A Survival Guide for Iranian Women Prescribed by Iranian Women: Participatory AI to Investigate Intimate Partner Physical Violence in Iran

Adel Khorramrouz, Mahbeigom Fayyazi and Ashiqur R. KhudaBukhsh
Rochester Institute of Technology
ak8480@rit.edu, mf4559@g.rit.edu, axkvse@rit.edu

Abstract

Intimate Partner Violence (IPV) is a global problem affecting more than 2 billion women worldwide. Our paper makes two key contributions. First, via a substantial corpus of 53,220 comments to 1,563 Intimate Partner Physical Violence (IPPV) posts gleaned from more than 10 million comments posted on 523,232 posts on a popular parental health website in Iran, we present the first-ever computational analysis of user comments on accounts of IPPV in Iran. We harness large language models and participatory AI and tackle extreme class imbalance and other linguistic challenges that arise from tackling low-resource languages to shed light on the gender struggles of a country with documented stark gender inequality. With active input from a woman with a history of advocacy for social rights and grounded in Iranian culture, we characterize comments on IPPV into three broad categories: empathy, confront, and conform, and analyze their distribution. Second, we release an important dataset of 3,400 comments on IPPV posts.

1 Introduction

Intimate Partner Violence (IPV) against women is a global problem. 30% women aged 15 or above have experienced physical and/or sexual IPV worldwide [Devries et al., 2013; Pill et al., 2017; Fairchild, 2019; Jewkes, 2002]. In the United States alone, by an intimate partner, 9.4% of women have been raped, 16.9% of women have experienced sexual violence other than rape, and 24.3% of women have experienced severe physical violence [Basile et al., 2011]. In several non-Western countries, the problem is even far more severe [Duvvury et al., 2013; Gage, 2005; Moshtagh et al., 2021; Boy and Kulczycki, 2008]. For instance, in Iran, the prevalence of domestic violence is estimated to be 66% [Hajnasiri et al., 2016].

Of the four different broad categories of IPV against women as outlined by WHO [2012], two involve physical and sexual violence (referred to as Intimate Partner Physical Violence, IPPV in this paper). While surveys and interviews are standard instruments to gather data about IPPVs [Nagae and Dancy, 2010; Taherkhani et al., 2014; Razaghi et al., 2021; Shorey et al., 2014], collecting data through surveys is a time and resource-consuming process and participation rate is often a concern as victims face social barriers and stigma to openly discuss their traumatic experience. Country-specific cultural norms may also act as barriers for women to come forward and share their struggles. Online forums can complement these surveys by allowing a safe space for the victims to voice their concerns [Hassan et al., 2020]. Extant research has thus focused on popular social interaction forums such as Reddit [Schrading et al., 2015; Homan et al., 2020] and Twitter [Salehi et al., 2023; McCauley et al., 2018; Al-Garadi et al., 2022; Hassan et al., 2020] to analyze domestic abuse and IPV.

Existing literature on domestic abuse in social media primarily focus on the detection of posts indicating abuse [Schrading et al., 2015; Salehi et al., 2023; Hassan et al., 2020; Farrokh-Eslamlou et al., 2014]. Our focus in this paper is different. Via a substantial corpus from a popular parenting health website in Iran, we study how society responds to first-person accounts of IPPV. Our study thus seeks to better understand a dimension seldom studied in traditional surveys – a collective societal position on IPPV in a country with stark gender inequality.
With active input from a woman with a history of advocacy for social rights and grounded in Iranian culture, we characterize comments on IPPVs into three broad categories: (1) empathy and support (critical of the perpetrator and empathetic to the victim); (2) conform or adjust or victim shaming (urging the victim to adjust or conform to avoid IPPV or shaming the victim); and (3) confront or retaliate (encouraging the victim to have more agency and resist IPPV and seek outside help). Aided by sophisticated natural language processing methods, we seek to estimate the distribution of user comments on these aforementioned categories to understand the overall societal makeup of IPPV in Iran.

Our paper considers social web expressions in Persian, a language not as resource-rich as English or Spanish. With the advent of powerful generative AI methods such as GPT-3, a pertinent question arises how would these technologies aid (or hinder) cultural diversity in computational social science research [Ziens et al., 2023]. Our work harnesses LLM advancements [Brown et al., 2020], well-known Persian NLP resources [Farahani et al., 2021], active sampling methods to tackle class imbalance [Palakodety et al., 2020; KhudiBukhsh et al., 2020], and most importantly, a participatory AI [Birhane et al., 2022; Harrington et al., 2019] framework to work with Iranian women in our AI-building process to understand a social justice question of import.

To summarize, our contributions are the following:

- **Social**: We analyze social responses to IPPV in Iran, a country with stark gender inequality. To our knowledge, domestic violence through the lens of bystanders’ responses has not been studied in the context of Iran.
- **Resource**: We release a novel dataset of 3,400 comments on IPPV posts1.
- **Participatory AI**: Our work is grounded in lived experiences; our key annotator is a woman who has lived for more than a decade in Iran and has a history of advocacy for social rights. Beyond annotation, she has also developed the categorization of comments on IPPV.

### 2 Dataset: Ninisite

We consider Ninisite.com (Nini in Persian means baby) (Ninisite from hereafter), one of the most popular websites in Iran focusing on health and parental issues faced by women [Jadesi, 2022]. According to Similarweb2, Ninisite is the #1 social interaction website in Iran with more than 50 million user visits per month. Ninisite’s tagline indicates that women are its primary audience (Please participate in the Ninisite exchange and connect with thousands of other mothers). With the majority of Ninisite users identifying as female [Jadesi, 2022], Ninisite offers a discussion forum to talk about issues such as family, motherhood, fertility, infertility, pregnancy, childbirth, and childcare [Rahbari, 2021]. In contrast with traditional social networking websites, the primary focus of Ninisite is not networking. User accounts are mostly anonymous [Rahbari, 2021; Jadesi, 2022] where they post queries seeking health and/or life advice starting with a unique post thread for a given query.

<table>
<thead>
<tr>
<th>Sub-forum</th>
<th># comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>problem - solution</td>
<td>6,333,333</td>
</tr>
<tr>
<td>spouse</td>
<td>4,644,661</td>
</tr>
<tr>
<td>meeting memories</td>
<td>1,244,661</td>
</tr>
<tr>
<td>wedding preparation and dowry</td>
<td>757,601</td>
</tr>
<tr>
<td>legal consultancy</td>
<td>284,824</td>
</tr>
</tbody>
</table>

Table 2: Count of comments from individual Ninisite sub-forums.

Other users (also anonymous accounts) respond to these posts. Prior literature has released a community question answering dataset [Boroujeni et al., 2022] from Ninisite with no mention or focus on IPPV.

The exchange of opinions section on Ninisite consists of sub-forums about various issues. Each user can go to each of these sub-forums and create a post. Every other user can see the post and write their comments on that specific post. For our project, we collect all posts from the following sub-forums: طرح مسئله - راه حل (problem - solution); خاطرات آشنایی (meeting memories); جهیزه و مقدمات ازدواج (wedding preparation and dowry); and مشاوره حقوقی (legal consultancy). We select these sub-forums because these sub-forums are likely to contain IPPV posts as opposed to sub-forums such as (art and literature). We obtain 275,370 (problem - solution), 160,731 (spouse), 41,830 (meeting memories), 300,90 (wedding preparation and dowry), and 15,211 (legal consultancy) posts from these forums. Overall, we collect 523,232 posts (denoted by \( D_{\text{post}} \)) created by more than 100,000 unique users (100,104) from August 2011 to June 2023. Each of these posts attracted 26 comments on average.

We next collect all comments on the aforementioned posts. Overall, we obtain 13.2 million comments (\( D_{\text{comments}} \)) posted by 240,519 unique users. Table 2 breaks down comments counts from individual sub-forums. We notice that compared to problem - solution, legal consultancy receives considerably less user engagement.

### 3 Background

We present a brief background on gender equality gap and information access barriers in Iran to sensitize and contextualize a global audience about the impact of this research.

#### 3.1 Gender Equality Gap in Iran

As per the Global Gender Gap Report3 published by the World Economic Forum in 2023, Iran ranks among the top 10 countries with the highest gender inequality. As per the report, only three countries have worst gender equality gap than in Iran: Algeria, Chad, and Afghanistan. Virginity is still a commodity in marital settings in Iran. Women often need certificates to prove they are virgin before marriage, a demand the World Health Organization (WHO) deems to be against human rights 4.

Gender struggles in Iran have witnessed self-sacrifice in

---

1Available on our project web page.
2Similarweb.com
protest in the recent past. Following Mahsa Amini’s death in police custody in 2022, the country witnessed an unprecedented protest concerning the moral policing of women wearing hijabs (headscarves) [Khorramrouz et al., 2023]. The hashtag #MahsaAmini was one of the most shared hashtags ever in the history of the Persian language Twitter.

3.2 Information Barrier

Popular social interaction platforms such as Twitter, Telegram, Facebook, and Instagram are blocked in Iran [Kargar and McManamen, 2018]. Technically savvy Iranians access such platforms through VPNs [Kermani, 2023]. Ninisite is accessible both from within Iran and outside Iran. Hence, the platform makes a wider cross-section of the Iranian population visible who does not have the means to bypass state surveillance and engage in sociopolitical discourse in global forums blocked in Iran. Ninisite also makes conscious efforts to stay as apolitical as possible [Rahbari, 2021]. It almost seems that the Twitter-sphere and Ninisite are two separate universes that seldom collide. While the Persian language Twitter-sphere was strife with the hashtag #MahsaAmini during the tumultuous time in 2022 [Kermani, 2023], the entire Ninisite forum records only one locked thread that mentions this hashtag underscoring Ninisite’s intention to steer clear of political controversies through firm moderation policies.

4 Related Work

Detecting posts expressing domestic abuse and violence from social web data has a vast body of prior literature [Schraed et al., 2015; Homan et al., 2020; Salehi et al., 2023; McCauley et al., 2018; Al-Garadi et al., 2022; Hassan et al., 2020]. Detecting criticality of abuse posts in Persian twitter [Salehi et al., 2023] has also been previously studied. Our work contrasts with existing literature in the following key ways. Instead of the victim’s account, we analyze how the audience responds to such first-person accounts of IPPV, an aspect (to our knowledge) that has never been studied in the context of Iran.

Through the active involvement of an Iranian woman in this AI-building process, our work resonates with the growing literature on participatory AI [Harrington et al., 2019; Delgado et al., 2022; Bondi et al., 2021; Birhane et al., 2022; Ovalle et al., 2023]. Annotated datasets form the core of supervised machine learning solutions. However, while addressing issues faced by vulnerable communities, the datasets often end up being annotated by annotators with little or no documentation [Guest et al., 2021; Ramesh et al., 2022]. Recent counter-examples include: (1) Ramesh et al. [2022] that present a lexicon of queer-related inappropriate words where one of the annotators identifies as queer; and (2) Guest et al. [2021] that present a misogyny dataset where the majority of the annotators identify as women. Our work draws inspiration from the aforementioned efforts.

Finally, our work draws from a diverse set of ML subfields [Palakodety et al., 2020; KhudaBukhsh et al., 2020] and NLP advancements [Bojanowski et al., 2017; Farahani et al., 2021; Brown et al., 2020] to tackle extreme class imbalance in a low-resource language task.

5 Modeling

Ninisite discusses a broad range of topics of which IPPV is expected to be a tiny fraction. We first need a reliable content classifier for IPPV given a post. Once we can detect IPPV posts with high precision, our next goal is to build a robust classifier to characterize the comments on these posts.

5.1 Detecting Posts Recounting IPPV

We first seek to narrow our search down to threads potentially discussing physical abuse. Given that random sampling of posts is expected to yield very few instances of physical violence let alone IPPV, following prior literature on mining rare class instances [Palakodety et al., 2020; KhudaBukhsh et al., 2020], we conduct nearest neighbor sampling in the document embedding space. We obtain 1,110 posts using our nearest neighbor sampling method (denoted by $T_{neighbors}$). As a control group, we randomly sample 1,110 posts from the set $D_{posts} - T_{neighbors}$. We annotate these 2,220 instances using two human annotators and a large language
The human annotators obtain near-perfect agreement (Cohen’s $\kappa$ score 0.92) and through an adjudication step the remaining few disagreements were resolved. However, between the lead human annotator and GPT-3.5, the agreement was substantially less (Cohen’s $\kappa$ score 0.37). The confusion matrix between GPT-3.5 labels and the human labels is presented in Table 3. We observe that when GPT-3.5 predicts no, almost on all occasions, humans also verify that the true label is no. Hence, for a low-resource language setting, GPT-3.5 can be used as a first line of defense to eliminate posts that are unlikely to indicate physical violence. The lone exceptions are scenarios that require a humane understanding of appropriate behavior. For instance, in a post, the user described that the user squeezed the hands of a mentally challenged child very hard. Both annotators marked this example as exhibiting physical violence where GPT-3.5 failed to catch that humane nuance. On several occasions, GPT-3.5 predicts the presence of physical violence that is refuted by both human annotators. We observe that GPT-3.5 often confuses with posts indicating screaming or shouting as instances of physical violence.

### Lead human annotator

<table>
<thead>
<tr>
<th>GPT 3.5</th>
<th>no</th>
<th>yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>no</td>
<td>96.55%</td>
<td>3.45%</td>
</tr>
<tr>
<td>yes</td>
<td>60.38%</td>
<td>39.62%</td>
</tr>
</tbody>
</table>

Table 3: Confusion matrix between human labels and GPT-3.5 on the task of annotating physical violence.

Our key takeaway is while zero-shot GPT-3.5 may present a cheap way to annotate examples, for tasks as sensitive as detecting IPPV, human annotations are more reliable. Overall, we obtain 430 posts indicating physical violence of which 345 posts describe IPPV.

In addition, we also observe that GPT-3.5 could not reliably identify victims or perpetrators. To be cost-efficient, we use a well-known LLM ParsBERT and build an IPPV classifier that achieves a performance of 82.36 F1 score.

We use this classifier to identify 2,137 IPPV posts from $D_{post} (0.4\%$ of all posts in $D_{post}$). A manual inspection of the high-confidence predictions indicates that our classifier performs reliably in the wild. Furthermore, to ensure we do not have any false positives, following Halterman et al. [2021], the lead annotator annotated all of the 2,137 data points and filtered out non-IPPV posts. This step yields 1,643 posts indicating an IPPV.

On average, we notice that IPPV posts attract 46.34 comments. However, from the same sub-forums, non-IPPV posts attract 26.29 comments. Hence, IPPV posts attract more user engagement on average. This result indicates that even though IPPV posts are rare, they elicit more engagement from Ninisite users. Overall, we obtain 53,220 comments on the IPPV posts. Table 5 presents the distribution of IPPV posts across different sub-forums.

Before characterizing the IPPV comments, we present an analysis of the distribution of victims and perpetrators based on human annotation because we feel that this would shed critical insights into understanding how deep-seated physical violence is in Iranian households [Tsang and Stanford, 2006]. Figure 2 presents the distributions of victims and perpetrators in IPPV threads as annotated by humans. We note that women are the overwhelming majority of the victims while men are the overwhelming perpetrators. A primary scenario when men get identified as victims is when women decide to counter physical violence with physical violence. Most IPPV threads are first-person accounts while a handful describes accounts of a close family member (e.g., a daughter recounting how her father beats up her mother). When a woman is a victim, not always the husbands are the sole perpetrator. A woman also faces violence from her in-laws and her father [Kousha, 2002]. Extant literature indicates that women often grow up getting socially conditioned that it is acceptable to be at the receiving end of violence [Rahmatian, 2009].

#### 5.2 Characterizing IPPV Comments

Our lead annotator is a person who (1) identifies as a female; (2) has fluency in Persian and Dari; and (3) has a record of active participation in advocacy for human rights. We first conduct open coding [Hancock et al., 2001] where our lead annotator examines 200 comments on IPPV threads and observes that the following broad themes emerge (illustrative examples are listed in Table 4):

1. **Escalate (familial):** comments that urge the victim to reach out to senior members of the family.
2. **Escalate (legally):** comments that urge the victim to seek legal action (e.g., gather forensic evidence of bruises, consult a lawyer for a divorce, complain to police about domestic violence, etc.).
3. **Confront:** comments that urge the victim to assume more agency and resist IPPV.
4. **Retaliate:** comments that urge the victim to counter physical violence with physical violence or by non-cooperation.
5. **Conform or adjust:** comments that urge the victim to give in to the demands of the perpetrator or alter the behavior of the victim to avoid IPPV (e.g., if talking to the victim’s family is causing tension resulting in physical violence, advising not talking to the family).
6. **Victim-blame:** comments that suggest that the victim is to be blamed for their misfortune.
7. **Empathize or support:** comments either are critical of the perpetrator or empathetic with the victim but provide no solution (e.g., offering prayers for the victim).
8. **Unrelated or other**

We collapse these fine-grained classes into the following three broad categories:

1. **conform or adjust or victim-blame (conform):** these comments put the onus on the victim.
2. **confront or retaliate or escalate (confront):** these comments urge the victim to assume more agency and seek outside help.
<table>
<thead>
<tr>
<th>Theme</th>
<th>Comments on IPPV</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Escalate (familial)</td>
<td>بیحود میکنی بزن مثله هر کی عصبانی میشه باید سر نزن و بچش خالی که باید یا برادرت یا باغبان هحف بزنن...</td>
<td>It is pointless for him to resort to violence. Just because someone gets angry doesn’t mean they should take it out on their wife and children. Tell your father or brother to talk to him.</td>
</tr>
<tr>
<td>Escalate (legal)</td>
<td>طلاق آخر دنبال نیست. ازش جدا شو و زندگی تو از تو بیان بزن. فکر خودت باش و بیگانگ. مختلف کن با زورگوگوشی‌تان.</td>
<td>Divorce is not the end of the world. Separate from him and rebuild your life anew. Prioritize your own peace of mind over people’s words. Focus on yourself and stand up for what you believe in. Resist their bullying.</td>
</tr>
<tr>
<td>Confront</td>
<td>تونه دوتا بزن و سطح پاش تا جیب بکشه مثل اسب بفهمه دست بردن یعنی چی.</td>
<td>You also hit him twice in the middle of his legs until he screams like a horse, so he understands what beating means.</td>
</tr>
<tr>
<td>Retaliative</td>
<td>تنها کاری که میتونی همیشه دست بزنی حتی جیب گردد. وقتی هر کسی، شما نیاز به شدید تحسین کنی.</td>
<td>The only thing you can do is to love more even in spite of being beaten and to control your tongue. It’s for your child’s sake, bear with him. …</td>
</tr>
<tr>
<td>Conform or adjust</td>
<td>این بچه چیکارش میکنه که می‌زنند خوژه رو دوسته کردن کی. بیشتر باید شد بهتر تشکر کنی.</td>
<td>…Look at what you’re doing to provoke him. If you don’t bother the snake, it won’t bite you. …</td>
</tr>
<tr>
<td>Victim-blame</td>
<td>آبنی خوهد جیکارتش میکنه که می‌زنند به مار… آسیب کارتی نداشته باشی که به نیست نمی‌زنی.</td>
<td>…I feel so sorry for you sister…</td>
</tr>
<tr>
<td>Empathize and support</td>
<td>دل کرد کرده چکار برات حواهر…</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Illustrative examples from our dataset categorized into seven broad themes as observed by our lead annotator.

![Figure 2: Distribution of victims and perpetrators in IPPV posts.](image)

Table 5: Relative distribution of IPPV posts across sub-forums. Of all the IPPV posts identified, 55.0% of posts come from the spouse forum.

<table>
<thead>
<tr>
<th>Sub Forums</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spouse</td>
<td>55.00%</td>
</tr>
<tr>
<td>Problem Solution</td>
<td>40.54%</td>
</tr>
<tr>
<td>Legal Consultancy</td>
<td>3.12%</td>
</tr>
<tr>
<td>Meeting Memories</td>
<td>0.83%</td>
</tr>
<tr>
<td>Wedding Preparation</td>
<td>0.51%</td>
</tr>
</tbody>
</table>

3. **empathize and support (empathize):** these comments are either critical of the perpetrator or empathetic to the victim but do not recommend any actionable path.

5.3 **Annotating IPPV Comments**

**Annotators** Our annotations were conducted by two annotators. Both are fluent speakers of Persian and have undergraduate degrees. Our lead annotator identifies as a cisgender woman, is grounded in Iranian culture, and has a record of advocacy for human rights. The other annotator identifies as a cisgender man.

**Disagreement Resolution** Extant literature indicates diverse approaches to resolve inter-annotator disagreements (e.g., majority voting [Davidson et al., 2017; Wiegand et al., 2019] or third objective instance [Gao and Huang, 2017]). After every round of independent annotation, we resolve disagreements through an adjudication step. The remaining unresolved instances are resolved by an expert computational social scientist (who identifies as a cisgender male) who does not know Persian. The lead annotator translated the instances in question into English.

**Inter-rater Agreement** After first round of annotation, we obtain Cohen’s $\kappa$ score of 0.56. After the adjudication step, we obtain a Cohen’s $\kappa$ score of 0.68.

5.4 **Active Learning Pipeline**

For better model accuracy, we break each comment into chunks using hazm \(^8\) (a Persian NLP tool) and label each chunk.

---

\(^8\)https://github.com/roshan-research/hazm
chunk separately. We initially annotate 1,500 chunks randomly sampled from 1,348 comments on 562 IPPV posts. Upon annotation and disagreement resolution, we obtain 388 empathize, 479 confront, 344 conform, and 289 others. The others are broadly direct responses to a question (e.g., how many days should I wait before hearing from a legal counselor), irrelevant comments, or misinformation (e.g., recommending visiting a magician or psychic).

Active learning is a well-established supervised machine learning technique where models (aka the learner) actively request labels for unlabeled instances [Settles, 2009]. Since our goal is to use the trained model for a social inference task, it is important to rectify high-confidence misclassifications. Minority class certainty sampling has found its use in rectifying high-confidence misclassifications and addressing class imbalance for classifying short documents such as movie reviews [Sindhwani et al., 2009; Attenberg et al., 2010], comments on YouTube videos [Palakodety et al., 2020], and search queries [KhudaBukhsh et al., 2015]. Our sampling strategy resembles ensemble sampling strategies described in Palakodety et al. [Palakodety et al., 2020] where we conduct (1) one round of certainty sampling for the classes confront and conform to address our initial class imbalance and (2) one round of uncertainty sampling. Overall, our final dataset (denoted by $D_{\text{final}}$) consists of 742 empathize, 1,158 confront, 879 conform, and 621 others instances.

During the annotation and adjudication process, our annotators realized that distinguishing between empathize and other was particularly challenging. Since our main goal is to understand social conditioning in Iran and a natural human response to stories of trauma would be empathy, our primary interest is to study the relative distribution of confront and conform. We thus combine the empathy class with other and report the performance of a three-way classifier in Table 6.

We again train a ParsBERT classifier to classify IPPV responses. Since many comments are arbitrarily long (79.19 ± 26.08 tokens), we observe that the performance slightly improves if we use GPT-3.5 to generate a summary of the posts first and then append the summarized post to the comment. Table 6 summarizes the performance of our machine learning models. We observe that the model trained on the final dataset with comments appended with summarized posts yielded the best performance.

6 Results: Analyzing IPPV Responses

We use our best-performing model and run inference on all comments on IPPV posts. Figure 3 indicates that between conform and confront, confront is the more common suggestion. Nonetheless, confront has a considerable presence. It is unclear what is the gender distribution of the responders. However, given that the primary audience of this forum is women [Jadesi, 2022], our results perhaps indicate that women often hesitate to advise other women to assume more agency possibly due to social conditioning that violence against women is acceptable [Rahmatian, 2009; Ofei-Aboagye, 1994]. When we investigate the temporal trend of IPPV comments, we notice a slight increase in the relative proportions of confront over the years. In 2018, confront represented 28.2±0.31 of the comments while in 2022, it increased to 35.2±0.37.

7 Conclusion and Discussion

This paper marks the first attempt (to our knowledge) to understand social web comments to (mostly) first-person accounts of IPPV in a country with stark gender inequality. Our analyses have the following takeaways. First, we note that IPPV posts elicit more user engagement perhaps indicating that Ninisite audience are receptive and are willing to engage with IPPV posts. Second, beyond IPPV, we note instances of a broad range of other types of domestic violence (e.g., daughters beaten by their fathers). We also observe that women constitute the lion’s share of the victims while men represent the overwhelming majority of perpetrators. Third, although confront outnumbered conform among comments, a considerable slice of comments still suggest conform. Finally, we observe a trend of a slight increase in confront comments over the years which perhaps indicates an improvement, albeit excruciatingly slow, in Iran’s gender struggles.

Methodologically, we address extreme class imbalance to form meaningful datasets for robust detection of IPPV posts and classification of IPPV comments. Our research draws from a broad range of NLP [Bojanowski et al., 2017; Farahani et al., 2021; Brown et al., 2020] and ML subfields [Palakodety et al., 2020; KhudaBukhsh et al., 2020] to develop our robust pipeline. That said, we believe one of our key contributions is our participatory AI design. With the advent of the new age LLMs, a key question in CSS and broadly AI research would be how well are we integrating the voices of the real stakeholders in our AI system-building processes. In that sense, our paper makes a small contribution to the worldwide inclusive AI effort.

Our research leads to the following points to ponder upon. Other categories of IPV and broader study on domestic violence. Of the four categories of IPV against women, we only focus on physical violence and sexual coercion. Our annotators observed rampant examples of emotional abuse (e.g., humiliating and belittling in front of the victims’ family members) and controlling behaviors (e.g., husbands requesting wives to share mobile passwords). This research will open the gates for broader study considering these additional aspects of IPV to form a more nuanced understanding of dif-

![Figure 3: Distribution of comments on IPPV.](image-url)
different IPV categories and societal responses. Also, Figure 2 reveals that Ninisite can provide important social web data on other forms of domestic violence. Contrastive analyses of societal response to different forms of domestic violence could be interesting follow-on research.

**Contrasting other cultures and countries** How do other societies or cultures respond to first-person accounts of IPV? Our study presents an important analysis of IPPV in Iran. Contrasting our findings with Western countries and other countries or regions with documented gender inequality merits deeper exploration.

**Limitations**

Prior literature indicates that Ninisite is a heavily moderated forum. Censorship practices can affect distributional properties in discourse [Bamman et al., 2012]. Censorship practices in Ninisite can thus skew the distribution of comments.

Our study on binary gender inequality runs the risk of oversimplifying gender, which lies on a spectrum. Further, same-sex marriage is yet not legal in Iran. Further nuances will be needed to extend our work to other cultures allowing same-sex marriages. We are also sensitive to previous studies that point out the potential harm of erasure of gender and sexual minorities [Dev et al., 2021]. We train our models on top of a well-known language model ParsBERT. It is possible that our analysis was influenced by inherent biases present in ParsBERT as documented in prior literature on LLMs [Bender et al., 2021].

While extant literature reports that the majority of Ninisite participants are women, the demographics of the audience are unclear. We do now know how many of them come from big cities, or how many of them have college education. Having more accurate information about the audience will make our study more meaningful. Finally, there could be many more women in Iran who would be afraid to voice their grievances in Ninisite. Our work does not cover their stories.

**Ethics Statement**

Ninisite is a publicly accessible forum where most accounts are anonymous with a low probability of doxing [Rahbarg2021]. Publicly available datasets from this forum has previously been featured in published datasets. We further conduct aggregate analysis on the comments refraining from user-focused analyses. Hence, we do not see any ethical concern. Rather, we believe our findings and methods can benefit policymakers and social scientists.

**Acknowledgements**

We thank Terrance Martin, a therapist with a medical record of treating trauma and abuse patients for more than a decade for his input.

**Contribution Statement**

Ashiqur R. KhudaBukhsh is the corresponding author. Adel Khorramrouz and Mahbeigom Faiyyazi are equal-contribution first authors.

**References**


[Kargar and McManamen, 2018] Simin Kargar and Keith McManamen. Censorship and collateral damage: Ana-


