover, Bender et al. (2021) advocate for focused data annotation over the expansion of language models, which can perpetuate biases and environmental concerns. Our study aligns with this approach, prioritizing meticulous curation to help AI discern ideological perspectives beyond sentiment analysis and sarcasm detection.

Limitations

Our study acknowledges several limitations in terms of methodology, data representation, and ethical considerations.

A primary limitation concerns annotator homogeneity. All annotators were female, native Arabic speakers with connections to the first author. This demographic uniformity potentially introduced a shared socio-cultural perspective that may have influenced the interpretation of subjective labels such as sarcasm, offensiveness, and politeness. However, this arrangement facilitated consistent communication during the annotation process—annotators could readily seek clarification about guidelines, which helped ensure consistency and reduce ambiguity.

A more diverse annotator pool might capture a broader spectrum of perspectives; however, it could also introduce

methodological challenges. Diverse annotators might hesitate to seek clarification, especially if they lack regular or informal contact with the guideline developer. In such cases, variance in inter-annotator agreement results may reflect communication gaps rather than genuine conceptual disagreement. Annotators uncomfortable with requesting clarification might misinterpret specific labels, affecting consistency. Future research should balance annotator diversity with robust training protocols to ensure both perspective breadth and methodological consistency.

Another limitation is the exclusive focus on YouTube comments, which may not comprehensively represent the full spectrum of discourse on the Hijab. We also recognize the inherent challenges in categorizing nuanced online discussions, which can lead to over-simplification of complex viewpoints.

Regarding data ethics, we implemented several measures to protect user privacy. All user data were anonymized, and usernames were not disclosed in the paper. Data collection was conducted in compliance with ethical standards, ensuring that no personally identifiable information (PII) was included in the analysis. Furthermore, the dataset was securely stored and processed, with access restricted to authorized researchers only, minimizing the risk of data breaches or misuse.

Despite these limitations, this dataset contributes to research on Arabic digital discourse, particularly regarding culturally sensitive topics. It supports advancement in hate speech detection, argument mining, gendered discourse analysis, and digital polarization studies in non-Western contexts—areas traditionally underrepresented in NLP research. Additionally, it offers a methodological benchmark for developing annotation schemes that balance cultural specificity with cross-contextual validity.

Discussion of Potential Misuse

While this research aims to contribute positively to understanding digital polarization, there is a risk that the findings could be misused. For instance, the insights gained from this analysis could potentially be used to reinforce negative stereotypes or to develop algorithms that profile individuals based on their online expressions. To mitigate these risks, we emphasize the importance of using this research responsibly and in a way that promotes constructive dialogue and understanding. The study's findings should be applied with caution, ensuring they are not used to further entrench biases or for discriminatory purposes. By outlining these potential misuse scenarios, we aim to raise awareness of the ethical implications and encourage the responsible use of this research.

Conclusion

This study examines digital polarization in the Arab world through YouTube comments on the Hijab. Analyzing a 10K-comment dataset reveals rhetorical strategies and gender dynamics in this debate. Findings indicate prevalent support for the Hijab, with sarcasm and offensive language more common among detractors. Supporters often maintain a hopeful and less offensive tone. The research enriches our understanding of digital discourse on cultural and religious topics, with implications for social media policy, content moderation, and AI tool development. Acknowledging limitations, further research should include diverse annotators and multiple platforms to better understand global digital polarization. This study advocates for a nuanced approach to digital discourse, promoting healthier online dialogue.

The analyses highlight the polarized nature of online discourse on the Hijab, with distinct differences in tone, argumentation, and sentiment across different stances. Supporters of the Hijab tend to use more respectful and hopeful language, while opponents are more likely to use sarcasm and offensive language and express pessimism. These findings provide valuable insights into the dynamics of digital polarization and the rhetorical strategies used in online debates on contentious social issues. The results underscore the importance of understanding the different ways in which people engage in online discussions and the potential impact of these interactions on broader societal debates.

Acknowledgments

The authors would like to acknowledge the support of the Data for Social Good Research Group, of which they are both members. This work was partially supported by the Qatar Research Development and Innovation Council (QRDI) through the Qatar National Research Fund (QNRF) grant NPRP14C-0916-210015.

References

Abbes, I., Zaghouani, W., El-Hardlo, O., & Ashour, F. 2020. Daict: A Dialectal Arabic Irony Corpus Extracted From Twitter. In Proceedings of the Twelfth Language Resources and Evaluation Conference, 6265-6271.

Ahmed, A. M., Zhang, X., Rezk, L. M., & Zaghouani, W. 2022. Building an Annotated L1 Arabic/L2 English Bilingual Writer Corpus: The Qatari Corpus of Argumentative Writing (QCAW). *Corpus-Based Studies Across Humanities*. https://catalog.ldc.upenn.edu/LDC2022T04.

Almanea, D., & Poesio, M. 2022. ArMIS—The Arabic Misogyny and Sexism Corpus with Annotator Subjective Disagreements. In Proceedings of the Thirteenth Language Resources and Evaluation Conference, 2282-2291.

Al-Mulla, S., & Zaghouani, W. 2020. Building a Corpus of Qatari Arabic Expressions. In Proceedings of the LREC 2020 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection, (pp. 24-31).

Al-Shammari, N. B. N. 2021. *من العلمانية إلى الخلقانية* (From Secularism to Creationism). مؤسسة وعي للدراسات والأبحاث.

Azmi, A. M., & Alzanin, S. M. 2014. Aara—a System for Mining the Polarity of Saudi Public Ppinion Through E-Newspaper Comments. *Journal of Information Science,* 40(3): 398-410. doi.org/10.1177/016555151452467.

Bosc, T., Cabrio, E., & Villata, S. 2016. Tweeties Squabbling: Positive and Negative Results in Applying Argument Mining on Social Öedia. In Computational Models of Argument, 21-32. IOS Press

Bothe, C., & Wermter, S. 2022. Conversational Analysis of Daily Dialog Data Using Polite Emotional Dialogue Acts. arXiv:2205.02921.

Bouamor, H., Habash, N., Salameh, M., Zaghouani, W., Rambow, O., Abdulrahim, D., Obeid, O., Khalifa, S., Eryani, F., Erdmann, A., & Others. 2018. The MADAR Arabic Dialect Corpus and Lexicon. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation Conference, 3387-3395.

Cabrio, E., & Villata, S. 2018. Five Years of Argument Mining: A Data-Driven Analysis. In Proceedings of IJCAI 18, 5427–5433.

Center for Humane Technology. (2022). The wisdom gap. Retrieved from https://www.humanetech.com/insights/the-wisdom-gap. Accessed: 2025-04-30.

Chakravarthi, B. R. 2022. Hope speech detection in YouTube comments. *Social Network Analysis and Mining,* 12(1): 75. https://doi.org/10.1007/s13278-022-00901-z

Charfi, A., Ben-Sghaier, M., Atalla, A., Akasheh, R., Al-Emadi, S., & Zaghouani, W. 2024. MARASTA: A Multi-Dialectal Arabic Cross-Domain Stance Corpus. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), 11060-11069.

Charfi, A., Zaghouani, W., Mehdi, S. H., & Mohamed, E. 2019. A Fine-Grained Annotated Multi-Dialectal Arabic Corpus. In Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019), 198-204.

Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. 2021. The Echo Chamber Effect on Social Media. In Proceedings of the National Academy of Sciences, 118(9): e2023301118. https://doi.org/10.1073/pnas.2023301118.

Clark, H. H., & Gerrig, R. J. 1984. On the Pretense Theory of Irony. *Journal of Experimental Psychology: General*, 113(1): 121–126. https://doi.org/10.1037/0096-3445.113.1.121

Filatova, E. 2012. Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing. In Proceedings of the Eighth International Conference on Language Resources and Evaluation Conference, 392-398

Dusmanu, M., Cabrio, E., & Villata, S. 2017. Argument Mining on Twitter: Arguments, Facts and Sources. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2317-2322.

Elfardy, H., & Diab, M. 2016. Addressing Annotation Complexity: The Case of Annotating Ideological Perspective in Egyptian Social Media. In Proceedings of the 10th Linguistic Annotation Workshop Held in Conjunction with ACL 2016 (LAW-X 2016), 79-88.

Hamdi, S. A. 2022. Mining Ideological Discourse on Twitter: The Case of Extremism in Arabic. *Discourse & Communication,* 16(1): 76-92.

Jalil, M. M. 2023. Should Liberal Feminists Support Hijab Ban in the West? *Public Integrity*, 26(4): 1-12. https://doi.org/10.1080/10999922.2023.2198333.

Jasim, K., Sadiq, A. T., & Abdullah, H. S. 2019. A Framework for Detection and Identification of the Components of Arguments in Arabic Legal Texts. In 2019 First International Conference of Computer and Applied Sciences (CAS), 67-72.

Joshi, A., Tripathi, V., Bhattacharyya, P., Carman, M., Singh, M., Saraswati, J., & Shukla, R. 2016. How Challenging is Sarcasm versus Irony Classification?: A Study With a Dataset from English Literature. In Proceedings of the Australasian Language Technology Association Workshop 2016, 123-127.

Khader, M., Al-Sharafi, A., Al-Sioufy, M. H., Zaghouani, W., & Al-Zawqari, A. 2024. Munazarat 1.0: A corpus of Arabic Competitive Debates. In Proceedings of the 6th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT) Co-located with the 2024 International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), 20-30.

Laabar, S., & Zaghouani, W. 2024. Multi-Dimensional Insights: Annotated Dataset of Stance, Sentiment, and Emotion in Facebook Comments on Tunisia's July 25 Measures. In Proceedings of the Second Workshop on Natural Language Processing for Political Sciences, co-located with the 2024 International Conference on Computational Linguistics, Language Resources, and Evaluation (LREC-COLING 2024), 22-32.

Lytos, A., Lagkas, T., Sarigiannidis, P., & Bontcheva, K. 2019. The Evolution of Argumentation Mining: From Models to Social Media and Emerging Tools. *Information Processing & Management,* 56(6): 102055.

Miras, K. M. K., & Al-Warikat, A. A. 2022. النماذج المعرفية المركبة لدى عبد الوهاب المسيري والنظرية الاجتماعية: دارسة تأصيلية. *المجلة الأردنية للعلوم الاجتماعية*, 15(3), 223–238.

Mubarak, H., Rashed, A., Darwish, K., Samih, Y., & Abdelali, A. 2020. Arabic Offensive Language on Twitter: Analysis and Experiments. arXiv:2004.02192.

Muhsen, D. K., Ali, S. M., Zaki, R. M., & Ahmed, A. A. 2021. Arguments Extraction for E-Health Services Based on Text Mining Tools. *Periodicals of Engineering and Natural Sciences,* 9(3): 309–316.

Nath, T., Singh, V. K., & Gupta, V. 2023. BongHope: An Annotated Corpus for Bengali Hope Speech Detection. *International Journal of Information Technology*. https://doi.org/10.1007/s41870-025-02484-2.

Pew Research Center. (2011). Future of the Global Muslim Population: Regional Middle East-North Africa. Pew Research Center's Forum on Religion & Public Life. Retrieved from https://www.pewresearch.org/religion/2011/01/27/future-of-the-global-muslim-population-regional-middle-east/. Accessed: 2025-04-30.

Uryupina, O., Plank, B., Severyn, A., Rotondi, A., & Moschitti, A. 2014. SenTube: A Corpus for Sentiment Analysis on YouTube Social Media. In Proceedings of the Ninth International

Conference on Language Resources and Evaluation Conference, 4244–4249.

Rahbari, L., Dierickx, S., Coene, G., & Longman, C. 2021. Transnational Solidarity with Which Muslim Women? The Case of the My Stealthy Freedom and World Hijab Day Campaigns. *Politics & Gender,* 17(1): 112–135.

Röchert, D., Neubaum, G., Ross, B., Brachten, F., & Stieglitz, S. 2020. Opinion-Based Homogeneity on YouTube: Combining Sentiment and Social Network Analysis. *Computational Communication Research,* 2(1): 81–108.

Mahmood, S. 2006. Feminist Theory, Agency, and the Liberatory Subject: Some Reflections on the Islamic Revival in Egypt. *Temenos-Nordic Journal of Comparative Religion,* 42(1): 111-152.

Schaefer, R., & Stede, M. 2020. Annotation and Detection of Arguments in Tweets. In Proceedings of the 7th Workshop on Argument Mining, 53–58.

Schaefer, R., & Stede, M. 2022. GerCCT: An Annotated Corpus for Mining Arguments in German Tweets on Climate Change. In Proceedings of the Thirteenth Language Resources and Evaluation Conference, 6121–6130.

Schultes, P., Dorner, V., & Lehner, F. 2013. Leave a Comment! An In-Depth Analysis of User Comments on YouTube. In Proceedings of Wirtschaftsinformatik 2013, 659-673.

Siersdorfer, S., Chelaru, S., Nejdl, W., & San Pedro, J. 2010. How Useful Are Your Comments? Analyzing and Predicting YouTube Comments and Comment Ratings. In Proceedings of the 19th International Conference on World Wide Web, 891–900.

Thelwall, M. 2018. Social Media Analytics for YouTube Comments: Potential and Limitations. *International Journal of Social Research Methodology,* 21(3), 303–316.

Vecchi, E. M., Falk, N., Jundi, I., & Lapesa, G. 2021. Towards Argument Mining for Social Good: A Survey. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 1338–1352.

Weber, I., Garimella, V. R. K., & Batayneh, A. 2013. Secular vs. Islamist Polarization in Egypt on Twitter. In Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 290–297.

Zaghouani, W., & Awad, D. 2016. Toward an Arabic Punctuated Corpus: Annotation Guidelines and Evaluation. In Proceedings of the 2nd Workshop on Arabic Corpora and Processing Tools 2016 Theme: Social Media, 22.

Zaghouani, W., & Charfi, A. 2018a. Guidelines and Annotation Framework for Arabic Author Profiling. In Proceedings of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools.

Zaghouani, W., & Charfi, A. 2018b. Annotation Guidelines for Text Analytics in Social Media. In Qatar Foundation Annual Research Conference Proceedings, 2018(3), IC-TPD879.

Zaghouani, W., Bouamor, H., Hawwari, A., Diab, M., Obeid, O., Ghoneim, M., Alqahtani, S., & Oflazer, K. 2016. Guidelines and Framework for a Large-Scale Arabic Diacritized Corpus. In Proceedings of the Tenth International Conference on Language Resources and Evaluation, 3637–3643.

Zaghouani, W., Habash, N., Bouamor, H., Rozovskaya, A., Mohit, B., Heider, A., & Oflazer, K. 2015. Correction Annotation for Non-Native Arabic Texts: Guidelines and Corpus. In Proceedings of the 9th Linguistic Annotation Workshop, 129–139.

Paper Checklist

(a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes

(b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes

(c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes

(d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes, especially with regards to gender distribution.

(e) Did you describe the limitations of your work? Yes, see the Limitations section.

(f) Did you discuss any potential negative societal impacts of your work? Yes

(g) Did you discuss any potential misuse of your work? Yes,

(h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes

(i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes

2. Additionally, if your study involves hypotheses testing...
(a) Did you clearly state the assumptions underlying all theoretical results? N/A

(b) Have you provided justifications for all theoretical results? N/A

(c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? N/A

(d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? N/A

In this research, we aimed to advance scientific understanding while respecting ethical considerations and social contracts. Our research question avoids violating privacy norms, unfair profiling, and disrespecting societies or cultures. The abstract and introduction accurately reflect our contributions and scope, with a clear explanation of the methodological approach. We also identified potential artifacts in the data, particularly focusing on gender distribution.

We discussed the potential negative societal impacts and outlined the limitations of our work, including measures like data anonymization and responsible data release. While we didn't extensivly cover potential misuse, privacy concerns were addressed. Compliance with ethical standards was ensured by reviewing and adhering to ethics guidelines. Since our study does not involve hypothesis testing, we maintained a focus on methodological transparency and ethical rigor.