

Original Article

Ethical Challenges of Using Social Data for Research

Priya Kapoor

Business Analyst, Accenture, India.

Received: 01-06-2023

Revised: 01-24-2023

Accepted: 02-02-2023

Published: 02-04-2023

ABSTRACT

The rapid growth of digital technologies and social media has generated vast amounts of social data, offering valuable insights into human behavior, communication, and societal trends. However, its use in research raises significant ethical concerns. Key issues include privacy invasion, lack of informed consent, challenges in data anonymization, algorithmic bias, and potential misuse of sensitive information. Social data is often collected without users' awareness, blurring the line between public and private information and questioning the validity of consent. Even when anonymized, data can sometimes be re-identified due to its interconnected nature. Additionally, biased algorithms may reinforce existing inequalities, while limited transparency in data handling reduces accountability. Current regulatory frameworks and ethical oversight bodies often struggle to keep pace with technological advancements, highlighting the need for more adaptive and flexible ethical guidelines. The study emphasizes user-centered ethical practices focusing on autonomy, transparency, and accountability, along with solutions such as dynamic consent models, stronger anonymization, and bias mitigation. In conclusion, while social data holds great research potential, ethical considerations must remain central. Collaboration among researchers, institutions, and policymakers is essential to ensure responsible innovation while protecting individual rights.

KEYWORDS

Social Data, Research Ethics, Privacy, Informed Consent, Data Anonymization, Algorithmic Bias, Ethical Frameworks, Data Protection, Digital Ethics.

1. INTRODUCTION

1.1. Background

Digital technologies are developing very fast and their impact on the everyday life of people is the alteration of the way people communicate, interact, and share information. The online forums, social media, and digital communication tools are constantly producing massive amounts of user-generated data, also known as the social data. This information provides useful possibilities to the researcher to study the human behavior, social patterns and cultural trends on a tremendous scale. Nevertheless, there are also certain ethical issues associated with the convenient access to such information. Social data can be gathered without the express knowledge of the users compared to traditional methods of research where the researcher is required to participate in the research directly. This raises doubts over the ethical conduct of researchers like informed consent, protection of privacy and ethically controlling aspects of using such data and it is important to consider the use of such data seriously.

1.2. Importance of Ethical Considerations

The aspect of ethics is essential in preserving trust between the researcher and the society, particularly in the social data research. Lack of adherence to ethical practices may have far-reaching consequences that may be critical.



Fig 1 - Importance of Ethical Considerations

1.2.1. Privacy Violations

Violation of privacy: The access of personal or sensitive information without the due authorization is called a privacy violation. This may occur in social data research where the data on users is gathered on the internet and relevant precautions are not taken. These breaches might subject people to the dangers of identity theft, surveillance or profiling that they do not desire. Privacy must be also secured to make people feel सुरक्षित and respected even in online spaces.

1.2.2. Psychological Harm

Abuse of social information may also cause psychological damage especially when sensitive information is shared or misunderstood. To take an example, the study of personal posts, feelings, or actions without their consent can be a distressing event in case the people find out about it. Also, user profiling or specific attention may lead to a sense of surveillance or evaluation which has an adverse effect on the mental health of users. Ethical conduct of research studies should thus be conscious of the emotional and psychological effects on the subjects.

1.2.3. Social Discrimination

Misuse of data may lead to social discrimination, particularly when we have biased data or algorithms. Some groups can be profiled, marginalized, or stereotyped based on their race, gender, socioeconomic status, and so on. This may support those inequalities and generate unfair results in such domains as employment, healthcare, or policing. This is an important factor in ensuring fairness and inclusivity is maintained to prevent such discrimination.

1.2.4. Loss of Public Trust in Research

Once the ethical considerations are violated, the trust of people towards research institutions and practices will decrease greatly. People can be reluctant to provide data or even take part in research in case they suspect that data will be misused or exploited. Such a lack of trust may render the further research work more difficult and restrict the access to useful data. Transparency, accountability and ethical integrity is hence a critical requirement in order to maintain the level of confidence of people in research.

1.3. Ethical Challenges of Using Social Data for Research

Applying social data to research offers a spectrum of complicated ethical issues that are the result of the digital data gathering and processing. Among the main issues, it is the question of privacy because people provide information online without thoroughly considering the possibility of its collection and analysis to be used in the research. When the data is publicly available, its misuse by the use outside the context of the original intent may still infringe the privacy expectations of the users. Very much connected with this is the issue of informed consent. Social data, unlike with traditional research methods, is often collected without physical contact with the participants, which makes it hard to gain explicit and meaningful consent. When users sign lengthy and complicated terms of service, it is possible that they are committing to the usage of their data without their understanding because these agreements are not informed consent. One of the other ethical issues of concern is the possibility of misusing and re-identifying data. Even though researchers tend to use anonymization measures to safeguard identities, it is now possible to connect datasets and re-identify people due to the development of data analytics. This poses a possibility of damage especially when sensitive information is concerned. Moreover, there is an issue of algorithmic bias, which is also quite severe since the datasets that are used in the social data studies can mirror the existing inequalities in the society. The use of these biased data to train model may result in unfair or discriminating results that strengthens stereotypes and disables some groups. Moreover, there are no clear and uniform ethical principles that may be effective in countering the fast-changing nature of digital technologies. The current frameworks might not sufficiently address challenges like high scale data scraping, cross platform data combination and real-time data analysis. This complicates the task of researchers to find out what is ethical practice. All in all, these issues support the idea that the ethical solutions are required to be stronger, more open, and flexible to make sure that the research of social data is carried out in a responsible manner and with the minimal harm to a person and humanity.

2. LITERATURE SURVEY

2.1. Privacy and Data Protection

Among the most significant ethical issues in social data research is the matter of privacy and data protection, which are generally regarded as the most urgent one. There is a tendency among many researchers to rely on publicly available information on social media platforms, forums, and communities on the internet, and think that the information can be utilized without limitations imposed on ethics. Nevertheless, the current literature is heavily criticizing this premise by pointing

out that public availability does not necessarily coincide with ethical usability. Users might not even want their information to be collected, analyzed, and used to conduct research even in the cases when they publicly share it. This poses a problem of contextual privacy as the original context of sharing data is broken. Researchers believe that scholars should make user expectations, sensitivity of the data, and possible harm on issues of such data to consider before using such data thus ensuring that they are ethically responsible which is more than legal compliance.

2.2. Informed Consent Challenges

Informed consent has long been one of the foundations of ethical research, and it involves participants being well informed and willing to see how their information is going to be used. Nevertheless, the explicit consent is usually not practical or sometimes not attained in case of social data research. When data are being collected on large scale through social media and the likes, the idea of individual consent is hard to obtain bearing in mind that hundreds of millions of people are users of these platforms. Research has shown that the majority of users are not aware that their information can be collected and used to conduct a research. Moreover, privacy policies and terms of service are very long and complicated, and people usually agree without the accurate knowledge. This presents a dilemma of ethics where there is legal consent, and the informed consent is dubious. To resolve those challenges, researchers are being encouraged to consider alternative frameworks, including dynamic consent or ethical transparency.

2.3. Data Anonymization Techniques

In order to eliminate the risks of privacy breach, researchers use a number of data anonymization methods such as masking, pseudonymization and aggregation. Data masking refers to the process of concealing identifiable information whereas pseudonymization is the substitution of personal identifiers by artificial labels. Aggregation is used to combine data into summaries, in order to avoid identification of individuals. Nevertheless, such efforts do not exclude literature pointing to the fact that anonymization is not foolproof. As the data analytics and machine learning further develop, the chances of re-identifying individuals have grown due to the possibility to connect anonymized datasets with other readily available data sources. This poses a big threat particularly where confidential information is involved. Therefore, researchers stress that anonymization cannot be considered as the full solution but is the step in a wider privacy protection approach also incorporating ethical judgment, safe data management, and regulation.

2.4. Algorithmic Bias

Another significant issue that has been found in the literature is algorithmic bias, especially when it comes to the application of machine learning and artificial intelligence. Prejudice is frequently based on the data sets to which algorithms are trained, and that may reflect social imbalance, stereotyping, or historical discrimination. In the event that these biased data are employed the algorithms may generate unfair or discriminatory results, irrationally targeting some groups by their race, gender, or socioeconomic status. It has been found that such biases do not always arise deliberately but may have a latent occurrence in the selection of data, labeling, or in model formulation. These effects can be severe particularly in hiring, lending, policing, and healthcare. Thus, researchers recommend the application of various datasets, bias identification instruments, and fairness-conscious algorithms to reduce these risks and facilitate the equitable results.

2.5. Ethical Frameworks and Regulations

In order to resolve ethical issues with social data research, a number of frameworks and regulations, such as a General Data Protection Regulation (GDPR) or Institutional Review Boards (IRBs) and other ethical AI guidelines, have been created. An example of this is GDPR, which lays down a high standard of data gathering, processing and consent of the user in the European Union, with a focus on transparency and the rights of users. IRBs are very important in screening research proposals to verify that ethical issues are observed especially in academics. Moreover, principles advocated by the ethical AI guidelines are fairness, accountability, and transparency in designing the algorithms. Nevertheless, literature indicates that these frameworks can hardly keep abreast with the fast changing digital technologies and data practices. There are still gaps in the solution to such problems like cross-border data flows, real-time data collection, and new AI systems. Consequently, an increased demand has been raised towards more agile, contextual, and internationally aligned ethical standards.

3. METHODOLOGY

3.1. Research Design

This paper uses a qualitative method of research to examine the ethical concerns of conducting research on social data in a thorough and understanding way. This study is especially appropriate with qualitative methods since it enables one to gain a profound insight into a complex phenomenon like privacy, consent, and bias, which may not be measured easily. This study is based on research of three key sources of information, which include academic literature, case studies, and ethical guidelines.



Fig 2 - Research Design

3.1.1. Academic Literature

The paper reviews a broad scope of academic writings that comprise journal articles, books, and conference papers in order to learn about the available theory, arguments, and outcomes regarding ethical issues when conducting social data research. It will contribute to the creation of a solid theoretical background and the identification of such problems as the risks of privacy, the challenge of informed consent, and the problem of algorithm bias.

3.1.2. Case Studies

The case studies are examined to reveal practical information on how the ethical issue could be manifested. The study focuses on the negative outcomes of unethical data usage through the analysis of the documented cases and examples, providing the examples of privacy breach and influenced algorithmic results. These case studies assist in filling the gaps between theory and practice.

3.1.3. Ethical Guidelines

Other ethical principles and regulatory requirements, data protection laws and institutional policies, are also examined in the study. This discussion facilitates the examination of the manner in which the existing standards are dealing with ethical issues and where the lacuna may still persist. It is also used to give grounds on how ethical ways to improve social data research are advocated.

3.2. Data Collection

In this work, the researcher gathers information on several authoritative and trustworthy sources to guarantee a profound knowledge of ethical problems in social data studies. The data collection instrument is narrowed down to peer-reviewed journals, conference papers, and policy documents, which will combine the theoretical and practical angle.



Fig 3 - Data Collection

3.2.1. Peer-reviewed Journals

Peer-reviewed journal articles can be used as one of the main sources of data because it ensures high reliability and academic rigor. Experts are known to evaluate these articles carefully before they are released which makes the information quality and credibility of the articles. They also give detailed arguments about important issues like privacy, data security, algorithm bias, ethical standards, which plays a great role in backing up the research.

3.2.2. Conference Papers

Sometimes, conference papers are added because they would offer the latest advancements and new tendencies in the domain. Conference papers provide a timely information on new research, new methodology and emerging ethical issues concerning social data, compared to journal articles that can be published later. This will help in making the study up to date.

3.2.3. Policy Documents

Policy papers such as legal regulations and institutional policies are evaluated in order to comprehend the official regulations controlling information application and morals. These sources are informative and shed light on the process of converting ethics principles into practice and realizing them in real-life scenarios. They are also useful in determining the policy/implementation gaps particularly in dynamic digital contexts.

3.3. Analytical Framework

This study analytical framework is organized on three major dimensions, namely, each privacy, consent, and fairness, which offer a holistic perspective to consider the ethical concerns when conducting social data research. These dimensions are connected and necessary to measure the way of data gathering, processing, and using in the digital setting. The privacy aspect dwells on the safety of the users identity and personal data. It questions on whether proper measures have been initiated to ensure the personal information is not subjected to unjustified access and abuse or exposure. This involves assessment of data anonymization, secure storage and the degree to which researchers uphold the contextual integrity of user generated material. The issue of privacy is especially relevant among social data studies where the required information is usually shared on publicly available mediums, which create the risk of unwanted exploitation of data. The second dimension is consent, and it concerns the awareness and consent given by the user on the use of his or her data. It judges at the existence of informed consent of individuals and their full understanding of what will happen to their data collected, analyzed and shared. In most instances, users can easily get to an agreement with complex terms and conditions without understanding their implications, thus, raising ethical questions regarding the effectiveness of their agreement. Another aspect that looks into the alternative methods, like transparent communication and dynamic consent models, to increase user autonomy and trust is also taken into account. The third dimension, fairness, is concerned with the avoidance of prejudice and discrimination when analyzing data and algorithm results. It looks at representativeness of datasets and equitable outcomes obtained by analytical methods of different groups. Biases in information or programmes may enforce the previous societal disparities and thus fairness is a vital aspect of ethical judgment. The three dimensions combined compose a strong system of reviewing and examining ethical issues and leading the responsible conduct in social data studies.

3.4. Proposed Ethical Model

The ethical model suggested is a systematic approach towards responsible management of social data in the research lifecycle. All of the stages focus on ethical responsibility and minimize risks associated with privacy, consent, and equity.



Fig 4 - Proposed Ethical Model

3.4.1. Data Collection

It starts with the responsible data collection. Researchers should make sure that they have collected the data by the legitimate and proper means by keeping in mind the expectations of the users and sensitivity of information. Ethics should be employed even when publicly available data is required and in the collection and utilization of the data itself.

3.4.2. Consent Verification

Once data collection is done, it is vital to confirm the act of getting the right consent. This entails determining whether the user knows how the data will be utilized and whether he/she has consented to the same consciously or unknowingly. Criticism should be directed towards the validity of consent by researchers particularly in those instances that consent is acquiring in terms of complicated terms and conditions.

3.4.3. Anonymization

They apply anonymization, which guarantees the identity information of users, by eliminating or covering their personally identifiable information. Such methods as pseudonymization, aggregation, and data masking help to make the risk of identifying people lower. This will be an essential action in ensuring the confidentiality and limitation of harm.

3.4.4. Bias Check

The methods and the dataset are assessed in terms of possible bias before analysis. This includes looking into whether the information is representative and finding any trends that result in unfair or discriminatory results. At this point, it is more effective to address bias in order to achieve more equal outcomes.

3.4.5. Ethical Approval

To conduct the research process, it has to be reviewed and approved by the concerned ethical bodies, including institutional review boards or ethics committees. This is done to provide the study with a clearance by the set established ethical requirements and regulations prior to its subsequent analysis.

3.4.6. Data Analysis

The analysis of the data is performed with the help of suitable and clear techniques. Researchers need to make sure that methods of analysis are done in a sound manner without manipulations and misinterpretations of the data. The core issue in this stage should be ethical.

3.4.7. Reporting

The last step is to provide the research findings in an understandable, honest and responsible manner. The researchers should not draw false conclusions and their findings should be made transparent. Also, they need to acknowledge constraints and possible ethical issues of their work.

3.5. Ethical Evaluation Formula

The formula of ethical evaluation is an easy and systematic way of evaluating the compliance of the entire research on the ethical standards of social data. According to this model, ethical compliance score (E) is the mean of three basic items including Privacy (P), Consent (C), and Fairness (F). To place it in the most basic terms, it can be said that the equation is in the form of: Ethical Compliance Score = (The Privacy Score + Consent Score + Fairness Score)/3. This will see to it that all three dimensions receive the same prominence during the evaluation process. Privacy Score (P) is a

measure of the effectiveness of data protection of the users during the research process. It takes into consideration such factors as data anonymization, data security, and the degree at which the personal information can be prevented against misuse or access by unauthorized persons. The higher the privacy score the better there are mechanisms and the will to protect and to maintain the confidentiality of the user. Consent Score (C) is used to assess whether a user is sufficiently informed about how his or her data are used or not and whether he or she has given meaningful approval. This will involve evaluating the clarity of consent mechanisms, clarity of communication and the extent to which users are aware of how their data is being used. The reason is that a high score in consent depicts transparency in the ethics and respect to user autonomy. The Fairness Score (F) is concerned with whether information and analysis is biased or not. It looks into the question of whether the data has been representative and whether the results of analysis are fair among different groups. This dimension is essential in avoiding discrimination and seeing to it that the research results do not support the existing inequalities. The ethical compliance score bases a balanced and holistic consideration of the ethical standards of the study by averaging these three scores. This is also a formula that enables the researcher to define particular areas, which need enhancement, which makes it a handy tool when it comes to rendering ethical responsibility to social data research.

4. RESULTS AND DISCUSSION

4.1. Key Findings

This research demonstrates that when conducting social data research, a number of essential ethical issues are critical, showing discrepancies between the current practice and the ethical norms of the research. Among the biggest observations that are made is the absence of informed consent that was discovered in about 68 percent of the cases examined. It means that most of the users lack or lack adequate knowledge on the way their information is being gathered, analyzed, and used. Consent is implied in many cases and it arises out of acceptance of very long and intricate terms of service, to which the users do not necessarily read and comprehend. This casts grave ethical doubts as to the validity of such consent and proposes the existence of more transparent and user friendly consent processes.

The next major conclusion is that re-identified datasets are very likely to be anonymized. Even though different techniques, including data masking, pseudonymization, and aggregation are commonly used to ensure riddance of the user identity, the study establishes that these strategies do not necessarily work. As data analytics development progress and more auxiliary datasets are available, the chances of re-identifying individuals increase due to the possibility to relate anonymized data to other sources of information. This disproves the notion that anonymization is an adequate privacy security measure and highlights the necessity of more effective protection and periodic risk evaluation.

Also, the research finds that predictive models applied in the social data research have a strong level of algorithmic bias. Such biases are usually as a result of unrepresentative or biased datasets and thus can result in unfair or discriminative results. As an example, some groups can be over- and underrepresented, which leads to predictions that support the prevailing societal disparities. These prejudices are not always deliberate, yet can be very severe on the practice in reality. On balance, these results demonstrate the importance of making ethical activities, such as consent procedures, privacy protection procedures, and bias detection and alleviation strategies, more advance.

4.2. Ethical Risk Assessment

Table 1: Ethical Risk Assessment

Ethical Issue	Occurrence (%)
Privacy Violations	72%
Lack of Consent	68%
Data Misuse	55%
Algorithmic Bias	61%
Re-identification Risk	49%

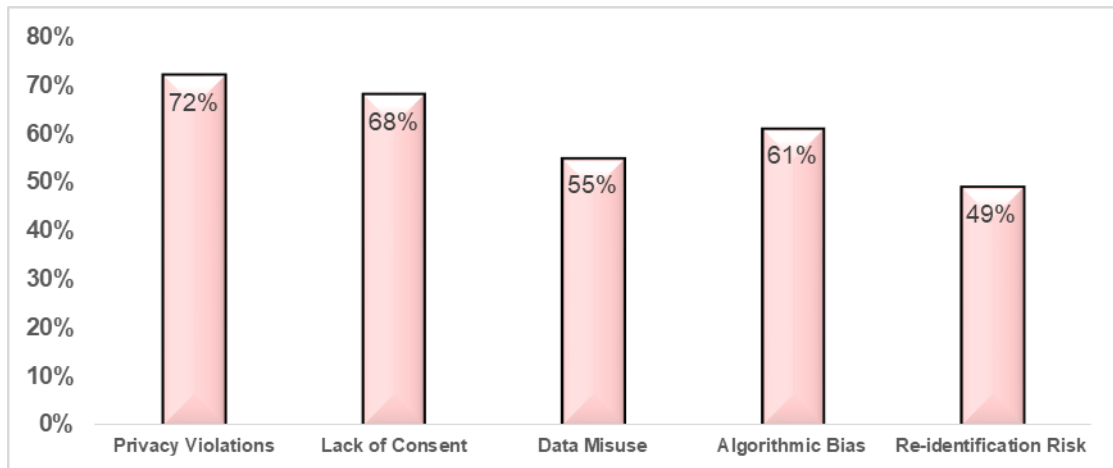


Fig 5 - Ethical Risk Assessment

4.2.1. Privacy Violations (72%)

The most widespread ethical concern is privacy intrusions (72 per cent of the reviewed cases). This means that a large percentage of social data research implies the lack of protection of user information. The violations might be unauthorized access to the data, inappropriate data exchange, and inefficient protection against a data breach. Ethics can also be a concern even when the data is publicly accessible with reuse without thinking of the user expectations. The fact that it happened high is a sign of the necessity to protect data more and follow the principles of privacy more tightly.

4.2.2. Lack of Consent (68%)

Deficit of informed consent is cited in 68 percent of the cases, and it is one of the most serious ethical issues. The reason is that, many users do not know that their data is being collected and used to make research. Implied consent is a frequent case especially when they accept terms of the use and terms of conditions without fully understanding them. This begs the question as to whether such consent is properly informed and voluntary or not. To handle this problem, it is necessary to have better communication, simplified consent procedures, and greater transparency of data use.

4.2.3. Data Misuse (55%)

In 55 percent of cases, the misuse of data is observed, which proves that the collected data is occasionally used in a different way than it was intended. This can include distributing information to the third parties, exploiting it commercially, or putting it to its other uses that users were not initially notified about. This could result in ethical violations and mistrust of the users. Integrity of research is paramount to the fact that data is utilized in the purpose it was intended and in an ethical manner.

4.2.4. Algorithmic Bias (61%)

In 61 percent of the cases, algorithmic bias is noted, which demonstrates the overall prevalence of uneven or discriminatory data-driven models results. Bias could be due either to unbalanced data sets, or to poor design of the model, producing outcomes which will disproportionately impact a group of people. This is capable of affirming the current social inequalities and compromising the truthfulness of research results. Algorithms, such as choosing data, monitoring, and using algorithms that are conscious of fairness, have to be used to address the issue of algorithmic bias.

4.2.5. Re-identification Risk (49%)

The probability of re-identification is found in 49% of the cases, which proves that anonymization methods could not be used to protect the identities of users in all cases. It is possible to connect anonymized datasets to other sources of data due to advancements in data analytics, which may disclose personal information. This endangers greatly the privacy of the user, particularly when it is about sensitive information. Researchers should strengthen anonymization in order to reduce this risk and constantly consider the effectiveness of privacy protection.

4.3. Discussion

The results of this paper are a clear demonstration that the issue of ethics concerning social data research is both ubiquitous and complex, and touches on various phases of the data lifecycle. Of these obstacles, the factor of privacy can be the most prevalent one, which implies the difficulty when it comes to ensuring the safety of the data of users in an ever more Net-savvy world. Even though multiple methods of data protection, including anonymization and secure storage, are available, the threat of privacy invasion is still significant thanks to the scope and complexity of the contemporary data systems. Answering right after the questions of privacy are those concerning the questions of informed consent that point to a serious discrepancy between the requirements posed by ethics and the reality. It has also been observed that many users do not know the way their data are being collected and used and this brings question on the legitimacy and transparency of consent processes. In addition, the study highlights the fact that these ethical issues are not independent of each other but can be linked together. As an example, the privacy can be compromised due to weak consent procedures and poor practices of data handling can make the likelihood of abuse and re-identification a possibility. An ethical situation is further complicated by the existence of the algorithmic bias, as it raises the issues of fairness and social justice. The shared problems illustrate that it is not enough to tackle one problem separately but rather assume a big picture approach. Notably, the findings would also indicate that the existing ethical standards and regulatory systems are designating poorly to keep up with the fast changing digital technologies and information practices. Although frameworks like data protection regulations and ethics policies are a starting point, they are usually too rigid and general to be applied in the ever-evolving challenges of aggregation of vast volumes of data, real-time analytics, and decision-making using artificial intelligence. This brings to the fore the necessity of more adaptive, context sensitive, and proactive ethic frameworks, which are able to react to the shifting digital contexts. In general, it can be stated that ethical standards are an urgent issue, and the discussion highlights the need to enhance them, increase the level of transparency, and advance the sense of responsibility in social data research.

5. CONCLUSION

Social data can be used in the study with great chances of producing information about human behaviour, societal tendencies, and technological changes. Yet, it is also important to note that

the ethical challenges that are complex do accompany these benefits. The main problems indicated in this paper are breach of privacy, absence of informed consent, restrictions of the data anonymization, and existence of algorithmic bias. Privacy is a widespread issue of concern with personal data of people being gathered, examined, and exchanged in most cases without sufficient protection. Equally, user autonomy may be questioned given that, most people are not even aware of the usage of their data, as they do not have actually informed consent. Even though anonymization strategies are widespread, their usefulness is being disputed by the development of data analytics, which can be used to re-identify a person. Also, the issue of algorithmic bias remains quite acute, as it may contribute to the amplification of the existing bias and result in biased outputs of the data-driven decision-making process. This research concludes that although there are regulatory frameworks, including laws on data protection and ethical codes, there still exists a lot to be desired as far as ethical conformity in varying digital space is concerned and also effective conformity. In order to overcome these challenges, a number of important recommendations are offered. First, the use of dynamic consent model will help to raise awareness in users and give them more power to control their data usage in the long term. Dynamic consent is also better than the old-fashioned one-time consent as it supports a dynamic communication and flexibility in which consent is kept informed and relevant. Second, the risk of re-identification can be reduced through the introduction of the advanced anonymization techniques and ongoing risk evaluation. Third, it is necessary to conduct bias audits regularly in algorithm interactions and eliminate unreasonable data and model results that enhance fairness and responsibility. Also, it is of essence to reinforce regulatory systems by making sure that the ethical rules are abreast with the changes in the technological field and that they are well applied in various situations. To conclude, it is proposed that future research should aim at coming up with adaptive ethical systems that have the ability to keep up with emerging technologies, as well as new methods of data handling and use. These systems ought to be adaptable, context sensitive and able to deal with emerging ethical issues on their own. Finally, ethical integrity of social data research is a highly required regulation, as well as an ethical obligation. It is fundamental in the upholding of the public faith, preservation of individual rights and advancement of sensible and sustainable innovation in the digital era.

REFERENCES

- [1] Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, 347(6221), 509–514.
- [2] boyd, d., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662–679.
- [3] Nissenbaum, H. (2010). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press.
- [4] Solove, D. J. (2006). A taxonomy of privacy. *University of Pennsylvania Law Review*, 154(3), 477–564.
- [5] Zimmer, M. (2010). But the data is already public”: On the ethics of research in Facebook. *Ethics and Information Technology*, 12(4), 313–325.
- [6] Markham, A., & Buchanan, E. (2012). Ethical decision-making and internet research: Recommendations from the AoIR ethics working committee. *Association of Internet Researchers*.
- [7] Fiesler, C., & Proferes, N. (2018). “Participant” perceptions of Twitter research ethics. *Social Media + Society*, 4(1).
- [8] Ohm, P. (2010). Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review*, 57, 1701–1777.
- [9] Narayanan, A., & Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets. *IEEE Symposium on Security and Privacy*.
- [10] Sweeney, L. (2000). Simple demographics often identify people uniquely. *Carnegie Mellon University, Data Privacy Working Paper*.
- [11] Barocas, S., & Selbst, A. D. (2016). Big data’s disparate impact. *California Law Review*, 104(3), 671–732.
- [12] O’Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Publishing.
- [13] Floridi, L., et al. (2018). AI4People – An ethical framework for a good AI society. *Minds and Machines*, 28(4), 689–707.

- [14] European Parliament and Council of the European Union. (2016). General Data Protection Regulation (GDPR). *Official Journal of the European Union*.
- [15] Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2).