No Humans Here: Ethical Speculation on Public Data, Unintended Consequences, and the Limits of Institutional Review

JESSICA PATER, Parkview Research Center, USA
CASEY FIESLER, University of Colorado, Boulder, USA
MICHAEL ZIMMER, Marquette University, USA

Many research communities routinely conduct activities that fall outside the bounds of traditional human subjects research, yet still frequently rely on the determinations of institutional review boards (IRBs) or similar regulatory bodies to scope ethical decision-making. Presented as a U.S. university-based fictional memo describing a post-hoc IRB review of a research study about social media and public health, this design fiction draws inspiration from current debates and uncertainties in the HCI and social computing communities around issues such as the use of public data, privacy, open science, and unintended consequences, in order to highlight the limitations of regulatory bodies as arbiters of ethics and the importance of forward-thinking ethical considerations from researchers and research communities.

CCS Concepts: • **Social and professional topics → Codes of ethics**.

Additional Key Words and Phrases: design fiction; institutional review boards; IRB; research ethics

1 AUTHOR'S INTRODUCTORY NOTES

> "Victor Frankenstein did not have the foresight that writing an IRB proposal might have forced him to summon, and so he was consumed by his scientific passion and curiosity to the point that he lost sight of his work's horizon—its potential risks and unintended consequences." [28]

Our computational capacity has greatly increased in the last 20 years, but new capabilities in fields like AI and machine learning are changing the nature of research and leaving significant gaps in both ethical norms and oversight [25, 53]. How often are researchers in this field, like Victor Frankenstein, "consumed by scientific passion" such that they might lose sight of potential risks?

In Harrison and Gannon's imaginary institutional review board (IRB) proposal based on the scientific research to animate life from the dead in Mary Shelley's 1818 novel, they use this fictional example to interrogate the benefits and limitations of the research ethics review process and how harmful outcomes might be averted by a scientist more carefully considering ethical obligations

38

Authors' addresses: Jessica Pater, jessica.pater@parkview.com, Parkview Research Center, IN, USA; Casey Fiesler, casey.fiesler@colorado.edu, University of Colorado, Boulder, CO, USA; Michael Zimmer, michael.zimmer@marquette.edu, Marquette University, WI, USA.

and possible consequences [28]. Similarly, HCI researchers have used methods from design fiction to encourage reflection about the potential downsides or social impacts of technology design and practice [47, 55], create fictional studies to provide space for critique [5, 36], and illustrate the importance of forward-thinking ethical considerations [17].

The recent prominence of discussions around unintended consequences in relation to research ethics [7, 30, 49] showcases the importance of speculation as one way to imagine if not anticipate potential harms [18]. In the fiction that follows, we first use creative speculation as a tool to demonstrate its own utility: if we can imagine a set of fictional "unintended" consequences of a research project, then might the researchers themselves be able to do the same?

Second, like the fictional Frankenstein proposal, we use an imaginary scenario to interrogate the boundaries of ethical regulation. The computing research community is struggling to form ethical norms, standards, or even best practices around work that does not fit traditional paradigms of human subjects research. Examples of these include the collection and use of public data, sharing and combining datasets, identifying harms beyond the individual subject, and consideration for potential downstream effects.

Despite the growing awareness of ethical issues within HCI and the potential limitations of IRBs to address them [20, 37, 53, 56, 59], both researchers and reviewers often pass responsibility onto ethical review bodies such as Institutional Review Boards (IRBs) in the United States. However, the purview of IRBs is typically constrained to the oversight of "human subjects research" as narrowly defined by government regulations. As a result, researchers might over-rely on this convenient external review process that might determine that a project is "not human subjects research" or "exempt"–and inappropriately conclude that the project therefore must not have ethical concerns. For example,in a recent high-profile example, a post-hoc IRB determination cleared a research study from the purview of their review, but many members of the community it impacted considered it to be harmful [11].

The fact that an IRB determines a project to be exempt does not mean that the project was deemed ethical - it simply means that it is exempt from continued review and monitoring by the review body [15]. Further, an IRB (or even an individual researcher's) determination that the project does not fit the federal definition of "human subjects research" provides no information about the potential ethical implications of the project. There could also be a contextual element to the IRB itself in that they may have the expertise to determine the immediate risk to the participant but may lack the domain context to understand more implicit or latent harms beyond the scope of their direct knowledge as Vitak et al. found when assessing IRBs and how they handle certain types of social computing research [52].

Presented as a U.S. university-based fictional memo describing a post-hoc IRB review of a research study about social media and public health, this design fiction draws inspiration from current debates and uncertainties in the HCI and social computing communities around issues such as the use of public data, privacy, open science, and unintended consequences, in order to highlight the limitations of regulatory bodies as arbiters of ethics and the importance of forward-thinking ethical considerations from researchers and research communities. Though the following illustrative example is fictional, it is inspired by real examples as well as current ethical debates within the HCI community.

2 FICTIONAL MEMO: IRB POST-HOC REVIEW OF PROTOCOL #460520

The Lincoln University IRB was asked to conduct this post-hoc review of protocol #460520, submitted by Dr. Smith. Dr. Smith was approached by the Jackson County Department of Health in April of 2020 to help identify COVID hotspots through the use of computational analysis of relevant public

data. It is a common practice for researchers to use public data sets for public health purposes [10, 13, 26, 31, 34] and specifically related to COVID-19 [16, 21, 41, 46].

Though Dr. Smith did not initially submit an IRB proposal for this work due to their own determination that it does not fall under IRB definitions of human subjects research (and/or that the exemption for public health surveillance may apply), reviewers of a manuscript based on the research asked for a statement about IRB approval. Therefore, Dr. Smith requested this post-hoc review process.

2.1 Background

Dr. Smith is a world-renowned machine learning expert. Specifically, their lab is known for creating sophisticated and highly accurate classifiers to infer identity traits (e.g., sexual orientation, political leanings, health conditions) [2, 12, 24, 33, 42] or emotional state [1, 35] from social media data. At the beginning of the COVID-19 pandemic, Dr. Smith's lab was approached by the Jackson County Department of Health. They wanted to know if the lab could help them identify COVID hotspots through the collection and analysis of public data. They were not sure what they were looking for or where they should be looking, but knew they wanted to be more proactive in their approach. Ideally, Dr. Smith's lab could help Jackson County officials target their limited coalition of resources to the locales that were in the greatest need as the response to this point had been *ad hoc* due to the strain on the small team.

The team from the county met with Dr. Smith's lab to discuss the community needs, the constrained resources, and the desire to move from a reactive to a proactive approach. Dr. Smith mentioned that their lab had been working with a public dataset from RoundHere. RoundHere is a neighborhood-based social platform that connects people and businesses based on location. Similar to other social media platforms like NextDoor, RoundHere allows users to establish a profile, send messages, and sell items on a marketplace, and has a news feed where you can view and interact with posts made by neighbors. Like other platforms, the username is a pseudonym created by the person who owns the account. The group discussed how they could use the API to pull data each week and then run the data through the classifier, looking specifically for discussions around COVID within the neighborhoods. By focusing more directly on acute symptoms and active phrases related to the contracting of COVID or the magnitude of dealing with a COVID diagnosis, they could potentially identify specific neighborhoods for the Jackson County Department of Health to target their limited resources in real-time.

2.2 The Study

The purpose of the study was to give public health officials a hyperlocal assessment of COVID within the neighborhoods of Jackson County, the second-largest metropolitan area in the state. The research question was: can COVID-related discussions on RoundHere be used to identify communities of need faster than traditional public health surveillance methods?

Dr. Smith's lab met several times in the months preceding the start of the project. As a lab, they discussed the project design extensively - mainly focusing on the refinement of the research questions and the data analysis approach. Dr. Smith's team determined the project did not constitute "human subjects research" as defined by the IRB because they were not engaging people directly and were following all the proper procedures per the site API to pull the data. Thus, no submission to the IRB was on file.

The researchers used the RoundHere API to pull data weekly to run through their classifier built to detect COVID. The RoundHere API allowed for the following data to be extracted: profile (username, neighborhood, household members, age), posts, comments and engagement, marketplace connections, and groups followed.

To detect COVID-related content, the researchers used terminology associated with symptoms, diagnoses, and emotive action language as they parsed the RoundHere data each week. These terms were derived from the NIH's National COVID Cohort Collaborative (NC3) Phenotype Data Acquisition. Hospitals from around the country submit a range of data surrounding COVID positive patients to the NC3 weekly. This includes criteria that would be queryable in most clinical common data models (CDMs), including OMOP, PCORnet, ACT, and TriNetX.

The research team used supervised learning methods, informed by the NC3 data, to analyze the data weekly, looking for patterns associated with COVID. They compared neighborhood results from the past several weeks, measuring increases in the volume of COVID discussions - both on the individual and community level. This approach was successful and was utilized by the Jackson County Public Health Department for over a year - allowing them to pinpoint where limited resources needed to be deployed.

Dr. Smith's team wrote up the research results for submission to the Annals of Machine Learning under the title: "A Real-time Machine Learning Approach to tracking COVID-19 at the Neighborhood Level." In parallel with this effort, a Master's student (C.M.) in Dr. Smith's lab took the lead on applying their existing identity and emotion inference models, which had been created based on Twitter and Reddit data, to this new dataset. Their master's thesis, titled "Undisclosed traits and emotional states predictable from local social media posts," was published on the open repository ArXiV along with the submitted COVID-related paper.

During the review of the "COVID-19 at the Neighborhood Level" paper, the journal asked that the dataset be published - a best practice in the field. The team agreed since the dataset only included publicly available information from the RoundHere API along with their COVID classifiers, and published the data to GitHub. Because the "Undisclosed traits" paper would also soon be under review, they included the identity and emotional state labels as a secondary dataset as well. The published datasets were posted to the r/Datasets Reddit page to increase their availability to the research community. A Reddit user asked whether the research went through a formal IRB review, and a member of Dr. Smith's team responded in a thread noting that "the data was already public, and thus not considered human subjects research."

"A Real-time Machine Learning Approach to Tracking COVID-19 at the Neighborhood Level" was published and won the journal's "Excellence in Research" award for the publication year 2020, garnering attention within the machine learning community. The lab currently has no ongoing projects related to this data collection. However, the "Undisclosed traits" student paper is currently under review at a social computing conference that meets biannually at a beach resort. Due to concerns raised by the reviewers, Dr. Smith has asked this IRB for a post-hoc review of the research for issues of human subjects compliance. It is possible that reviewer concerns were exacerbated by media attention for a series of events that followed the publication of the pre-print papers.

2.3 Subsequent Events

For the purposes of a post-hoc review, we must consider only the research as described above, as if the protocol had been submitted beforehand and without knowledge of what came to pass after the project had been completed. However, in the interests of transparency we will make note here of additional known outcomes of the research.

Following attention within the research community for the original research paper published in Annals of Machine Learning, a graduate student (J.L.) at another university was inspired by the paper. This student had recently developed a novel classifier that could predict political party affiliation based on latent language in general posts, and they needed a large dataset from another platform to validate the model. It also was helpful that it was an election year, and in addition to COVID, political discussions were plentiful within the RoundHere data, allowing for secondary
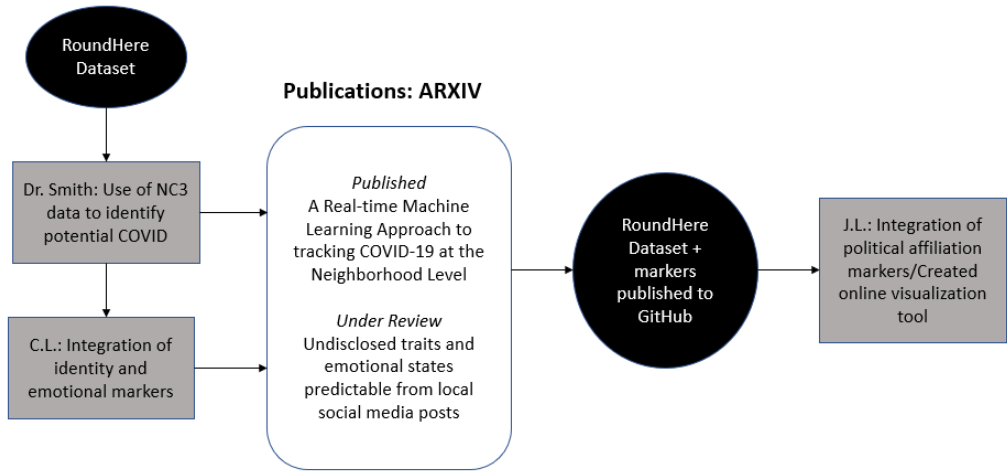
Fig. 1. How data was used and the outputs from the data

validation of the model outputs. In addition to using the public dataset released by Dr. Smith's lab to validate their model (finding that it performed better than the classifier from Dr. Smith's lab for political leanings), the student also created an online map-based visualization that allowed people to drill down into the neighborhood level to look at predicted political party affiliations. Figure 1 shows the process in which data was utilized and the products created from the data.

When the visualization tool was posted to the r/dataisbeautiful subreddit, the post received a great deal of engagement, particularly after it made it to Reddit's front page. As the events described below unfolded, J.L. eventually responded to a number of questions about the ethical implications of this tool, with the following justification: (1) they did not create the original dataset; (2) all of the data was publicly available and anyone could have done these visualizations for themselves; (3) using an existing dataset was the quickest way to get out another first-author paper before their job applications were submitted; and (4) the topic (politics) and the visualization tool increased the chance of the work getting media attention, which was also important because they were on the job market.

The attention from the Reddit post resulted in additional national press attention for both the tool and the original dataset and research. From the inquiries sent to our office from both reporters and the general public, we are aware of the following events:

1. Another Reddit user extended J.L.'s code (published to GitHub) to create a new visualization that included not only predicted party affiliations, but all of the attributes included in Dr. Smith's lab's original research and dataset from C.M.'s "Undisclosed traits" paper. With this new tool, anyone could look at neighborhood-level predictions for percentages of gender, race, sexual orientation, educational level, mental health conditions, and emotional states.

2. A group of users on the 8chan message board began digging into the original dataset, which included post content along with neighborhood-level geographic information and usernames, as well as labels from the identity inference models. They created a list of RoundHere users with certain combinations of identity attributes (e.g. sexual orientation and political orientation) and were able to re-identify individuals by connecting usernames to profiles on other platforms, resulting in

"doxing" of a significant number of RoundHere users.

3. A number of journalists identified individual RoundHere users in the data (by messaging them on RoundHere via their usernames) and interviewed them as part of news items about the research and the doxing. A number of users were unaware of the controversy or their presence in the dataset until they were contacted by journalists.

4. A number of bloggers, pundits, and news outlets with specific agendas began to use the combination of COVID and identity-based predictions to make sweeping assumptions about certain protected groups. Some businesses also began to refuse service to neighborhoods that the model identified as having high rates of COVID infections.

5. Following this national attention, RoundHere users began to file complaints both to the platform and to this board and J.L.'s university. RoundHere released a public statement noting they are investigating whether Dr. Smith's data collection was a violation of their terms of service, and threatened possible legal action via the Computer Fraud and Abuse Act. Several users have also filed defamation lawsuits based on identify labels applied to their accounts in the dataset. This office has received over 100 complaints from RoundHere users about being part of research without their consent.

6. There is currently a widely circulated petition to the White House with over 100,000 signatures citing an erosion of public trust in scientists and asking for tighter federal research ethics regulations, citing Dr. Smith's research and ensuing events as one example.

2.4   Review & Determination

It is our understanding that the media attention for this research occurred after the "Undisclosed traits" paper had been submitted, but that it is likely that reviewers are aware of the evolving situation. Though Dr. Smith informed us after we were asked by the lead author C.M. to provide this determination that they plan to withdraw the paper from consideration, we feel that it is appropriate to provide this review given the number of requests this office has received for comment. We also felt as if these requests and their context obligated us to comment on the events that unfolded after the research had been conducted despite the fact that, as noted previously, our review is based on how the research would have been described to us in a pre-project proposal.

As a reminder, the scope of our review is framed by the Federal Policy for the Protection of Human Subjects (45 CFR §46) – commonly known as the "Common Rule" – the baseline standard of research ethics by which nearly all U.S. academic institutions hold their researchers. According to §46.101:

> [T]his policy applies to all research involving human subjects conducted, supported, or otherwise subject to regulation by any Federal department or agency that takes appropriate administrative action to make the policy applicable to such research... and institutional review boards (IRBs) reviewing research that is subject to this policy must comply with this policy.

Definitions provided in 45 CFR §46.102 (quoted here verbatim) further clarify the meaning of this rule:

- *Research* means a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge.

- *Human subject* means a living individual about whom an investigator (whether professional or student) conducting research: (1) Obtains information or biospecimens through intervention or interaction with the individual, and uses, studies, or analyzes the information or biospecimens; or (2) Obtains, uses, studies, analyzes, or generates identifiable private information or identifiable biospecimens
- *Private information* includes information about behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place, and information that has been provided for specific purposes by an individual and that the individual can reasonably expect will not be made public (e.g., a medical record).
- *Identifiable private information* is private information for which the identity of the subject is or may readily be ascertained by the investigator or associated with the information.

First, we note that the research underlying both the COVID tracking and the undisclosed traits papers did constitute systematic investigation to contribute to generalizable knowledge. Though the initial research did have a component specific to the organizational functions of the Jackson County Health department specifically, the conclusions of the two papers confirm that Dr. Smith and his students intended to make claims about predictive power of RoundHere data beyond just that local context. However, in order to fall under IRB jurisdiction, the project must not be just research, but *human subjects* research.

Therefore, second, we conclude that the research does not involve an intervention or interaction with individuals, nor was identifiable private information collected. The data used by Dr. Smith's lab was openly visible online and publicly available via RoundHere's API without a need for interacting with individual research subjects. Moreover, due to the use of pseudonyms and the lack of other identifying information in the dataset (e.g., birthdates), users' identities could not be "readily ascertained" and therefore the data was also not identifiable. We recognize that third parties were indeed able to identify some individuals in the data, however, we contend that the measures they went to do do so do not constitute "readily" and also note that in the research actually described and conducted by Dr. Smith, no such measures were planned or taken. Additionally, according to the definition, data must be both identifiable and private. RoundHere users were not interacting in a space where they could reasonably expect that observation and recording was not taking place (and indeed, RoundHere's privacy policy makes this very clear) and therefore the information in the dataset was not "private."

As a result, if Dr. Smith's research had been submitted to this IRB in advance of the project starting, our determination would have been that it does not fit the regulatory definition of "human subjects research" and therefore is outside our review authority. This aligns with Dr. Smith's own determination prior to launching the project. Further, based on our available information, Dr. Smith followed standard research best practices throughout the project, including adhering to the rules of RoundHere's publicly available API and releasing their dataset without any personally identifiable information. We do not see evidence of any violation of the university's academic integrity policies.

We understand this might be an unsatisfactory determination given the subsequent events that have occurred since the research has been completed and published. The committee spent considerable time discussing the outcomes of the research at each step, assessing the assumptions taken, the responsibilities of the primary researcher for secondary analyses, the ethical implications, and the role of this regulatory body with respect to these. However, federal regulations strongly discourage IRBs from considering concerns about potential long-term social implications when reviewing research protocols. Thus, even if we were able to foresee these subsequent events, we would be required to remove them from our official determination. Further, university policy restricts the IRB from providing general ethical advice outside of our limited mandate of enforcing

the federal research guidelines. We regret we cannot provide additional ethical guidance to the research community regarding this matter.

3 AUTHORS' CONCLUDING REMARKS

We present this design fiction – a speculative IRB post-hoc review – to help the HCI and social computing communities consider the limitations of relying on IRB reviews as a sufficient proxy for broader ethical review, as well as more generally the potential downstream consequences and unintended uses of our research, particularly as related to the artifacts that we create (e.g. datasets, online tools). To do this, we have proposed to connect the faceless "public data" to the person. We drew inspiration from recent controversies such as the OKCupid dataset that was released without anonymization [57], researchers experimenting with the insertion of malicious code in the Linux kernel [11], and similar cases where one's research negatively impacts a community or is built upon by others in unexpected - and perhaps unapproved - ways.

This example draws attention to different types of harms and unintended consequences in HCI and social computing research. In this example, we are considering the various layers of harm similar to those found within ecological models [6, 22].

- Individual harms: people who were doxed, people who were upset to find out they were part of a research study without their consent [19, 38], individuals who were harmed by seeing the labels on their identity or emotional state [1, 48]
- Community harms: the online neighborhood communities on RoundHere, demographic groups and neighborhoods that were inappropriately singled out through analyses
- Institutional harms: trust in science, trust in the university, trust in ethical review, and trust in the RoundHere platform

It is also worth noting that when research ethics controversies happen, there is often reputational harm to the researchers, including graduate students. These situations are not necessarily a moral failure on the part of the researcher, but rather indicative of a lack of training, guidance, and clear norms. Next, we point to some of these pain points and reflect on how we might improve them.

Our fictional example draws particular attention to unintended consequences in the context of curated datasets based on "public" data and how humans are labeled and represented in data [48] and the trustworthiness of this data [43]. Public datasets, whether released by a company or published by a researcher, are important for many reasons, including transparency in the interests of open science, the removal of barriers and increasing visibility and publicity of work [54]. However, the scenario outlined here highlights several of the risks associated with public datasets. So if public datasets are important but also potentially problematic, how do we address this tension between openness and privacy/protection? One example of a best practice is to publish the dataset, but with permission. It provides the researcher to vet what the user is wanting to do with the data and offers a point of reflection [58]. However, if this system is abused it could lead to gate keeping - a serious issue within the scientific community [45].

Of course, this tension between open science and privacy is not the only value tension within the social computing research community. For a number of years there has been a recognition of a lack of strong norms around issues such as what constitutes "public" data, when consent is appropriate or necessary, and whether content that is sensitive or comes from vulnerable populations should be handled differently [19, 53]. However, there have been strong attempts to support norm setting and community standards, including research ethics workshops at conferences like GROUP [4, 20, 39], the work of the SIGCHI research ethics committee [32], and calls for changes in reviewing processes [30]. More recently in the AI research community, the NeurIPS conference instituted a requirement that all papers include a statement of the "potential broader impact of their work, including its

ethical aspects and future societal consequences" and provided guidance for authors on how to do so, including highlighting uncertainties [23].

From the review board perspective, there are arguments for more community presence and involvement among research review in order to prevent harms to communities. For example community advisory boards (CABs) within tribal nations [3] or low income or low resource settings [44]. CABs provide a formalized way for community members voice concerns, influence priorities based on institutional knowledge, and advise the research team on processes that are respectful and acceptable within the community. Social computing researchers have also noted the ethical importance of understanding the norms and practices of communities studied, rather than just "parachuting in" and taking data without giving anything back to the community [14].

There are also concerns about IRB "mission creep" that focus largely on its role for compliance, e.g. "focusing more on procedures and documentation than difficult ethical questions" and "efforts to protect against lawsuits" [27]. As a consequence, there is some concern about the expansion of IRBs in terms of misdirecting their energies and drawing resources from intervening in the riskiest research [27]. It is therefore important to keep in mind that, according to the U.S. Code of Federal Regulations, an "IRB should not consider possible long-range effects of applying knowledge gained in the research (for example, the possible effects of the research on public policy) as among those research risks that fall within the purview of its responsibility." The further assessment of long-term, potentially speculative harms to populations could not only put additional strain on this limited resource, but also could have a chilling effect on research "while simultaneously distracting an IRB from the important oversight issues that do fall under its purview" [29].

With respect to the computing domain, researchers may not see how their research aligns with other human subjects research and the ethical concerns related to those that work directly with human subjects. As noted by Buchanan et al. in their discussion of the role of human subjects review in security research, "CS research, broadly conceived, is concerned with systems and processes, and the connection to individuals may not be directly evident. The connection between a disparate piece of data and a human subject may, indeed, be a stretch" [8]. Part of this disconnect is the doubt that a connection can easily be made between disparate data and an individual. However, as technology advances, what we thought once impossible becomes fully realized and thus nullifying the skepticism. An example of this is the ability to conduct a reverse image search. This may require a fundamental change in the definition of a human subject or ethics reviews be handled by a third party that is not the IRB [8]. For example, assistance from research-community-based groups not mandated by regulatory statute could potentially provide a more safe space for researchers. The primary point that we hope our fictional study conveys is that IRBs are limited in scope and therefore cannot be relied upon to be the arbiters of ethics for all research projects; just because something does not fall under the definition of human subjects research does not mean that there is not potential for harm–to individuals, to communities, or to society. Therefore, it is imperative that individual researchers, even when their work does not involve directly interacting with people, are considering possible harms and implications, and that our research communities are doing everything they can to help support those researchers in these considerations.

Finally, this design fiction was written from a U.S. university perspective. How would the outcomes of this research been different if the lab was located within Europe, thus falling under GDPR regulations? What if they were located within an industry where researchers and the research were located in several countries? For example, with respect to the data processing that was presented in this design fiction, Article 6 of the GDPR states this is lawful under a certain set of conditions that this study might not align with. Underlying the processing of the data is the assumption that we have the ability to understand and assess the potential adverse implications for the rights of the data subjects [40]. This design fiction strikes at the heart of this assumption, as

the culminating issue was a secondary use to the initial research which was facilitated by current industry norms - i.e. the sharing of deidentified datasets with publications.

4 CONCLUSION

Like other design fictions that make use of fictional research scenarios to imagine possible harms or social impacts [17, 36, 55], we have constructed a scenario intended to tease out vulnerabilities in our current practices and to encourage forward-thinking ethical considerations. Moreover, this exercise in itself was an example of how researchers and technologists might imagine the possible downstream consequences of well-intentioned work–because we did exactly that.

Through our reflections we do not intend to place blame on any one party, nor do we present this as a call-to-action for reconfigurations of ethics review or a need to redefine "human subjects." We also recognize that while this design fiction focuses on a US-focused regulatory environment, the broader concerns apply more broadly to our global research community. Indeed, beyond one's geographic location, we suggest that researchers need to attend to a broader set of ethical questions regardless of what their local regulatory frameworks prescribe.

It is our intention that the social computing and HCI community reflect on the evolution of events in the design fiction and the subsequent IRB determination. What other actions or processes could have resulted from this situation? What protections should the researchers at the various levels have taken? What should our field embrace outside of the IRB review to ensure we are considering the overall ethical nature of our research as well as the potential downstream impacts? As raised at a community panel in 2019 [9], are there research subjects we should not be engaged with because of the potential downstream negative effects? At what point does downstream use of one's data or research move so far afield that it is no longer under the realm of responsibility for the researcher generating it? We can look to parts of our community for potential methods or pathways forward. For example, the participatory design and co-design communities' use of institutioning (engagement with institutions) and commoning (engagement with grassroots communities) to start the process of challenging the status quo [50] and developing ways reflect between common practice and various institutions [51].

As computational applications continue to evolve how the data or research can be further manipulated, remixed, or transformed, our understanding and best practices of ethical treatment and approaches will also need to evolve. Finally, people and institutions are imperfect. A different decision at any point in the process presented would have changed the overall outcome. The beauty of the design fiction mechanism is it allows the reader to explore where in this scenario the researchers and IRB could have made a different decision or could have done better.

5 ACKNOWLEDGEMENTS

REFERENCES

[1] Nazanin Andalibi and Justin Buss. 2020. The Human in Emotion Recognition on Social Media: Attitudes, Outcomes, Risks. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3313831.3376680

[2] Nazanin Andalibi, Pinar Rzturk, and Andrea Forte. 2017. Sensitive Self-Disclosures, Responses, and Social Support on Instagram: The Case of #Depression. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 1485–1500. https://doi.org/10.1145/2998181.2998243

[3] Jyoti Angal, Julie M. Petersen, Deborah Tobacco, and Amy J. Elliott. 2016. Ethics review for a multi-site project involving tribal nations in the northern plains. *Journal of Empirical Research on Human Research Ethics* 11, 2 (2016), 91–96.

[4] Pernille Bjorn, Casey Fiesler, Michael Muller, Jessica Pater, and Pamela Wisniewski. 2018. Research Ethics Town Hall Meeting. In *Proceedings of the 2018 ACM Conference on Supporting Groupwork* (Sanibel Island, Florida, USA) *(GROUP '18)*. Association for Computing Machinery, New York, NY, USA, 393–396. https://doi.org/10.1145/3148330.3154523

[5] Mark Blythe. 2014. Research through Design Fiction: Narrative in Real and Imaginary Abstracts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 703–712. https://doi.org/10.1145/2556288.2557098

[6] Urie Brofenbrenner. 1992. *Ecological systems theory*. Jessica Kingsley Publishers.

[7] Amy Bruckman. 2020. 'Have You Thought About…': Talking about Ethical Implications of Research. *Commun. ACM* 63, 9 (Aug. 2020), 38–40. https://doi.org/10.1145/3377405

[8] Elizabeth Buchanan, John Aycock, Scott Dexter, David Dittrich, and Erin Hvizdak. 2011. Computer science security research and human subjects: Emerging considerations for research ethics boards. *Journal of Empirical Research on Human Research Ethics* 6, 2 (2011), 71–83. https://doi.org/10.1525/jer.2011.6.2.71

[9] Stevie Chancellor, Shion Guha, Jofish Kaye, Jen King, Niloufar Salehi, Sarita Schoenbeck, and Elizabeth Stowell. 2019. The Relationships between Data, Power, and Justice in CSCW Research. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing* (Austin, TX, USA) *(CSCW '19)*. Association for Computing Machinery, New York, NY, USA, 102–105. https://doi.org/10.1145/3311957.3358609

[10] Lauren E. Charles-Smith, Tera L. Reynolds, Mark A. Cameron, Mike Conway, Eric HY Lau, Jennifer M. Olsen, and Julie A. Pavlin. 2015. Using social media for actionable disease surveillance and outbreak management: a systematic literature review. *PloS one* 10, 10 (2015), e0139701. https://doi.org/10.1371/journal.pone.0139701

[11] Monica Chin. 2021. How a University got Itself Banned from the Linux Kernel. (2021). https://www.theverge.com/2021/4/30/22410164/linux-kernel-university-of-minnesota-banned-open-source

[12] Munmun De Choudhury, Scott Counts, and Eric Horvitz. 2013. Social Media as a Measurement Tool of Depression in Populations. In *Proceedings of the 5th Annual ACM Web Science Conference* (Paris, France) *(WebSci '13)*. Association for Computing Machinery, New York, NY, USA, 47–56. https://doi.org/10.1145/2464464.2464480

[13] Kerstin Denecke, Peter Dolog, and Pavel Smrz. 2012. Making Use of Social Media Data in Public Health. In *Proceedings of the 21st International Conference on World Wide Web* (Lyon, France) *(WWW '12 Companion)*. Association for Computing Machinery, New York, NY, USA, 243–246. https://doi.org/10.1145/2187980.2188019

[14] Brianna Dym and Casey Fiesler. 2020. Ethical and privacy considerations for research using online fandom data. *Transformative Works and Cultures* 33 (2020).

[15] Elisabeth S. Faller and Claudia Norman. 2015. *Ethics & the IRB review process: A guide for SoTL researchers at UC*. https://www.uc.edu/content/dam/uc/cetl/docs/IRB|%20for|&20SoTL.pdf

[16] S. Fazeli, D. Zamanzadeh, A. Ovalle, T. Nguyen, and G. Gee adn M. Sarrafzadeh. 2021. COVID-19 and Big Data: Multi-faceted Analysis for Spatio-temporal Understanding of the Pandemic with Social Media Conversations. *arXiv preprint arXiv:2104.10807* (2021).

[17] Casey Fiesler. 2019. Ethical Considerations for Research Involving (Speculative) Public Data. *Proc. ACM Hum.-Comput. Interact.* 3, GROUP, Article 249 (Dec. 2019), 13 pages. https://doi.org/10.1145/3370271

[18] Casey Fiesler. 2021. Innovating Like an Optimist, Preparing Like a Pessimist: Ethical Speculation and the Legal Imagination. *Colorado Technology Law Journal* 19, 1 (2021).

[19] Casey Fiesler and Nicholas Proferes. 2018. "Participant" perceptions of Twitter research ethics. *Social Media+Society* 4, 1 (2018).

[20] Casey Fiesler, Pamela Wisniewski, Jessica Pater, and Nazanin Andalibi. 2016. Exploring Ethics and Obligations for Studying Digital Communities. In *Proceedings of the 19th International Conference on Supporting Group Work* (Sanibel Island, Florida, USA) *(GROUP '16)*. Association for Computing Machinery, New York, NY, USA, 457–460. https://doi.org/10.1145/2957276.2996293

[21] Max Fleming, Priyanka Dondeti, Caitlin N. Dreisbach, and Adam Poliak. 2021. Fine-Tuning Transformers for Identifying Self-Reporting Potential Cases and Symptoms of COVID-19 in Tweets. *arXiv preprint arXiv:2104.05501* (2021).

[22] Centers for Disease Control and Prevention. 2015. The social-ecological model: A framework for prevention. (2015). https://www.cdc.gov/violenceprevention/about/social-ecologicalmodel.html

[23] Centere for the Governance of AI. 2020. A Guide to Writing the NeurIPS Impact Statement. (2020). https://medium.com/@GovAI/a-guide-to-writing-the-neurips-impact-statement-4293b723f832

[24] Eduard Fosch-Villaronga, A. Poulsen, Roger Andre Søraa, and B. H. M. Custers. 2021. A little bird told me your gender: Gender inferences in social media. *Information Processing & Management* 58, 3 (2021), 102541. https://doi.org/10.1016/j.ipm.2021.102541

[25] Phoebe Friesen, Rachel Douglas-Jones, Mason Marks, Robin Pierce, Katherine Fletcher, Abhishek Mishra, Jessica Lorimer, Carissa Véliz, Nina Hallowell, Mackenzie Graham, Mei Sum Chan, Huw Davies, and Taj Sallamuddin. 2021. Governing AI-Driven Health Research: Are IRBs Up to the Task? *Ethics & human research* 43, 2 (2021), 35–42. https://doi.org/10.1002/eahr.500085

[26] Venkata Rama Kiran Garimella, Abdulrahman Alfayad, and Ingar Weber. 2016. Social Media Image Analysis for Public Health. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5543–5547. https://doi.org/10.1145/2858036.2858234

[27] C. Kristina Gunsalus, Nicholas C. Burbules Edward M. Bruner, Leon Dash, Matthew Finkin, Joseph P. Goldberg, and William T. Greenough. 2007. The Illinois white paper: Improving the system for protecting human subjects: Counteracting IRB "mission creep". *Qualitative Inquiry* 13, 5 (2007), 617–649.

[28] Gary Harrison and William L. Gannon. 2015. Victor Frankenstein's Institutional Review Board Proposal, 1790. *Science and Engineering Ethics* 21, 5 (2015), 1139–1157. https://doi.org/10.1007/s11948-014-9588-y

[29] Molly Havard and David Magnus. 2011. Beyond the IRB: Local Service Versus Global Oversight. *The American Journal of BioEthics* 11, 5 (2011), 1–2. https://doi.org/10.1080/15265161.2011.576932

[30] Brent Hecht. 2018. It's time to do something: Mitigating the negative impacts of computing through a change to the peer review process. (2018). https://acm-fca.org/2018/03/29/negativeimpacts/

[31] Liaquat Hossain, D. Kam, F. Kong, R. T. Wigand, and Terence Bossomaier. 2016. Social media in Ebola outbreak. *Epidemiology & Infection* 144, 10 (2016), 2136–2143.

[32] Matthew Hutson. 2021. Who Should Stop Unethical A.I.? (2021). https://www.newyorker.com/tech/annals-of-technology/who-should-stop-unethical-ai

[33] Michal Kosinski. 2021. Facial recognition technology can expose political orientation from naturalistic facial images. *Scientific reports* 11, 1 (2021), 1–7. https://doi.org/10.1038/s41598-020-79310-1

[34] Yubo Kou, Xinning Gui, Yunan Chen, and Kathleen Pine. 2017. Conspiracy Talk on Social Media: Collective Sensemaking during a Public Health Crisis. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 61 (Dec. 2017), 21 pages. https://doi.org/10.1145/3134696

[35] Adam DI Kramer, Jamie E. Guillory, and Jeffrey T. Hancock. 2014. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences* 111, 24 (2014), 8788–8790. https://doi.org/10.1073/pnas.1320040111

[36] Joseph Lindley and Paul Coulton. 2016. Pushing the Limits of Design Fiction: The Case For Fictional Research Papers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 4032–4043. https://doi.org/10.1145/2858036.2858446

[37] Cosmin Munteanu, Heather Molneaux, Wendy Moncur, Mario Romero, Susan O'Donnell, and John Vines. 2015. Situational Ethics: Re-Thinking Approaches to Formal Ethics Requirements for Human-Computer Interaction. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 105–114. https://doi.org/10.1145/2702123.2702481

[38] Jessica Pater, Oliver L. Haimson, Nazanin Andalibi, and Elizabeth D. Mynatt. 2016. "Hunger Hurts but Starving Works": Characterizing the Presentation of Eating Disorders Online. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (San Francisco, California, USA) *(CSCW '16)*. Association for Computing Machinery, New York, NY, USA, 1185–1200. https://doi.org/10.1145/2818048.2820030

[39] Jessica Pater, Alicia Nobles, Sarah Gilbert, Michael Zimmer, and Casey Fiesler. 2020. Data is People: New Challenges and Tensions in Social Media Research. In *14TH International Conference on Web and Social Media (ICWSM)*. AAAI.

[40] EU: General Data Protection Regulation. 2021. Chapter 2, Article 6: Data Processing. (2021). https://gdpr-info.eu/art-6-gdpr/

[41] Abeed Sarker, Sahithi Lakamana, Whitney Hogg-Bremer, Angel Xie, Mohammed Ali Al-Garadi, and Yuan-Chi Yang. 2020. Self-reported COVID-19 symptoms on Twitter: an analysis and a research resource. *Journal of the American Medical Informatics Association* 27, 8 (2020), 1310–1315.

[42] Judy Shen and Frank Rudzicz. 2017. Detecting anxiety through reddit. In *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology—From Linguistic Signal to Clinical Reality*. 58–65.

[43] Katie Shilton, Emanuel Moss, Sarah A. Gilbert, Matthew J. Bietz, Casey Fiesler, Jacob Metcalf adn Jessica Vitak, and Michael Zimmer. 2021. Excavating awareness and power in data science: A manifesto for trustworthy pervasive data research. *Big Data & Society* 8, 2 (2021).

[44] Kafuruki Shubis, Omar Juma, Rahiya Sharifu, Brandt Burgess, and Salim Abdulla. 2009. Challenges of establishing a Community Advisory Board (CAB) in a low-income, low-resource setting: experiences from Bagamoyo, Tanzania. *Health Research Policy and Systems* 7, 1 (2009), 1–5.

[45] Kyle Siler, Kirby Lee, and Lisa Bero. 2015. Measuring the effectiveness of scientific gatekeeping. *Proceedings of the National Academy of Sciences* 112, 2 (2015), 360–365.

[46] Ruba Skaik and Diana Inkpen. 2020. Using Social Media for Mental Health Surveillance: A Review. *ACM Comput. Surv.* 53, 6, Article 129 (Dec. 2020), 31 pages. https://doi.org/10.1145/3422824

[47] Robert Soden, Michael Skirpan, Casey Fiesler, Zahra Ashktorab, Eric P.S. Baumer, Mark Blythe, and Jasmine Jones. 2019. CHI4EVIL: Creative Speculation on the Negative Impacts of HCI Research. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3290607.3299033

[48] Eric P.S. Baumer Stevie Chancellor and Munmun De Choudhury. 2019. Who is the "Human" in Human-Centered Machine Learning: The Case of Predicting Mental Health from Social Media. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 147 (Nov. 2019), 32 pages. https://doi.org/10.1145/3359249

[49] Harini Suresh and John V. Guttag. 2019. A framework for understanding unintended consequences of machine learning. *arXiv preprint arXiv:1901.10002* (2019).

[50] Maurizio Teli, Marcus Foth, Mariacristina Sciannamblo, Irina Anastasiu, and Peter Lyle. 2020. Tales of Institutioning and Commoning: Participatory Design Processes with a Strategic and Tactical Perspective. In *Proceedings of the 16th Participatory Design Conference 2020 - Participation(s) Otherwise - Volume 1* (Manizales, Colombia) *(PDC '20)*. Association for Computing Machinery, New York, NY, USA, 159–171. https://doi.org/10.1145/3385010.3385020

[51] Maurizio Teli, Peter Lyle, and Mariacristina Sciannamblo. 2018. Institutioning the Common: The Case of Commonfare. In *Proceedings of the 15th Participatory Design Conference: Full Papers - Volume 1* (Hasselt and Genk, Belgium) *(PDC '18)*. Association for Computing Machinery, New York, NY, USA, Article 6, 11 pages. https://doi.org/10.1145/3210586.3210590

[52] Jessica Vitak, Nicholas Proferes adn Katie Shilton, and Zahra Ashktorab. 2017. Ethics regulation in social computing research: Examining the role of institutional review boards. *Journal of Empirical Research on Human Research Ethics* 12, 5 (2017), 372–382. https://doi.org/10.1177/1556264617725200

[53] Jessica Vitak, Katie Shilton, and Sahra Ashktorab. 2016. Beyond the Belmont Principles: Ethical Challenges, Practices, and Beliefs in the Online Data Research Community. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (San Francisco, California, USA) *(CSCW '16)*. Association for Computing Machinery, New York, NY, USA, 941–953. https://doi.org/10.1145/2818048.2820078

[54] Chat Wacharamanotham, Lukas Eisenring, Steve Haroz, and Florian Echtler. 2020. Transparency of CHI Research Artifacts: Results of a Self-Reported Survey. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376448

[55] Richard Y. Wong, Ellen Van Wyk, and James Pierce. 2017. Real-Fictional Entanglements: Using Science Fiction and Design Fiction to Interrogate Sensing Technologies. In *Proceedings of the 2017 Conference on Designing Interactive Systems* (Edinburgh, United Kingdom) *(DIS '17)*. Association for Computing Machinery, New York, NY, USA, 567–579. https://doi.org/10.1145/3064663.3064682

[56] Michael Zimmer. 2010. "But the data is already public": on the ethics of research in Facebook. *Ethics and information technology* 12, 4 (2010), 313–325. https://doi.org/10.100z.252Fs10676-010-9227-5

[57] Michael Zimmer. 2016. OkCupid Study Reveals the Perils of Big-Data Science. (2016). https://www.wired.com/2016/05/okcupid-study-reveals-perils-big-data-science/

[58] Michael Zimmer. 2021. The Zuckerberg Files. (2021). https://zuckerbergfiles.org/

[59] Michael Zimmer and Edward Chapman. 2020. Ethical Review Boards and Pervasive Data Research: Gaps and Opportunities. *AoIR Selected Papers of Internet Research* (2020).