









# Comparing User Perspectives in a Virtual Reality Cultural Heritage Environment

Luana Bulla<sup>1,3</sup> , Stefano De Giorgis<sup>2</sup> , Aldo Gangemi<sup>1,2</sup> ,  
Chiara Lucifora<sup>1,2</sup>  , and Misael Mongiovi<sup>1</sup> 

<sup>1</sup> ISTC - National Research Council, Rome and Catania, Italy  
{luana.bulla,aldo.gangemi,chiara.lucifora,misael.mongiovi}@istc.cnr.it

<sup>2</sup> University of Bologna, Bologna, Italy  
{stefano.degiorgis,aldo.gangemi,chiara.lucifora}@unibo.it

<sup>3</sup> University of Catania, Catania, Italy

**Abstract.** Virtual reality enables the creation of personalized user experience that brings together people of different cultures and ethnicity. We consider a novel concept of virtual reality innovation in museums, which is cognitively grounded and supported by data and their semantics, to enable users sharing their experiences, as well as to take the perspective of other users, with the ultimate goal of increasing social cohesion. The implementation of this scenario requires an autonomous artificial system that detects emotions and values from a dialogue involving museum visitors who express their personal point of view, listen to those from other visitors, and possibly take the perspective of others.

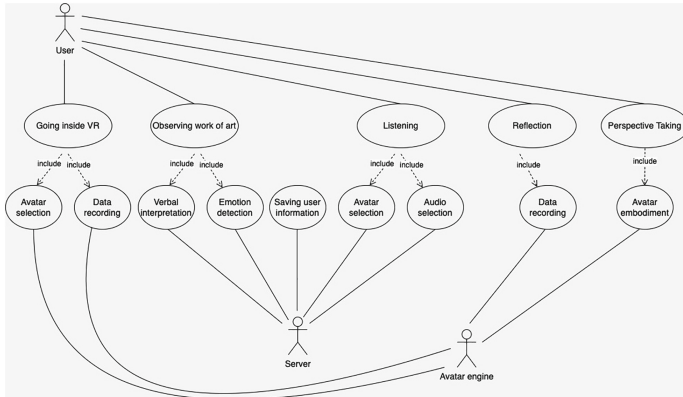
An important feature of this system is the ability of detecting similarity and dissimilarity between user perspectives expressed in speech, when exposed to artworks. This ability helps defining an effective strategy for sharing diverse user perspectives for increasing social cohesion. Moreover, it enables an unbiased quantification of the success of the interaction in terms of change in the user perspective. Based on results from previous work, we employ the Ekman's emotion model and Haidt's moral value model to extract emotional and moral value profiles from user descriptions of artworks. We propose a novel method for measuring the similarity between user perspectives by comparing emotional and moral value profiles. Our results show that the employment of unsupervised text classification models is a promising research direction for this task.

**Keywords:** Text similarity · Emotion detection · Moral value detection · Virtual reality

## 1 Introduction

Virtual Reality (VR) is an innovative technology that, by isolating the perceptive channels of users, makes them feel immersed in a virtual environment, through the sensations of virtual embodiment [39]. Its effectiveness is mainly due to its imaginative power, which allows users to experience situations as they were real

[26]. This sensation of reality is given by the system’s ability to process the information received, and to offer visual and sound feedback in real time [40]. Today VR is applied in many domains related to cognitive sciences, including social technology, which facilitates interactions by promoting the psychological well-being of people in social contexts. As a countermeasure, as a part of a wider project<sup>1</sup>, which aims to ensure social cohesion, participation and inclusion through cultural engagement, we use virtual reality technology to build an innovative system, in which users are subjected to the Interpretation-Reflection Loop (IRL) [8] in an immersive environment that makes them share their personal interpretations with others, listen to others’ interpretations given by different people, and take another point of view through virtual embodiment. We are implementing a system that supports continuous communication between its constituent parts, i.e., a VR interface jointly with an external server that detects emotions and moral values from user speech, and the user itself. The system selects a specific avatar for each user, based on their personal characteristics such as age, gender and nationality; subsequently, following the user’s first interpretation, the external server detects emotions and moral values to select an appropriate avatar that will interact with the user. All users’ interpretations are recorded and associated with specific avatars. These data are saved in an external dataset that the server can access for further reuse (Fig. 1).



**Fig. 1.** A use case diagram of the mixed VR-AI system for emotion-value-based perspective taking.

An essential feature of the system is the ability to quantify similarity between user perspectives. For instance, presenting perspectives that are different from those of users can make them reflect about their standpoint, and encourage them to adopt a wider perspective, ultimately increasing social cohesion. The

<sup>1</sup> SPICE: Social Cohesion, Participation and Inclusion through Cultural Engagement – EC Grant Agreement number 870811.

experiments conducted until now focus on the emotional and axiological (moral) meaning of those perspectives, which will be eventually combined with other semantic features in future work. The task can also be useful for comparing perspectives of a user at two different points in time, e.g. before and after an interaction, to quantify the propensity of people to change their point of view under certain circumstances (e.g., when impersonating an avatar holding a different standpoint). Considering the large volume of data to process, and the necessity of performing the task on-line and autonomously, it is essential that the system is able to reason with user perspectives and quantify their similarity.

We adopt a novel method to quantify the similarity between user perspectives through emotional and moral values lenses [25]. We leverage text classifiers to extract emotional profiles and moral value profiles from text obtained after speech-to-text conversion. We then compare profiles from different speeches by means of cosine similarity. To evaluate our method, we collected data from our VR prototype, obtaining a total of 396 interpretations on 6 works of art associated to 6 avatars. We compared some state-of-the-art supervised or unsupervised classifiers based on Transformers [1] and adopt the top-performing one for computing the similarity between perspectives.

The remainder of the paper is structured as follows. Next, we present some background on social and technological aspects of our work (Sect. 2). In Sect. 3 we present our VR scenario used for data collection. Section 4 introduces our method for evaluating and comparing user perspectives. We report our results in Sect. 5, and then conclude in Sect. 6.

## 2 Background

We provide some background on social (Sect. 2.1) and technological (Sect. 2.2) aspects related to this work. First, we briefly overview recent work on social cohesion, which motivates the importance of implementing suitable strategies and developing effective tools for favouring it. Second, we discuss the reference models we employ and evaluate for automatically detecting emotions and moral values evoked by people’s speech. We also describe state-of-the-art metrics for computing the semantic similarity between natural language sentences, which we employ as a starting point for our method.

### 2.1 Improving Social Cohesion

In the field of cognitive science, social cohesion represent an important topic that needs to be investigated since, as shown in recent studies [21, 22, 34, 35], the lack of social cohesion, which increases social isolation, can have a strong impact on people’s health, impacting life expectancy and increasing the risk of death, as well as the risk of developing dementia. In general, social exclusion is mainly characterized by rejection (i.e. being denied) and ostracism (i.e. being ignored) [41], that lead to what social psychologists call social pain that has a strong impact on people’s emotions and behaviours [34]. Previous studies have shown

that virtual embodiment enables empathetic sharing of personal experiences that boosts social inclusion [6, 39] as well as a reduction of prejudices towards different people [43]. For instance, Slater et al. [38] have shown that adopting another point of view allows the user to be more sensitive to violent behavior, leading to a reduction in racial aggression [23] and a decrease in child maltreatment [19]. In this line, in our previous work (Lucifora et al., under revision), we used an ecological paradigm in virtual reality to promote social cohesion through the virtual embodiment. Our results show that the percentages of people who change their interpretation (totally or partially) when they embody the avatar are equal to or greater than the ones who do not change their interpretation at all (Chi square between subjects p-level 0.018; Chi-square within subjects group n.1 p-level 0.007; Chi-square within subjects group n.2 p-level 0.016). This result allows us to state that the use of virtual embodiment can improve social cohesion and reduce ostracism and prejudice.

## 2.2 Reference Models

The following paragraphs describe the two theoretical frameworks adopted to extract users' emotion and value profile, in turn used to measure social cohesion.

*Basic Emotions Theory.* Originally developed by Paul Ekman [12], the Basic Emotions (BE) theory adopted in this work is a 2018 revised version [13], available online<sup>2</sup>. In this version of the theory, complex emotions are traced back to 6 main emotions, used here for classification tasks: Anger, Fear, Sadness, Disgust, Enjoyment, and Surprise. The Basic Emotions theory is adopted in this context because, among models that consider emotions as discrete categories, it is a widely used framework in well-known datasets used for tasks of automatic extraction of emotional content from natural language text [3, 24, 37], and thus lends itself to meeting the intentions of the work.

*Moral Foundations Theory.* The theoretical framework adopted to map moral values is the Moral Foundations Theory (MFT) [15], by Graham and Haidt. This framework organizes values in six "foundations", namely dyads of positive pole (value) vs negative pole (violation). The six dyads are:

- *Care/Harm:* caring versus harming behavior, it is grounded in mammals attachment system and cognitive appraisal mechanism of dislike of pain. It grounds virtues of kindness, gentleness and nurturance.
- *Fairness/Cheating:* it is based on social cooperation and typical nonzero-sum game theoretical situations based on reciprocal altruism. It underlies ideas of justice, rights and autonomy.
- *Loyalty/Betrayal:* it is based on tribalism tradition and the positive outcome coming from cohesive coalition, as well as the ostracism towards traitors.

---

<sup>2</sup> The Atlas of Emotions developed from a revised version of Ekman's Basic Emotions theory is available here: <https://atlasofemotions.org/>.

- *Authority/Subversion*: social interactions in terms of societal hierarchies, it underlies ideas of leadership and deference to authority, as well as respect for tradition.
- *Purity/Degradation*: derived from psychology of disgust, it implies the idea of a more elevated spiritual life; it is expressed via metaphors like “the body as a temple”, and includes the more spiritual side of religious beliefs.
- *Liberty/Oppression*: it expresses the desire of freedom and the feeling of oppression when it is negated.

MFT is adopted in this work since it is described as “nativist, cultural-developmental, intuitionist, and pluralist approach to the study of morality” [15]. It is “Nativist”, due to its neurophysiological grounding; “cultural-developmental” because it considers environmental variables in the morality-building process [16]; “intuitionist” in asserting that there is no unique moral or non-moral trigger, but rather many co-occurring patterns resulting in a rationalized judgment [17]; “pluralist” in considering that more than one narrative could fit the moral explanation process [18]. This final aspect of perspective pluralism is particularly relevant according to the this work’s aim of fostering social cohesion.

Previous work on moral values and emotions detection focuses mainly on machine learning techniques and quantitative or semantic text document analysis [1, 28, 29, 32].

For the purposes of our study, we selected three different methods based on natural language models to classify the emotional and moral perspective of users based on their natural language descriptions of artworks<sup>3</sup>. The first is based on Asprino et al. [5], which employs a zero-shot classification in an unsupervised way, drawing inspiration from Yin et al. [44]. This approach is based on the use of a model trained for a natural language inference (NLI) task, where a *premise* is determined whether or not it entails a *hypothesis*. Starting from the input text document as a premise, the system recognizes if one or more taxonomic labels, provided as hypotheses, are semantically coherent with the input. The advantage of this method is that it does not require training on an annotated dataset and hence it is directly applicable in our scenario, where we cannot rely on a large annotated dataset. We evaluate two additional supervised BERT-based machine learning models trained on available datasets. Both systems take textual data as input to perform a multi-class classification task based on Ekman’s emotion taxonomy. [20, 30]. The former is trained on a set of tweets, while the latter has a greater heterogeneity, as it is fine-tuned on a large number of different sources such as Twitter, Reddit, student self-assessments and television dialogue expressions. Both these methods suffer from the data shift problem, i.e. their performances degrade when applied to data with different characteristics than the training set [42].

To measure the change of user perspective and compare perspectives of different users, we consider as a baseline a state-of-the-art technique for semantic




























---

<sup>3</sup> All methods take as input text documents, which can be generated by a speech-to-text tool.

similarity among text documents, consisting on computing a vector representation of each text document and compare them by means of cosine similarity. We employ Sentence-BERT [33] for computing the text representation. Sentence-BERT is a variant of the pre-trained BERT model [11] that generates semantically significant sentence embeddings using siamese and triplet network structures. The limit of this method in our scenario is that it considers the overall semantic similarity and cannot focus on a specific lens (emotional or value) to compare perspectives.

### 3 Data Collection

We applied an ecological paradigm in VR, made up by a helmet of virtual reality (Oculus Quest 2), equipped with a headphone that provides 3D audio effects, and an immersive and realistic virtual museum based on Blue Dot Studios- Art Gallery, developed with Unity Engine 3D 2020.3.33 as software. Inside the virtual museum, we included some artworks of the Modern Art Gallery in Turin (Italy) and avatars, which we built using Ready-player.me and Mixamo.com for the animations. To have a representative sample of the world population we used three main variables, that are i) age (children, adults, elderly), ii) gender (male, female, no gender), iii) nationality (European, African, Asian). In total we built 27 avatars, multiplying 3 variables to 9 categories (Fig. 2).

	Adulthood			Childhood			Seniors		
African									
Asiatic									
European									
	Male	Female	No-Gender	Male	Female	No-Gender	Male	Female	No-Gender

**Fig. 2.** Avatar generation based on age, gender and nationality.

We carried out an experimental study between subjects, on a total sample of 44 subjects, divided into two experimental groups ( $N = 22$ ) in order to balance our stimuli in relation to the main characteristics of the avatars and art movements. So far, we recording 396 verbal interpretations (in Italian language) on 6 works of art, related to 6 avatars. We have prepared an interview based on four questions based on widely used theories for emotions [12] and moral values [15]. Two questions are related to the user’s feeling (Q1 “How does this work of art make you feel?”; Q2 “What moral value does this work of art inspire you?”), while two questions are related to the specific categorization of emotions and values (Q3 “Among these emotions (Ekman’s model) which do you think the work

of art arouses?"; Q4 "Of these values (Haidt's model) which one do you think the work of art arouses?"). Our procedure is based on a 4-step timeline: At time 0 the user provides its personal interpretation of the work, based on an interview given by the researcher. At time 1 the user is invited to listen to the (verbally expressed) interpretation provided by the avatar who shares the scene with him. At time 2, users give their personal reflection, discussing their interpretations against those of the avatar. No specific questions are asked here. At the last time (time 3), the user is invited to respond again to the structured interview, given by the researcher, but this time users take the role of the avatar through virtual embodiment. Time 2 and time 3 were reversed in the two experimental groups in order to understand how much virtual embodiment can influence the interpretation-reflection loop. We have recorded 3 verbal interpretations (time 0, time 2, time 3) for each participant.

## 4 Comparing Perspectives by Emotion and Moral Value Classification

In order to detect the emotions and moral values from human-avatar dialogues, we analyzed the answers to Q1 and Q2 at time 0 and time 3, plus the free reflection at time 2, using the supervised and unsupervised AI systems for natural language understanding discussed in Sect. 2.2. Following Asprino et al. [5], we apply a pre-trained Natural Language Inference system as ready-made zero-shot sequence classifier (namely NLI-based). We use the framework to categorize emotions and moral values, developing two different system configurations based respectively on the taxonomies of Ekman and Haidt. We compared the results with two distinct BERT-based supervised models for emotion detection: E-BERTweet [30] and E-DistilRoBERTa [20] trained on annotated benchmark data [4, 9, 10, 27, 31, 36].

We propose a novel method to determine how emotionally and morally comparable the text documents are. Specifically, we compare the answers to Q1 and Q2 at time 0 and time 3 to understand whether the users changed their perspective after the virtual embodiment (time 3) with respect to their first interpretation of the artwork (time 0). Since to the best of our knowledge there is no approach available to measure the similarity of perspective through an emotional and/or value lens, we consider as a baseline the cosine similarity between Sentence-BERT text representations (Sect. 2.2).

Our approach considers the output function of the (emotion or moral value) classifier, typically a probability distribution over the set of classes computed by a softmax normalization function, of each text document and compute their cosine similarity. Specifically:

$$emotional\_similarity(t_1, t_2) = \frac{\mathbf{f}_e(t_1) \cdot \mathbf{f}_e(t_2)}{\|\mathbf{f}_e(t_1)\| \|\mathbf{f}_e(t_2)\|} \quad (1)$$

$$value\_similarity(t_1, t_2) = \frac{\mathbf{f}_v(t_1) \cdot \mathbf{f}_v(t_2)}{\|\mathbf{f}_v(t_1)\| \|\mathbf{f}_v(t_2)\|} \quad (2)$$

where  $\mathbf{f}_e(t)$  and  $\mathbf{f}_v(t)$  are the output vectors of the emotional and value classifiers on text  $t$ , respectively;  $\cdot$  is the dot product between vectors and  $\|\mathbf{x}\|$  is the 2-norm of vector  $\mathbf{x}$ .

The result is a value between 0 and 1, where 0 represent completely different perspectives, while 1 represent perfectly coincident perspectives. For example, let’s assume that the emotional profile of a user is equal to “joy” at time 0 and “sadness” at time 3. The emotional similarity score will therefore be equal to a 0.13 point in percentage, which indicates a sharp change in his emotional state. The method can be applied even when the output does not represent a probability distribution, provided that it is given as a vector of  $N$  non-negative values, where  $N$  is the number of classes. In our experiments (Sect. 5) we consider the output of the zero-shot NLI-based emotion and moral value classifiers normalized by softmax.

## 5 Results and Evaluation

We consider two basic experimental frameworks dealing with emotional and value-based detection and similarity in text, respectively. Next we describe the setup and results of detection, then we discuss the results concerning similarity.

**Emotion and Moral Detection.** For the moral and emotion classification we considered the answers to questions Q1 and Q2 at time 0 (t0), time 2 (t2) and time 3 (t3) for both experimental groups (see Sect. 3) and divided our sample in three categories, that are: (i) *irrelevant* answers, which do not contain any emotions or moral values; (ii) *incoherent* answers, which contain emotions incompatible with their annotations; (iii) *relevant* (and coherent) answers, which express and annotate clear emotions or moral values. The splitting process was done manually, with the support of human experts. We consider the user’s annotations of emotions and moral values (answers to questions Q3 and Q4, respectively) as the ground truth and compare the system’s predictions with them to measure performances. For the emotion classification task we compare the unsupervised NLI-based model [5], and the supervised E-BERTeet [30] and E-DistilRoBERTa [20] models, introduced in Sect. 2.2. We adopt the unsupervised NLI-based [5] method for detecting moral values.

We report the results in terms of weighted F1-score. We compute the outcomes considering six basic emotions and six value dimensions, respectively. The results show that the models are able to detect the emotions and values from relevant (and coherent) answers with significant accuracy, for any step (Table 1 and Table 2). As expected, the performances degrade for irrelevant and incoherent answers, since in that case user annotations are not representative of the emotions and moral values expressed in the free answer.

With the exception of time t0 of the second trial, the NLI-based model [5] performs noticeably better in both the first and second group for the emotion detection task. In this scenario, the F1-score ranges from 0.56 to 0.72. For the task of identifying moral values the F1 score fluctuates between 32% and 49% on

relevant answers. Considering the difficulty of the task, because of subjectivity and ambiguity of moral values expressed in natural language statements, we consider this result promising.

**Table 1.** Emotion detection results for the unsupervised zero-shot model (NLI-based) and supervised BERT-based models (E-BERTweet and E-DistilRoBERTa) in terms of weighted F1-score. In relevant (and coherent) answers, NLI-based outperforms supervised methods (which have been trained on third-part datasets). As expected, the F1-score is low for irrelevant and incoherent answers since the users’ annotations do not represent the real emotions expressed in the free answers.

Category	Model	Group n. 1			Group n. 2		
		t0	t2	t3	t0	t2	t3
Relevant	NLI-based	<b>.72</b>	<b>.62</b>	<b>.70</b>	.56	<b>.56</b>	<b>.47</b>
	E-BERTweet	.56	.34	.47	.36	.36	.22
	E-DistilRoBERTa	.67	.48	.61	<b>.66</b>	.55	<b>.47</b>
Irrelevant	NLI-based	.26	.05	.26	.11	.20	.09
	E-BERTweet	.11	.01	.01	.42	.10	.04
	E-DistilRoBERTa	.44	.01	.06	.36	.01	.04
Incoherent	NLI-based	.19	.18	.13	.06	.25	.20
	E-BERTweet	.00	.00	.00	.00	.00	.00
	E-DistilRoBERTa	.00	.23	.22	.08	.00	.00

**Table 2.** Moral value detection results for the unsupervised NLI-based zero-shot model in terms of weighted F1-score. As for the emotion detection case, the F1-score is low for irrelevant and incoherent answers since the users’ annotations do not represent the real moral values expressed in the free answers.

Category	Group n. 1			Group n. 2		
	t0	t2	t3	t0	t2	t3
Relevant	.48	.49	.35	.38	.32	.29
Irrelevant	.09	.30	.42	.18	.30	.33
Incoherent	.27	.30	.33	.18	.32	.19

**Emotional and Moral Similarity.** To evaluate the emotional and moral similarity scores, we consider the portion of cases of our sample that are consistent and relevant with user annotations. Two annotators evaluated the statements given at time 0 and time 3 to determine whether there is a change in the user’s emotional and moral perspective. Annotators specifically highlighted sentences that expressed distinctions with a “yes” (corresponding to non similar perspectives) and coherent text with a “no” (corresponding to similar perspectives). We

compare the scores with the conforming annotations to see how well our emotional and moral similarity score can predict a change in perspective by applying a threshold and predicting a change when the similarity is below the threshold.

We use the Area Under the Curve (AUC) measure to evaluate the performances of our change-in-perspective classifier. AUC provides an aggregate measure of performances against all possible similarity thresholds. AUC is also recommended in an unbalanced dataset framework when comparing continuous and binary variables. Results are reported in Table 3. The outcomes show that our tool consistently produces AUC comparable or higher than the reference BERT-based semantic similarity system Sentence-BERT [33] discussed in Sect. 2.2. We obtained an AUC score of 81% and 70% for emotional and moral semantic similarity, respectively.

**Table 3.** AUC of the similarity metrics for predicting a change in perspective. We evaluate emotional, moral, and semantic similarity for coherent statements between answers at time 0 and time 3.

Task	Emotional Similarity	Moral Similarity	SBert	Support
Emotional validation	<b>0,81</b>		<b>0,81</b>	25
Moral validation		<b>0,70</b>	0,66	19

In summary, NLI-based emotion and moral detection systems show do not match the annotation), significant performances when the user annotation is coherent with the free description. Furthermore, the NLI model outcomes for the emotion detection task are superior to those of supervised systems trained on third-part datasets. This confirms the versatility of unsupervised architectures in out-of-domain tasks. The results obtained from the emotional and moral similarity system show an overall positive trend in capturing the emotional and value nuances of sentences.

## 6 Conclusion

In order to improve social cohesion, we implement an integration of VR and AI to deploy perspective taking into museums, and automatically update VR environment objects, works of art, and user metadata. To this end, we have recorded a total of 396 verbal interpretations from 44 subjects, using an experimental procedure based on a 4-step workflow, employed on an immersive VR system enriched with automated extraction of emotions and moral values. Our results lead to two important considerations. The first one is related to the utility of VR interfaces to promote social cohesion. In this sense, in line with recent literature [2], our results show that virtual embodiment is able to increase prosocial behavior and arouse empathy among users. These results motivate us to investigate perspective taking in immersive environments for different contexts e.g., including political or religious conflicts.

Our data relating to the higher number of empathic people compared to non-empathic people during virtual embodiment confirms previous studies that show that the first perspective in an immersive virtual reality leads to a strong feeling of embodiment towards an artificial body [7], which on its turn allows users to be more empathetic in relation to the feeling of other people [14]. The second result of our research is related to the use of natural language processing models. Our results show that zero-shot learning with transformers is able to detect emotions and moral values that are evoked by the personal interpretation of artworks. Furthermore, we proposed a method for determining the degree of emotional and moral coherence between two statements, and show that it is useful to infer the change of user’s perspectives.

In conclusion, our results point to the feasibility of building an innovative VR system based on data semantics, as a valid tool to sensitize people to other interpretations, improving social cohesion through virtual embodiment.

**Acknowledgements.** This work is supported by the H2020 projects TAILOR: Foundations of Trustworthy AI - Integrating Reasoning, Learning and Optimization – EC Grant Agreement number 952215 – and SPICE: Social Cohesion, Participation and Inclusion through Cultural Engagement – EC Grant Agreement number 870811, as well as by the Italian PNRR MUR project PE0000013-FAIR.

## References

1. Acheampong, F.A., Nunoo-Mensah, H., Chen, W.: Transformer models for text-based emotion detection: a review of BERT-based approaches. *Artif. Intell. Rev.* **54**(8), 5789–5829 (2021)
2. Ahn, S.J., Le, A.M.T., Bailenson, J.: The effect of embodied experiences on self-other merging, attitude, and helping behavior. *Media Psychol.* **16**(1), 7–38 (2013)
3. Alm, E.C.O.: *Affect in\* text and speech*. University of Illinois at Urbana-Champaign (2008)
4. del Arco, F.M.P., Strapparava, C., Lopez, L.A.U., Martín-Valdivia, M.T.: Emo-event: a multilingual emotion corpus based on different events. In: *Proceedings of the 12th Language Resources and Evaluation Conference*, pp. 1492–1498 (2020)
5. Asprino, L., Bulla, L., De Giorgis, S., Gangemi, A., Marinucci, L., Mongiovi, M.: Uncovering values: detecting latent moral content from natural language with explainable and non-trained methods. In: *Proceedings of Deep Learning Inside Out (DeeLIO 2022): The 3rd Workshop on Knowledge Extraction and Integration for Deep Learning Architectures*, pp. 33–41 (2022)
6. Blascovich, J., Loomis, J., Beall, A.C., Swinith, K.R., Hoyt, C.L., Bailenson, J.N.: Immersive virtual environment technology as a methodological tool for social psychology. *Psychol. Inq.* **13**(2), 103–124 (2002)
7. Casula, E.P., et al.: Feeling of ownership over an embodied avatar’s hand brings about fast changes of fronto-parietal cortical dynamics. *J. Neurosci.* **42**(4), 692–701 (2022)
8. Daga, E., et al.: Integrating citizen experiences in cultural heritage archives: requirements, state of the art, and challenges. *ACM J. Comput. Cult. Herit. (JOCCH)* **15**(1), 1–35 (2022)

9. Dan-Glauser, E.S., Scherer, K.R.: The difficulties in emotion regulation scale (DERS). *Swiss J. Psychol.* (2012)
10. Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., Ravi, S.: Goemotions: a dataset of fine-grained emotions. arXiv preprint [arXiv:2005.00547](https://arxiv.org/abs/2005.00547) (2020)
11. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint [arXiv:1810.04805](https://arxiv.org/abs/1810.04805) (2018)
12. Ekman, P.: Basic emotions. In: *Handbook of Cognition and Emotion*, vol. 98, no. 45–60, p. 16 (1999)
13. Ekman, P.: *Atlas of emotion* (2018)
14. Fusaro, M., Tieri, G., Aglioti, S.M.: Seeing pain and pleasure on self and others: behavioral and psychophysiological reactivity in immersive virtual reality. *J. Neurophysiol.* **116**(6), 2656–2662 (2016)
15. Graham, J., et al.: Moral foundations theory: the pragmatic validity of moral pluralism. In: *Advances in Experimental Social Psychology*, vol. 47, pp. 55–130. Elsevier (2013)
16. Graham, J., Nosek, B.A., Haidt, J.: The moral stereotypes of liberals and conservatives: exaggeration of differences across the political spectrum. *PLoS ONE* **7**(12), e50092 (2012)
17. Haidt, J.: The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychol. Rev.* **108**(4), 814 (2001)
18. Haidt, J.: *The righteous mind: why good people are divided by politics and religion*. Vintage (2012)
19. Hamilton-Giachritsis, C., Banakou, D., Garcia Quiroga, M., Giachritsis, C., Slater, M.: Reducing risk and improving maternal perspective-taking and empathy using virtual embodiment. *Sci. Rep.* **8**(1), 1–10 (2018)
20. Hartmann, J.: Emotion english distilroberta-base (2022). <https://huggingface.co/j-hartmann/emotion-english-distilroberta-base/>
21. Holt-Lunstad, J., Smith, T.B., Baker, M., Harris, T., Stephenson, D.: Loneliness and social isolation as risk factors for mortality: a meta-analytic review. *Perspect. Psychol. Sci.* **10**(2), 227–237 (2015)
22. Holt-Lunstad, J., Smith, T.B., Layton, J.B.: Social relationships and mortality risk: a meta-analytic review. *PLoS Med.* **7**(7), e1000316 (2010)
23. Kishore, S., Spanlang, B., Iruretagoyena, G., Halan, S., Szostak, D., Slater, M.: A virtual reality embodiment technique to enhance helping behavior of police toward a victim of police racial aggression. *PRESENCE: Virtual Augmented Reality* **28**, 5–27 (2022)
24. Li, Y., Su, H., Shen, X., Li, W., Cao, Z., Niu, S.: Dailydialog: a manually labelled multi-turn dialogue dataset. arXiv preprint [arXiv:1710.03957](https://arxiv.org/abs/1710.03957) (2017)
25. Lieto, A., Pozzato, G.L., Striani, M., Zoia, S., Damiano, R.: Degari 2.0: a diversity-seeking, explainable, and affective art recommender for social inclusion. *Cogn. Syst. Res.* **77**, 1–17 (2023)
26. Lucifora, C., et al.: Cyber-therapy: the use of artificial intelligence in psychological practice. In: Russo, D., Ahram, T., Karwowski, W., Di Bucchianico, G., Taiar, R. (eds.) *IHSI 2021. AISC*, vol. 1322, pp. 127–132. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-68017-6\\_19](https://doi.org/10.1007/978-3-030-68017-6_19)
27. Mohammad, S., Bravo-Marquez, F., Salameh, M., Kiritchenko, S.: Semeval-2018 task 1: affect in tweets. In: *Proceedings of the 12th International Workshop on Semantic Evaluation*, pp. 1–17 (2018)

28. Nandwani, P., Verma, R.: A review on sentiment analysis and emotion detection from text. *Soc. Netw. Anal. Min.* **11**(1), 1–19 (2021)
29. Pacheco, M.L., Goldwasser, D.: Modeling content and context with deep relational learning. *Trans. Assoc. Comput. Linguist.* **9**, 100–119 (2021)
30. Pérez, J.M., Giudici, J.C., Luque, F.: pysentimiento: a python toolkit for sentiment analysis and socialnlp tasks. arXiv preprint [arXiv:2106.09462](https://arxiv.org/abs/2106.09462) (2021)
31. Poria, S., Hazarika, D., Majumder, N., Naik, G., Cambria, E., Mihalcea, R.: Meld: a multimodal multi-party dataset for emotion recognition in conversations. arXiv preprint [arXiv:1810.02508](https://arxiv.org/abs/1810.02508) (2018)
32. Priniski, J.H., et al.: Mapping moral valence of tweets following the killing of George Floyd. arXiv preprint [arXiv:2104.09578](https://arxiv.org/abs/2104.09578) (2021)
33. Reimers, N., Gurevych, I.: Sentence-BERT: sentence embeddings using Siamese BERT-networks. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics (2019). <https://arxiv.org/abs/1908.10084>
34. Riva, P., Eck, J.: The many faces of social exclusion. In: *Social Exclusion: Psychological Approaches to Understanding and Reducing Its Impact*, pp. ix–xv (2016)
35. Riva, P., Wirth, J.H., Williams, K.D.: The consequences of pain: the social and physical pain overlap on psychological responses. *Eur. J. Soc. Psychol.* **41**(6), 681–687 (2011)
36. Saravia, E., Liu, H.C.T., Huang, Y.H., Wu, J., Chen, Y.S.: Carer: contextualized affect representations for emotion recognition. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 3687–3697 (2018)
37. Scherer, K.R., Wallbott, H.G.: Evidence for universality and cultural variation of differential emotion response patterning. *J. Pers. Soc. Psychol.* **66**(2), 310 (1994)
38. Slater, M., et al.: Bystander responses to a violent incident in an immersive virtual environment. *PLoS ONE* **8**(1), e52766 (2013)
39. Slater, M., Spanlang, B., Sanchez-Vives, M.V., Blanke, O.: First person experience of body transfer in virtual reality. *PLoS ONE* **5**(5), e10564 (2010)
40. Steuer, J.: Defining virtual reality: dimensions determining telepresence. *J. Commun.* **42**(4), 73–93 (1992)
41. Wesselmann, E.D., Grzybowski, M.R., Steakley-Freeman, D.M., DeSouza, E.R., Nezelek, J.B., Williams, K.D.: Social exclusion in everyday life. In: *Social Exclusion: Psychological Approaches to Understanding and Reducing Its Impact*, pp. 3–23 (2016)
42. Wright, D., Augenstein, I.: Transformer based multi-source domain adaptation. arXiv preprint [arXiv:2009.07806](https://arxiv.org/abs/2009.07806) (2020)
43. Yee, N., Bailenson, J.N.: Walk a mile in digital shoes: the impact of embodied perspective-taking on the reduction of negative stereotyping in immersive virtual environments. *Proc. PRESENCE* **24**, 26 (2006)
44. Yin, W., Hay, J., Roth, D.: Benchmarking zero-shot text classification: datasets, evaluation and entailment approach (2019)