Investigating the Data Management and Practices of Computing Researchers in the Mental Health Space

William Eickman
Georgia Institute of Technology
Atlanta, Georgia, USA
weickman@gatech.edu

ABSTRACT

The integration of technology in mental health research holds significant promise but introduces complex ethical considerations around data management. This paper explores the current practices of computing researchers working with mental health data, particularly focusing on their data management approaches, methodological transparency, and institutional guidance. Through interviews with computing researchers who have developed predictive or generative systems for mental health applications, we identify key challenges including inconsistent IRB protocols, unclear guidelines for non-clinical mental health data, fragmentation across disciplines, and varied approaches to transparency. Our findings highlight opportunities for better collaborative data management between computing researchers and mental health professionals, as well as the need for clearer critical data education. This research contributes insights toward developing standardized practices in a rapidly evolving interdisciplinary field.

1 INTRODUCTION

According to the World Health Organization (WHO), mental illness stands as the leading cause of disability worldwide [26]. The integration of technological interventions in mental health research and treatment offers several potential benefits, including real-time monitoring capabilities and broader population reach. The use of predictive modeling and other machine learning (ML) applications has seen explosive growth in almost every area of research over the last ten years, and social computing and mental health studies are no exception. However, these advancements are accompanied by significant ethical concerns and privacy implications that warrant careful consideration [4].

Thieme et al. [25] provide a comprehensive review of the current state of machine learning in mental health, highlighting its utility in detection and diagnosis of illnesses (clinical depression, generalized anxiety disorder, etc.), assessing patient-clinician relationships, and increasing treatment scalability. However, they also emphasize that the field is still in its infancy. As the field continues to evolve, ongoing critical examination of research practices and their implications, as well as the development of supporting tools and frameworks will be essential to maintain the delicate balance between innovation and responsible, ethical research. Ajmani et al. [1] brings to light that there are currently significant inconsistencies in reporting practices in predictive mental health research, between both ethical requirements and data management methods.

The inconsistencies we observe could stem from several scenarios: 1) There may be a set of common research standards that circulate within this specific research community but are not being directly disclosed in publications. If this is the case, it poses a problem for newcomers to the field and interdisciplinary researchers who may not be aware of them. 2) The lack of consistency in practice could simply indicate that there does not exist a common set of research data management procedures for working with mental health data. 3) The possibility that while formal standards do exist, whether as an extension of HIPAA or some other source, they're being overlooked or disregarded in practice, creating a gap between expectations and actual implementation.

In an attempt to evaluate how these could be addressed, we conducted detailed semi-structured qualitative interviews with researchers who have published in HCI venues on mental health computing, structured around a contextual inquiry activity based on the Follow the Data method [20]. Our working hypothesis was that there is a need for standardized ethical guidelines, transparent data management, and potentially new specialized collaborative tools. All together, these three things will need to be combined to form new community-driven best practices.

2 LITERATURE REVIEW

2.1 Mental Health Surveillance and Sensitive Inferences

Traditionally, mental health surveillance relied heavily on clinical data collection and self-reporting mechanisms. Grover et al. [11] outline how e-mental health professionals have adapted their data handling practices to accommodate the shift toward digital methodologies. While digital surveillance methods offer unprecedented access to behavioral data, they also introduce new complexities in data management and ethical considerations [24].

Chancellor et al. [4] conducted a critical examination of the representations and language used in machine learning approaches to mental health prediction, utilizing discourse analysis. Foucault famously described discourse as an action of language "that systematically form[s] the objects of which they speak" [9]. Their analysis identified five key discursive representations of humans in human-centered machine learning: 1) Disorder/patient, 2) Social Media, 3) Scientific, 4) Data/ML, and 4) Person. The authors argue that calling non-mentally ill individuals "normal", referring to non-validated "controls", and reducing individuals to "accounts" or "posts" could lead to misrepresentation of findings or participants as well as potentially diminishing their agency and respect as research subjects.

Schoenebeck et al. [21] explores how sensitive inferences derived from research can be exploited in commercial contexts, particularly in targeted advertising. Their findings suggest that clickwrap contracts on social platforms are often used as legal permission for

corporations to exploit data, provided they disclose their practices and offer users some form of choice. This phenomenon appears to be confused with genuine consent, permeating in both commercial and in some research scenarios.

2.2 Predictive Mental Health Methods

Early attempts at using machine learning for mental health prediction, such as the work by O'Dea et al. [16] on detecting potential suicidality through users' Twitter posts, showed promise but faced significant limitations. On the technical side, their study achieved generally high precision (75%) but poor recall (36%) in their model. From an ethical standpoint, they faced criticism for not including a statement about IRB approval regarding the collection of user social media data and limited ethical considerations.

In more recent work, Sharma et al. [22] developed a reinforcement learning model called PARTNER for facilitating empathetic conversations in online mental health support. Their work demonstrated improved ethical considerations, including obtaining IRB approval, removing unsafe expressions from training data, and providing detailed disclosure of methods. In a subsequent study, Sharma et al. [23] explored the use of large language models (LLM) for cognitive reframing of negative thoughts, featuring enhanced ethical precautions including the refusal to collect identifiable data and implementation of risk mitigation through content filtering.

Thieme et al. [25] urge caution in presenting ML developments to avoid premature claims about potential usefulness and real-world impact. The authors identify several key challenges, including data quality issues and reliability concerns, generalizability problems across populations, and the need for more transparent and explainable AI in mental health applications.

2.3 Data Collection and Management Practices

Khan and Hanna [12] propose a framework for dataset accountability that emphasizes transparency and ethical considerations throughout the data lifecycle. Their work highlights the importance of documenting data provenance, collection methods, and potential biases inherent in the dataset.

With these frameworks in mind, it is also important to consider more specific challenges, such as the concern of data-driven violations of consent brought up by Pendse et al. [17] that arise from the combined histories of service-user exclusion from data practices and long-running issues of non-consensual data collection in mental health.

Zhang et al. [27] conducted an in-depth analysis of how data science workers collaborate, identifying key challenges in the field: 1) ambiguous role definitions among team members, 2) tool limitations that may not adequately support complex workflows, and 3) the need for established best practices that balance efficiency with ethical considerations.

Sands et al. [20] propose a "Follow the Data" framework, which, although originally developed for astronomy, provides valuable insights for mental health computing. Their framework emphasizes the actions of discovering, locating, retrieving, and storing data.

2.4 Ethics and Methodology Disclosure

Ajmani et al. [1] conducted a systematic review of ethics disclosures in predictive mental health research, revealing significant inconsistencies in reporting practices. They found high disclosure rates for data sources (70%) and personal data modification (81%), both of which are directly tied in most cases to the development of the predictive models used. Less quantitative aspects saw significantly lower rates, such as consent procedures (24%), data storage methods (7%), and anonymization techniques (33%).

A framework developed by Rocca et al. [19] for approaching research when dealing with LLM-based inferences regarding psychiatric conditions, ELLIPS, posits that current predictive modeling has remained relatively ineffective due to a lack of being informed by clinical practice and knowledge. Their toolkit identifies seven core principles of respect for a person's autonomy, beneficence, and justice, as well as responsible scientific inference, credit allocation, transparency, and social responsibility.

3 METHODS

To understand current practices and methodologies, six researchers were interviewed from January 2025 to March 2025. These individuals were academic researchers, graduate students, and other academic figures from various American universities. Our interview protocol was approved by our Institutional Review Board with all participants giving explicit consent in the study. Participants were also provided with a $20 Amazon gift card as a token of appreciation for their time.

3.1 Recruitment of Researchers for Interviews

The criteria for the selection process included individuals who were first or second authors on academic papers published in the last 5 years involving the use of machine learning or artificial intelligence in a mental health research context. Potential participants were identified through the authors' connections and network, as well as through having read their work in the process of building out the literature review for this project. Additional participants were found through snowball sampling.

3.2 Structure of Interviews

The first author performed semi-structured interviews virtually over Zoom. These interviews ranged from 30 to 60 minutes and focused on applying the Follow the Data [20] and ELLIPS [19] frameworks in a series of questions focused on assessing the following:

(1) What data management methods and flow information did you disclose?
(2) Was any form of machine learning/predictive or generative modeling used in this project? If so, how?
(3) What were the overall goals/aims of this project? Follow Up: How did these goals influence your high-level decisions regarding your data?
(4) What do you feel should be standard practice as a researcher handling data within your field or area of expertise?

Interviews were recorded and automatically transcribed, then manually reviewed for correctness.

3.3 Data Analysis of Interviews

The interview recordings and transcripts were reviewed by the first author, then the final transcripts were uploaded to a local host version of Taguette for data analysis. These transcripts were reexamined and analyzed using the open coding framework to identify and code for the major emergent themes. These themes were discussed and refined along with the other authors to finalize the code book. The interviews and transcripts were then re-reviewed and analyzed to be coded into these finalized themes, with further thematic analysis conducted to identify subthemes within the set coded categories.

4 FINDINGS

4.1 IRB and Other Institutional Inconsistencies

When asked about their experiences dealing with Institutional Review Boards (IRB) and other governing bodies that may dictate data usage as it pertains to the topics of mental health data and predictive or generative machine learning tools and systems, there were a wide variety of opinions. There seemed to be a fair amount of regulation on the matter, but inconsistency between different institutions as to what aspects were most important as well as the expansiveness of the protections or regulations in place. There were notable exceptions, such as one participant's (P2) response saying "Actually, there's no high-level institutional stuff." They further went on to discuss how the inconsistency surrounding topics such as LLM use, as well as approval of data management tools and platforms can be "the most frustrating part" of this research.

Despite frustrations amongst these researchers around IRB in many cases, typically the requirements for protection in studies seemed relatively standard. As P3 stated, "I think most of the concerns of IRB board is about if it is identifiable afterwards...", and typical protocol around things like pilot studies remained as spoken about by P4, "...we did not go for an IRB approval, we just tested the system with the members of the lab... Everyone was OK with participating in the study to see how the system works." The responses of all participants would suggest that the protections that are presently in place seem to be working to an extent. However, with the differing approaches to how they are enforced or carried out and the constant change and development being seen in the core technologies used for predictive work, there is certainly a possibility for issues to arise.

4.1.1 GPT and other LLM Use. One major area in which opinions seem divided, and frustrations are evident, is in policies regarding the use of publicly available large language models (LLMs). Services like ChatGPT, Claude, and DeepSeek have been the center of attention in the computing world for the better part of the last two years, and they continue to see more and more use amongst researchers in studies geared towards using and understanding them. Regarding IRB stances on when these tools are okay to use, one of our participants claimed they felt "...they are not keeping up with the development of LLMs..." (P2). This seems to support the notion that the recent explosive growth and improvement of LLMs has either been too fast for review boards, or that in an effort to limit or regulate their use in research settings, they are deliberately remaining stagnant.

Most of the confusion seems to stem from why certain uses are allowed and others are not. In certain cases, there just seems to be a perceived lack of continuity, with P2 recounting, "I don't think you can actually like under IRB to put this data to GPT....I guess they just don't want GPT to see the raw data or something like that. But actually, it's interesting 'cause later on I have a much heavier use of LLM on another project, and it's actually letting people input some like mood like diaries and to change it." For a junior researcher, or even those more experienced, not having consistency with when they can utilize certain tools or methods can stifle their work and make it difficult to make informed methodological decisions.

4.2 HIPAA Compliance

A majority of participants were concerned with how to classify and handle mental health data when it is collected outside of a clinical setting. In the case of our participants, the data being used did not require HIPAA protection. However, confusion can easily be created by a lack of clarity around mental health data. This is especially true for online forums or passive sensing data that is clearly outside the bounds of clinical care but still has elements of patient information.

P1 gave one such example from their project utilizing passively gathered physiology sensor data, in which they used data gathered from a study at a medical school. They explained further, stating "...because this was coming from a medical school, I think there is an initial read that maybe the data was covered under HIPAA and that was actually a mistake because yes, it was collected by a medical school, but it wasn't collected as a part of care it was purely as a part of research... It doesn't have to be perfectly HIPAA compliant. There's no medical practitioner involved." There are certainly questions about whether such information should be HIPAA compliant and where that line is drawn, but the key factor is the lack of clear direction for researchers who are attempting to follow best practices. As P1 stated, "It is always weird that you're like, yes, this is not health data, but it's health data."

P3 worked on an LLM-empowered assistant for supporting online peer counseling. Because platforms such as the one used for their training purposes are not considered an actual clinical setting, they stated they didn't have to worry about any HIPAA protections for gathering peer-to-peer conversations. "[website] is an online mental health peer support platform. It's somewhere in HIPAA guidelines; it's not official patient conversations." This particular site did have its own deidentification and user protection agreement in place through a data management agreement. However, not standardizing the handling of what is so close to therapeutic conversation logs under the same regulations as actual clinical conversation data presents an easy path to mismanagement.

It may not be necessary to instate the full breadth of HIPAA protections onto data outside of these true clinical scenarios, but as these standards are already effective in protecting patients and filtering out misconduct in research, either adapting those regulations or applying a similarly concrete set of guidelines more specific to non-clinical scenarios could be critical to enabling and encouraging better work.

4.3 Fragmentation of the Field

Many researchers we spoke with struggled with what they feel is the disconnected nature of the field as they have peers in computer science, psychology, and multiple other disciplines. This means that professionals with similar research foci are publishing and communicating in different venues, such as journals and conferences. As predictive mental health research has origins in both medical and computing fields, there are multiple communities of practice [13] with regards to both research methods and data management specifically.

Further, these distinct disciplines understand potential outcomes and the validity of different techniques in the context of their own field, which necessitates different data handling approaches. The researchers we spoke with often echoed sentiments of how this affects data practices and the systems they create, as one individual put it, "...there are computer scientists who are in CS Labs who are doing this, there are people who are in bioinformatics who are doing this, there are people who are in clinical applied psychology doing this, and they're all building their own things" (P4). Collaboration between these parties appears to be limited, and researchers on the computing side feel as though their lack of access to mental health professionals themselves, as well as their tools and systems, limits the effectiveness of their studies.

4.3.1 Access to Professionals. The connections required in the medical community to find and recruit psychotherapists and other mental health professionals are often difficult to establish, especially for junior researchers coming from a computer science background. As one participant, a second-year master's student, stated, "I do interviews with like end users and just like people in college, that kind of thing, but finding psychotherapists is actually challenging." (P2) This creates a barrier to entry into the space and also limits some of the work computing researchers can accomplish, lacking in some of the information, access, and expertise that these professionals have or can provide.

As P5 described, identifying usable, scientifically validated datasets can be difficult. This can be especially impactful in the case of projects such as therapeutic or counseling chatbots, where role-playing or informed empathetic language, often gathered from open social media platforms such as Reddit, can be used, but will not be as effective as having the input of an actual licensed professional. P4 was optimistic about the role that mental health experts can play in the field, particularly in sharing their domain expertise as well as the resources to which they have access.

4.3.2 Communication Across Disciplines. In any field that brings together a variety of disciplines such as predictive mental health, there is always going to be dissention within its practitioners regarding best practices, how information is discussed, what factors are most important, and who makes these decisions. The researchers we spoke to, all coming from a computing background of some kind, seem to be aptly aware of this and how it affects communication.

One such example from P1 was discussing the use of the ML data pre-processing method SMOTE (Synthetic Minority Oversampling Technique) and how it can change the balance of the dataset. While in some cases, this would be negligible as the shape of the data for classification and prediction purposes would remain relatively the same, but in healthcare where the context matters immensely, it may limit the interpretability of the model for practitioners. "...a very rare disease, saying that there's like a 20% chance that someone actually has a disease that has meaning to a clinician to say, well, that might be kind of high and maybe we should look into this, right?" (P1)

P4 brought up the issue of meeting the expectations of health science researchers when attempting to design more participatory studies with passive sensing data and allowing for information to be shared back with participants. "If you go to like the BioMed people who do these at population scale, they will actually say that we're tampering with the control condition right?" (P4) Incentivizing better communication is something that may have to be handled at a grant or institutional level.

4.4 Transparency and Disclosure

4.4.1 In Writing. While the variation among researcher backgrounds may lead to variation in how methods are written about and what aspects are considered most important to include for transparency purposes, most researchers seem to be attempting to operate as openly and ethically as possible. Despite this, as studies like the Ajmani et al. [1] have revealed, there has been a wide gap between information disclosed in ethics disclosure sections in papers and the practices and decisions of the researchers.

Methodologically for data management this is less of an issue of existence, as very rarely if ever will an HCI or ML paper not include a sufficient methods section. However, there remains an issue of specificity, as in our interviews we found there was always more information and insight to be gained by talking about a researcher's data management practices than included in writing in their work.

Table 1: Disclosure practices in participants' papers

	Participant				
Disclosure Element	P1	P2	P3	P4	P5
Explicit ethics section	No	No	Yes	No	No
IRB approval statement	Yes	Yes	Yes	Yes	No
Data handling details	Full	Brief	Full	Summary	None

4.4.2 In Practice. There is certainly a movement amongst the researchers we spoke to, most notably P3 and P4, to push for more participant inclusion throughout their research processes when handling information or making predictions that sensitively impact the individuals involved. As stated earlier, this process is somewhat complicated by IRB rulings on how information can be shared back to participants, and requires additional work as P3 describes saying, "I actually think that's a little better too, because it is burdensome, yes, but it gives more control and gives more transparency to the process than like just creating everything through and then making whatever assumption you want."

They acknowledged that in past studies such as the one we discussed mainly in interview dealing with an online mental health forum, participants are often signing away consent for their information to be used in ways they may not totally understand without

much feedback back to them. This platform had a data use agreement for its users that allowed conversational data to be used by researchers, but as P3 themselves stated, "I don't even know how hard it would be to actually, like if it's a massive, scroll down, scroll down, scroll down, but they are not going to read every line of it." This presents a unique issue as to how you can ensure these participants are informed.

In an effort to solve this issue, P4 discussed how they are reaching for more direct and continuous participation by keeping individuals whose data is being used personally updated as the project progresses. In their own words, "...we are asking them to sort of come in like as a come in and meet us for like a consultation every week and we not just show them the data, we actually kind of show them like a machine prediction or machine estimate..." This sort of structure for a study is cumbersome and more time consuming and certainly doesn't lean towards the big data style that has been the status quo but provides a more humanistic approach to very human research projects.

Furthermore, P4 spoke about enabling the participants in another study to have more control over the data collection process, by having them be responsible for downloading and sending their social media data at regular intervals with informed consent. P4 themselves did discuss that these methods are not purely being done out of desire for more ethical research, but also for more enriching data and using possibly incomplete predictions as probes to compare to the individuals' actual feelings or emotional state. However, the result of having more informed and involved participants is overall positive.

4.5 Public vs Private

4.5.1 Datasets. Sourcing or choosing a dataset to work with, whether that be a publicly available large collection from services like Kaggle or Hugging Face, or a smaller, privately developed dataset, is one of the most crucial aspects of any machine learning project. For predictive mental health, locating the best source of information to develop these models is incredibly context dependent and varied based on the approach and goals of the researchers.

While earlier projects in the field focused mainly on the use of privately gathered and stored social media data, the more recent studies of the researchers we spoke to primarily utilized already publicly available data. For some this shift is a direct response to ethical concerns, as explained by a quote from P4 stating, "We stay away from data that's private. Why manage things which can be harmful?" For others, such as P5, there is an element of practicality to it as well, as projects such as theirs require scientifically validated physiological sensing datasets. "We did not collect data ourselves any data. We used the data sets that were already there, so there are a lot of standardized wearable datasets and there are, and even the mental health data set that I used was posted on hugging face on if it's still there." (P5)

The assumption when using these already publicly available datasets is that they are safe and already vetted for possibly identifying or sensitive information, however this is difficult to maintain when dealing in data that is so inextricably linked to the individuals it represents, such as counseling conversations, sleep patterns, and physical stress indicators. This is something that some of the researchers we spoke to seem to be aware of, as P1 stated when describing their project which involved both public and privately gathered data, "I will say like if this was just the public data, we probably wouldn't have to do any of this, and you know we can have a conversation about whether all the data that was public, you know, should have been public."

For those working with private datasets, such as P3, there are other challenges that arise. While the data is certainly privatized for a reason, there are benefits to be had from having large public datasets published for career purposes for researchers looking for vertical mobility. The processes for companies involved in keeping these datasets private seems secure, as P3 described, "If someone wants that data, they have to send us a request, and we forward that request to [company] and then they consider their item data use agreements and stuff with their legal team."

4.5.2 Testing. An area where there is clearly still a divide amongst what should and should not be public, is in how the models and tools developed by these researchers are tested. It has been a point of criticism and concern toward prominent work in the field that testing systems with sensitive outcomes developed from sensitive data can pose a threat to participants. For some researchers we spoke to, this ethos was clearly informing their decisions, as P3 discussed that their model was never publicly released, similar to their dataset, as it was fine-tuned with real people's actual peer counselor messages. However, when speaking about their upcoming project working with AI patients and AI mentors for training counselors, they mentioned "We have this deployed in a classroom for testing, so we hope to get good feedback." This disconnect seems to point to a possible misunderstanding of what makes these systems so sensitive.

4.6 Data Management Education

Key to understanding the interviewees' methodology is to understand where it comes from. We asked our participants questions to better grasp what informed the decisions they were making regarding their data handling, interested in determining if their methodologies were being fully informed by institutional factors, more personal learned experienced, or intuition. The sources from which individuals have learned data management seem to be as varied as the individuals themselves.

4.6.1 Formal-ish. Being computing researchers and not medical professionals, the expectation would be that most of the training or education that the individuals we interviewed would have had would be more grounded in traditional methods of data management for computer science research. However, the experiences that have brought these researchers to work in such an interdisciplinary field has also provided them with unique learning experiences.

For example, P3 spoke about how much of their ethos around handling both mental health data such as the counselor conversations from the specific project we referenced, as well as other studies, was largely informed by their experience interning with an electronic healthcare record company. When asked about how they would suggest other researchers to improve their practices, they voiced support for more researchers exploring these trainings, stating "I think those companies do have very strong training on

Table 2: Participant data management approaches across the "Follow the Data" framework

	P1	P2	P3	P4	P5
Discover	Searching for publicly available sensing datasets	Using connections to identify relevant clinicians and medical school students	Partnered with peer counselor site	Searching for individuals U.S. who conduct information work	Using sites like Kaggle and Huggingface
Locate	Sampling criteria for two previously collected datasets described, private dataset collected from medical school	Snowball sampling at universities in the north and southeast from those previous connections	Classifier dataset from previous study; rest of data from peer counseling site	Recruitment was done through workplace emails, messaging boards, and newspaper advertisements, with a form link for participants to sign up	Searching for scientifically validated datasets
Retrieve	Downloaded data from public datasets	Semi-structured interviews and interactive prototyping	Chats were created by researchers	Multi-institutional collaboration to gather and analyze both the quantitative sensor information from three different devices and user feedback	Downloading publicly available datasets; performing any necessary preprocessing
Store	Stored internally on private server	Recordings downloaded and deidentified; stored in institutional drive	Last 5 utterances stored in AWS backend by assumption	Data stored in encrypted servers at main institution, with remote access	Unclear

how to manage data. I'm sure those trainings are publicly, widely available in some senses so, probably follow those." None of the interviewees cited IRB or other institutional policies as their primary source of guidance.

4.6.2 Experience. The remainder of the researchers we spoke to described their education in data handling to be based on experiential learning and copying their peers and mentors. P4 spoke about having always had someone to turn to in terms of what "was and was not allowed" for their data practices, a sentiment echoed by P2, who described primarily relying on their PI or senior lab members for guidance. This seems to be in line with the expected standard of how most labs would operate, as P1 stated, "It's easy enough to ask."

However this isn't always the case. Sometimes judgement on these issues seems to be more "incidental" as described by P4, with certain practices having come from past mistakes or blowback on previous studies that led to more consideration on aspects of the data storage or processing that have been shown provably to be more sensitive. Researchers like P5 also look to others' practices in their writing, choosing to identify and maintain the status quo. "I think my practices were mostly due to or based on the the papers that I read specially the scientific data sets that we have on how the researchers for those scientific data sets managed their data." (P5)

However, this sort of pseudo-standardization is contributing to the variety we see in how these methods propagate and are written about, as each individual is bringing their own interpretation and experiences to the table without one clear unifying set of directions. The fortunate outcome as of the time being is the intuition of HCI and ML researchers seems to be in line with protecting the sensitive nature of the information they're handling.

5 DISCUSSION

5.1 Enabling Better Collaboration with Psychiatrists and Other Mental Health Professionals

Our findings point to an ongoing disconnect between psychiatrists, researchers, and developers working in predictive mental health. Participants described a lack of shared language, tools, and routines across domains, resulting in misalignment around how data is collected, interpreted, and used in downstream decision-making. These disconnects have implications not only for collaboration but also for the long-term viability of predictive systems in clinical practice.

5.1.1 Designing Systems of Better Collaborative Data Management. One area of opportunity is the design of systems that support collaborative data management and interpretation. Participants described situations where data lived in siloed systems with unclear documentation or lacked features that could support collaborative engagement, such as annotation tools or versioning histories. A few referenced the potential value of tools that foreground the context of data collection and make it easier for others to interpret the data's meaning, quality, and limitations.

This echoes calls in HCI literature for infrastructure that foregrounds provenance and negotiation of meaning across stakeholders [8]; [15] et al., 2010). In predictive mental health, these negotiations are particularly critical because data is not just technical—it

carries clinical and ethical implications. Design should support traceability and the ability to surface "data stories" [18], ensuring that collaborators understand how and why data was generated. Systems that embed contextual metadata, allow collaborative tagging, or flag moments of ambiguity can make otherwise invisible assumptions legible across domains.

Such infrastructural interventions should also be flexible and co-adaptive. As research through design often reveals, knowledge and practice co-evolve [29]. What may begin as a simple annotation system could become a site of shared language-building or policy negotiation. Thus, rather than aiming for fixed standards, designers should support ongoing dialogue and co-interpretation, which align with a more participatory approach to system development.

5.1.2 Developing Workshops and Frameworks to Standardize Practices. In parallel, participants highlighted a desire for shared frameworks and routines that would help align practices across domains. Workshops, interdisciplinary team check-ins, and recurring design sessions were all proposed as strategies for building common ground. These structures could support not only mutual understanding but also sustained relationships that enable trust and iterative feedback. Participants noted that creating space for this kind of work was challenging but necessary for more grounded, inclusive, and effective design.

These insights build on participatory design traditions in HCI, which emphasize the need to create sustained spaces for co-construction [2]; [6] et al., 2012). Rather than attempting to encode all perspectives into a singular design, workshops and frameworks act as scaffolding for conversation, reflection, and negotiation. When used regularly, they create rhythms of engagement that foster shared responsibility across disciplinary boundaries.

Moreover, frameworks can serve as epistemic bridges. Participants often referenced a lack of shared concepts—what one group views as "predictive accuracy," another might interpret as "risk signaling." Frameworks like speculative enactments or co-specification activities [10] help surface these mismatches not as barriers, but as moments for mutual learning. The value of these interventions lies not in closure, but in the relationships they sustain over time.

5.2 Develop Clearer Critical Data Education for Approaching Mental Health Data

Another major theme that emerged from our interviews was a widespread uncertainty around how best to manage and interpret mental health data, particularly in research involving machine learning. Participants reported mixed understandings of institutional policies, regulatory frameworks, and best practices for data governance. This uncertainty was especially acute for researchers working with non-clinical data or novel computational tools.

5.2.1 Continuing Workshops for Data Management Education. Several participants expressed confusion about what constitutes identifiable mental health data, or how current laws like HIPAA applied to their projects. In many cases, they described relying on informal networks or trial-and-error to navigate questions of compliance. There was strong interest in clearer, more actionable guidance, especially for researchers working outside of traditional clinical settings. Participants also emphasized the value of workshops and shared documentation as a way to build local knowledge and reduce reliance on ad hoc decision-making.

This desire aligns with HCI's growing investment in critical data literacy [5]. Educational workshops must move beyond compliance checklists to include interpretive practices, like evaluating how assumptions travel from data collection to model development. Just as "data feminism" foregrounds context and power, mental health data practices require workshops that attend to histories of surveillance, diagnostic harm, and algorithmic opacity. These aren't ancillary topics—they are essential to designing ethically grounded systems.

Designing for this kind of learning means thinking through not just what knowledge is conveyed, but how and by whom. Peer-led sessions, case-based learning, and failure sharing were all mentioned by participants as more resonant than static training modules. These participatory formats reflect research through design values: learning emerges through doing, discussing, and iterating together.

5.2.2 Developing Workshops for IRB Standardization Regarding the Use of LLMs . In addition to regulatory uncertainty, participants also described challenges navigating IRB processes, particularly in projects involving generative AI or large language models (LLMs). Some had difficulty articulating the risks or novelty of their work in ways that IRBs could understand, while others noted that different IRBs offered conflicting advice. Standardized templates, cross-institutional working groups, and ongoing conversations with IRB representatives were proposed as ways to improve clarity and consistency.

This is a rapidly evolving concern. As LLMs become more integrated into research and clinical pipelines, IRBs face pressure to adjudicate technologies whose risks are still emergent. In computing research, calls have been made for IRBs to adopt anticipatory approaches [28]—recognizing that harm is not always visible at the point of approval, but often unfolds through use. Workshops that bring together researchers, ethicists, and IRB staff could support shared understanding of both technical affordances and social impacts.

These workshops also serve as sites of value negotiation. Rather than positioning IRBs as gatekeepers and researchers as applicants, co-designed processes could foster reciprocal accountability. Tools like ethical impact canvases or harm scenario planning [14] might help foreground nuanced concerns around consent, data representativeness, and secondary use—especially when data involves vulnerable populations.

5.2.3 Providing Clearer Language for HIPAA as It Relates to Non-Clinical Mental Health Data. Finally, participants called for more targeted educational resources that would help them build confidence in handling mental health data. These resources could take the form of training sessions, toolkits, or example protocols and should be grounded in the kinds of messy, real-world cases that researchers actually face. Importantly, participants did not just want to follow rules; they wanted to understand why those rules mattered and how to make thoughtful decisions in cases where rules were unclear. These findings point to an opportunity for more critical, contextualized forms of data education that recognize the complexities of working with mental health data in practice.

The lack of clarity around HIPAA and non-clinical data reflects a broader ambiguity in the regulatory landscape. As studies in HCI have shown, legal frameworks often lag behind technical capabilities [3]. This disjuncture places the burden of interpretation on individual researchers, which can lead to over-cautiousness, underreporting, or inconsistent data practices. Clearer interpretive materials—developed in collaboration with legal scholars, ethicists, and researchers—could fill this gap.

Moreover, these resources must be designed with usability in mind. Rather than dense legalese, participants wanted accessible, example-driven explanations of what compliance looks like in practice. Literature on explainability [7] underscores that interpretation is not just about access to information—it's about sensemaking. Resources that integrate storytelling, visualization, or interactive walkthroughs could better support nuanced, context-specific understanding of privacy requirements.

Together, these interventions move toward a broader vision of critical data education—one that centers not only knowledge, but care, context, and collective responsibility in data work.

6 LIMITATIONS

Consistent with traditional limitations of interview-based work, we report on interviewees' own experiences and perspectives in their own words, which may not be a perfect representation of how they actually behave in situ [21]. Our sample size of six researchers, while sufficient for qualitative insights, may not capture the full diversity of perspectives in this interdisciplinary field. Additionally, all participants were from American institutions, which may limit the generalizability of our findings to other cultural or regulatory contexts [26].

7 CONCLUSION

This research investigated the data management practices of computing researchers working with mental health data through interviews with six active researchers in the field. Our findings revealed several key challenges: inconsistent IRB protocols, unclear guidelines for non-clinical mental health data, fragmentation across disciplines, and varied approaches to transparency. We identified opportunities for improvement in three main areas: (1) better collaborative data management between computing researchers and mental health professionals, (2) clearer critical data education, and (3) more standardized ethical guidelines.

The rapid evolution of predictive mental health technologies makes addressing these challenges particularly urgent. While our participants demonstrated thoughtful approaches to data management, the lack of consistent standards creates unnecessary barriers and potential risks. Future work should focus on developing the collaborative frameworks, educational resources, and standardized guidelines identified in our discussion, with particular attention to emerging technologies like large language models.

Mental illness remains one of the most pressing issues of our time, and technological interventions hold great promise for improving mental health outcomes. However, as with any emerging medical practice, proper care and safety for all involved must remain the utmost priority. By addressing the data management challenges identified in this research, we can help ensure that predictive mental health technologies develop in a way that is both innovative and ethically responsible.

ACKNOWLEDGMENTS

We thank our interview participants for sharing their time and insights. This work was supported in part by the National Science Foundation under Grant No. 2048123. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] L. H. Ajmani, S. Chancellor, B. Mehta, C. Fiesler, M. Zimmer, and M. De Choudhury. 2023. A Systematic Review of Ethics Disclosures in Predictive Mental Health Research. *2023 ACM Conference on Fairness, Accountability, and Transparency* (2023), 1311–1323. https://doi.org/10.1145/3593013.3594082

[2] S. Bødker and K. Grønbæk. 1993. Cooperative prototyping: users and designers in mutual activity. *International Journal of Man-Machine Studies* 39, 5 (1993), 815–838. https://doi.org/10.1006/imms.1993.1085

[3] R. Calo. 2012. Against notice skepticism in privacy (and elsewhere). *Notre Dame Law Review* 87, 3 (2012), 1027–1072.

[4] S. Chancellor, E. P. S. Baumer, and M. De Choudhury. 2019. Who is the "Human" in Human-Centered Machine Learning: The Case of Predicting Mental Health from Social Media. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–32. https://doi.org/10.1145/3359249

[5] C. D'Ignazio and L. F. Klein. 2020. Data Feminism. *MIT Press* (2020).

[6] C. DiSalvo, P. Sengers, and H. Brynjarsdóttir. 2012. Mapping the landscape of sustainable HCI. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2012), 1975–1984. https://doi.org/10.1145/2207676.2208347

[7] F. Doshi-Velez and B. Kim. 2017. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608* (2017). http://arxiv.org/abs/1702.08608

[8] P. Dourish. 2006. Implications for design. *Proceedings of the SIGCHI conference on Human Factors in computing systems* (2006), 541–550. https://doi.org/10.1145/1124772.1124855

[9] Michel Foucault. 1972. *The Archaeology of Knowledge*. Pantheon Books, New York.

[10] W. Gaver, A. Boucher, S. Pennington, and B. Walker. 2004. Cultural probes and the value of uncertainty. *interactions* 11, 5 (2004), 53–56. https://doi.org/10.1145/1015530.1015555

[11] S. Grover, S. Sarkar, and R. Gupta. 2020. Data Handling for E-Mental Health Professionals. *Indian Journal of Psychological Medicine* 42, 5_suppl (2020), 85S–91S. https://doi.org/10.1177/0253717620956732

[12] M. Khan and A. Hanna. 2022. The Subjects and Stages of AI Dataset Development: A Framework for Dataset Accountability. (2022). https://doi.org/10.2139/ssrn.4217148 SSRN Scholarly Paper 4217148.

[13] J. Lave and E. Wenger. 1998. Communities of Practice: Learning, Meaning, and Identity. *Cambridge University Press* (1998).

[14] M. A. Madaio, L. Stark, J. Wortman Vaughan, and H. Wallach. 2020. Co-designing checklists to understand organizational challenges and opportunities around fairness in AI. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020), 1–14. https://doi.org/10.1145/3313831.3376445

[15] M. J. Muller and S. Neuwirth. 2010. Provenance and the use of process documentation in collaborative work. *Proceedings of the ACM 2010 conference on Computer supported cooperative work* (2010), 271–280. https://doi.org/10.1145/1718918.1718969

[16] B. O'Dea, S. Wan, P. J. Batterham, A. L. Calear, C. Paris, and H. Christensen. 2015. Detecting suicidality on Twitter. *Internet Interventions* 2, 2 (2015), 183–188. https://doi.org/10.1016/j.invent.2015.03.005

[17] S. R. Pendse, L. Stapleton, N. Kumar, M. De Choudhury, and S. Chancellor. 2024. Advancing a consent-forward paradigm for digital mental health data. *Nature Mental Health* (2024). https://doi.org/10.1038/s44220-024-00330-1

[18] K. H. Pine and M. Liboiron. 2015. The politics of measurement and action. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (2015), 3147–3156. https://doi.org/10.1145/2702123.2702298

[19] R. Rocca, G. Pistilli, K. Maheshwari, and R. Fusaroli. 2024. Introducing ELLIPS: An Ethics-Centered Approach to Research on LLM-Based Inference of Psychiatric Conditions. *arXiv preprint arXiv:2409.15323* (2024). http://arxiv.org/abs/2409.15323

[20] A. Sands, C. L. Borgman, L. Wynholds, and S. Traweek. 2012. Follow the data: How astronomers use and reuse data. *Proceedings of the American Society for Information Science and Technology* 49, 1 (2012), 1–3. https://doi.org/10.1002/meet.14504901341

[21] S. Schoenebeck, C. Goray, A. Vadapalli, and N. Andalibi. 2024. Sensitive Inferences in Targeted Advertising. *Northwestern Journal of Technology and Intellectual Property* 21, 2 (2024). https://scholarlycommons.law.northwestern.edu/njtip/vol21/iss2/1

[22] A. Sharma, I. W. Lin, A. S. Miner, D. C. Atkins, and T. Althoff. 2021. Towards Facilitating Empathic Conversations in Online Mental Health Support: A Reinforcement Learning Approach. *arXiv preprint arXiv:2101.07714* (2021). http://arxiv.org/abs/2101.07714

[23] A. Sharma, K. Rushton, I. W. Lin, D. Wadden, K. G. Lucas, A. S. Miner, T. Nguyen, and T. Althoff. 2023. Cognitive Reframing of Negative Thoughts through Human-Language Model Interaction. *arXiv preprint arXiv:2305.02466* (2023). http://arxiv.org/abs/2305.02466

[24] R. Skaik and D. Inkpen. 2021. Using Social Media for Mental Health Surveillance: A Review. *Comput. Surveys* 53, 6 (2021), 1–31. https://doi.org/10.1145/3422824

[25] A. Thieme, D. Belgrave, and G. Doherty. 2020. Machine Learning in Mental Health: A Systematic Review of the HCI Literature to Support the Development of Effective and Implementable ML Systems. *ACM Transactions on Computer-Human Interaction* 27, 5 (2020), 1–53. https://doi.org/10.1145/3398069

[26] World Health Organization. 2019. Mental Disorders. https://www.who.int/news-room/fact-sheets/detail/mental-disorders

[27] A. X. Zhang, M. Muller, and D. Wang. 2020. How do Data Science Workers Collaborate? Roles, Workflows, and Tools. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–23. https://doi.org/10.1145/3392826

[28] M. Ziewitz. 2016. Governing algorithms: Myth, mess, and methods. *Science, Technology, & Human Values* 41, 1 (2016), 3–16. https://doi.org/10.1177/0162243915608948

[29] J. Zimmerman, J. Forlizzi, and S. Evenson. 2007. Research through design as a method for interaction design research in HCI. *Proceedings of the SIGCHI conference on Human factors in computing systems* (2007), 493–502. https://doi.org/10.1145/1240624.1240704