Sympathy as a Lens for Human–Robot Interaction: Analysing YouTube Responses to Robot Abuse

Vlatka Tolj
vlatka.tolj@student.kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Caterina Neef
caterina.neef@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Barbara Bruno
barbara.bruno@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

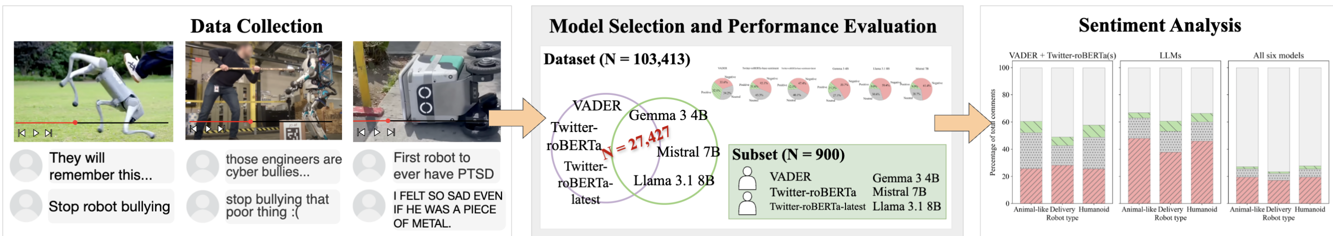

Figure 1: Overview of the sentiment analysis pipeline

Abstract

When witnessing the abuse of others, humans generally exhibit emotional responses. A large number of studies in Human-Robot-Interaction (HRI) have shown that humans also react with sympathy when robots are abused, but most of these insights come from controlled laboratory studies using short videos and student samples. To complement existing research with observations drawn from real-world online discussions, this paper presents a sentiment analysis of 103,413 YouTube comments on videos depicting abuse of animal-like, humanoid, and cart-shaped robots. To validate our sentiment classification, we analysed the comments using a lexicon-based tool, two fine-tuned language models, and three general-purpose state-of-the-art large language models (LLMs). The comparison yielded interesting results: LLMs generally classified science-fiction–related comments, e.g., references to dystopian TV shows, as negative, while lexicon and fine-tuned models mainly labelled them as neutral. The six models agreed on the classification of a total of 27,427 comments, which we used to explore the sentiment expressions occurring across videos featuring robots with different physical forms. Our findings provide large-scale, ecologically valid insights into how emotional responses to robot abuse are expressed and analysed in online video platforms.

CCS Concepts

• **Human-centered computing → Empirical studies in HCI**; **Social media**; • **Information systems → Sentiment analysis**.

Keywords

Affective computing, Online discourse, Robot abuse

HRI '26, Edinburgh, Scotland, UK

ACM ISBN 979-8-4007-2128-1/2026/03
https://doi.org/10.1145/3757279.3785595

ACM Reference Format:
Vlatka Tolj, Caterina Neef, and Barbara Bruno. 2026. Sympathy as a Lens for Human–Robot Interaction: Analysing YouTube Responses to Robot Abuse. In *Proceedings of the 21st ACM/IEEE International Conference on Human-Robot Interaction (HRI '26), March 16–19, 2026, Edinburgh, Scotland, UK*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3757279.3785595

1 Introduction

"No man is devoid of a heart sensitive to the suffering of others..." This quote from Confucian philosopher Mencius poetically summarises how witnessing someone being abused (nearly) always evokes an emotional response in the observer, regardless of their age, the context in which the violence occurs and whether the victim is a human or an animal [20, 32, 44].

A large and growing body of studies in Human-Robot Interaction (HRI) suggests that observing the mistreatment of robots elicits an emotional response in humans [2, 16, 41]. More specifically, research has shown that witnessing robot abuse can trigger increased physiological arousal, negative affect, and expressions of sympathy [41, 42]. Experiments have also shown that although participants will comply with instructions to harm a robot, they will often show visible discomfort while doing so, and this discomfort tends to be stronger when the robot is perceived as more intelligent and lifelike [2]. More recent work indicates that such reactions are additionally shaped by social and demographic factors such as the observer's gender, prior experiences with abuse, and personal biases [16]. Finally, various studies and sources suggest that the physical appearance of a robot and its level of anthropomorphism influence people's emotional responses toward witnessing it being abused [22, 30, 31, 39], with humanoid robots likely to evoke more empathetic responses due to their human-like characteristics [39]. While remarkably insightful, most findings to date stem from controlled laboratory settings with short videos and student samples, leaving the field with a lack of studies on what emotional responses to robot abuse occur outside controlled laboratory environments.

In an effort towards bridging this gap, we present a descriptive and exploratory approach that complements controlled studies for investigating emotional responses to robot abuse, illustrated in Figure 1. Our contributions are two-fold: (1) We conduct a large-scale sentiment analysis of viewers' comments on videos of robot abuse publicly available on a popular video sharing platform. (2) We evaluate three approaches to this sentiment analysis, leveraging a lexicon-based tool, fine-tuned language models and general-purpose state-of-the-art large language models (LLMs), and discuss their effectiveness in capturing the nuances of human emotional expression. Considering the exploratory nature of this work, the analysis presented in this paper addresses the following research question (RQ):

Which emotional responses are expressed in YouTube comments on robot-abuse videos, and how do sentiment-analysis models map these responses across animal-like, humanoid, and cart-shaped robots?

The remainder of this paper is structured as follows: In Section 2, we review HRI research investigating empathetic responses to robot abuse and briefly survey the academic discourse on the concept of empathy and sympathy in the field of psychology, as well as recent HRI work that uses online video platforms as a data source. In Section 3, we outline the data collection and our proposed sentiment analysis approach, which leverages both lexicon-based tools and language models. We then present (Section 4) and discuss (Sections 5, 6) the results of the sentiment analysis in relation to our RQ, along with the limitations of our work. Finally, the Conclusion summarises our findings.

2 Related work

2.1 Emotional responses toward robots

Early research on people's reaction to robot abuse mainly built on results from the field of Human-Computer Interaction, and in particular the *Media Equation theory* [38], which suggests that humans tend to respond to media and technology as if they were real social actors, unconsciously applying social norms and expectations to machines during interactions. Bartneck and Hu [2] tested the Media Equation theory in morally questionable situations inspired by Milgram's obedience experiment, in which participants were asked to administer electric shocks or "kill" a robot. The participants showed less concern for harming the robot than a human, indicating limits to the Media Equation in abuse contexts. Their findings also suggest that the robot's perceived intelligence influenced participants' willingness to harm it.

To explore human emotional responses toward witnessing, rather than directly inflicting, robot abuse, Rosenthal-von der Pütten et al. [41] exposed participants to videos in which the dinosaur-like robot Pleo was either treated gently or abused by a human actor. Abuse elicited higher arousal and more negative affect, indicating an empathetic response toward the robot. A follow-up study [42] found similar responses to affectionate treatment of humans and robots, but significantly stronger neural and emotional responses when witnessing abuse directed at the human. Similarly, Mattiassi et al. [30] investigated how affective empathy varies across a spectrum of social actors (a human, a cat, an anthropomorphic robot, an animal-like robot, a machine-like robot, a smartphone, and a Rubik's cube). Participants reported the highest empathy for a cat, moderate empathy for humanoid and animal-like robots, and little to none for machine-like robots and objects. A subsequent study [31] confirmed stronger negative affect and arousal toward the cat, suggesting that perceived vulnerability drives empathetic responses. Robots evoked moderate emotional responses, with human-like and animal-like robots evoking greater empathetic concern than machine-like robots. An earlier study by Riek et al. [39], which examined how empathy toward robots changes along the anthropomorphic spectrum, further corroborates the hypothesis of a correlation between the human-likeness of a robot and the strength of the empathetic response it evokes. Drawing on Simulation Theory, they found that participants reported more empathetic emotional responses to abuse of human-like robots and less toward mechanical robots.

Expanding the focus of analysis beyond immediate emotional reaction, Küster et al. [22] investigated how YouTube videos of robot abuse influence people's perceptions of robots' minds, morality, and fear of robots. They found that functional abilities may influence mind perception more strongly than appearance: the animal-like robot opening a door in the video was rated higher in experience, agency, and consciousness than a humanoid robot. Appearance, however, evoked greater emotional concern, with the humanoid robot seen as more deserving of compassion despite being perceived as less conscious. The study also highlights the power of online media in shaping public opinions toward robots and provides a basis for our work and methodology.

Except for Bartneck and Hu's study [2], the previously mentioned studies relied on videos, often from YouTube, to examine emotional responses to robot abuse. These studies typically compare reactions to different agents or robot types and consistently show that humanoid robots elicit stronger emotional responses than less anthropomorphic ones. To complement these findings in an uncontrolled, real-world setting, we analyse online viewers' reactions to publicly available YouTube videos. Sample sizes in the works outlined above vary widely (from 14 [42] to 163 participants [30, 31]), reflecting the constraints of controlled laboratory studies. By contrast, analysing YouTube comment sections allows us to observe emotional reactions to robot abuse on a much larger scale, albeit at the expense of participants' information. Nevertheless, such naturally occurring responses can offer valuable insights into how people perceive robot abuse in everyday contexts and contribute to understanding broader public attitudes toward robots [22].

2.2 Robots, empathy and sympathy

As the previous section shows, most HRI research on emotional responses to robot abuse is framed around the concept of *empathy*. However, Quick [37] recently raised concerns about the conceptual clarity of empathy in HRI research, arguing that many HRI studies rather investigate *sympathy*, thereby introducing ambiguity. As an example, the author argues that the question "How sorry do you feel for the protagonist?" used by Riek et al. [39] to measure people's empathy for the robot, actually captures a feeling of sympathy, rather than empathy. To clarify the scope and foundations of this work and support interdisciplinary research on the topic, we briefly report the definitions of key concepts that we follow.

As pointed out by Darwall [8], unlike words like *compassion* and *sympathy*, *empathy* only entered the English language in 1909 when Titchener coined it to translate the German word *Einfühlung* ("feeling-into"), in an attempt to distinguish it from the already existing term *Mitgefühl* ("feeling-with"), translated in English as sympathy. Since then, numerous psychologists have investigated the various words and underlying constructs, aiming at providing clear and valid definitions and frameworks. Two schools of thought can be found: one which identifies sympathy as part of empathy [3, 9, 10] and another, more recent, which clearly distinguishes empathy and sympathy as separate emotional processes [7, 14].

In a review on empathy research, Cuff et al. [7] explain that empathy involves "feeling as" another person, while sympathy refers to a more detached response, such as "feeling for" another person, or another entity. They summarise empathy as an emotional response "similar to one's perception...and understanding...of the stimulus emotion, with recognition that the source of the emotion is not one's own." For sympathy, they draw on Eisenberg et al. [13], who define it as "a vicarious emotional reaction based on the apprehension of another's emotional state or situation, which involves feelings of sorrow or concern for the other." In light of these distinctions, Quick [37] questions the appropriateness of the concept of *empathy* in HRI and rather calls for the use of *sympathy*. Accordingly, this work focuses on and discusses the *sympathetic* responses expressed by people commenting on YouTube videos displaying robot abuse, since these reactions are less about sharing the robots' presumed emotions and more about expressing concern, pity, or moral condemnation for their mistreatment.

2.3 Online video platforms as a data source

Recent HRI research has often turned to online video platforms as data sources to investigate human-robot interactions. Since our work similarly draws on YouTube data, we consider it important to situate our approach within this growing body of HRI research.

Yu et al. [48] used 117 TikTok videos to examine real-world encounters with delivery robots, identifying public perceptions that would be difficult to observe in controlled studies. Their analysis of 2,067 comments revealed that viewers frequently anthropomorphised delivery robots, attributing thoughts, feelings, and social roles to them, often describing them as cute or novel. Comments were largely positive, with many expressing affection towards the robots, interest in the technology, or even sympathy for robots that appeared stuck or struggling. Strait et al. [45] analysed 1,200 YouTube comments across 24 videos of machine-like and highly human-like robots, showing that online reactions closely mirror the uncanny valley effect, as highly anthropomorphic robots elicited significantly more discomfort for the viewers. Their analysis further revealed that fears of replacement or a robot takeover were not shaped by the robots' appearance, and that female-gendered robots were disproportionately sexualised in the comments. Hover et al. [18] analysed 1,788 YouTube comments on videos of humanoid and android robots to investigate public attitudes towards different levels of human-likeness and gender. They found that highly human-like robots evoked more negative reactions, were more frequently described as uncanny, and that female-gendered robots were sexualised significantly more often than male robots. These too are reactions that are difficult to capture in controlled studies.

Beyond analysing viewer responses, prior work has also used user-generated videos as a data source. Nielsen et al. [35] conducted a digital ethnographic study of 104 YouTube videos showing unguided, in-the-wild interactions with service robots in public spaces. They found that users frequently struggled due to touchscreen or voice input issues, navigation errors or unexpected stops, and interruptions from bystanders. Their findings illustrate the inherently messy and dynamic nature of real-world HRI and highlight the value of online videos for capturing naturally occurring behaviours that are difficult to reproduce in controlled laboratory settings.

Together, these studies show that YouTube and similar online video platforms offer complementary insights to controlled studies, capturing public perceptions of robots as well as spontaneous reactions and interaction patterns. However, using online video data also comes with important limitations, including the skewed demographics of platform users, the performative nature of recorded content and commentary, and the biases involved in selecting videos for analysis. These and further limitations are discussed in Section 6.

3 Methodology

3.1 Data collection

YouTube, the world's largest video-sharing platform and second-largest social media platform[1], was used as the primary platform for the collection of robot abuse videos for our analysis. The following inclusion/exclusion criteria were considered:

- Included videos should attract a large audience engagement (e.g., typically exceeding one million views) to ensure a sufficient number of comments for the analysis.
- The physical, verbal, or sexual abuse of real robots by real humans should be clearly visible in each included video.
- Videos showing scenes taken from other (included) videos and videos which contain AI-generated content instead of real robots and humans are excluded.
- Videos featuring clips from other media, e.g. scenes from the movie *Terminator*, are excluded.
- Videos that include background music which can potentially manipulate the viewers' emotional state, e.g., music with suggestive lyrics such as "Tubthumping (I get knocked down)" by Chumbawamba, are excluded.
- In contrast, original audio elements such as humans verbally abusing the robot were retained, since they are part of the actual abusive context. Other incidental background sounds (e.g., news reports commentary) were not regarded as a possible exclusion criterium.

A total of 13 videos, collected on three separate occasions in May 2025, were selected for the sentiment analysis based on their relevance to robot abuse scenarios and the above criteria:

- The queries *"robot abuse"* and *"robot mistreated"*, along with the YouTube recommended videos section were first used to gather data. This resulted in the identification of two relevant videos: "Testing Robustness" [12] published by Boston

[1] https://thedailyjuice.net/the-largest-video-sharing-platforms-in-2025-ranked/

Dynamics and "Robot Tries to Escape from Children's Attack" [43] published by IEEE Spectrum.
- Additional searches for *"robot harassed"* and *"robot bullied"* yielded four more relevant videos: "PUNKING MARCUS: Bullying Food Delivery Robots is Actually Good" [26] by Film The Robots LA, "This robot getting bullied by stranger" [40] by ripL, "Kai Cenat & Fanum Bully $70,000 Robot!" [5] by Successful Celebrity and "Atlas, The Next Generation" [11] again by Boston Dynamics. A further search with the term *"robot kicked"* yielded two more relevant videos: "Robot dog keeps its balance in extreme training" [36] by Tech Pandaren and "Watch robot dog 'Spot' run, walk...and get kicked" [34] by On Demand News.
- Five additional videos were identified by specifically browsing the YouTube channel "Film The Robots LA", which exclusively features videos of robots being abused: "ROBOT VIOLATED: Austin the Food Delivery Robot Treated Poorly" [27], "DELIVERY ROBOT LOOTED: Austin Got Robbed" [23], "MESSY ROBOT: Haiden the Food Delivery Robot Has All Types of Things Going On" [24], "OTTO's ESCAPE: Food Delivery Robot Flees From Kids" [25] and "Saving Austin the Food Delivery Robot From Himself" [28].

Table 1 summarises the characteristics of the videos included in our analysis. The videos cover a range of robot types: animal-like [12, 34, 36], humanoid [5, 11, 43], delivery [23–28], and non-humanoid bipedal [40] robots, as shown in Figure 2. The videos were uploaded by different publishers, including a robotics manufacturing company (Boston Dynamics), a tech organisation (IEEE Spectrum), news and tech content channels, a fan channel featuring various internet personalities, and private individuals. Abuse was mainly either physical (e.g., kicking, pushing, hindering the robot in task execution), or verbal (mocking, insulting), and in some cases a combination of both. One video [27] depicted a partially clothed individual straddling the robot, which we classified as sexual abuse. The context of the videos varied considerably. Some were created explicitly for the purpose of recording an episode of robot bullying, e.g., videos made by an individual following and filming delivery robots while verbally abusing them [23–26]. One video [5] featured an internet personality and other individuals deliberately kicking and mocking a humanoid robot. Other videos [11, 12, 34, 36], depicted "abuse" in the context of demonstrations intended to showcase the robot's capabilities. For example, the Atlas video [11] did not only show the robot being pushed with a hockey stick, but also included scenes of non-abusive tasks.

To extract the comments from the videos we used the youtube-comment-downloader[2]. Grouping videos according to the type of robot they feature, we obtained 17,394 comments for videos showing animal-like robots, 27,193 for delivery robots, 58,826 for humanoid robots, and 169 for non-humanoid bipedal robots, resulting in a total of 103,582 comments. Due to the disproportionately small number of comments (n = 169), the non-humanoid bipedal robot group was excluded from the sentiment analysis. As a result, only the animal-like, delivery, and humanoid robot categories were included in the sentiment analysis, amounting to 103,413 comments.

[2]https://github.com/egbertbouman/youtube-comment-downloader

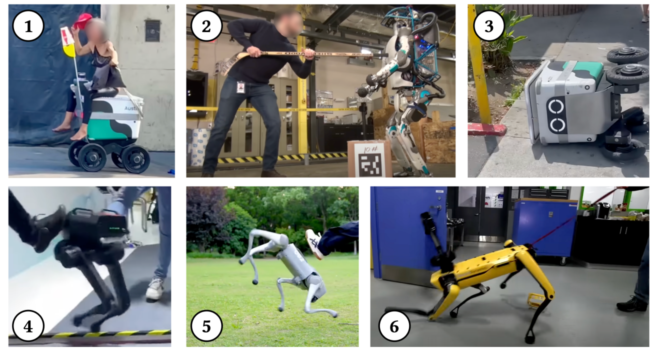

Figure 2: Images of the different robot types in the collected videos: (1, 3) delivery [23, 27], (2) humanoid [11], (4) non-humanoid bipedal [40], (5, 6) animal-like [12, 36].

3.2 Sentiment analysis

3.2.1 Model selection. To verify inter-model agreement and ensure a more robust analysis, we rely on 6 models for sentiment analysis. The first selected model is VADER, a lexicon-based sentiment analysis tool specifically optimised for short, informal social media texts that can handle emojis and slang in online discourse [19]. The second and third model are Twitter-roBERTa-base-sentiment and Twitter-roBERTa-base-sentiment-latest, which are roBERTa-base transformer models for sentiment analysis specifically fine-tuned on Twitter content, trained on 58 million and 124 million tweets, respectively [1]. We argue that the short-text style of YouTube comments likely shares similar characteristics to tweets (i.e., posts on X, formerly Twitter).

To complement the above solutions, which were specifically designed or fine-tuned for sentiment analysis, we also considered three LLMs, namely Gemma 3 4B, Mistral 7B, and Llama 3.1 8B. Gemma 3 4B, developed by Google DeepMind (England), supports multilingual understanding and long-context reasoning [47], and can be run efficiently on local hardware thanks to its limited size. Mistral 7B [21], developed by Mistral AI (France), is a mid-sized, open-source solution that surpasses larger models such as Llama 2 13B on a wide range of benchmarks, including reasoning, while remaining computationally affordable [21]. Finally, Llama 3.1 8B, developed by Meta AI (USA), is a larger open-source model with multilingual understanding [17]. All three LLMs include sentiment analysis among their possible use cases and support multilingual understanding. We argue that the frequent presence of sci-fi and other cultural references in the comments on the selected videos calls for sentiment analysis tools that can leverage significant general-purpose knowledge. The chosen LLMs might thus enable a more accurate assessment of the comments and their emotional content.

3.2.2 Human annotators baseline. To preliminarily validate the classification performance of the 6 considered models, establish a reference for inter-model agreement and collect examples for the few-shot prompting of the LLMs, two human annotators independently and intuitively labelled a subset of 900 comments drawn from the most-liked comments of the collected videos: three humanoid robot videos [5, 11, 43] (100 comments each), three animal-like

Table 1: Metadata and descriptive statistics of videos

Video	Publisher	Robot Type	Setting	Abuse Type	Abuser	Length	Views	Comments	Upload Date
[12]	Robotics company	Animal-like	Lab	Physical	Adult (individual)	01:03	34,215,515	16,018	20.02.2018
[43]	Tech organisation	Humanoid	Urban	Physical	Children (group)	02:04	608,736	703	06.06.2017
[26]	Private person	Delivery	Street, sidewalk	Verbal	Adult (individual)	00:10	60,262	102	25.08.2023
[40]	Private person	Bipedal	Exhibition hall	Physical	Adults (individual turns)	00:40	155,326	169	07.07.2024
[5]	Fan/clip channel	Humanoid	Private home	Physical, verbal	Adults (group)	00:15	18,485,213	11,029	29.01.2025
[11]	Robotics company	Humanoid	Lab	Physical	Adult (individual)	02:41	40,738,208	47,104	24.02.2016
[36]	Tech content channel	Animal-like	Outdoor	Physical	Adult (individual)	00:19	643,759	1,071	20.06.2024
[34]	News channel	Animal-like	Outdoor, indoor	Physical	Adult (individual)	01:11	485,449	305	11.02.2015
[27]	Private person	Delivery	Street, sidewalk	Sexual (molesting)	Adult (individual)	00:13	1,637,330	6,331	07.07.2023
[23]	Private person	Delivery	Street, sidewalk	Verbal, physical	Adult (individual)	00:59	10,503,655	16,112	03.07.2023
[24]	Private person	Delivery	Street, sidewalk	Verbal	Adult (individual)	00:36	1,021,208	699	04.07.2023
[25]	Private person	Delivery	Street, sidewalk	Verbal	Adult (individual)	00:29	1,560,870	1,204	28.06.2023
[28]	Private person	Delivery	Street, sidewalk	Verbal	Adult (individual)	00:59	3,213,217	2,745	25.06.2023

robot videos [12, 34, 36] (100 comments each), and six delivery robot videos [23–28] (50 comments each). This produced a balanced dataset including 300 comments per robot type category. The annotators achieved a global inter-rater agreement of 75.78%. Considering the robot type categories, agreement was highest for humanoid robots (81.67%), followed by animal-like robots (75.33%) and delivery robots (68.00%). Details on the annotators' labels can be found in the Supplementary Material, Figure 5. The two annotators were both female, of similar age (32–33 years), and had similar educational backgrounds in higher education, with one at the Master's level and the other at the Postdoctoral level. Discrepancies in their labelling highlight the subtle and nuanced nature of sentiment analysis and further support our decision to use only data where multiple models provided the same label.

3.2.3 LLM prompting. Following manual labelling, we defined the sentiment annotation guidelines, which were integrated into the LLM prompt, shown in Appendix B in the Supplementary Material. Recurring phrases such as robots "remembering the abuse" and "getting their revenge" reflect common dystopian sci-fi tropes, which shape public fear of robots [4, 29, 33]. Given their context, such responses were therefore labelled as negative sentiment. Negative sentiment also included expressions of sympathy, pity, sadness, discomfort, moral condemnation, or anger at the mistreatment. Positive labels were assigned to comments expressing enjoyment, praise, amusement, or humour directed at the ab use. Neutral sentiment included comments with descriptive observations, remarks without emotional tone, or discussions unrelated to the content of the video.

3.3 N-gram analysis of sentiment

To better understand what constitutes positive, neutral, and negative sentiment in this context, we conducted an n-gram analysis (considering 1-word, 2-words and 3-words) for each robot type, using the natural language processing toolkit NLTK[3]. The n-gram analysis identifies commonly used phrases and counts their occurrence across sentiment labels. Its results are further discussed in Section 5. Detailed results are provided in Tables 5, 6, and 7 in the Supplementary Material. As viewers often express their opinions by liking existing comments rather than posting their own, we also provide an overview with the top five most-liked comments grouped by robot type in Table 8 in the Supplementary Material.

[3] https://www.nltk.org/

4 Results

4.1 Classification performance on annotated dataset

When evaluated against the subset of manual annotations where both annotators agreed on the same sentiment label (N = 682), LLMs outperformed VADER and Twitter-roBERTa models by a large margin, as shown in Table 2. Gemma 3 4B achieved the highest F1-score (0.80), followed by Llama 3.1 8B (0.77) and Mistral 7B (0.75). In comparison, VADER and Twitter-roBERTa were below 0.62 in F1. These findings indicate that the chosen LLMs enable a more accurate assessment of the comments and their emotional content compared to VADER and Twitter-roBERTa models.

Table 2: Performance of sentiment analysis models on the subset of manually labelled aligned annotations (N = 682)

Model	Accuracy	Precision	Recall	F1
VADER	0.625	0.588	0.682	0.589
roBERTa	0.696	0.661	0.633	0.618
roBERTa-latest	0.716	0.651	0.629	0.620
Gemma 3 4B	**0.846**	0.786	**0.815**	**0.797**
Llama 3.1 8B	0.839	**0.802**	0.744	0.767
Mistral 7B	0.821	0.759	0.754	0.754

4.2 Agreement between models

Figure 3 reports the sentiment label distribution of the six considered models over the 103,413 comments considered in our analysis.

4.2.1 Agreement between VADER and Twitter-roBERTa models. VADER and Twitter-roBERTa-base-sentiment agree on 63,531 comments (61.45%), while agreement between VADER and Twitter-roBERTa-base-sentiment-latest is slightly lower at 62,193 comments (60.15%). As expected, the two Twitter-roBERTa models show high consistency, assigning identical labels to 89,523 comments (86.59%). All three models agree on 57,766 comments (55.86%), including 10,523 related to animal-like robots, 13,305 to delivery robots, and 33,938 to humanoid robots. As illustrated in Figure 3, VADER assigns positive, neutral, and negative labels in roughly equal proportions, whereas the Twitter-roBERTa models predominantly classify comments as neutral or negative.

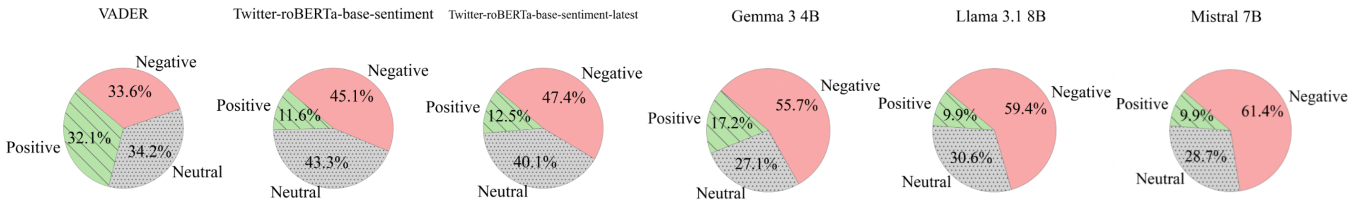

Figure 3: Sentiment label distribution across models for all comments (N = 103,413)

4.2.2 Agreement within large language models (LLMs). Gemma 3 4B and Llama 3.1 8B show the highest agreement (78.59%, 81,263 comments), followed by Llama 3.1 8B and Mistral 7B (76.04%, 78,632 comments) and Gemma 3 4B and Mistral 7B (73.07%, 75,584 comments). All three models agree on 67,164 comments (64.93%), including 11,634 animal-like robot, 16,499 delivery robot, and 39,031 humanoid robot comments, reflecting a distribution comparable to that of the other three models. As shown in Figure 3, all three LLMs predominantly assign negative sentiment. Notably, Gemma 3 4B classifies a higher proportion of comments as positive (17.2%) than Llama 3.1 8B and Mistral 7B (9.9% each).

4.2.3 Agreement across all six models. All six models agree on a total of 27,427 comments (26.52%), including 4,725 comments for animal-like, 6,389 for delivery, and 16,313 for humanoid robots.

4.3 Sentiment and robot type

To determine whether the distribution of sentiment differs across the three different robot types, we conducted a Chi-Square χ^2 test of independence on contingency tables reporting the counts of negative, positive, and neutral comments for each of the three robot categories (delivery, animal-like, humanoid). Three contingency tables were designed and tested:

- (1) considering the 57,766 labels on which VADER and the two Twitter-roBERTa models agree (see Section 4.2.1);
- (2) considering the 67,164 comments and labels on which the three LLMs agree (see Section 4.2.2);
- (3) considering the 27,427 comments and labels on which all six models agree (see Section 4.2.3).

As discussed above, the three subsets are similarly spread over the robot categories, with an equal number of comments for delivery and animal-like robots, and a number of comments for humanoid robots which is approx. 2.5 times larger than the other categories.

Table 3 presents the results of the three Chi-square χ^2 tests: All tests indicate a statistically significant association (p-value < 0.05) between robot type and sentiment distribution. However, effect size estimates suggest that the strength of these associations is negligible to small in all cases (Cramér's V < 0.1). The six-models set yielded the weakest association (V = 0.032), the LLM-only analysis produced a slightly stronger association (V = 0.061), and the VADER + Twitter-roBERTa analysis showed the largest, though still small, effect (V = 0.085). Taken together, these results suggest that although minor distributional differences between robot types were detectable thanks to the large sample sizes, the practical impact of robot type on emotional responses was minimal. The corresponding contingency tables can be found in Table 4 in the Supplementary Materials.

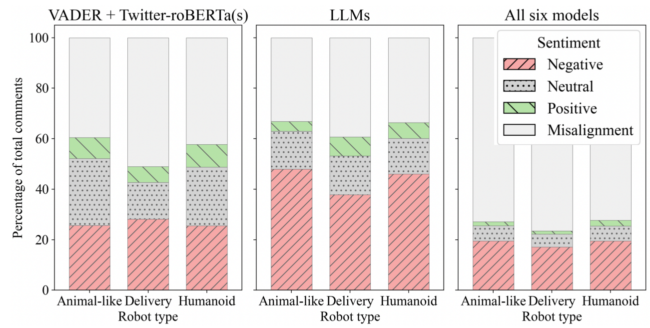

Figure 4: Sentiment distribution across robot type categories

Table 3: The Chi-square test results for all three approaches

Analysis method	**χ^2 (df=4)**	**p-value**	**Cramér's V**
(1) VADER + roBERTa	842.49	< .001	0.085
(2) LLM-only	507.26	< .001	0.061
(3) All 6 models	57.46	< .001	0.032

5 Sentiment analysis of comments on robot abuse

As Figure 4 shows, sentiment towards robot abuse videos is predominantly negative regardless of robot type, and the Chi-square test (see Section 4.3) likewise shows no evidence that the robot type influences the sentiment expressed in the comments. These findings suggest that the robot's physical form does not play a strong role in influencing sympathetic responses to YouTube videos.

Considering the subset of 57,766 aligned labels between VADER and the two Twitter-roBERTa models (see Figure 4 – left), humanoid robot abuse videos elicited a predominantly negative emotional response, although neutral sentiment was nearly as frequent. For animal-like robots, neutral sentiment even slightly exceeded negative sentiment. Notably, only delivery robots showed a clearly higher proportion of negative sentiment compared to neutral or positive sentiment. The n-gram analysis reveals several recurring phrases across robot categories. Explicit expressions of sympathy are common in negative comments, including phrases such as "stop robot abuse" (30 instances for animal-like robots, 52 for humanoids) and "poor robot" (93 instances for animal-like, 379 for humanoid, and 88 for delivery robots). Beyond sympathetic responses, many comments reference pop culture and media. The most frequent phrase within positive sentiment was the *Simpsons* quote "I, for one, welcome our new overlords", appearing 12 times for animal-like robots and 68 times for humanoids. Neutral comments often referenced dystopian sci-fi media, such as *Skynet* from *Terminator* (365 instances in the humanoid group) and the show *Black Mirror* (138 instances for animal-like robots). The phrase "robot lives matter" appeared 15 times for delivery robots, 35 for humanoids, and 15 for animal-like robots within neutral sentiment.

In the sentiment analysis of the 67,164 comments on which all three LLMs agree, strong negative sentiment was observed across all robot categories (see Figure 4 – centre). This is likely a direct result of the prompt engineering strategy (see Section 3.2.3), which defines expressions of sympathy and sci-fi references as negative, as also reflected in the n-gram analysis. For example, "stop robot abuse" occurred 38 times for animal-like robots and 56 times for humanoids. Other recurring terms within negative sentiment included *Black Mirror* (218 instances for animal-like robots) and "robot lives matter" (26 instances for animal-like, 63 for humanoid, and 35 for delivery robots). By contrast, VADER and the Twitter-roBERTa models classified these references as neutral, highlighting a key difference in labeling strategies and the importance of general knowledge for sentiment analysis of context-dependent text such as YouTube comments.

Finally, the analysis of the 27,427 comments on which all models agree (see Figure 4 – right) shows a clear predominance of negative sentiment across all robot types, with neutral and positive labels occurring far less frequently. The n-gram analysis indicates that negative sentiment is dominated by sympathetic expressions such as "feel bad robot" (59 instances for animal-like robots, 289 for humanoids, 21 for delivery robots), "stop bullying" (191 for humanoids, 50 for animal-like robots), and "poor Austin" (224 instances for delivery robots, referring to the robot named Austin in videos [23], [27], [28]). As this subset includes only comments with full model agreement, no science-fiction references appeared within negative sentiment. Interestingly, the term "metalhead" occurred three times within neutral sentiment for animal-like robots, referring to the *Black Mirror* episode *Metalhead*, which features an animal-like robot in a dystopian setting [33]. This was likely labelled as neutral due to limited contextual understanding, again underscoring the importance of general knowledge in sentiment interpretation. For full n-gram details, see Tables 5, 6, and 7 in the Supplementary Materials.

Given social media platforms' predominance as a medium for individuals to express their views and engage with others, analysing and understanding the discourse occurring around robot abuse therein provides an important perspective on how people perceive robots, that naturally complements, and could even inform, lab studies. It is noteworthy that the majority of the analysed comments contain expressions of concern for the robots' well-being. In a time characterised by widespread social fears and negative opinions about robots and the influence of AI [29], our findings suggest that public attitudes toward robots are remarkably nuanced.

6 Discussion and limitations

6.1 Data collection

6.1.1 Videos and robot visibility. Our data collection focused on YouTube videos depicting robot abuse and included robots with different physical forms, namely delivery, humanoid, and animal-like robots. Notably, the dataset did not contain highly anthropomorphic social robots, despite our search strategy (see Section 3) not explicitly excluding them. Instead, we relied on general search terms related to robot abuse, allowing robot types to emerge organically from the platform. The absence of videos showing abuse of highly anthropomorphic social robots (e.g., Ameca or Sophia) suggests that such robots may be less frequently abused in publicly shared online videos, an observation we consider noteworthy. One possible explanation is that these robots are expensive and typically deployed in controlled or institutional settings, reducing their exposure in everyday public or private environments. Similarly, we did not encounter videos depicting abuse of commonly used social robots such as NAO or Pepper. While this does not imply that these robots are never abused, it reflects the lower likelihood of surfacing videos of social robot abuse through general robot abuse-related queries and calls for dedicated studies. Consequently, online video datasets may over-represent abuse of publicly accessible, non-social robots while under-representing abuse of highly anthropomorphic social robots. The robot types included in our dataset are therefore influenced by YouTube's search and recommendation mechanisms.

Finally, despite a systematic search strategy, our dataset remains limited to 13 videos, which represents a relatively small sample.

6.1.2 Video content. Understanding emotional responses to real-world robot abuse in online videos offers insight into societal acceptance of robots. However, the content and framing of these videos may influence the results, as we could not control for their purpose or context. The analysed videos varied widely in intent and setting, ranging from public encounters in which individuals deliberately followed and verbally harassed delivery robots, to private-home videos of internet personalities kicking or mocking humanoid robots. Other videos showed abusive actions within demonstrations intended to showcase robot capabilities, such as the Atlas video [11], which combined scenes of physical "abuse" with non-abusive task execution. These contextual differences likely shaped viewers' interpretations and emotional responses. Abuse framed as public bullying may be perceived as morally troubling, whereas similar actions in demonstration settings may be seen as necessary testing. Thus, even when comment patterns appear similar, the underlying context and intent of the behaviour may elicit different emotional reactions.

6.1.3 Viewer comments. Beyond the video content itself, our analysis relies on viewer comments as proxies for emotional responses, raising ethical and methodological concerns common to YouTube-based research. YouTube users are not demographically representative of the general population, with a stronger presence of younger users and engagement shaped by cultural, geographic, and socio-economic factors [15, 46]. The lack of socio-demographic metadata further limits our ability to assess how such factors influence expressed sentiment. Commenting behaviour on YouTube is additionally shaped by platform-specific social norms. Early dominant tones can influence later comments through social conformity [6], and sarcasm, irony, or performative expression may amplify or distort sentiment. As a result, the sentiment observed may partly reflect surrounding discourse rather than the video content alone.

Our analysis also captures reactions from viewers who are detached from the in-situ interaction with the robot. However, this limitation is shared with prior video-based studies discussed in Section 2, where participants responded to mediated rather than first-hand experiences. We therefore do not consider this detachment a limitation unique to our study design.

Finally, our dataset is imbalanced toward humanoid robots, which received approximately 2.5 times more comments than animal-like or delivery robots. This disparity in the volume of comments, while deserving of further analyses, may have introduced additional bias into the sentiment analysis.

6.2 Sentiment analysis

Our sentiment analysis approach involved a deliberate trade-off between robustness and representativeness. To make the sentiment labels as reliable as possible, we prioritised comments for which multiple models agreed on the same label, resulting in a smaller but more consistent dataset. This decision significantly reduced the size of the analysed corpus and likely led to the exclusion of comments expressing more nuanced and context-dependent sentiment. However, we favoured higher confidence in sentiment labels over having sentiment labels with uncertain reliability.

Human annotators were introduced as a reference point to calibrate model prompting and to assess inter-model consistency on a subset (N = 900) of the data. Inconsistently labelled comments in this subset frequently contained sarcasm, pop-culture references, emojis, or complex affective expressions (e.g., "I'm dying"), which proved challenging for both the models and human annotators to classify consistently. As a result, our data-cleaning procedure likely resulted in an under-representation of such nuanced emotional responses, which we explicitly acknowledge as a major limitation of our approach.

We adopted a three-category sentiment labelling approach (positive, neutral, negative), a common practice in sentiment analysis, which allowed us to leverage established models designed specifically for this purpose (e.g., VADER). However, this limited our ability to capture more context-dependent pop-culture-related emotional responses. To address this limitation, we compared traditional sentiment models with state-of-the-art LLMs that we explicitly prompted to account for contextual nuance while still producing three-category sentiment labels. This comparison reduced certain misclassifications, but the fundamental constraints of our approach still remain.

These limitations highlight the inherent challenges of applying sentiment analysis to complex emotional responses. Nevertheless, we argue that the current study retains strong intrinsic value, as it lays the groundwork for future research. Building on the study presented in this paper, in our future work, we intend to move beyond three-class sentiment labels in favour of more nuanced representations of emotional responses.

6.3 Discussion of results

In relation to prior work, our findings highlight the importance of considering delivery robots as enablers of a form of HRI that already unfolds in real-world public spaces. As such, emotional responses to their abuse offer valuable insights into societal acceptance of robots. Notably, even minimally anthropomorphic, cart-shaped delivery robots – featuring only subtle cues such as LED eyes or human names (e.g., Austin, Marcus, Haiden) – elicited sympathetic responses, suggesting that even limited anthropomorphic features can evoke strong emotional reactions.

Previous studies [30, 31] also relied on YouTube videos – specifically videos such as those featuring Atlas and Spot [11, 12] – but analysed only short excerpts showing robot abuse within broader demonstrations of robot capabilities. In contrast, our analysis includes the same videos in their original length and context. Despite the abuse being performed by the robot creators and framed as part of technical demonstrations, viewer comments (see Table 8 in the Supplementary Material) overwhelmingly focused on the brief abusive moments. This suggests that abusive scenes leave a stronger impression on viewers than surrounding non-abusive content, underscoring the need for further research on how contextual framing shapes perceptions of robot abuse.

7 Conclusion

In this paper, we analysed the sentiment of 103,413 YouTube comments on 13 robot abuse videos using six models: VADER, Twitter-roBERTa-base-sentiment, Twitter-roBERTa-base-sentiment-latest, and the LLMs Gemma 3 4B, Mistral 7B, and Llama 3.1 8B. The models showed notable discrepancies in sentiment annotation, underscoring the importance of general knowledge (e.g., references to sci-fi films or popular TV shows) in sentiment analysis of social media content while also highlighting the promise of general-purpose LLMs with few-shot prompting for this task.

The goal of our analysis was to capture emotional responses expressed in YouTube comments on robot abuse videos and to examine how these responses are distributed across sentiment categories and robot physical forms. The results show that viewers predominantly express negative emotional responses across all robot types, largely independent of physical appearance, and point to new opportunities for studying robot abuse in real-world contexts.

Acknowledgments

This work was partially supported by the Baden-Württemberg Ministry of Science, Research and Art (MWK) through the Margarete von Wrangell programme.

References

[1] Francesco Barbieri, Jose Camacho-Collados, Leonardo Neves, and Luis Espinosa-Anke. 2020. TweetEval: Unified benchmark and comparative evaluation for tweet classification. *arXiv preprint arXiv:2010.12421* (2020).

[2] Christoph Bartneck and Jun Hu. 2008. Exploring the abuse of robots. *Interaction Studies* 9, 3 (2008), 415–433.

[3] C. Batson, Jim Fultz, and Patricia Schoenrade. 1987. Distress and Empathy: Two Qualitatively Distinct Vicarious Emotions with Different Motivational Consequences. *Journal of personality* 55 (1987), 19–39. doi:10.1111/j.1467-6494.1987.tb00426.x

[4] Duan Bo, Aini Azeqa Ma'rof, and Zeinab Zaremohzzabieh. 2024. The Influence of Negative Stereotypes in Science Fiction and Fantasy on Public Perceptions of

Artificial Intelligence: A Systematic Review. *Studies in Media and Communication* 13 (12 2024), 180. doi:10.11114/smc.v13i1.7212

[5] Succesful Celebrity. 2025. *Kai Cenat & Fanum Bully $70,000 Robot!* Retrieved May 6, 2025 from https://www.youtube.com/shorts/UgXsvCwgfz8 YouTube, Jan. 29, 2025.

[6] Junhan Chen and Shilin Xia. 2024. Are online users influenced by what other users say? Meta-analyzing the cognitive, emotional, and behavioral impact of online comment valence. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 18 (11 2024). doi:10.5817/CP2024-5-8

[7] Benjamin M.P. Cuff, Sarah J. Brown, Laura Taylor, and Douglas J. Howat. 2016. Empathy: A Review of the Concept. *Emotion Review* 8, 2 (2016), 144–153. doi:10.1177/1754073914558466

[8] Stephen Darwall. 1998. Empathy, sympathy, care. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 89, 2/3 (1998), 261–282.

[9] Cynthia M. Davis. 1990. What Is Empathy, and Can Empathy Be Taught? *Physical Therapy* 70, 11 (1990), 707–711. doi:10.1093/ptj/70.11.707

[10] Mark H. Davis. 1996. *Empathy: A Social Psychological Approach* (1 ed.). Routledge. doi:10.4324/9780429493898

[11] Boston Dynamics. 2016. *Atlas, The Next Generation.* Retrieved May 6, 2025 from https://www.youtube.com/watch?v=rVlhMGQgDkY YouTube, Feb. 24, 2016.

[12] Boston Dynamics. 2018. *Testing Robustness.* Retrieved May 5, 2025 from https://www.youtube.com/watch?v=aFuA50H9uek YouTube, Feb. 20, 2018.

[13] Nancy Eisenberg, Cindy Shea, Gustavo Carlo, and George Knight. 1991. Empathy-related responding and cognition: A "chicken and the egg" dilemma. *Handbook of Moral Behavior and Development* 2 (1991).

[14] Jakob Eklund and Martina Meranius. 2021. Toward a consensus on the nature of empathy: A review of reviews. *Patient Education and Counseling* 104, 2 (2021), 300–307. doi:10.1016/j.pec.2020.08.022

[15] GMI Research Team. 2025. *YouTube Statistics 2025 (Demographics, users by country & more).* Global Media Insight. https://www.globalmediainsight.com/blog/youtube-users-statistics/ Accessed: December 15, 2025.

[16] Hideki Garcia Goo, Katie Winkle, Tom Williams, and Megan K Strait. 2023. Victims and observers: How gender, victimization experience, and biases shape perceptions of robot abuse. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2439–2446.

[17] Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, and et al. 2024. The Llama 3 Herd of Models. arXiv:2407.21783 [cs.AI] https://arxiv.org/abs/2407.21783

[18] Quirien R. M. Hover, Ella Velner, Thomas Beelen, Mieke Boon, and Khiet P. Truong. 2021. Uncanny, Sexy, and Threatening Robots: The Online Community's Attitude to and Perceptions of Robots Varying in Humanlikeness and Gender. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction* (Boulder, CO, USA) *(HRI '21)*. Association for Computing Machinery, New York, NY, USA, 119–128. doi:10.1145/3434073.3444661

[19] Clayton Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, Vol. 8. 216–225.

[20] Gregory R Janson, JoLynn V Carney, Richard J Hazler, and Insoo Oh. 2009. Bystanders' reactions to witnessing repetitive abuse experiences. *Journal of Counseling & Development* 87, 3 (2009), 319–326.

[21] Albert Q. Jiang and et al. 2023. Mistral 7B. arXiv:2310.06825 [cs.CL] https://arxiv.org/abs/2310.06825

[22] Dennis Küster, Aleksandra Swiderska, and David Gunkel. 2021. I saw it on YouTube! How online videos shape perceptions of mind, morality, and fears about robots. *New Media & Society* 23, 11 (2021), 3312–3331. doi:10.1177/1461444820954199

[23] Film The Robots LA. 2023. *DELIVERY ROBOT LOOTED: Austin Got Robbed*. Retrieved May 15, 2025 from https://www.youtube.com/shorts/qDWmArnak18 YouTube, July 3, 2023.

[24] Film The Robots LA. 2023. *MESSY ROBOT: Haiden the Food Delivery Robot Has All Types of Things Going On.* Retrieved May 15, 2025 from https://www.youtube.com/shorts/EwJAW5GWw-8 YouTube, July 4, 2023.

[25] Film The Robots LA. 2023. *OTTO's ESCAPE: Food Delivery Robot Flees From Kids.* Retrieved May 15, 2025 from https://www.youtube.com/shorts/ZRiE46Wj6oo YouTube, June 28, 2023.

[26] Film The Robots LA. 2023. *PUNKING MARCUS: Bullying Food Delivery Robots is Actually Good*. Retrieved May 6, 2025 from https://www.youtube.com/shorts/K_esB9BvqvY YouTube, Aug. 25, 2023.

[27] Film The Robots LA. 2023. *ROBOT VIOLATED: Austin the Food Delivery Robot Treated Poorly*. Retrieved May 15, 2025 from https://www.youtube.com/shorts/NIPwTGf-TWA YouTube, July 7, 2023.

[28] Film The Robots LA. 2023. *Saving Austin the Food Delivery Robot From Himself*. Retrieved May 15, 2025 from https://www.youtube.com/shorts/CZab9LqkhlY YouTube, June 25, 2023.

[29] Yuhua Liang and Seungcheol Lee. 2017. Fear of Autonomous Robots and Artificial Intelligence: Evidence from National Representative Data with Probability Sampling. *International Journal of Social Robotics* 9 (06 2017). doi:10.1007/s12369-017-0401-3

[30] Alan D. A. Mattiassi, Mauro Sarrica, Filippo Cavallo, and Leopoldina Fortunati. 2019. *Degrees of Empathy: Humans' Empathy Toward Humans, Animals, Robots and Objects: Italian Forum 2017*. 101–113. doi:10.1007/978-3-030-04672-9_7

[31] Alan D. A. Mattiassi, Mauro Sarrica, Filippo Cavallo, and Leopoldina Fortunati. 2021. What do humans feel with mistreated humans, animals, robots, and objects? Exploring the role of cognitive empathy. *Motivation and Emotion* 45, 4 (Aug. 2021), 543–555. doi:10.1007/s11031-021-09886-2

[32] Larry Morton and Beth Daly. 2008. Empathic correlates of witnessing the inhumane killing of an animal: An investigation of single and multiple exposures. *Society & Animals* 16, 3 (2008), 243–255.

[33] Khaled Nader, Paul Toprac, Samuel Scott, and Stuart Baker. 2022. Public understanding of artificial intelligence through entertainment media. *AI & Society* (2022), 1–14. doi:10.1007/s00146-022-01427-w

[34] On Demand News. 2015. *Watch robot dog 'Spot' run, walk...and get kicked*. Retrieved May 6, 2025 from https://www.youtube.com/watch?v=aR5Z6AoMh6U YouTube, Feb. 11, 2015.

[35] Sara Nielsen, Mikael B. Skov, Karl Damkjær Hansen, and Aleksandra Kaszowska. 2023. Using User-Generated YouTube Videos to Understand Unguided Interactions with Robots in Public Places. *J. Hum.-Robot Interact.* 12, 1, Article 5 (Feb. 2023), 40 pages. doi:10.1145/3550280

[36] Tech Pandaren. 2024. *Robot dog keeps its balance in extreme training*. Retrieved May 6, 2025 from https://www.youtube.com/shorts/0O-8Nd9x63I YouTube, June 20, 2024.

[37] Oliver Quick. 2022. Empathizing and Sympathizing With Robots: Implications for Moral Standing. *Frontiers in Robotics and AI* 8 (2022). doi:10.3389/frobt.2021.791527

[38] Byron Reeves and Clifford Nass. 1996. The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. *Bibliovault OAI Repository, the University of Chicago Press* (1996).

[39] Laurel D. Riek, Tal-Chen Rabinowitch, Bhismadev Chakrabarti, and Peter Robinson. 2009. Empathizing with robots: Fellow feeling along the anthropomorphic spectrum. In *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*. 1–6. doi:10.1109/ACII.2009.5349423

[40] ripL. 2024. *This robot getting bullied by stranger*. Retrieved May 6, 2025 from https://www.youtube.com/shorts/3y2C5zqK-Ls YouTube, July 7, 2024.

[41] Astrid M Rosenthal-von der Pütten, Nicole C Krämer, Laura Hoffmann, Sabrina Sobieraj, and Sabrina C Eimler. 2013. An experimental study on emotional reactions towards a robot. *International Journal of Social Robotics* 5, 1 (2013), 17–34.

[42] Astrid M. Rosenthal-von der Pütten, Frank P. Schulte, Sabrina C. Eimler, Sabrina Sobieraj, Laura Hoffmann, Stefan Maderwald, Matthias Brand, and Nicole C. Krämer. 2014. Investigations on empathy towards humans and robots using fMRI. *Computers in Human Behavior* 33 (2014), 201–212. doi:10.1016/j.chb.2014.01.004

[43] IEEE Spectrum. 2017. *Robot Tries to Escape from Children's Attack*. Retrieved May 5, 2025 from https://www.youtube.com/watch?v=CuJT9EtdETY YouTube, June 6, 2017.

[44] Melissa M Stiles. 2002. Witnessing domestic violence: The effect on children. *American Family Physician* 66, 11 (2002), 2052–2067.

[45] Megan K. Strait, Cynthia Aguillon, Virginia Contreras, and Noemi Garcia. 2017. The public's perception of humanlike robots: Online social commentary reflects an appearance-based uncanny valley, a general fear of a "Technology Takeover", and the unabashed sexualization of female-gendered robots. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 1418–1423. doi:10.1109/ROMAN.2017.8172490

[46] Wuyou Sui, Anna Sui, and Ryan E Rhodes. 2022. What to watch: Practical considerations and strategies for using YouTube for research. *DIGITAL HEALTH* 8 (2022), 20552076221123707. doi:10.1177/20552076221123707

[47] Gemma Team. 2025. Gemma 3 Technical Report. arXiv:2503.19786 [cs.CL] https://arxiv.org/abs/2503.19786

[48] Xinyan Yu, Marius Hoggenmüller, Tram Thi Minh Tran, Yiyuan Wang, and Martin Tomitsch. 2024. Understanding the Interaction between Delivery Robots and Other Road and Sidewalk Users: A Study of User-generated Online Videos. 13, 4 (2024). doi:10.1145/3677615

Received 2025-09-30; accepted 2025-12-01