Analyzing Femicide Reactions in YouTube Comments: a Comparative Study of Giulia Cecchettin and Carol Maltesi

Sveva Silvia Pasini[1,*,†], Marco Madeddu[2,†], Chiara Ferrando[2,†], Chiara Zanchi[1] and Viviana Patti[2]

[1]University of Pavia, Department of Humanities, Strada Nuova 65, 27100 Pavia, Italy
[2]University of Torino, Computer Science Department, Corso Svizzera 185, 10149 Torino, Italy

Abstract

Nowadays, Gender-Based Violence (GBV) has undergone a normalization process, whereby violent behaviors, by being justified as normal, have become subtle and difficult to recognize. In NLP, GBV has been investigated within the broad topic of Hate Speech detection, distinguishing between the different targets of hateful contents. Considering the pervasiveness of GBV and its media representation in our society, the main goal of our research is to explore people's reactions to femicide events, considered the most brutal expression of GBV. In particular, we collected 932 YouTube comments in response to the news regarding Giulia Cecchettin's femicide and we proposed an annotation task through a fine-grained annotation schema that builds upon Ferrando et al. [1] with some modifications. The qualitative analysis of the annotated comments revealed some differences from the GBV-Maltesi dataset [1], especially regarding misogyny, aggressiveness and responsibility attribution. We tested different LLMs, investigating their ability to recognize the presence of aggressiveness and responsibility in both Maltesi and Cecchettin datasets and to indicate their target, using different prompts.
Warning: This paper contains examples of offensive content.

Keywords

Hate Speech, Femicide, Responsibility framing, Social media

1. Introduction

A 2024 survey from the EU, involving 114,023 women aged between 18 and 74, revealed that one out of three women experienced some form of violence starting from the age of 15[1]. Taking into account the alarming situation, in this contribution we intend to investigate and analyze the perception of Gender-Based Violence (GBV), which can be defined as a form of violence directed against a person caused by the person's gender or that affects persons of a particular gender disproportionately[2]. Nowadays, GBV has undergone a process of normalization that has made the physical, sexual, psychological and economic harms more subtle and difficult to recognize, spreading cultural beliefs and values that support and justify the perpetration of GBV by presenting it as a normal component of relationships [3]. In addition, in online contexts, GBV includes a broad range of behaviors which are facilitated through a range of digital technologies [4]. These practices are expanding continuously and include non-consensual sharing of images and videos, deepfakes, social media-based harassment, and the dissemination of private information [5]. In Natural Language Processing (NLP) field, GBV is part of the broad topic of Hate Speech (HS) detection. Several studies investigate GBV by analyzing specific misogynistic [6, 7, 8], homophobic and transphobic [9, 10, 11], or sexist discourses [12, 13] depending on the target affected by the hateful contents.

It is essential to emphasize that GBV is understood as a continuum of violence with a pyramidal structure, in which each layer of the pyramid both contributes to and stems from a culture (often referred to as "Rape Culture") that normalizes sexist behavior within society [14]. From the base upward, each act of violence is a direct consequence of the previous ones, up to the apex of the pyramid which consists of femicide, i.e. the intentional elimination of a person for gender-related motivation[3].

Considering the pervasiveness of GBV, our research consists of an analysis of its public perception, carried out by collecting people's reactions to femicide news on YouTube.

Building on the assumption that certain sociodemographic characteristics of the victims might have an im-

CLiC-it 2025: Eleventh Italian Conference on Computational Linguistics, September 24 — 26, 2025, Cagliari, Italy [2]
*Corresponding author.
†These authors contributed equally.
svevasilvia.pasini01@universitadipavia.it (S. S. Pasini); marco.madeddu@unito.it (M. Madeddu); chiara.ferrando@unito.it (C. Ferrando); chiara.zanchi@unipv.it (C. Zanchi); viviana.patti@unito.it (V. Patti)
0009-0004-5620-0631 (M. Madeddu); 0009-0000-9593-2510 (C. Ferrando); 0000-0001-9216-9986 (C. Zanchi); 0000-0001-5991-370X (V. Patti)

[1]https://eige.europa.eu/publications-resources/publications/eu-gender-based-violence-survey-key-results
[2]https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/gender-equality/gender-based-violence/what-gender-based-violence_en
[3]https://www.unwomen.org/en/articles/explainer/five-essential-facts-to-know-about-femicide

pact on determining the lesser or greater spread of the news [15] and its perception, this study aims to compare two cases of femicide involving victims who differ in key characteristics that shape public perceptions of the event. We intend to do so by adopting the same methodology previously developed by [1] (whose annotating scheme is reported in Appendix A): in their contribution, the authors analyzed YouTube comments reacting to Carol Maltesi's femicide news, a 26-year-old single mother and sex worker who was brutally killed by her ex partner, Davide Fontana. The same methodology will be adopted in the case of Giulia Cecchettin, a 22-year-old university student killed by her former partner, Filippo Turetta, in November 2023 in Padua, Italy. Although Giulia Cecchettin and Carol Maltesi share some common features, such as age, skin color and origin, they differ significantly in others, such as motherhood and job. From an intersectional perspective, in which different axes of identity such as gender, ethnicity, sexuality, class, and ability intersect [16] and create different degrees of discrimination [17], these sociodemographic dimensions may be relevant in influencing different news perceptions in users and merit further explorations in our study.

In addition, the Cecchettin case has been selected because of its significant media resonance (due in part to the young age of both the perpetrator and the victim), and the widespread public and social engagement it generated, largely due to the active involvement of Giulia Checchettin's family. Furthermore, the dataset has been created to allow a diachronic analysis. In fact, comments have been extracted throughout to cover the entire sequence of events that preceded the discovery of the body, i.e. the kidnapping of the victim and her disappearance for a week, elicited strong emotional responses from the public.

The most significant contributions of this work are detailed below:

- The creation of the GBV-Cecchettin corpus[4]: a collection of 932 annotated YouTube comments responding to news coverage of the Cecchettin femicide extracted from 33 videos (Section 3). This corpus proposes itself as a valid resource for both computational and social studies purposes.
- An analysis of the GBV-Cecchettin corpus, including a comparison of the main similarities and differences with GBV-Maltesi dataset (Section 4).
- Experiment specifically aimed at analyzing the automatic detection of aggressiveness and responsibility attribution in YouTube comments, performing both quantitative and qualitative analysis of the results (Section 5). These tasks can be useful for automatically assessing the impact of news framing.

[4] https://github.com/madeddumarco/GBV-Cecchettin

2. Related Works

In recent years, the escalation of GBV has made femicide a topic of daily discussion[5], exposing people to news and contents related to this extreme form of violence. Several researches pointed out the role of media in (mis)representing femicides, analyzing the topic from different points of view. On the one hand, previous studies investigated the narrative strategies adopted in news reporting these kinds of events, while on the other hand, they focused on humans' perceptions and opinions about femicides that emerge from and are co-constructed by news coverage.

Regarding the narrative strategies adopted in presenting femicides, it has been noticed that news media typically cover the killing of women as isolated incidents rather than as parts of a broader context [18] that stems from the pyramid of GBV. This narration can be damaging as people exposed to these forms of media may struggle to recognize it as a part of a widespread social problem [19], causing the persistence of violence.

Several linguistic studies focused on GBV representation in Italian media, creating corpora [20], and emphasizing dominant strategies and narrative patterns [21, 22, 23]. In this context, Mandolini [19] conducted a qualitative discourse analysis focused on journalistic narratives of femicide in newspapers which describe different attitudes in the portrayal of femicide. In particular, the author highlights discursive strategies that (directly or indirectly) blame the victim and implicitly excuse the perpetrator, referring to gender stereotypes and romantic love rhetoric. Moreover, other studies focused on the responsibility framing in GBV [24, 25, 26], specifically identifying lexical choices and syntactic constructions that overshadow the agentivity and responsibility of femicide perpetrators in Italian news [27]. Considering that different linguistic choices trigger different perceptions and responsibility attributions [24], Minnema et al. [28] involved human annotators to ascribe a degree of perceived responsibility to the perpetrator, to the victim, or to an abstract concept (such as jealousy). They also conducted experiments highlighting that such perception can be modeled automatically. Finally, in Minnema et al. [28], the authors introduced a new task of responsibility perspective transfer, exploring the challenge of rewriting descriptions of GBV to increase the perceived level of responsibility attributed to the perpetrator. This is particularly relevant to our contribution, as it highlights the crucial role of linguistic structures and narrative patterns in assigning different degrees of responsibility to different event participants.

[5] As noted by the national observatory managed by "Non Una Di Meno" association (https://osservatorionazionale.nonunadimeno.net/anno/), Italy is consistently affected by GBV, reporting 120 femicides in 2023, 115 in 2024, and 48 until June 2025.

To our knowledge, Ferrando et al. [1] is the only study that suggests a shift in paradigm and methodology, trying to emphasize the importance of analyzing the spontaneous users' perception of femicide news in social media. This contribution focused on the collection of YouTube comments and personal opinions manually annotated with an ad hoc annotation scheme. It resulted in the release of the GBV-Maltesi dataset containing YouTube comments related to Carol Maltesi's femicide. In particular, the authors proposed a fine-grained annotation scheme, able to investigate different aspects that are relevant to femicide events, noting the presence of empathy, misogyny, aggressiveness, responsibility, humor and other dimensions that are thoroughly described in the original paper and briefly discussed in Section 3.

3. Dataset

3.1. Data Collection

To build the GBV-Cecchettin dataset, we extracted 9440 comments from 33 different YouTube videos uploaded to the platform between November 11th 2023, the day Giulia Cecchettin and Filippo Turetta went missing, and December 7th, two days after Giulia Cecchettin's burial. Considering the great quantity of media content released regarding this femicide, we mainly selected videos uploaded by nationally relevant news broadcasters (e.g., *La Repubblica*,*Rai*, *Fanpage.it*), moved by three main motivations: avoid subjective interpretations and favor factual information, take advantage of the broad spectrum of users that navigate the platform daily, and mimic GBV-Maltesi data collection process [1] to ensure continuity with the previous research. Within the already mentioned time frame, we also identified three subsections for investigating how the users' perception shifted over the days. We aimed to analyze them separately at first and compare them later to detect any differences. The first time-section (or phase) contains 10 videos published from November 11th to 17th, when the case was still defined as a missing person case; for the second time-section we collected 13 videos uploaded between November 18th and November 25th, isolating the first reactions when learning about the femicide; the third and last section, which included 10 videos released between November 26th and December 7th, gathered users' last considerations and comments related to the funeral.

For each of the 33 videos we collected all first level comments. For the annotation phase we extracted 1,500 examples from the gathered comments, maintaining the original balance between time-sections resulting in 195, 1,073 and, 232 respectively for the first, second and third time-phases. The selection has been made using BERTscore [29], aiming to maximize the differences within the corpus. Adopting an intra-section approach, we compared the cosine similarity of every entry in each section and then selected the least similar on average to other comments. This method was applied with a two-fold objective: firstly, considering annotation to be a time-consuming task, we decided to avoid annotators labeling very similar or repetitive texts (e.g., RIP); secondly, various types of entries were needed as training sets for the experimental phase.

3.2. Annotation Scheme's Revisions

Bearing in mind a conjoined work with the two corpora, GBV-Cecchettin was labeled following Maltesi's annotation scheme (See Appendix A), proposed in Ferrando et al. [1]. However, trying to explore the phenomenon with even more accuracy, we partially altered the scheme with the following innovations:

- *Support*: originally labeled as *empathy towards the event*, encompassing any form of empathy shown towards the victims, their families, or the event in general, the category was renamed *support*, to better capture its broader emotional and ideological dimensions. Moreover, annotators were asked to specify the intended targets (could be multiple) of the support, indicating whether it was directed to the victim, the perpetrator, the victim's social network (VSN), the perpetrator's social network (PSN), the female population, or the male population.
- *Misogyny*: we added social and economic class to the already available intersectional labels.
- *Aggressiveness*: we added institutions, male population, and split social network into VSN and PSN to the already present targets.
- *Responsibility Attribution*: we added institutions, male population, physical and psychological factors, and split social network into VSN and PSN to the already present targets.
- *Humor*: we added the possibility to indicate a target among victim, perpetrator, VSN, PSN, media, rape culture/femicide.

In addition, we decided to include two new dimensions, *Topic* and *Extenuations*, the former motivated by the interest to monitor which aspects of the case were more discussed within the comments and the latter introduced to identify examples of the two common tendencies when dealing with GBV, which are the justification of the masculine and the victim blaming [14]. Topic proposed nine selectable options (Victim, Perpetrator, Victim and Perpetrator, Perpetrator and Victim, VSN, PSN, Media/Information, Rape Culture, Femicide and Other) while Extenuations presented four labels to choose from:

Dimension	GBV-Maltesi		GBV-Cecchettin	
	Yes %	No %	Yes %	No %
Misogyny	9.03%	90.97%	1.93%	98.07%
Intersectionality	4.63%	95.36%	0.54%	99.46%
Aggressiveness	24%	76%	21.57%	78.43%
Agg. Perpetrator	19.19%	80.81%	12.66%	87.34%
Agg. Victim	1.23%	98.77%	0.00%	100.0%
Agg. Social Network	0.88%	99.11%	2.47%	97.53%
Agg. Perpetrator Social Network	-	-	0.97%	99.3%
Agg. Victim Social Network	-	-	1.50%	98.50%
Agg. Male Population	-	-	0.64%	99.36%
Agg. Media	2.73%	97.27%	4.94%	95.06%
Agg. Institutions	-	-	1.72%	98.28%
Agg. Rape Culture	0.41%	99.59%	0.11%	99.89%
Responsibility	32.89%	67.11%	24.03%	75.86%
Resp. Perpetrator	22.09%	77.91%	7.94%	92.06%
Resp. Victim	6.55%	93.45%	1.61%	98.39%
Resp. Social Network	2.11%	97.89%	5.58%	94.42%
Resp. Perpetrator Social Network	-	-	3.86%	96.14%
Resp. Victim Social Network	-	-	1.72%	98.28%
Resp. Male Population	-	-	1.07%	98.93%
Resp. Media	0.99%	99.01%	0.97%	99.03%
Resp. Institutions	-	-	8.26%	91.74%
Resp. Rape Culture	4.06%	95.94%	1.07%	98.93%
Resp. Psycho-fisical Factor	-	-	1.50%	98.50%
Empathy towards the event/ Support	28.25%	71.75%	36.16%	63.84%
Sup. Perpetrator	-	-	0.97%	99.03%
Sup. Victim	-	-	22.75%	77.25%
Sup. Social Network	-	-	18.78%	81.22%
Sup. Perpetrator Social Network	-	-	2.25%	97.75%
Sup. Victim Social Network	-	-	16.52%	83.48%
Sup. Male Population	-	-	0.75%	99.25%
Sup. Female Population	-	-	1.07%	97.42%
Humor	3.14%	96.86%	1.29%	98.71%
Macabre	3.27%	96.72%	0.0%	100.0%
Context	97.51%	2.49%	1.60%	98.40%

Table 1
Distribution of all dimensions across the GBV-Maltesi and GBV-Cecchettin dataset.

- *Victimization of the perpetrator*: to be selected when comments highlight external factors that portray the perpetrator as a victim of the circumstances.
- *Psychologization*: to be selected when the perpetrator is described using terms or attributes that justify the killing because of a psychological instability.
- *Victim blaming*: to be selected when, although the perpetrator is held responsible for the killing, certain assumptions or claims are presented that partially or completely deny the victim's status as a victim.
- *Dehumanization of the perpetrator*: to be selected when the perpetrator's humanity (or part of it) is denied or when the perpetrator is diminished or ridiculed based on psychological or physical characteristic, particularly those irrelevant to the case.

4. Corpus Analysis

In the annotation phase, we involved 10 students from a Master's Degree course in Linguistics, 7 of whom self-identified as women and 3 as men, mostly interested in GBV-related matters. Each participant annotated 750 comments, with all examples being annotated 5 times each. All people involved participated voluntarily. Throughout the process, we held meetings with the annotators to clarify any doubt about the scheme.

We excluded all comments that were annotated as not classifiable by at least one annotator, ending up with 932 comments. All examples were aggregated via majority voting between annotators.

We report all statistics of the corpus in Table 1. The dimensions with the most positive examples are: Support (36.2% of the corpus), Responsibility Attribution (24%), and Aggressiveness (21.6%). We also report the statistics regarding the different time-parts in Appendix B. Analyzing the three different time-phases, we found that during the first week, Support was mainly directed to

the VSN (66.7%), while less attention was given to Giulia Cecchettin, the victim (36.9%), due to her unknown condition as a missing person. Although Turetta was also intended as such, the perpetrator was already perceived accountable for the femicide (50%), sharing the responsibility with his parents (PSN, 22.2%), blamed for how they educated their son. Turetta's accountability also explains the aggressive manifestations directed to him (61.9%). Once the femicide had been uncovered, the Support was re-oriented towards the victim (67%, while towards VSN it was reduced to 43%). In this second time-section, the institution was perceived as accountable as the perpetrator (institutions and perpetrator: both 33.1%) due to a detail that emerged from the reconstructions: regardless of a witness reporting Turetta's aggression, the police did not intervene. Aggressive comments against Turetta increased (66.4%) because he was both confirmed as the perpetrator and also because he was found still alive despite theories regarding a possible suicide after committing the femicide. The media also become a target of users aggression (16.4%), who found the pervasiveness of the report tactless and disrespectful. Moreover, users found it inappropriate to dedicate such great attention to a single victim or a case of femicide, arguing that many other events deserve the same visibility. The third and last time-section presents similar results regarding Support (victim: 75%; VSN: 33.3%) while it shows interesting outcomes for Responsibility Attribution and Aggressiveness; in both dimensions, the perpetrator (R: 21.4%; A: 34.8%) had been overlooked in favor of, respectively, institutions (64.3%) and media (45.7%), indicating the users overcoming the specific Cecchettin case to reflect on the role of the State and the media in the GBV phenomenon.

The results recorded for Responsibility Attribution and Aggressiveness, especially in the third time-section, highlight how in the last week the users started questioning and discussing the wider problem of GBV, going beyond the specific Cecchettin case to reflect on the role of the State and the media in both preventing, punishing and narrating the femicides. As the third time-section regards the broadcast of the victim's burial, aggressiveness towards the media is mostly a condemnation for exploiting both Cecchettin's murder and her relatives' pain for their gain.

4.1. Divergences and Similarities with GBV-Maltesi

In this Section, we present a comparative analysis of the reactions to the femicides of Maltesi and Cecchettin, highlighting similarities and divergences.

As mentioned above, the selection of the cases was guided by an intersectional approach, focusing on victims who presented diverse sociodemographic traits. Among those traits, motherhood and profession appear to be particularly influential in shaping user responses, with Maltesi being a single mother and sex worker, and Cecchettin being a student with no children. In fact, despite the common brutal nature of both femicides, the corpora statistics reveal notable differences in the expression of misogyny, empathy, aggressiveness, and the attribution of responsibility within the comments, proving how these characteristics are very influential in how online users perceive the two femicides.

To be more specific, Misogyny is present in only 1.9% of the GBV-Cecchettin entries, compared to 9.03% in GBV-Maltesi. These results are understandable considering the two victims' profiles: while Cecchettin was a young university student close to graduation and, therefore, harder to blame, Maltesi was a single mother and a sex worker, details often mentioned in the comments. This can also be noticed in the Intersectionality label, present in 0.5% of the comments in the former and 4.63% in the latter. Consequently, the empathy expressed towards the events had also been affected, since we register higher support for Cecchettin (36.2%) compared to Maltesi (28.5%). The lower empathy shown towards Maltesi causes users to be more ironic when discussing this case (3.14%), while less humor is shown in GBV-Cecchettin (1.3%). GBV-Cecchettin recorded few instances for Responsibility Attribution and Aggressiveness towards the victims: the former accounted for only 1.6% of the corpus while the latter was entirely absent. In contrast, GBV-Maltesi revealed a higher degree of Responsibility Attribution directed at the victim (6.55%), largely caused by her occupation as a sex worker. This aspect led many users to perceive her as partially responsible for the violence, therefore, justifying the perpetrator's aggression. In addition, her status as a single mother living apart from her child intensified both aggressive and victim-blaming narratives within the comments, as shown from the entries (e.g., *Ha abbandonato il figlio per darsi al porno, un rifiuto umano giustamente smaltito.* **English translation:** She abandoned her son to turn to porn, a human waste rightly disposed of.). Finally, these findings support our claim by illustrating how victims' sociodemographic traits influence users reactions to femicide news and shape their perceptions of blame attribution.

Considering other targets of Aggressiveness, Cecchettin's PSN is explicitly attacked when his family takes his side, trying to excuse Turetta for the crime. In particular, we observed more Aggressiveness towards Turetta's family, because since the perpetrator was a young student still living in the household his parents are perceived as partially involved in the crime. In GBV-Maltesi, the authors did not report any examples of attacks towards Fontana's family, probably because he was a 44 years old man, responsible for his own actions. On the contrary, users wrote aggressive comments against Maltesi's parents for not supporting their daughter or looking for

her for several months, e.g., *Caspita 3 mesi e nessuno si è insospettito che non rispondeva* (**English translation**: Wow 3 months and no one got suspicious that she didn't answer).

Among the various differences, the two corpora also present some similarities: Media were rarely the target of Responsibility attribution, and Aggression toward Rape Culture and Victim also show similar outcomes in both corpora, with a 0.41% in GBV-Maltesi and 0.11% in GBV-Cecchettin for the former and a 1.23% in GBV-Maltesi and absent in GBV-Cecchettin for the latter.

5. Experiments

In this Section, we report the experiments we conducted to demonstrate applications of the resource in NLP. All experiments have been carried on both GBV-Cecchettin and GBV-Maltesi. We used LM-Eval-Harness [30] to generate all outputs.

We focused on the categories of Aggressiveness and Responsibility Attribution as these dimensions are particularly susceptible to the narrative framing of news coverage. Thus, automatic analysis of users comments can offer a deeper understanding of how specific narratives influence public perception.

First, we explain the experimental setting by describing the tasks, listing all models and prompts used. Then, in Section 6, we report and analyze the results obtained from the various models across all tasks.

5.1. Tasks

We carried the following four tasks:

- **Aggressiveness Detection** (Agg_{binary}), which is a binary classification task on the presence of aggressiveness in a comment. The task is carried out in a multiple-choice setting.
- **Responsibility Detection** ($Resp_{binary}$), which is a binary classification on the presence of responsibility in a comment. The task is carried out in a multiple-choice setting.
- **Target of Aggressiveness Recognition** (Agg_{target}), in which the model is given a comment and asked to list all targets of aggressiveness. The task is carried out in a generation setting, meaning that we had to post-process the outputs to extract the targets detected.
- **Target of Responsibility Recognition** ($Resp_{target}$), in which the model is given a comment and asked to list all targets that are attributed Responsibility. The task is carried out in a generation setting, meaning that we had to post-process the outputs to extract the targets detected.

5.2. Prompts

As LLMs can be sensitive to different formulation of prompts [31, 32], we designed four different prompt structures:

- *P1*, which is structured as following: first we explain the type of input (a comment), then we briefly describe the task to carry, we list all possible answers and the format we require.
- *P2*, which is structured as following: the description of the femicide case that can be found on Corriere della sera femicides observatory LaVentisettesimaOra [6] and then *P1*.
- *P3*, which is structured as following: a definition of the term 'femicide' and then *P1*.
- *P4*, which is structured as following: the definition of femicide, the importance of femicide awareness, the description from LaVentisettesimaOra and then *P1*.

For an example see Appendix C. We used these four prompt structures for all four tasks by just changing the description of the task. The description from LaVentisettesimaOra varied according to the corpora we used (Maltesi's description for GBV Maltesi and the same for Cecchettin).

5.3. Data Splits and Few-Shot

For Agg_{binary} and $Resp_{binary}$, we tested the models on the entirety of GBV-Maltesi and GBV-Cecchettin. Meanwhile, for Agg_{target} and $Resp_{target}$, we only interrogated models on examples that presented at least one target for the respective dimension. For all tasks, we tested the models in both a zero-shot and few-shot setting (in our case five examples). We did not perform any fine-tuning of the models.

Note that, GBV-Cecchettin and GBV-Maltesi have different lists of possible targets. Thus, we change the target list given inside the prompt depending on the dataset used.

5.4. Models

As we are testing LLMs on tasks that concern texts in Italian, we selected the five best unique models based on the LLM-Evalita leaderboard [33]. The models are the following: phi-4, gemma-2-9b-it, LLaMAntino-3-ANITA-8B-Inst, Qwen2.5-14B-Instruct and Llama-3.1-8B-Instruct.

6. Results Analysis

We report the results for the binary detection tasks, $Resp_{binary}$ and Agg_{binary}, in Table 2. Meanwhile, in Table 3

[6] https://27esimaora.corriere.it/la-strage-delle-donne/

Task	Few-shot	Dataset	Phi	Qwen	ANITA	Gemma	Llama
Agg_{binary}	0	Maltesi	0.45	**0.65**	0.49	0.42	0.29
		Cecchettin	0.37	**0.63**	0.49	0.42	0.27
	5	Maltesi	**0.62**	0.6	0.59	0.57	0.55
		Cecchettin	**0.68**	**0.68**	0.55	0.59	0.52
$Resp_{binary}$	0	Maltesi	0.58	0.55	0.58	**0.6**	0.52
		Cecchettin	0.53	**0.66**	0.59	0.57	0.53
	5	Maltesi	0.63	0.58	0.59	**0.66**	0.65
		Cecchettin	0.62	**0.68**	0.59	0.67	0.64

Table 2
F1-Macro (average scores across all prompts) for the binary tasks

Task	Few-shot	Dataset	Phi	Qwen	ANITA	Gemma	Llama
Agg_{target}	0	Maltesi	0.21	0.29	0.28	**0.45**	0.22
		Cecchettin	0.16	0.24	0.25	**0.36**	0.21
	5	Maltesi	0.43	0.51	0.47	0.51	**0.53**
		Cecchettin	0.38	0.43	0.42	**0.47**	0.43
$Resp_{target}$	0	Maltesi	0.21	**0.36**	0.33	0.33	0.3
		Cecchettin	0.19	0.29	0.24	**0.39**	0.26
	5	Maltesi	0.44	0.55	0.52	**0.56**	0.49
		Cecchettin	0.36	**0.48**	0.46	**0.48**	**0.48**

Table 3
F1-Macro (average scores across all prompts) for the target detection tasks

we reported the results for the target recognition tasks, $Resp_{target}$ and Agg_{target}. All results reported are the average of the F1-macro obtained by models on all prompts introduced in Section 5.2.

Overall Considerations In general, as expected, models performance improves in the few-shot setting compared to the zero-shot approach. The impact of few-shot varies depending on the model and task, e.g., LLama in Agg_{binary} performs very poorly when prompted in zero-shot but is aligned with other models in few-shot. Overall, models do not show noticeable differences in performance when being tested on the two different datasets. This could indicate that the Aggressiveness and Responsibility, shown in reaction to femicides, present similar traits across various cases. Moreover, this can be taken as a positive indication that the annotation process has been consistent across the two datasets despite involving different annotators.

Binary Classification Almost always, models in a zero-shot setting perform better in $Resp_{binary}$ compared to Agg_{binary}. We found it interesting as aggressive, and more generally abusive language, is usually a well studied phenomenon, meanwhile, responsibility detection is rather new. Analyzing the outputs, we found that is caused by the fact that models in the zero-shot setting for Agg_{binary} were generally biased towards the positive label (e.g., Llama predicting the positive label 90% of the times on average). The factors causing this behavior could be many, for example, the dramatic context of femicide and comments often citing the violence committed during the crime. Also, it could be hypothesized that most of the models taken in examination have gone through a post-processing phase where they are instructed to not generate aggressive or abusive text [34], thus creating strong biases towards certain terms that can be seen as aggressive.

Targets Recognition Moving on to the tasks focused on determining the targets of Aggressiveness and Responsibility ($Resp_{target}$ and Agg_{target}), we find that these tasks show lower scores than the detection tasks. This is not surprising as this is a multi-label classification task (compared to a binary) and models were interrogated in a generative setting instead of multiple-choice. This last point is important as models had to recognize how to format their output in a correct manner, which did not always happen. For these tasks, we do not see the trend of models performing better for Responsibility Attribution over Aggressiveness that we observed in the binary setting. In fact, different models performed better for one dimension and others performed better in the other. Also, we observe a sharper increase in performance when switching to the few-shot approach compared to the binary tasks. In fact, the majority of models gain 0.2 in F1-macro score.

Analysis of Recognition for Specific Targets We analyzed the outputs of Agg_{target} and $Resp_{target}$ to understand

Task	FS	Data	Vict.	Perp.	VSN	Media	Rape	PSN	M. Pop.	Instit.	Fact. F-P
Agg_{target}	0	Malt.	0.58	0.72	0.6	**0.79**	0.58	-	-	-	-
		Cecc.	0.42	0.67	0.61	**0.83**	0.48	0.67	0.57	0.62	-
	5	Malt.	0.64	**0.88**	0.69	0.84	0.55	-	-	-	-
		Cecc.	0.47	**0.87**	0.75	0.85	0.5	0.85	0.62	0.7	-
$Resp_{target}$	0	Malt.	0.51	0.46	0.6	**0.63**	0.54	-	-	-	-
		Cecc.	0.55	0.33	0.66	0.71	0.55	0.65	0.58	**0.74**	0.61
	5	Malt.	0.69	**0.77**	0.76	0.68	0.68	-	-	-	-
		Cecc.	0.63	0.73	0.71	0.8	0.54	0.79	0.55	**0.84**	0.69

Table 4
F1-Macro (average scores across all prompts) for target detection tasks. Scores for Gemma only. Missing values are due to the Cecchettin schema having more targets and physical and physiological factor not being a possible target of aggressiveness.

their performances on each possible target. In addition, we performed a qualitative analysis, investigating model behaviors.

First, we calculated F1-macro score (averaged across all prompts) for single targets, casting each target as a single binary label. As the number of possible combination between tasks, Few-shot, subset, models and list of targets is very large, we decided to only focus on the model that had the best overall performance across Agg_{target} and $Resp_{target}$, which is Gemma.

We reported the results in Table 4. Focusing on Agg_{target}, the model shows the best performance for specific targets, mostly the Perpetrator, PSN and the Media with F1-macro averages reaching 0.88. From a qualitative perspective, this can be caused by the explicitness of the comments that are aimed towards the perpetrator, with users expressing hatred towards him and often invoking a punishment consisting of a life-sentence. For instance, *Uccisa da un miserabile vigliacco . Essere orrendo.Giulia abbiamo perso tanto. RIP* (**English translation**: Killed by a miserable coward. Horrible person. Giulia, we lost a lot. RIP).

Meanwhile, $Resp_{target}$ shows different patterns, with the model not correctly recognizing the Perpetrator responsibility in the zero-shot setting. This can be due to the fact that Responsibility Attribution is more subtle and difficult to identify compared to Aggressiveness. In particular, the attribution of responsibility is not always conveyed through explicit and direct expressions, but it is often deduced from the context or the femicide event itself.

Model performs well on the institutions label, as comments explicitly attribute the responsibility to Italy's lack of severe punishments, even going so far as to invoke physical punishment or death for the perpetrator.

In both tasks, we observe that the model does not perform well in detecting the crucial victim label. In many cases, the reference to the victim's sphere was enough to recognize her as the target of aggressiveness, even if the intention was completely different. For instance, e.g., *Poverina aveva bisogno aiuto anche lei ascoltate le sue parole quel delinquente andremmo ammazzato mi dispiace* (**English translation**: Poor girl, she needed help too listen to her words that criminal should be killed I'm sorry) and *Non ci sono parole per descrivere questo schifoso. Mi dispiace tantissimo per lei e per la sua famiglia, pregherò per lei* (**English translation**: There are no words to describe this lousy guy. I feel so sorry for her and her family, I will pray for her). This may be due to the presence of certain terms or the recognition of the main referent in the comments without an understanding of the overall meaning of the sentence. In fact, in several cases it seems that the feminine form of adjective was sufficient to recognize the victim as target, even though the intention was to support her and take her side. For GBV-Maltesi dataset, the model recognizes the Responsibility attributed to the victim, specifically in comments that blame her for her own death, citing her life choices, her status as a mother living apart from her child, and her job, e.g., *Purtroppo le scelte di vita sbagliate e le sue abitudini la hanno esposta al male e a tanti rischi* (**English translation**: Unfortunately, her wrong lifestyle choices and habits have exposed her to evil and many risks) Notably, the ethical judgment commonly related to "she was asking for it" does not appear in the Cecchettin case.

7. Conclusion ans Future Works

In this paper, we present the GBV-Cecchettin dataset, which collects people's reactions to the news of Giulia Cecchettin's femicide. We chose the topic because of the pervasiveness of Gender-Based Violence in our society. We further improved the the fine-grained annotation schema proposed in Ferrando et al. [1] and applied it to a new femicide case. The GBV-Cecchettin is composed by 932 comments, annotated by 10 master students.

The annotated corpus shows interesting insights, revealing both similarities and divergences between GBV-Maltesi and GBV-Cecchettin. In particular, our analysis

focused on the different perceptions related to misogyny, aggressiveness, and the attribution of responsibility, emphasizing the role of victims' sociodemographic traits in shaping those perceptions. In the experimental phase, we tested several LLMs with four different prompts to both GBV-Maltesi and GBV-Cecchettin, to investigate their ability to detect the presence of aggressiveness and responsibility (binary classification task) and to identify their target from a fixed list (recognition task). The results reveal that the former task is easier than the latter. Aggressiveness binary detection seems to be a harder task then Responsibility detection given the violent nature of the femicide context. In the target recognition task, we found that some targets are easier than others. For example, aggressiveness towards the perpetrator is easier due to the explicitness of the comments directed towards him.

Despite its contributions, this study has several limitations that should be considered when interpreting the results. First, we reckon the comments selection procedure, although being motivated (see Section 3.1), can be considered inadequate to capture the users' perception of the Cecchettin femicide. Second, we acknowledge that involving solely Linguistics Master's Degree students as annotators might lead to biases in the annotated data.

The last two limitations we identify in this research lay the foundation for future work. Having only investigated data drawn from YouTube and recognizing its limitations, we aim to expand our data source in future work, wanting to gather entries from different platforms. Lastly, we are interested in exploring other languages and not limiting ourselves to Italian, adapting the fine-grained annotation schema in a multilingual study to develop a more global perspective on how GBV is perceived.

Considering the power of news media in making a difference for human rights in general and women's rights in particular [18], we strongly advocate the urgency of focusing on how different framing of news can lead to different online reactions. Therefore, as future work, we plan to study how specific narratives (e.g., terms used by the media) can directly influence users perception.

Ethical Consideration

The Cecchettin dataset was created in accordance with YouTube's Terms of Service. Among the 10 people involved in the annotation phase, 8 of them are Italian, one Russian and one US-American, all enrolled in a Italian MA Linguistics course. 9 of them claimed to be interested in GBV-related matters, and 5 had already taken part to GBV-related projects. All the annotators involved in this study participated voluntarily, without any incentives or obligation. From the beginning, we met with them several times to ensure that the topic did not disturb them psychologically or emotionally, offering support and help if they need it. This approach continued throughout all stages of the research.

Acknowledgments

We would like to thank all the students involved in the annotation task, who worked with professional attitude on a difficult topic like GBV.

The work has been partially supported by "HARMONIA" project - M4-C2, I1.3 Partenariati Estesi - Cascade Call - FAIR - CUP C63C22000770006 - PE PE0000013 under the NextGenerationEU programme.

References

[1] C. Ferrando, M. Madeddu, V. Patti, M. Lai, S. Pasini, G. Telari, B. Antola, Exploring YouTube comments reacting to femicide news in Italian, in: F. Dell'Orletta, A. Lenci, S. Montemagni, R. Sprugnoli (Eds.), Proceedings of the 10th Italian Conference on Computational Linguistics (CLiC-it 2024), CEUR Workshop Proceedings, Pisa, Italy, 2024, pp. 356–365. URL: https://aclanthology.org/2024.clicit-1.43/.

[2] C. Bosco, E. Ježek, M. Polignano, M. Sanguinetti, Preface to the Eleventh Italian Conference on Computational Linguistics (CLiC-it 2025), in: Proceedings of the Eleventh Italian Conference on Computational Linguistics (CLiC-it 2025), 2025.

[3] M. Rodelli, K. Koutra, K. B. Thorvaldsdottir, H. Bilgin, N. Ratsika, I. Testoni, D. M. Saint Arnault, Conceptual development and content validation of a multicultural instrument to assess the normalization of gender-based violence against women, Sexuality & Culture 26 (2022) 26–47.

[4] N. Suzor, M. Dragiewicz, B. Harris, R. Gillett, J. Burgess, T. Van Geelen, Human rights by design: The responsibilities of social media platforms to address gender-based violence online, Policy & Internet 11 (2019) 84–103.

[5] G. Abercrombie, A. Jiang, P. Gerrard-Abbott, I. Konstas, V. Rieser, Resources for automated identification of online gender-based violence: A systematic review, in: 7th Workshop on Online Abuse and Harms 2023, Association for Computational Linguistics, 2023, pp. 170–186.

[6] P. Zeinert, N. Inie, L. Derczynski, Annotating online misogyny, in: C. Zong, F. Xia, W. Li, R. Navigli (Eds.), 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), Association for Computational Linguistics, Online,

2021, pp. 3181–3197. URL: https://aclanthology.org/2021.acl-long.247/. doi:10.18653/v1/2021.acl-long.247.

[7] A. Muti, F. Ruggeri, C. Toraman, L. Musetti, S. Algherini, S. Ronchi, G. Saretto, C. Zapparoli, A. Barrón-Cedeño, Pejorativity: Disambiguating pejorative epithets to improve misogyny detection in italian tweets, arXiv preprint arXiv:2404.02681 (2024).

[8] E. Guest, B. Vidgen, A. Mittos, N. Sastry, G. Tyson, H. Margetts, An expert annotated dataset for the detection of online misogyny, in: P. Merlo, J. Tiedemann, R. Tsarfaty (Eds.), Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, Association for Computational Linguistics, Online, 2021, pp. 1336–1350. URL: https://aclanthology.org/2021.eacl-main.114/. doi:10.18653/v1/2021.eacl-main.114.

[9] B. R. Chakravarthi, P. Kumaresan, R. Priyadharshini, P. Buitelaar, A. Hegde, H. Shashirekha, S. Rajiakodi, M. Á. García, S. M. Jiménez-Zafra, J. García-Díaz, R. Valencia-García, K. Ponnusamy, P. Shetty, D. García-Baena, Overview of third shared task on homophobia and transphobia detection in social media comments, in: B. R. Chakravarthi, B. B, P. Buitelaar, T. Durairaj, G. Kovács, M. Á. García Cumbreras (Eds.), Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity, Inclusion, Association for Computational Linguistics, St. Julian's, Malta, 2024, pp. 124–132. URL: https://aclanthology.org/2024.ltedi-1.11/.

[10] D. Nozza, et al., Nozza@ lt-edi-acl2022: Ensemble modeling for homophobia and transphobia detection, in: Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion, Association for Computational Linguistics, 2022.

[11] J. Vásquez, S. Andersen, G. Bel-enguix, H. Gómez-adorno, S.-l. Ojeda-trueba, HOMO-MEX: A Mexican Spanish annotated corpus for LGBT+phobia detection on Twitter, in: Y.-l. Chung, P. R{\"ottger}, D. Nozza, Z. Talat, A. Mostafazadeh Davani (Eds.), The 7th Workshop on Online Abuse and Harms (WOAH), Association for Computational Linguistics, Toronto, Canada, 2023, pp. 202–214. URL: https://aclanthology.org/2023.woah-1.20/. doi:10.18653/v1/2023.woah-1.20.

[12] W. Lei, N. A. S. Abdullah, S. R. S. Aris, A systematic literature review on automatic sexism detection in social media, Engineering, Technology & Applied Science Research 14 (2024) 18178–18188.

[13] F. Rodríguez-Sánchez, J. Carrillo-de Albornoz, L. Plaza, Detecting sexism in social media: an empirical analysis of linguistic patterns and strategies, Applied Intelligence 54 (2024) 10995–11019.

[14] C. Vagnoli, Maledetta sfortuna, Rizzoli, 2021.

[15] P. Lalli, L'amore non uccide, Femminicidio e discorso pubblico: cronaca, tribunali, politiche (2021).

[16] K. Crenshaw, Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics, The University of Chicago Legal Forum 140 (1989) 139–167.

[17] H.-W.-S. Bao, P. Gries, Intersectional race–gender stereotypes in natural language, British Journal of Social Psychology (2024).

[18] C. Bouzerdan, J. Whitten-Woodring, Killings in context: An analysis of the news framing of femicide, Human Rights Review 19 (2018) 211–228.

[19] N. Mandolini, Femminicidio, prima e dopo. un'analisi qualitativa della copertura giornalistica dei casi stefania noce (2011) e sara di pietrantonio (2016), Problemi dell'informazione 45 (2020) 247–277.

[20] E. Cappuccio, B. Muscato, L. Pollacci, M. Marchiori Manerba, C. Punzi, C. Mala, M. Lalli, G. Gezici, M. Natilli, F. Giannotti, Beyond headlines: A corpus of femicides news coverage in Italian newspapers, in: F. Dell'Orletta, A. Lenci, S. Montemagni, R. Sprugnoli (Eds.), Proceedings of the 10th Italian Conference on Computational Linguistics (CLiC-it 2024), 2024.

[21] S. Abis, P. Orrù, et al., Il femminicidio nella stampa italiana: un'indagine linguistica, gender/sexuality/italy 3 (2016) 18–33.

[22] L. Busso, C. R. Combei, O. Tordini, Narrating gender violence a corpus-based study on the representation of gender-based violence in italian media, in: Language, Gender and Hate Speech: A Multidisciplinary Approach, 2020.

[23] F. Formato, Gender, discourse and ideology in Italian, Springer, 2019.

[24] C. Meluzzi, E. Pinelli, E. Valvason, C. Zanchi, Responsibility attribution in gender-based domestic violence: A study bridging corpus-assisted discourse analysis and readers' perception, Journal of pragmatics 185 (2021) 73–92.

[25] E. Pinelli, C. Zanchi, Gender-based violence in italian local newspapers: How argument structure constructions can diminish a perpetrator's responsibility, in: Discourse Processes between Reason and Emotion: A Post-disciplinary Perspective, Springer, 2021, pp. 117–143.

[26] P. Orrù, Femminicidio e violenza di genere nella stampa on-line: un'analisi quantitativa, Lingue e culture dei media 8 (2024) 175–187.

[27] G. Minnema, S. Gemelli, C. Zanchi, V. Patti, T. Caselli, M. Nissim, et al., Frame semantics for

social nlp in italian: Analyzing responsibility framing in femicide news reports, in: CEUR Workshop Proceedings, volume 3033, CEUR-WS, 2021, pp. 1–8.
[28] G. Minnema, H. Lai, B. Muscato, M. Nissim, Responsibility perspective transfer for italian femicide news, arXiv preprint arXiv:2306.00437 (2023).
[29] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, Y. Artzi, Bertscore: Evaluating text generation with bert, arXiv preprint arXiv:1904.09675 (2019).
[30] L. Gao, J. Tow, B. Abbasi, S. Biderman, S. Black, A. DiPofi, C. Foster, L. Golding, J. Hsu, A. Le Noac'h, H. Li, K. McDonell, N. Muennighoff, C. Ociepa, J. Phang, L. Reynolds, H. Schoelkopf, A. Skowron, L. Sutawika, E. Tang, A. Thite, B. Wang, K. Wang, A. Zou, The language model evaluation harness, 2024. URL: https://zenodo.org/records/12608602. doi:10.5281/zenodo.12608602.
[31] F. M. Polo, R. Xu, L. Weber, M. Silva, O. Bhardwaj, L. Choshen, A. F. M. de Oliveira, Y. Sun, M. Yurochkin, Efficient multi-prompt evaluation of llms, in: A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, C. Zhang (Eds.), Advances in Neural Information Processing Systems, volume 37, Curran Associates, Inc., 57 Morehouse Ln, Red Hook, NY 12571, United States, 2024, pp. 22483–22512.
[32] M. Sclar, Y. Choi, Y. Tsvetkov, A. Suhr, Quantifying language models' sensitivity to spurious features in prompt design or: How i learned to start worrying about prompt formatting, 2024. URL: https://arxiv.org/abs/2310.11324. arXiv:2310.11324.
[33] B. Magnini, R. Zanoli, M. Resta, M. Cimmino, P. Albano, M. Madeddu, V. Patti, Evalita-llm: Benchmarking large language models on italian, arXiv preprint arXiv:2502.02289 (2025).
[34] Y. Bai, A. Jones, K. Ndousse, A. Askell, A. Chen, N. DasSarma, D. Drain, S. Fort, D. Ganguli, T. Henighan, et al., Training a helpful and harmless assistant with reinforcement learning from human feedback, arXiv preprint arXiv:2204.05862 (2022).
[35] V. Basile, C. Bosco, E. Fersini, D. Nozza, V. Patti, F. M. Rangel Pardo, P. Rosso, M. Sanguinetti, SemEval-2019 task 5: Multilingual detection of hate speech against immigrants and women in Twitter, in: J. May, E. Shutova, A. Herbelot, X. Zhu, M. Apidianaki, S. M. Mohammad (Eds.), Proceedings of the 13th International Workshop on Semantic Evaluation, Association for Computational Linguistics, Minneapolis, Minnesota, USA, 2019, pp. 54–63. URL: https://aclanthology.org/S19-2007. doi:10.18653/v1/S19-2007.

A. GBV-Maltesi Scheme

Here we report the GBV-Maltesi guidelines that served as the starting point for the scheme used for GBV-Cecchettin. All dimension were used in GBV-Cecchettin, except for the ones that received changes as noted in 3.2.

- *Non classifiable:* if the comment cannot be analysed because it is not written in Italian, because it consists only of emojis, because it is not comprehensible or not relevant to the topic (any comment that was marked as NC from at least 1 annotator was removed from the corpus);
- *Empathy:* whether, in the comment, there are expressions of empathy in support of the victim, her family or the event in general (i.e., condolences);
- *Misogyny:* whether, in the comment, there is a presence of discriminatory expression against women, including blaming, objectifying, discriminatory and sexist practices used towards them and their life choices. If misogyny is present, we asked annotators to indicate its target (group or individual) based on [35]. Moreover, we asked to specify if the expressed misogyny contained intersectionality traits and to select from a list what other dimensions were involved: age, religion, job, nationality, skin color, class, sexual orientation, gender, physical condition, educational background, language and culture;
- *Aggressiveness:* whether there is aggressiveness in the comment and to whom it is directed (allowing multiple choices): victim, perpetrator, social network (family, friends, colleagues), media, rape culture;
- *Responsibility:* if there is explicit attribution of responsibility for the murder in the text, state who is blamed (allowing multiple choices): victim, perpetrator, social network (family, friends, colleagues), media, rape culture;
- *Humor:* specify whether the text conveys humorous content through irony, sarcasm, word games or hyperbole;
- *Macabre:* specify whether there are macabre aspects detailing how the victim was killed;
- *Context:* indicate whether the context was helpful to better understand the meaning of the comments;
- *Notes:* free space for suggestions, observations or doubts.

B. GBV-Cecchettin Statistics for Time-Parts

In Table 5 we report the various statistics about the different time parts of GBV-Cecchettin. In Table 6 we report

Dimension	Part 1	Part 2	Part 3
Number of Examples	189	544	199
Misogyny	0.53%	2.02%	3.02%
Intersectionality	0.00%	0.55%	1.01%
Aggressiveness	11.11%	24.63%	23.12%
Responsibility	9.52%	32.72%	14.57%
Support	34.39%	36.76%	36.18%
Humor	2.65%	0.37%	2.51%
Macabre	0.00%	0.00%	0.00%
Context	2.12%	1.10%	2.51%

Table 5
Statistics about presence of dimensions (positive label only) in the different time-phases

Dim.	Target	Part 1	Part 2	Part 3
Aggressiveness	Perpetrator	61.9%	66.4%	34.8%
	Victim	0.00%	0.00%	0.00%
	PSN	9.5%	5.2%	0.00%
	VSN	14.3%	4.5%	10.9%
	Male Pop.	4.8%	3.0%	2.2%
	Media	14.3%	16.4%	45.7%
	Inst.	0.00%	9.0%	8.7%
	Rape Cult.	0.00%	0.7%	0.00%
Responsibility	Perpetrator	50.0%	33.1%	21.4%
	Victim	16.7%	6.2%	3.6%
	PSN	22.2%	16.9%	7.1%
	VSN	0.00%	7.3%	10.7%
	Male Pop.	5.6%	5.1%	0.00%
	Media	5.6%	3.4%	7.1%
	Institutions	0.00%	33.1%	64.3%
	Rape Culture	0.00%	5.6%	0.00%
	P-F Factor	5.6%	7.3%	0.00%
Support	Perpetrator	10.8%	1.00%	0.00%
	Victim	36.9%	67.0%	75.0%
	PSN	12.3%	5.5%	2.8%
	VSN	67.7%	43.0%	33.3%
	Male Pop.	0.00%	2.5%	2.8%
	Female Pop.	4.6%	10.0%	1.4%

Table 6
Statistics about the target of Aggressiveness, Responsibility, and Support in the different time-phases

the various statistics of the targets relative to Aggressiveness, Responsibility and Support for the different time parts of GBV-Cecchettin. Note that they can sum up to more than 100% as annotators could select more than one.

C. Prompts

In Table 7, we report the prompts for Agg_{binary}. Other tasks present the layout but slightly change in task description and possible values.

Num	Text	English Translation
p1	Dato un commento riguardo un femminicidio, stabilisci se esso contiene aggressività, scegliendo tra: Vero, Falso. [Commento]: [Risposta]:	Given a comment about a femicide, determine whether it contains aggressiveness, choosing between: True, False. [Comment]: [Answer]:
p2	Leggi il seguente testo: {***Descrizione del caso di femminicidio in questione proposta da LaVentisettesimaOra***} {***p1***}	Read the following text: {***LaVentisettesimaOra's description of the femicide case in question***} {***p1***}
p3	Il termine femminicidio viene usato per indicare l'uccisione di una persona di genere femminile nell'ambito di una relazione affettiva o familiare. Il femminicidio costituisce l'atto finale di violenze fisiche e psicologiche ripetute nel tempo. È molto importante essere consapevoli della gravità del fenomeno, soprattutto dal momento che circa ogni due giorni, in Italia, viene uccisa una donna. {***p1***}	The term femicide is used to refer to the killing of a female gender person in an emotional or familiar relationship. Femicide is the final act of physical and psychological violence repeated over time. It is very important to be aware of the seriousness of the phenomenon, especially because a woman is killed approximately every two days in Italy. {***p1***}
p4	{***p3***} Il seguente commento riguarda il caso di femminicidio di **Nome della vittima**. Descrizione del caso: ***Descrizione del caso di femminicidio in questione proposta da LaVentisettesimaOra***. {***p1***}	{***p3***} The following comment refers to the femicide case of **Name of the victim**. Case description: {***LaVentisettesimaOra's description of the femicide case in question***} {***p1***}

Table 7
The four different prompt structures we used with the English translation.

Declaration on Generative AI

During the preparation of this work, the author(s) did not use any generative AI tools or services.