Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

HITeC

COYPU

Angelie Kraft and Ricardo Usbeck

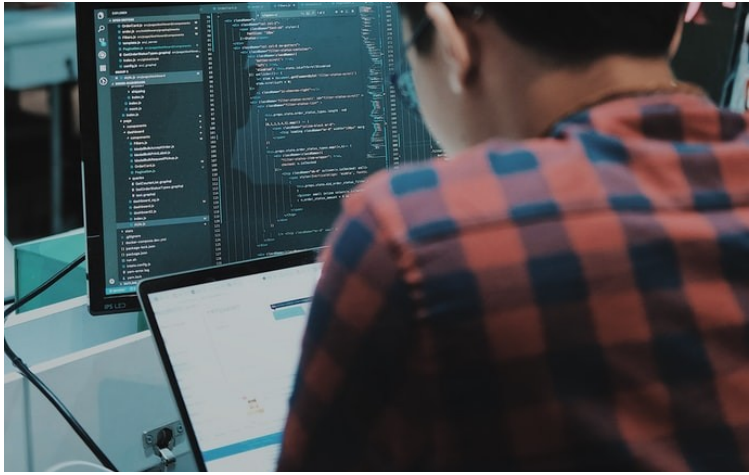
The Ethical Risks of Analyzing Crisis Events on Social Media with Machine Learning

Social Media as a Crisis Warning System

- Social media posts provide
  - Crisis characterizations,
  - Geo- and temporal information,
  - Emotional indicators,
  - Realtime updates
- Patterns like posting dynamics (e.g. bursts) are helpful to determine urgency

Machine Learning Use Cases

- NLP (Text):
  - Distinguish informative from uninformative posts
  - Classify the crisis type
  - Classify information type (e.g. warnings, utilities, or needs)
  - Classify sentiment (emotions)
- CV (Images):
  - Determine locations
  - Compute level of flooding



What Are Ethical Pitfalls?

- Representativeness
- Misinformation
- Privacy
- Algorithmic Bias
- Availability
- Transparency

Representativeness

- Twitter not among the top-10 most popular platforms but still most researched!
- Users
  - Mainly from U.S. or Japan
  - 38.5% aged 25 - 34, 21% aged 35 – 49
- Tweets stem from non-representative sample of people

(Information source: statista.com)

Representativeness

- Tweets echo particular groups' interests and opinions
- Amplification through echo chambers (DiFranzo & Gloria-Garcia, 2017. Filter bubbles and fake news.)
- Attention dynamics correspond to global power gradients
  - For example: More coverage of Ukraine crisis than genocide in Ethiopia (https://www.npr.org/sections/goatsandsoda/2022/03/04/1084230259)

Representativeness

- Instead of samples biased towards a niche group of young people from developed countries, we should research
  - A sample that is balanced across different attributes (ethnicity, age, socioeconomic status, educational backgeround, etc.)
  - In this context, we could alternatively consider over-emphasizing marginalized groups that are more affected by crises

Misinformation

- Intentional & unintentional spread of false or inaccurate information
- False rumors transmit “farther, faster, deeper, and more broadly than the truth in all categories of information” (Vosoughi et al., 2018. The spread of true and false news online.)

Misinformation

- Can be an obstacle to establishing containment measures
- Can trigger public fear

Example: Lockdown rumors in U.S. caused panic buying, leading to demand-supply gaps

(Tasnim et al., 2020. Impact of rumor and misinformation on COVID-19 in social media.)

Misinformation

- Must be detected and removed to avoid consolidation when modeled with ML algorithms
- Side note on *automation bias*: humans overestimate the truthfulness of algorithmic outputs (Mosier & Skitka, 2018. Human Decision Makers and Automated Decision Aids: Made for Each Other?)

Privacy

- Availability of personal information on the web (e.g. social media profiles) does not obviate the need for unambiguous consent
- Often users are not aware what their consent entails when clicking on „Agree“
  (Hemphill, 2021. Saving social media data: Understanding data management practices among social media researchers and their implications for archives.)
- GDPR requires data to be retractable: often not practical
  - public corpora quickly duplicated, sample texts quoted in publications
- Texts might allow retracing users via online search

Privacy

- During times of crisis, the data shared publicly on social media is especially personal
  - Location details
  - Description of individuals (incl. names, images, ...)
  - Descriptions of physical and psychological harm, emotions like grief and fear

Algorithmic Bias

- Numerous NLP models reproduce social biases
  - Sentiment classifiers consider statements about/by certain groups more likely as negative
  - Approaches based on large-scale foundation models are generally at risk of bias

(Bender et al., 2021. On the dangers of stochastic parrots: Can language models be too big?)

Algorithmic Bias

- CV applications for analyzing images of humans are also biased
  - E.g. perform worse for people of color and women
    (Buolamwini & Gebru, 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification.)

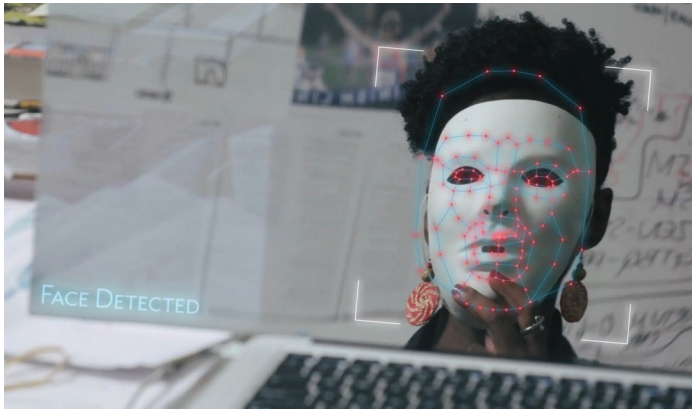


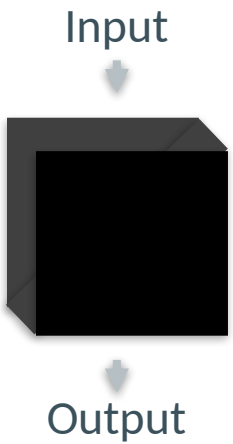
Image: Still from movie “Coded Bias”

Availability

- Training corpora and ML models are available only for a small set of languages
  - Neglect of "low-resource languages“
- Those who are disadvantaged to begin with & would particularly benefit from disaster prevention are not at the focus of innovation!
  - Reification of social inequalities

Transparency

- ML models are non-transparent decision makers
  - Irregularities, like bias or lack of factuality, are not easily spotted
  - Risky especially in high-stakes situation
- Explainable methods, open-source and open-data practices needed
  - But with hightened privacy efforts



Take away

- Crisis informatics can safe lives and economies & social media data analyzed with ML is effective and swift
- There are a number of risks on a data and algorithmic level that affect already disadvantaged people disproportionally
- In the context of social media & crisis, developers/researchers must be especially cautious to protect those that are vulnerable

Thank you for your attention!

Questions?

Image source if not otherwise stated: unsplash.com