

# The Ethical Risks of Analyzing Crisis Events on Social Media with Machine Learning

Angelie Kraft<sup>1,2,\*</sup>, Ricardo Usbeck<sup>1,2</sup>

<sup>1</sup>Universität Hamburg, Department of Informatics, Vogt-Kölln-Straße 30, 22527 Hamburg

<sup>2</sup>Hamburger Informatik Technologie-Center e.V. (HITeC), Vogt-Kölln-Straße 30, 22527 Hamburg

## Abstract

Social media platforms provide a continuous stream of real-time news regarding crisis events on a global scale. Several machine learning methods utilize the crowd-sourced data for the automated detection of crises and the characterization of their precursors and aftermaths. Early detection and localization of crisis-related events can help save lives and economies. Yet, the applied automation methods introduce ethical risks worthy of investigation — especially given their high-stakes societal context. This work identifies and critically examines ethical risk factors of social media analyses of crisis events focusing on machine learning methods. We aim to sensitize researchers and practitioners to the ethical pitfalls and promote fairer and more reliable designs.

## Keywords

crisis informatics, machine learning, artificial intelligence, social media, ethics, risks

## 1. Introduction

Social media platforms are a bottom-up community-driven means for real-time information exchange during crisis events [1]. They are an important tool in keeping citizens and authorities up-to-date in urgent situations [2, 3]. The shared information can help to establish precautionary measures, organize humanitarian aid, or keep track of missing people. Algorithmic approaches are used to efficiently filter, condense, and extract large amounts of social media posts [4, 5]. Respective systems nowadays largely rely on deep learning (DL) methods for natural language processing (NLP) [6], computer vision (CV) [7], or multimodal techniques [8].

The COVID-19 pandemic is a contemporary example where privacy and personal liberties were sacrificed for the quick development of new technologies [9]. Although crisis events ask for fast responses, the innovation process must not happen at the cost of ethical considerations. In this paper, we identify the main ethical risks when analyzing social media content via machine learning (ML) to detect and characterize crises. To scrutinize ethical aspects of technology, we take on a sociotechnical view [10]: We consider algorithms, their in-, and output data, as well as the social system within which these are embedded. At the heart of this assessment is the

---

*International Workshop on Data-driven Resilience Research 2022, July 6, 2022, Leipzig, Germany*


\*Corresponding author.


✉ [angelie.kraft@uni-hamburg.de](mailto:angelie.kraft@uni-hamburg.de) (A. Kraft); [ricardo.usbeck@uni-hamburg.de](mailto:ricardo.usbeck@uni-hamburg.de) (R. Usbeck)

🌐 <https://krangelie.github.io/> (A. Kraft);

<https://www.inf.uni-hamburg.de/en/inst/ab/sems/people/ricardo-usbeck.html> (R. Usbeck)

🆔 0000-0002-2980-952X (A. Kraft); 0000-0002-0191-7211 (R. Usbeck)

 © 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

potential long-term impact on people’s well-being, values, expectations, and fair treatment, and ultimately on whom a computer system serves and whom it harms. We elaborate on each of the risks to sensitize practitioners and researchers developing and deploying respective systems.

## 2. Related Work

For several years now, ML methods have been used for the analysis of social media posts regarding various types of natural disasters, like floods, hurricanes, earthquakes, fires, and draughts around the globe [5]. Systems have been developed to facilitate early warnings and to support disaster responses or damage assessments [4]. NLP methods can help to distinguish informative from uninformative texts posted on social media, classify the type of crisis event the text belongs to [6, 11], or the type of crisis-related content that is discussed (e.g., warnings, utilities, needs, affected people [4]). The same can be done based on photos through CV approaches [8]. The semantic content of posts can be further leveraged with spatial and/or temporal information to facilitate crisis mapping. For the Chennai flood in 2015, Anbalagan and Valliyammai [2] built a crisis mapping system that classified related tweets regarding their content type (e.g., requests for help, sympathy, warnings, weather information, infrastructure damages, etc.). This information was combined with the geographic coordinates derived from textually mentioned locations via geoparsing. Tools like this which can identify and locate a crisis-related event can help emergency responders navigate complex information streams.

In 2015, Crawford and Finn [12] outlined different classes of limitations of using social media data in crisis informatics. **Ontological limitations:** Social media activities spike around more sensational instances, although crises onsets are oftentimes followed-up by long-term effects. So, the time frame of a virtual **discourse is not representative** of the actual crisis timeline. Further, applications for humanitarian aid have in the past demonstrated a risk of reifying **power imbalances:** “Although crowdsourcing projects can allow the voices of those closest to a disaster to be heard, some projects most strongly enhance the agency of international humanitarians” (p. 495, [12]). **Epistemological limitations:** The interpretability of social media data is limited by the role that platforms play in shaping the data. Recommendation systems determine what users get to see and share. Moreover, a platform can be seen as a cultural context, with its trends and communicative patterns. Contents may exaggerate real events and be charged with opinion and emotion. Finally, distinguishing between human- and bot-generated messages is not always feasible. **Ethical issues:** The main point here is the issue of **privacy**. Personal statements of users are gathered at a time in which they are especially vulnerable. Their posts oftentimes include sensitive information about location or well-being and the needs of themselves or others. Crawford and Finn [12] claim that consent must not be sacrificed for “the greater good”.

The privacy issue was also listed as one ethical risk factor by Alexander [13], alongside the loss of discretion caused by a tendency for sharing intimate details. Moreover, the author pointed out that especially wealthy and technologically literate individuals benefit from digital means of disaster management. This adds to the previously mentioned reification of power imbalances. Finally, the spread of rumors and misinformation through users, as well as ideology-driven governance of platforms affect the reliability of details and can cause an overall

**misrepresentation of crises** and their causes.

Regarding the use of artificial intelligence (AI) in crisis informatics, Tzachor et al. [9] highlight issues of the **disparate impact** of algorithmic outputs, as well as the lack of **transparency** and **trustworthiness** of AI models. The authors demand a principle of **ethics with urgency** [9] which entails (1) “**ethics by design**” to consider ethical risks throughout the development process and foresee broader societal impacts, (2) validated **robustness of AI systems**, and (3) building **public trust** through independent oversight and transparency.

### 3. Ethical Risks

The presented work consolidates previous ethical risk assessments of crisis informatics with social media data (Section 2) with an emphasis on ML methods. We expand on previous works by examining recent technological advancements and newer insights on their potential risks. For a better overview, the following sections are sorted by data- and algorithm-related concerns. Please note that there is a conceptual overlap between some of the issues mentioned: e.g., limited representativeness of data is problematic because algorithms capture and reproduce biases [14]. However, awareness of the problem layers allows for an in-depth understanding and faceted scrutiny of future software.

#### 3.1. Limited Representativeness

To understand who communicates and receives information on social media, it is necessary to take a disaggregated look at user demographics. In 2020, there were more than 3.6 billion social media users worldwide.<sup>1</sup> Facebook ranks first amongst the most popular platforms, with 2.9 billion users as of January 2022.<sup>2</sup> Even though Twitter did not make the top ten list with only 426 million users, it is still the most researched social media platform [4]. The reason for this might be its easily accessible API for researchers, allowing them to analyze its full stream of posts. By far margin, the majority of Twitter users come from the United States or Japan (India ranked third with less than half of the amount of users in Japan, as of January 2022).<sup>3</sup> In April 2021, 38.5% of all Twitter users ranged between ages 25 and 34, and 21% were between 35 and 49 years old.<sup>4</sup> These numbers indicate that most research done on Twitter corpora is based on the **perceptions of a non-representative sample of people**. Here, perception relates to both the reality witnessed by individuals due to spatio-temporal factors, and also to belief and ideology – especially in the context of crisis [15].

Social media platforms use recommendation systems to display content that echoes users’ interests and opinions. The **filter bubble hypothesis** states that this mechanism leads to isolated **echo chambers** and polarization of social networks [16]. Regarding the attention dynamics on social media, some voices recently argued that the Twitter community paid more attention to the 2022 Ukraine crisis than other wars and genocides happening in the

---

<sup>1</sup><https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>

<sup>2</sup><https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>

<sup>3</sup><https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/>

<sup>4</sup><https://www.statista.com/statistics/283119/age-distribution-of-global-twitter-users>

meantime.<sup>5</sup> They claim that such phenomena stem from and reinforce global power inequalities. Social media attention propagates to mainstream media and governments, and affects decisions regarding humanitarian aid.

Following the principle of equity, **we suggest an over-emphasis on minority and disadvantaged groups** during software development, instead of targeting a representative sample. After all, we should focus on those who rely most on humanitarian aid and crisis relief.

### 3.2. Misinformation

The impact that misinformation can have on societies became evident in the COVID-19 pandemic. As mentioned in [17], false social media rumors about a lockdown in the United States inviting civilians to stockpile certain products – a behavior that affects supply chains and causes demand-supply gaps [18]. While households with higher socioeconomic status are more able and likely to stockpile [19], low-income households are prone to food insecurity due to decreased availability and increased prices [20].

Misinformation comprises different forms of **intentionally and unintentionally false or inaccurate information** [21]: e.g., disinformation (intentional), rumors, fake news, urban legends, spamming and trolling. Through an analysis of ca. 126,000 news items shared via Twitter between 2006 and 2017, Vosoughi et al. [22] found false rumors transmitted “farther, faster, deeper, and more broadly than the truth in all categories of information” (p. 2). During the COVID-19 pandemic, inaccurate social media posts about infection prevention were shared more often than accurate posts [23]. Thus, misinformation can be an obstacle to establishing containment measures [17]. Moreover, it can unnecessarily trigger public fear. In the 2014 Ebola epidemic, a majority of the misinforming tweets exaggerated the spread and fatality of the disease [24].

Automated analysis systems must incorporate a mechanism that discards false information to **avoid its consolidation and dissemination**, as well as the conjuring of inappropriate coping measures. While the identification of misinformation is mainly done through expert coding, ML solutions are on the rise [25]. Detecting misinformation directly through visual or textual content is difficult. This is why some works also incorporate contextual information, like temporal patterns of posting (posts published in bursts), propagation through social networks, or hashtags [21].

### 3.3. Privacy

The availability of personal information on the web does not obviate the **need for unambiguous consent** regarding its collection and use through a third party [26]. Informed consent about third-party use is usually provided by the user as a prerequisite to using the platform. But even then, users might not be fully aware of what their consent entails [27]. Furthermore, the willingness to consent is subject to change and, according to GDPR law, the agreement is retractable by the user. Hence, published corpora – such as CrisisLex [28] – must be taken down or altered if users wish to remove their data. Yet, that is in practice hardly feasible: ML training corpora are downloaded, copied, and shared, ML models are trained on them, and certain posts

---

<sup>5</sup><https://www.npr.org/sections/goatsandsoda/2022/03/04/1084230259>

might be cited in publications. Direct quotes in public datasets may be traceable and allow identification of the author's identity.<sup>6</sup> The Chennai flood crisis mapping system [2] mentioned earlier geoparsed user posts to derive geographic coordinates from textual location mentions. Even though these were at the time posted to alert readers regarding events happening at certain places, it is questionable whether consent was provided for the type of post-processing done by the authors.

**During times of crisis, the data shared publicly on social media are especially personal.** The very content of crisis-related information spans from the date and location details, characterizations of individuals including names and imagery and reports about physical and psychological harms to utterances of grief and fear [13]. At trying times, people are at their most vulnerable. Collecting their crisis-related posts without dedicated consent and careful consideration, thus, severely violates individual privacy [12, 9].

### 3.4. Algorithmic Bias

Crises affect vulnerable groups disproportionately and biased automation systems risk exacerbating this dynamic. In the recent COVID-19 health crisis, for instance, the hasty development and non-peer-reviewed publishing of socially **biased algorithms have contributed to social inequalities** between black and white communities in the United States [29].

In the context of social media, sentiment classifiers have been used to determine the emotional state of the public during crises, estimate overall social impact, or filter individual posts for urgency [30, 31]. However, available sentiment analysis systems are to a large extent socially biased [32]. It was pointed out by Yang et al. [33] that this is one of the major bias-introducing factors in disaster informatics. Not only content but also dialect can yield bias [34]: for example, hate speech detectors may overestimate the toxicity of sentences in African-American Vernacular English [35]. Most large-scale language models like BERT [36], GPT [37], and their successors have been shown to reproduce a variety of biases and stereotypes [14]. Their extensive use across NLP tasks makes **social bias a general issue in this domain**.

Another source of bias identified by Yang et al. [33] is yielded by the positive correlation between socioeconomic status and tweet density [38]. Some crisis event detection systems consider temporal patterns, like bursts of tweets. Such systems might detect events disproportionately more in wealthier areas [33]. We earlier mentioned the privacy risk of geoparsing. Besides that, another ethical limitation arises from biases in the technologies used in geoparsing pipelines [33]. Named-entity recognition (NER) systems detect names of places, subjects, and the like in texts. These were found to exhibit binary gender bias, i.a. the accuracy with which female entities are identified was lowered. Yang et al. [33] suspect that NER systems might also be prone to other types of sociodemographic biases. Finally, the authors claim that CV systems for disaster characterization, like flood depth estimators [39], might yield disparate outcomes. That is because some of these systems use humans for scale reference. However, models for the recognition of human features output differently accurate results for different social groups [40, 41]. To avoid the disparate impact of biased algorithms, ML systems must undergo in-depth auditing, for example, via disaggregated performance evaluation [40].

---

<sup>6</sup><https://www.ucl.ac.uk/data-protection/sites/data-protection/files/using-twitter-research-v1.0.pdf>

### 3.5. Availability of Machine Learning Technologies

Modern ML technologies mostly serve developed countries [4], both in terms of availability and fit. As mentioned earlier, algorithms are often biased or inaccurate for whole demographic groups [14]. This in effect not only derogates people but also prevents them from benefiting from technological progress.

During our research, we noticed that non-English crisis corpora are not easily found [42]. The same counts for well-performing language models and other NLP applications. The **neglect of so-called “low-resource” languages** (language for which digital textual content is less available or has not been systematically gathered) is a widely discussed issue. 88% of all languages are completely ignored by NLP research, with no hope for change anytime soon [43] despite a growing amount of (European) platforms for NLP systems [44]. Those who suffer most from crises and would particularly benefit from support and prevention systems are least likely to be considered during development. With this, social inequalities are further reified.

### 3.6. Lack of Transparency

We have discussed different factors affecting the reliability of ML systems: unrepresentative and nonfactual data (from intentional and unintentional misinformation), algorithmic bias, and a lack of fit to most language or cultural regions. To complicate matters further, most ML and AI applications are **non-transparent decision makers**. So, irregularities are not easily spotted by non-technical personnel. The black-box nature of these complex models is a restricting factor, especially in a high-stakes crisis.

To improve the transparency of research and development, open-source and open-data practices have emerged. Public availability of training data facilitates scrutiny of a model’s potential limitations. However, this habit conflicts with the fact that social media data created during emergencies are particularly sensitive (see above). While reproducibility is certainly an important control mechanism, strict guidelines and compliance of practitioners are needed to ensure that heightened privacy needs are met.

Finally, we emphasize the need for explainable and interpretable ML methods. The inability to trace why a system suggests certain decisions **limits public control and legitimization** [9]. Authorities and civil persons should be able to comprehend the reasons behind algorithmic decisions – to act justifiably and not fall victim to an algorithmic fallacy. Hybrid AI methods - combining DL and Knowledge Graphs [45] - require less data, are explainable through their ground, and can therefore be used more effectively and efficiently in sensitive areas [46].

## 4. Future Directions

As claimed in [9], ML systems for crisis need “ethics with urgency”: (1) Ethical issues must be considered from the outset by foreseeing the system’s societal impact, (2) systems must be robust, and (3) public trust through independent oversight and transparency must be built. We suggest evaluating crisis technology as sociotechnical systems [10]: algorithms are embedded in social and political dynamics which pose ethical requirements to the data and algorithmic outcomes. Understanding the stakeholders’ values and needs, the long-term effects of the system

on society (and vice versa), and context-specific societal demands during a crisis becomes an essential element of software development.

Developers and researchers should consider where and how the data were collected and whose experiences and motivations they represent. “Datasheets for datasets” [47] can help guide in-depth examination of the data and shed light on potential risk factors. The resulting datasheet is an accompanying artifact to the developed system allowing for transparency and accountability later on. Similarly, we recommend the use of model cards [48] to transparently document how and on which data an ML model was developed and evaluated, what its technical and ethical limitations are, as well as its intended use. To circumvent disparate impact, data and algorithms should undergo bias auditing, for example through disaggregated performance analyses and the use of suitable fairness metrics. The choice of fairness metric again heavily relies on the social setting of the system [10]. We suggest putting disadvantaged groups at the focus of crisis technology to approach equity and help those particularly affected by crises. Moreover, we encourage examining whether or not a planned technical solution is appropriate in the given situation, to begin with, and avoid technosolutionism [10]. If ethical risks are inevitable, abandoning an idea must be considered as an option. All in all, given its context-specific nature, there is no standard solution for ethically developing crisis technology. This must be judged on a case-by-case basis.

## 5. Conclusion

ML-based analysis of social media streams can facilitate a swift aggregation and filtering of information during crises. This can support civilians, emergency responders, and authorities to cope quickly. However, the pairing of social media-sourced data and ML algorithms gives rise to several ethical risks. In this position paper, we addressed issues of representativeness, factuality, and privacy of social media data, ML algorithms’ proneness to reproduce bias, as well as their unavailability for many languages and cultures, and their lack of transparency. Furthermore, the vulnerability of social media users during crises is increased. This results in heightened ethical requirements for crisis informatics systems to secure people’s well-being. We conclude that the harms otherwise disproportionately affect already disadvantaged groups. Future work must focus on supporting these very groups to strive for equity.

To be able to fulfill the inherent goal of helping those in need, practitioners must examine all facets of the impact their software is going to have in the long run. Rapid development at the cost of ethics will else paradoxically defeat its purpose.

## Acknowledgments

The authors acknowledge the financial support by the Federal Ministry for Economic Affairs and Energy of Germany in the project CoyPu (project number 01MK21007[G]).

## References

- [1] L. Palen, Online social media in crisis events, *EDUCAUSE Quarterly Magazine* 31 (2008) 76 – 78. URL: <https://er.educause.edu/-/media/files/article-downloads/eqm08313.pdf>.
- [2] B. Anbalagan, V. Chinnaiah, #chennai floods: Leveraging human and machine learning for crisis mapping during disasters using social media, in: *HiPCW 2016*, IEEE Computer Society, 2016, pp. 50–59. doi:10.1109/HiPCW.2016.016.
- [3] C. Reuter, M.-A. Kaufhold, Fifteen years of social media in emergencies: A retrospective review and future directions for crisis informatics, *JCCM* 26 (2018) 41–57. doi:10.1111/1468-5973.12196.
- [4] L. Dwarakanath, A. Kamsin, R. A. Rasheed, A. Anandhan, L. Shuib, Automated machine learning approaches for emergency response and coordination via social media in the aftermath of a disaster: A review, *IEEE Access* 9 (2021) 68917–68931. doi:10.1109/ACCESS.2021.3074819.
- [5] R. I. Ogie, J. C. Rho, R. J. Clarke, Artificial intelligence in disaster risk communication: A systematic literature review, in: *ICT-DM, 2018*, pp. 1–8. doi:10.1109/ICT-DM.2018.8636380.
- [6] J. Liu, T. Singhal, L. T. Blessing, K. L. Wood, K. H. Lim, CrisisBERT: A robust transformer for crisis classification and contextual crisis embedding, in: *ACM HT, ACM*, 2021, pp. 133–141. doi:10.1145/3465336.3475117.
- [7] P. Dewan, A. Suri, V. Bharadhwaj, A. Mithal, P. Kumaraguru, Towards understanding crisis events on online social networks through pictures, in: *ASONAM '17, IEEE/ACM*, 2017, pp. 439–446. doi:10.1145/3110025.3110062.
- [8] M. Imran, F. Ofli, D. Caragea, A. Torralba, Using AI and social media multimodal content for disaster response and management: Opportunities, challenges, and future directions, *Information Processing & Management* 57 (2020) 102261. doi:10.1016/j.ipm.2020.102261.
- [9] A. Tzachor, J. Whittlestone, L. Sundaram, S. Ó. hÉigeartaigh, Artificial intelligence in a crisis needs ethics with urgency, *Nature Machine Intelligence* 2 (2020) 365–366. doi:10.1038/s42256-020-0195-0.
- [10] A. D. Selbst, D. Boyd, S. A. Friedler, S. Venkatasubramanian, J. Vertesi, Fairness and abstraction in sociotechnical systems, in: *FAT\* 2019, ACM*, 2019, pp. 59–68. doi:10.1145/3287560.3287598.
- [11] H. M. Zahera, R. Jalota, M. A. Sherif, A. N. Ngomo, I-AID: identifying actionable information from disaster-related tweets, *IEEE Access* 9 (2021) 118861–118870. doi:10.1109/ACCESS.2021.3107812.
- [12] K. Crawford, M. Finn, The limits of crisis data: analytical and ethical challenges of using social and mobile data to understand disasters, *GeoJournal* 80 (2015) 491–502. doi:10.1007/s10708-014-9597-z.
- [13] D. E. Alexander, Social media in disaster risk reduction and crisis management, *Science and Engineering Ethics* 20 (2014) 717–733. doi:10.1007/s11948-013-9502-z.
- [14] E. M. Bender, T. Gebu, A. McMillan-Major, S. Shmitchell, On the dangers of stochastic parrots: Can language models be too big? , in: *FAccT 2021, ACM*, 2021, pp. 610–623. doi:10.1145/3442188.3445922.
- [15] M. J. Landau, S. Solomon, J. Greenberg, F. Cohen, T. Pyszczynski, J. Arndt, C. H. Miller,



- D. M. Ogilvie, A. Cook, Deliver us from evil: The effects of mortality salience and reminders of 9/11 on support for President George W. Bush, *PSPB* 30 (2004) 1136–1150. doi:10.7282/T3NV9GMW.
- [16] D. DiFranzo, K. Gloria-Garcia, Filter bubbles and fake news, *XRDS: Crossroads, The ACM Magazine for Students* 23 (2017) 32–35. doi:10.1145/3055153.
- [17] S. Tasnim, M. M. Hossain, H. Mazumder, Impact of rumor and misinformation on COVID-19 in social media, *J. Prev. Med. Public Health* 53 (2020) 171–174. doi:10.3961/jpmph.20.094.
- [18] V. Sukhwani, S. Deshkar, R. Shaw, COVID-19 lockdown, food systems and urban–rural partnership: Case of nagpur, india, *IJERPH* 17 (2020) 5710. doi:10.3390/ijerph17165710.
- [19] M. O’Connell, Á. De Paula, K. Smith, Preparing for a pandemic: spending dynamics and panic buying during the COVID-19 first wave, *Fiscal Studies* 42 (2021) 249–264. doi:10.1111/1475-5890.12271.
- [20] S. Dasgupta, E. J. Robinson, Impact of COVID-19 on food insecurity using multiple waves of high frequency household surveys, *Scientific Reports* 12 (2022) 1–15. doi:10.1038/s41598-022-05664-3.
- [21] L. Wu, F. Morstatter, K. M. Carley, H. Liu, Misinformation in social media: Definition, manipulation, and detection, *ACM SIGKDD Explorations Newsletter* 21 (2019) 80–90. doi:10.1145/3373464.3373475.
- [22] S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online, *Science* 359 (2018) 1146–1151. doi:10.1126/science.aap9559.
- [23] J. Obiała, K. Obiała, M. Mańczak, J. Owoc, R. Olszewski, COVID-19 misinformation: Accuracy of articles about coronavirus prevention mostly shared on social media, *Health Policy and Technology* 10 (2021) 182–186. doi:10.1016/j.hlpt.2020.10.007.
- [24] T. K. Sell, D. Hosangadi, M. Trotochaud, Misinformation and the US Ebola communication crisis: analyzing the veracity and content of social media messages related to a fear-inducing infectious disease outbreak, *BMC Public Health* 20 (2020) 550. doi:10.1186/s12889-020-08697-3.
- [25] K. Hunt, P. Agarwal, J. Zhuang, Monitoring misinformation on Twitter during crisis events: A machine learning approach, *Risk Analysis* 00 (2020). doi:10.1111/risa.13634.
- [26] M. Zimmer, “But the data is already public”: On the ethics of research in Facebook, *Ethics Inf. Technol.* 12 (2010) 313–325. doi:10.1007/s10676-010-9227-5.
- [27] L. Hemphill, M. L. Hedstrom, S. H. Leonard, Saving social media data: Understanding data management practices among social media researchers and their implications for archives, *JASIST* 72 (2021) 97–109. doi:10.1002/asi.24368.
- [28] A. Olteanu, C. Castillo, F. Diaz, S. Vieweg, Crisislex: A lexicon for collecting and filtering microblogged communications in crises, in: *ICWSM 2014*, The AAAI Press, 2014. URL: <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8091>.
- [29] E. Rössli, B. Rice, T. Hernandez-Boussard, Bias at warp speed: how AI may contribute to the disparities gap in the time of COVID-19, *JAMIA* 28 (2021) 190–192. doi:10.1093/jamia/ocaa210.
- [30] H. J. Kaur, R. Kumar, Sentiment analysis from social media in crisis situations, in: *ICCCA*, 2015, pp. 251–256. doi:10.1109/CCAA.2015.7148383.
- [31] C. Zhang, W. Yao, Y. Yang, R. Huang, A. Mostafavi, Semiautomated social media analytics for sensing societal impacts due to community disruptions during disasters, *Comput.-Aided*

- Civ. Infrastruct. Eng. 35 (2020) 1331–1348. doi:10.1111/mice.12576.
- [32] S. Kiritchenko, S. Mohammad, Examining gender and race bias in two hundred sentiment analysis systems, in: \*SEM 2018, ACL, 2018, pp. 43–53. doi:10.18653/v1/S18-2005.
- [33] Y. Yang, C. Zhang, C. Fan, A. Mostafavi, X. Hu, Towards fairness-aware disaster informatics: an interdisciplinary perspective, IEEE Access 8 (2020) 201040–201054. doi:10.1109/ACCESS.2020.3035714.
- [34] S. L. Blodgett, B. O’Connor, Racial disparity in natural language processing: A case study of social media african-american english, CoRR abs/1707.00061 (2017). URL: <http://arxiv.org/abs/1707.00061>.
- [35] M. Sap, D. Card, S. Gabriel, Y. Choi, N. A. Smith, The risk of racial bias in hate speech detection, in: ACL, ACL, 2019, pp. 1668–1678. doi:10.18653/v1/P19-1163.
- [36] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: NAACL, ACL, 2019, pp. 4171–4186. doi:10.18653/v1/N19-1423.
- [37] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, Improving language understanding by generative pre-training, OpenAI blog (2018). URL: [https://cdn.openai.com/research-covers/language-unsupervised/language\\_understanding\\_paper.pdf](https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf).
- [38] L. Li, M. F. Goodchild, B. Xu, Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr, Cartogr. Geogr. Inf. Sci. 40 (2013) 61–77.
- [39] H. Bai, G. Yu, A Weibo-based approach to disaster informatics: incidents monitor in post-disaster situation via Weibo text negative sentiment analysis, Natural Hazards 83 (2016) 1177–1196.
- [40] J. Buolamwini, T. Gebru, Gender shades: Intersectional accuracy disparities in commercial gender classification, in: FAT\* 2018, PMLR, 2018, pp. 77–91. URL: <http://proceedings.mlr.press/v81/buolamwini18a.html>.
- [41] M. Du, F. Yang, N. Zou, X. Hu, Fairness in deep learning: A computational perspective, IEEE Intelligent Systems 36 (2021) 25–34. doi:10.1109/MIS.2020.3000681.
- [42] K. Rudra, N. Ganguly, P. Goyal, S. Ghosh, Extracting and summarizing situational information from the twitter social media during disasters, ACM Trans. Web 12 (2018) 17:1–17:35. doi:10.1145/3178541.
- [43] P. Joshi, S. Santy, A. Budhiraja, K. Bali, M. Choudhury, The state and fate of linguistic diversity and inclusion in the NLP world, in: ACL, ACL, 2020, pp. 6282–6293. doi:10.18653/v1/2020.acl-main.560.
- [44] G. Rehm, D. Galanis, P. Labropoulou, S. Piperidis, M. Weiß, R. Usbeck, J. Köhler, M. Deligianis, K. Gkirtzou, J. Fischer, C. Chiarcos, N. Feldhus, J. M. Schneider, F. Kintzel, E. Montiel-Ponsoda, V. Rodríguez-Doncel, J. P. McCrae, D. Laqua, I. P. Theile, C. Dittmar, K. Bontcheva, I. Roberts, A. Vasiljevs, A. Lagzdins, Towards an interoperable ecosystem of AI and LT platforms: A roadmap for the implementation of different levels of interoperability, in: IWLTP@LREC 2020, ELRA, 2020, pp. 96–107. URL: <https://aclanthology.org/2020.iwltlp-1.15/>.
- [45] A. Hogan, E. Blomqvist, M. Cochez, C. d’Amato, G. de Melo, C. Gutiérrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, A. N. Ngomo, A. Polleres, S. M. Rashid, A. Rula, L. Schmelzeisen, J. F. Sequeda, S. Staab, A. Zimmermann, Knowledge graphs, ACM Comput. Surv. 54 (2021) 71:1–71:37. doi:10.1145/3447772.

- [46] M. Ebrahimi, A. Eberhart, F. Bianchi, P. Hitzler, Towards bridging the neuro-symbolic gap: deep deductive reasoners, *Appl. Intell.* 51 (2021) 6326–6348. doi:10.1007/s10489-020-02165-6.
- [47] T. Gebru, J. Morgenstern, B. Vecchione, J. W. Vaughan, H. Wallach, H. D. III, K. Crawford, Datasheets for datasets, *Commun. ACM* 64 (2021) 86–92. doi:10.1145/3458723.
- [48] M. Mitchell, S. Wu, A. Zaldivar, P. Barnes, L. Vasserman, B. Hutchinson, E. Spitzer, I. D. Raji, T. Gebru, Model cards for model reporting, in: *FAT\* 2019*, ACM, 2019, p. 220–229. doi:10.1145/3287560.3287596.