

# Chapter 21

## Ethical Issues in Biomedical Text Mining



JangHyeon Lee, Vasundhara Dahiya, and Hazra Imran

**Abstract** This chapter discusses the ethical issues that must be inspected before deploying automated or smart healthcare systems, as discussed in the book. In biomedical text mining, the nature of the data, such as text, is much more objective. However, the sensitivity of such data is likely to be hampered as and when the data is used within different infrastructures, such as in the medical, technical, and social domains. The automation of natural language processing (NLP) techniques for such sensitive data can fall short of ethical, particularly social and technical, regarding data collection, modelling, and self-learning models. The user's agency needs to assure that the entire process during automation should ensure that ethics is followed. Informed consent for a patient's data becomes a sensitive area of consideration. While the identification of anonymization has been used to deal with data privacy, it is imperative to apply an ethical framework carefully. These aim to ensure the context in which the systems will be used. Despite these data issues, critical questions persist over recommender systems, information extraction systems, and automation of biased data with large models. Biases can creep into the system as easily from scientific objectivity as from social problems and systemic injustice. Which model is being deployed, in what ethical framework, and in what domain is only the starting point to dealing with such data? Hence, it is important to inspect such systems for their data, design, and the social environment in which they could be used. These socio-technical considerations, which form a humanistic approach, are necessary to ensure the smooth functioning of such powerful technology.

**Keywords** Ethical issues in biomedical domain · Data privacy · Socio-technical considerations · Bias reduction · Technical ethics design framework

---

J. Lee  
Simon Fraser University, Burnaby, BC, Canada  
e-mail: [jl465@sfu.ca](mailto:jl465@sfu.ca)

V. Dahiya · H. Imran (✉)  
Digital Humanities, Indian Institute of Technology, Jodhpur, Rajasthan, India  
e-mail: [himran@sfu.ca](mailto:himran@sfu.ca)

V. Dahiya  
e-mail: [vasundhra.dahiya@gmail.com](mailto:vasundhra.dahiya@gmail.com)

## 21.1 Introduction

This section will discuss some ethical considerations in biomedical text mining and guide researchers and practitioners to navigate these complex issues.

### *21.1.1 Importance and Challenges of Ethical Considerations in Biomedical Text Mining*

Ethics plays a crucial role in biomedical text mining (Mittelstadt and Floridi 2016; Beauchamp 2003; Hsiao and Lu 2019), as it involves handling sensitive and potentially personal data and ensuring that the research and applications derived from the analysis are fair, transparent, and respectful of individual rights. The automated extraction, analysis, and interpretation of data from sizable volumes of biomedical literature, clinical notes, and other healthcare-related papers is known as biomedical text mining. To protect individual privacy, guarantee the accuracy and dependability of results, and uphold public confidence, addressing ethical concerns in this field is imperative because it deals with sensitive and potentially life-altering information (Taylor 2019). The following are some of the major justifications for why ethical considerations are crucial and challenging in biomedical text mining:

1. **Data privacy and confidentiality:** Biomedical data often contains highly sensitive personal information, including medical history, diagnoses, and treatment plans. It is crucial to ensure that this information is anonymized and that patient identities are protected to prevent unauthorized access, use, or disclosure of patient data (Dalianis 2018). Patient privacy must be safeguarded to prevent potential harm to their personal or professional lives. Ensuring robust data security measures is essential for protecting patient privacy and maintaining trust in the field.
2. **Bias and fairness:** Biomedical text mining algorithms may inadvertently perpetuate or exacerbate existing biases in the data, leading to unfair treatment or outcomes for certain groups of patients (Celi et al. 2022). If patient information is disclosed without consent or is not adequately anonymized, it can lead to stigmatization and discrimination. Patients may be subjected to unwanted attention, judgement, and prejudice based on their health conditions or treatments. Ethical considerations must address potential biases and work towards developing more inclusive and equitable models.
3. **Quality and accuracy:** Biomedical text mining results can directly influence clinical decision-making and patient care. Therefore, ensuring the quality and accuracy of the extracted information is crucial. Researchers and practitioners must take steps to validate their findings, address potential errors, and continuously improve their methodologies to reduce the likelihood of misleading or incorrect results. However, ensuring the quality and accuracy of results in biomedical text mining is crucial but can be challenging due to the complexity and

variety of data sources and the potential for errors in data extraction or processing (Goodman 2020).

4. **Responsible innovation:** As new technologies and methodologies emerge in biomedical text mining, it is essential to consider the potential ethical implications of these advancements. Researchers must assess and address any positive and negative unintended consequences that may arise from using new tools and techniques. Ensuring that the methods and findings of biomedical text mining research are transparent and easily understandable is essential for maintaining public trust and promoting the responsible use of these technologies in healthcare (Inau et al. 2021). However, making biomedical text mining methods and results transparent and understandable can be challenging, particularly with the increasing complexity of machine learning algorithms. Researchers must develop ways to explain their models and findings in a way that is accessible to both experts and non-experts.
5. **Collaboration and data sharing:** Ethical considerations in biomedical text mining also include promoting a culture of collaboration and data sharing among researchers, institutions, and industry partners (Coetzee et al. 2021). This can help to maximize the benefits of biomedical text mining for healthcare and research while minimizing potential risks and negative consequences. However, encouraging a culture of collaboration and data sharing can be challenging due to concerns about privacy, intellectual property, and competitiveness. Developing frameworks and incentives for sharing data and knowledge while protecting the interests of all stakeholders is an ongoing ethical challenge in biomedical text mining.

Ethical considerations are crucial in ensuring biomedical text mining technologies' responsible and trustworthy use (Shah 2011). By addressing privacy, security, consent, fairness, transparency, accuracy, innovation, and collaboration concerns, researchers and practitioners can work together to advance the field while safeguarding the interests and well-being of patients and the broader community (Dalianis 2018). Also, addressing these ethical challenges requires a concerted effort from researchers, practitioners, policymakers, and other stakeholders to develop responsible and ethical practices in biomedical text mining. This will help to ensure that the field continues to advance in a way that respects the privacy, well-being, and rights of patients and the broader community.

## 21.2 Ethics

The ethics of biomedical text mining involves balancing the benefits of knowledge discovery with the potential risks to privacy, confidentiality, and informed consent of the individuals whose data is being used. This section discusses the key ethical considerations in biomedical text mining, including data privacy and confidentiality, informed consent, data ownership and access, and potential biases in data analysis. It

also highlights current guidelines and best practices researchers can follow to ensure ethical conduct when working with biomedical text data.

### ***21.2.1 Ethical Concerns at the Academia/Industry Level***

Ethical concerns arise at the academia and industry levels when dealing with data-driven applications, research, and technology development (Karimian et al. 2022). While some concerns overlap, specific challenges and priorities are unique to each domain.

#### **21.2.1.1 Concerns at the Academic Level**

1. **Research integrity:** Academic researchers must maintain high standards of research integrity, avoiding practices such as data fabrication and falsification.
2. **Ethical review and oversight:** Research projects involving human participants, animals, or sensitive data should be reviewed by independent ethics committees or institutional review boards (IRBs) to ensure compliance with ethical guidelines.
3. **Transparency and reproducibility:** Academic researchers should strive for transparency in their research methodologies, data handling, and results, ensuring their work is reproducible and can be independently verified.
4. **Open access and data sharing:** Encouraging open access to research findings and promoting data sharing can help advance scientific knowledge, but researchers must balance this with protecting participant privacy and adhering to relevant data protection regulations.
5. **Dual-use research:** Researchers should be aware of the potential dual-use implications of their work, ensuring that their research is not misused for harmful purposes.

#### **21.2.1.2 Concerns at the Industry Level**

1. **Commercial interests and biases:** Industry players must be aware of the potential biases that commercial interests can introduce into data collection, analysis, and decision-making, striving to maintain objectivity and prioritize public welfare.
2. **Data privacy and security:** Companies must protect user or customer data, ensuring that data is stored securely, used responsibly, and shared only with the appropriate consent and safeguards.
3. **Algorithmic fairness and accountability:** Industry players should prioritize developing algorithms and models that are fair, unbiased, and accountable, addressing any potential unintended consequences or discriminatory impacts that may arise from their deployment.

4. **Corporate social responsibility (CSR):** Companies should adopt CSR policies that incorporate ethical considerations into their business practices and decision-making processes, demonstrating a commitment to society's and the environment's well-being.
5. **Transparency and explainability:** Companies should strive for transparency in their data-driven technologies, ensuring that stakeholders, including customers and regulators, can understand how decisions are made and how data is used.

By addressing these concerns at both academia and industry levels, researchers, practitioners, and organizations can work together to develop and implement effective, ethical, and responsible data-driven technologies and applications, promoting the well-being of individuals and society as a whole.

### ***21.2.2 Ethical Issues with Data***

Using proprietary data for algorithm development may involve privacy and security risks, particularly when handling sensitive information (Dalianis 2018). A data breach occurs when unauthorized individuals gain access to sensitive, confidential, or protected information, often due to inadequate security measures or system vulnerabilities. The use of breached data, even if available as open-source information, raises significant ethical and legal concerns. Some of the issues related to using breached data include:

1. **Privacy violation:** Using breached data infringes on the privacy rights of individuals whose information has been compromised. Even if the data is publicly available, using it without the individuals' consent is a breach of their privacy.
2. **Legal repercussions:** Accessing, using, or disseminating breached data can lead to legal consequences, as it may violate data protection regulations, such as the General Data Protection Regulation (GDPR) in the European Union or the Health Insurance Portability and Accountability Act (HIPAA) in the United States.
3. **Loss of trust:** Utilizing breached data in research or business applications can lead to a loss of trust from stakeholders, including research participants, customers, and the general public. This can damage the reputation of researchers, companies, and institutions using the data.
4. **Potential harm to individuals:** The use of breached data can potentially cause harm to individuals whose information has been compromised, such as identity theft, financial loss, or reputational damage.
5. **Unreliable data quality:** The quality and accuracy of breached data may be questionable, as it may be incomplete, outdated, or intentionally manipulated. Using such data can lead to unreliable or misleading results in research or business applications.

Given the ethical and legal concerns associated with using breached data, utilizing such information in research or business applications is generally inappropriate.

Researchers and organizations should prioritize obtaining data through legitimate means, ensuring they have the necessary permissions and consent from individuals involved and comply with relevant data protection regulations and ethical guidelines.

### **21.2.2.1 Ethical Guidelines in Biomedical Research**

Ethical guidelines are essential in biomedical research to ensure that research is conducted responsibly and that the rights, safety, and well-being of participants and the broader community are protected (Goodman 2020; Zhang et al. 2019). Ethical guidelines outline the principles and best practices researchers must follow.

Biomedical research can potentially improve human health and well-being, but it must be conducted ethically and responsibly (Karimian et al. 2022). Researchers must assess the risk–benefit balance of their study carefully, minimizing potential harm or risks to participants while maximizing potential benefits. It is essential to ensure that the benefits and burdens of research are fairly distributed among participants and the broader society, with an equitable selection of participants and avoidance of unfair targeting or exclusion of vulnerable populations. To maintain ethical standards in biomedical research, researchers must comply with all relevant laws, regulations, and guidelines, including data protection, patient privacy, and biosafety (Ford et al. 2021). Adhering to professional codes of conduct and ethical guidelines specific to their discipline is also essential. Protecting the welfare of participants is a top priority, and researchers must take necessary precautions to ensure their safety and well-being. By upholding ethical principles, researchers can help promote trust and confidence in biomedical research and advance the understanding and treatment of human health conditions.

Therefore, protecting the confidentiality and privacy of study participants is crucial (Richterich 2018). Researchers must obtain informed consent from participants, ensuring they are fully aware of the study’s purpose, methods, risks, and potential benefits. Participants should have the right to ask questions, withdraw from the study at any time, and have their privacy and confidentiality protected. Researchers must respect participants’ autonomy and ensure they are free from coercion, manipulation, or undue influence. Authorized personnel must be the only ones who access data to safeguard the confidentiality and privacy of research participants, and adequate measures should be put in place to protect sensitive information. Researchers should take steps such as securely storing and anonymizing personal information and using it solely for the purposes it was collected to ensure that data is handled appropriately.

Additionally, researchers should inform participants about how their data will be used, stored, and shared and obtain their consent for these processes. It is crucial to ensure that personal information is kept confidential and that access is restricted only to authorized personnel (Inau et al. 2021). These measures will help build trust between participants and researchers and ensure ethical considerations are upheld throughout the process. By adhering to these ethical guidelines, biomedical researchers can ensure that their work is conducted responsibly and that the rights, safety, and well-being of research participants and the broader community

are protected. This, in turn, helps maintain public trust in biomedical research and promotes the responsible advancement of scientific knowledge.

### 21.2.2.2 Ethical Considerations in the Data Processing

The use of data in various applications, such as cancer research and drug development, raises several ethical issues and questions. Considering the ethical issues related to data collecting, data handling, as well as data modelling, the following points should be taken into account:

#### 1. Data collecting

- **Informed consent:** Ensure that participants are fully informed about the purpose, methods, risks, and potential benefits of data collection and that they can withdraw at any time.
- **Privacy:** Protect participants' privacy during data collection using techniques such as anonymization and encryption.
- **Fair representation:** Strive to collect data from diverse and representative samples to minimize biases and ensure that the resulting models are fair and generalizable.

In cancer research and drug development, informed consent is crucial for collecting patient data. Researchers must ensure that participants understand how their data will be used, stored, and shared and the potential risks and benefits involved (Dalianis 2018). Also, protecting patient privacy and maintaining confidentiality is essential in cancer research and drug development, as sensitive health information can be collected. Ensuring data security and implementing privacy-preserving techniques, such as anonymization and encryption, is critical.

#### 2. Data handling

- **Transparency:** Be transparent about the data handling and cleaning processes, clarifying how missing or erroneous data points are treated and how selection biases are addressed.
- **Minimize biases:** Identify and address potential selection biases that could impact the fairness or accuracy of the resulting models. This might involve oversampling underrepresented groups, re-weighting data points, or using other techniques to correct for biases.
- **Data provenance:** Maintain a clear record of data provenance to ensure that data sources, transformations, and cleaning processes are well-documented and traceable.

In the biomedical domain, data-driven algorithms can be prone to bias, leading to unfair treatment or outcomes for certain individuals or groups (Rios et al. 2020). Ensuring fairness and addressing potential biases in data collection, analysis, and algorithm development is essential. Here, transparency in data collection, analysis, and decision-making processes is crucial to maintaining trust in these applications.

### 3. Data modelling

- **Algorithmic fairness:** Choose models and techniques less prone to bias and promote fairness in predictions or decision-making. Consider using fairness metrics and techniques to assess and mitigate biases in the model.
- **Explainability and interpretability:** Prioritize models and techniques that are interpretable and explainable, enabling stakeholders to understand the rationale behind predictions or decisions.
- **Model validation:** Rigorously validate models using appropriate techniques, such as cross-validation or holdout validation, to ensure their accuracy, generalizability, and robustness.
- **Ethical AI:** Incorporate ethical considerations into the development and deployment of AI and machine learning models, addressing potential unintended consequences and ensuring transparency, accountability, and fairness.

Developers should strive to ensure that their algorithms and models do not perpetuate bias and discrimination, which can harm certain populations (Cirillo et al. 2020). This can involve assessing and correcting bias in training data, regularly monitoring performance for bias, and using fairness-aware machine learning techniques. Users should also ensure that their systems are transparent and explainable. This can involve providing detailed documentation of the data sources, algorithms, and methods used to generate results, enabling end-users to understand and interpret the results.

By carefully considering these aspects of data collection, handling, and modelling, researchers and practitioners can develop data-driven applications that are accurate and effective but also ethical and responsible, taking into account the rights and well-being of individuals and society.

## 21.3 Solutions for Ethical Issues with Data

There is a growing demand for robust predictive models in healthcare that can improve diagnosis, prognosis, and treatment while protecting patients' privacy (Minot et al. 2021). Balancing the need for accurate prediction models with privacy concerns can be challenging, but several strategies can help address these issues:

1. **Data anonymization:** Data anonymization techniques, such as data masking, pseudonymization, or aggregation, can help remove or obscure personally identifiable information (PII) from patient data while maintaining its utility for building predictive models. Ensuring anonymized data complies with relevant privacy regulations, such as GDPR or HIPAA, is crucial.
2. **Data minimization:** Collecting and processing only the minimum necessary data required for building predictive models can help reduce the risk of privacy breaches. This approach involves identifying the essential data elements for the specific use case and limiting the collection and processing of additional data.



3. **Differential privacy:** Differential privacy is a mathematical framework that adds carefully calibrated noise to the data or query results, making it difficult to re-identify individuals while preserving the overall utility of the data for analysis. This technique can enable the development of predictive models while maintaining strong privacy guarantees for individuals.
4. **Federated learning:** Federated learning is a decentralized approach to machine learning where models are trained on local datasets at multiple sites, and only model updates or aggregated parameters are shared, rather than raw patient data. This approach helps protect patient privacy while still enabling the development of robust predictive models.
5. **Privacy-aware evaluation metrics:** When evaluating the performance of predictive models, it is essential to consider both the predictive accuracy and the privacy risk. Developing and using privacy-aware evaluation metrics can help ensure that models balance these competing objectives correctly.

Addressing the ethical considerations of data privacy and confidentiality in biomedical text mining requires various solutions. By implementing these solutions, healthcare professionals and researchers can promote the responsible use of data and maintain patient trust.

### **Positive Aspects of Data Ethics**

Adhering to data ethics principles in using data and technology can positively impact individuals, organizations, and society (Lhotska 2022). First, good data ethics helps build trust and credibility among stakeholders, including research participants, customers, regulators, and the general public. This trust is crucial for maintaining a positive reputation and fostering long-term relationships. Compliance with relevant data protection regulations, such as GDPR or HIPAA, also helps organizations avoid legal issues, financial penalties, and reputational damage.

In addition, ethical data practices lead to higher-quality data and more reliable, unbiased algorithms and models, enabling better-informed decisions that benefit individuals, businesses, and society. Ethical considerations also promote fairness and equity in data-driven applications, ensuring that algorithms and models do not perpetuate biases or discrimination and that the benefits of these technologies are accessible to all (Cirillo et al. 2020). Furthermore, prioritizing protecting individual rights, such as privacy, autonomy, and consent, ensures that individuals' well-being and dignity are safeguarded.

A strong commitment to data ethics also promotes social responsibility, encouraging organizations to consider the broader social and environmental implications of their actions, promoting corporate social responsibility, and a commitment to the well-being of society and the environment. This, in turn, can foster a culture of innovation and collaboration as organizations and researchers work together to develop new methods, technologies, and applications that prioritize ethical considerations (Celi et al. 2022). Finally, addressing potential ethical risks and challenges proactively ensures the long-term sustainability of data-driven technologies and applications, avoiding potential pitfalls or controversies arising from unethical practices.

Data ethics principles benefit all stakeholders and contribute to a more responsible, equitable, and sustainable digital ecosystem.

## 21.4 Bias Cases

Discrimination and bias in biomedical text mining can take different forms, leading to healthcare research and delivery disparities. Racial and gender biases are two examples of how bias can emerge in biomedical text mining (Marshall et al. 2014; Flórez-Vargas et al. 2016; Bolukbasi et al. 2016).

Biased data sources, bias in clinical trials, bias in the literature, and algorithmic bias are potential sources of discrimination and bias in biomedical text mining for racial and gender minorities (Celi et al. 2022). For instance, if text mining data sources or clinical trials disproportionately represent one race or gender, the resulting models and analyses may not accurately represent underrepresented populations' experiences and health outcomes. Selective reporting of positive results or underrepresentation of research involving certain groups, such as minorities or women, in the biomedical literature can impact the quality and representativeness of data utilized in text mining (Bailey et al. 2022).

Additionally, machine learning models and natural language processing techniques used in biomedical text mining can inadvertently learn and perpetuate biases in the training data, resulting in discriminatory predictions or recommendations that disproportionately affect certain racial or gender groups (Zhao et al. 2019; Cirillo et al. 2020).

To mitigate discrimination and bias in biomedical text mining, researchers and practitioners should utilize diverse data sources that represent a range of racial, ethnic, and gender groups, address potential biases in the literature, utilize fairness-aware machine learning techniques to identify and mitigate biases in algorithms and models, and continuously monitor and evaluate model performance for potential disparities in their predictions or recommendations (Dalianis 2018). Additionally, collaborating with experts from diverse fields, such as gender studies or social sciences, can provide valuable insights into potential biases and challenges related to race and gender in biomedical text mining. By prioritizing equity and inclusivity in biomedical text mining, researchers and practitioners can contribute to more accurate and reliable healthcare research and delivery for individuals from all backgrounds.

## 21.5 Regulations

The regulation of data ethics in biomedical text mining is part of broader efforts to ensure algorithmic accountability and data protection in various domains, including the biomedical field (Inau et al. 2021). Here, we discuss these efforts and their implications for biomedical text mining in the US, EU, and India:

1. **United States:** Algorithm Accountability Act: Introduced in 2019, the Algorithmic Accountability Act requires companies to evaluate and address the impacts of their automated decision systems and data processing practices, including potential biases, inaccuracies, and privacy risks (Richterich 2018). While this act does not specifically target biomedical text mining, its implications extend to this field. Companies involved in biomedical text mining must ensure that their algorithms and data processing practices adhere to fairness, accuracy, and transparency principles and protect users' privacy.
2. **European Union:** General Data Protection Regulation (GDPR): The GDPR is a comprehensive data protection regulation that came into effect in 2018. It has significant implications for biomedical text mining, as it sets strict standards for data privacy, consent, and transparency. Organizations engaged in biomedical text mining must ensure that they obtain explicit consent for processing personal data, including health data, and must provide clear explanations of their data processing practices. The GDPR also mandates that organizations implement privacy by design and privacy by default, which requires considering data protection and privacy throughout developing and deploying text mining systems (Richterich 2018).
3. **India:** Department of Science and Technology (DST) and Department of Biotechnology (DBT): In India, agencies such as the DST and DBT are responsible for promoting and regulating research and innovation in various fields, including biotechnology and data science. Although these agencies do not have specific regulations targeted at biomedical text mining, they play a role in setting guidelines and promoting best practices for research and data management in the biomedical domain. Organizations engaged in biomedical text mining in India must adhere to the broader data protection and privacy guidelines established by these agencies and other relevant regulations, such as the Information Technology Act and the Personal Data Protection Bill (when enacted).

In conclusion, although no specific regulations exclusively target biomedical text mining, existing data protection and algorithmic accountability regulations in the USA, EU, and India have implications for this field. Organizations involved in biomedical text mining must ensure that their practices align with these regulations and prioritize data privacy, transparency, and fairness throughout the development and deployment of their systems.

## 21.6 Need for Standardization

There is a growing need for harmonized global regulation of ethics, particularly in the context of data-driven technologies such as biomedical text mining, AI, and machine learning. As these technologies become increasingly integrated into various aspects of society, their ethical implications must be addressed consistently and effectively

across different countries and regions (Inau et al. 2021). Some of the reasons for pursuing harmonized global regulation of ethics include the following:

1. **Cross-border data flows:** Data-driven technologies often involve data transfer across national borders. Inconsistent or conflicting regulations between countries can create barriers to data sharing and collaboration, hampering innovation and developing new technologies.
2. **Consistency and interoperability:** Harmonized global regulations can ensure consistency in the ethical principles and guidelines applied to data-driven technologies, making it easier for organizations to comply with regulations and fostering interoperability between systems developed in different countries.
3. **Preventing regulatory arbitrage:** Organizations may engage in regulatory arbitrage without harmonized global regulations, choosing to operate in jurisdictions with lax regulations to avoid more stringent ethical requirements. A global framework can help prevent such practices and ensure a level playing field for all organizations.

While achieving harmonized global regulation of ethics is a complex and challenging task, it is essential to ensure the responsible and equitable development and deployment of data-driven technologies, including biomedical text mining, AI, and machine learning. Multilateral organizations, such as the United Nations, World Health Organization, and international standards bodies, can play a crucial role in fostering dialogue and cooperation between countries to develop and implement a harmonized global ethical framework (Richterich 2018). Ultimately, harmonized global regulations can facilitate international collaboration in research and development, enabling researchers and organizations to work together more effectively to address ethical challenges and develop innovative solutions.

## 21.7 Different Treatments

Ethical concerns in the biomedical field cannot be treated with less regulation than in other domains. In fact, given the significant potential impact on human health and well-being, the biomedical field often requires more rigorous ethical oversight than other areas (Karimian et al. 2022). Ethical considerations in the biomedical field are particularly important because they involve human subjects, including patients, study participants, and healthcare providers. The potential risks of biomedical research and interventions can be severe, including physical harm, psychological trauma, and privacy infringement.

Moreover, the complexity and variability of biomedical data and research necessitate careful ethical considerations. Biomedical data may include sensitive information, such as genetic information, medical histories, and biometric data, which require strict safeguards to protect privacy and confidentiality (Dalianis 2018). Biomedical research involves complex ethical considerations, including issues related to

informed consent, the use of placebos, risk–benefit assessments, and the potential impact on vulnerable populations.

Therefore, regulatory bodies must ensure that biomedical research and interventions adhere to strict ethical standards to protect human subjects’ rights, dignity, and welfare. However, regulations in the biomedical field must be designed to foster innovation and progress while ensuring ethical considerations are met (Ford et al. 2021). Balancing innovation and progress with ethical considerations can be challenging, but it is essential to ensure that new technologies, such as biomedical text mining, are developed and deployed responsibly and ethically.

In summary, ethical concerns in the biomedical field cannot be treated with less regulation than in other domains. Instead, ethical considerations in the biomedical field often require more rigorous oversight and strict adherence to ethical standards to protect human subjects’ rights, dignity, and welfare. Regulations must balance fostering innovation and progress while ensuring that ethical considerations are met to promote the responsible development and deployment of new technologies.

## 21.8 Recommended Practice

A technical ethics design framework provides a structured approach for incorporating ethical considerations into developing, deploying, and evaluating technology, including data-driven applications and AI systems (Ford et al. 2021; Samberg et al. 2021). Here is a suggested framework that can be adapted to specific contexts and industries:

1. **Define ethical principles and values:** Begin by identifying the core ethical principles and values relevant to the technology being developed, such as fairness, accountability, transparency, privacy, and autonomy. These principles should align with industry standards, regulations, and societal expectations.
2. **Stakeholder analysis:** Identify all stakeholders who may be affected by the technology, including users, customers, employees, regulators, and the broader community. Assess their needs, interests, and concerns, and determine how the technology may impact them positively and negatively.
3. **Risk assessment:** Conduct a thorough risk assessment to identify potential ethical risks, harms, and unintended consequences that may arise from the technology. Consider privacy violations, algorithmic bias, discrimination, or data misuse.
4. **Ethical Design and Development:** Integrate ethical considerations into the design and development process, ensuring the technology aligns with the defined ethical principles and values. This may involve privacy-preserving data methods, fairness-aware machine learning, or explainable AI.
5. **Education and training:** Provide education and training for employees, developers, and other stakeholders on the ethical principles, guidelines, and policies related to the technology. This helps to ensure that ethical considerations are understood and applied consistently throughout the organization.

6. **Monitoring and evaluation:** Establish mechanisms for monitoring and evaluating the ethical performance of the technology, including ongoing assessments of its impacts on stakeholders, adherence to ethical guidelines and policies, and the identification and mitigation of new risks or harms. For example, regularly evaluate the performance and impact of the system, including its accuracy, sensitivity, and specificity. This can help identify potential biases, inaccuracies, or unintended consequences, enabling developers and users to make necessary adjustments.
7. **Continuous improvement and adaptation:** Regularly update and refine the technical ethics design framework based on feedback, new insights, or changes in regulations, industry standards, or societal expectations. This ensures the technology remains ethically aligned and adapts to evolving needs and concerns.
8. **Open science practice:** Ethical guidelines should promote collaboration and sharing of data, algorithms, and tools while respecting privacy and intellectual property rights. Open science practices can help accelerate discoveries and ensure biomedical text mining benefits the global community.

By following these practices, developers and users of biomedical text mining can contribute to the responsible and ethical development and deployment of this technology, promoting the advancement of biomedical research and ultimately improving healthcare outcomes for all.

## 21.9 Summary

In conclusion, ethical considerations play a critical role in shaping the future of biomedical text mining. As this field advances, researchers, practitioners, and other stakeholders must recognize the importance of addressing ethical concerns and challenges to ensure these powerful technologies' responsible development and use.

Throughout this book, we have explored the significance of ethical considerations in biomedical text mining, the challenges associated with algorithm development, data privacy, and potential solutions to mitigate these concerns. We have also discussed the need for ethical guidelines in biomedical research and the unique ethical issues in various applications.

It is vital to emphasize the importance of ethical practices at academic and industry levels, as a collaboration between these sectors is crucial in addressing data breaches and fostering positive aspects of data ethics. By implementing a technical ethics design framework, we can work to reduce biases, improve transparency, and ensure that the benefits of biomedical text mining are shared equitably among all stakeholders.

Regulations and standardization efforts are pivotal in guiding the responsible development of biomedical text mining technologies. The need for harmonized policies and guidelines becomes increasingly apparent as we navigate the complexities of different treatments and practices across various jurisdictions.

By fostering a deeper understanding of these issues and encouraging ongoing dialogue among researchers, practitioners, policymakers, and the public, we can collectively work towards a future where biomedical text mining technologies are developed and utilized responsibly, transparently, and equitably. As we embrace the potential of these innovations, let us remain vigilant in our commitment to upholding the highest ethical standards and ensuring the well-being of all individuals and communities affected by their applications.

## References

- Bailey AH, Williams A, Cimpian A (2022) Based on billions of words on the internet, people = men. *Sci Adv* 8(13):eabm2463
- Beauchamp TL (2003) Methods and principles in biomedical ethics. *J Med Ethics* 29(5):269–274
- Bolukbasi T, Chang KW, Zou JY, Saligrama V, Kalai AT (2016) Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In: *Advances in neural information processing systems*, p 29
- Celi LA, Cellini J, Charpignon ML, Dee EC, Dernoncourt F, Eber R, Yao S et al (2022) Sources of bias in artificial intelligence that perpetuate healthcare disparities—a global review. *PLOS Digit Health* 1(3):e0000022
- Cirillo D, Catuara-Solarz S, Morey C, Guney E, Subirats L, Mellino S, Mavridis N et al (2020) Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. *NPJ Digit Med* 3(1):81
- Coetzee T, Ball MP, Boutin M, Bronson A, Dexter DT, English RA, Terry SF et al (2021) Data sharing goals for nonprofit funders of clinical trials. *J Particip Med* 13(1):e23011
- Dalianis H (2018) Ethics and privacy of patient records for clinical text mining research. *Clin Text Min Second Use Electron Patient Rec* 97–108
- Flórez-Vargas O, Brass A, Karystianis G, Bramhall M, Stevens R, Cruickshank S, Nenadic G (2016) Bias in the reporting of sex and age in biomedical research on mouse models. *Elife* 5:e13615
- Ford E, Shepherd S, Jones K, Hassan L (2021) Toward an ethical framework for the text mining of social media for health research: a systematic review. *Front Digit Health* 2:592237
- Goodman KW (2020) Ethics in health informatics. *Yearb Med Inform* 29(01):026–031
- Hsiao YW, Lu TP (2019) Text-mining in cancer research may help identify effective treatments. *Transl Lung Cancer Res* 8(Suppl 4):S460
- Inau ET, Sack J, Waltemath D, Zeleke AA (2021) Initiatives, concepts, and implementation practices of FAIR (findable, accessible, interoperable, and reusable) data principles in health data stewardship practice: protocol for a scoping review. *JMIR Res Protoc* 10(2):e22505
- Karimian G, Petelos E, Evers SM (2022) The ethical issues of the application of artificial intelligence in healthcare: a systematic scoping review. *AI Ethics* 2(4):539–551
- Lhotska L (2022) Ethics in engineering education. In: *2022 31st annual conference of the European association for education in electrical and information engineering (EAEEIE)*. IEEE, pp 1–5
- Marshall IJ, Kuiper J, Wallace BC (2014) Automating risk of bias assessment for clinical trials. In: *Proceedings of the 5th ACM conference on bioinformatics, computational biology, and health informatics*, pp 88–95
- Minot JR, Cheney N, Maier M, Elbers DC, Danforth CM, Dodds PS (2021) Interpretable bias mitigation for textual data: reducing gender bias in patient notes while maintaining classification performance. *arXiv preprint arXiv:2103.05841*
- Mittelstadt BD, Floridi L (eds) (2016) *The ethics of biomedical big data*, vol 29. Springer
- Richterich A (2018) *The big data agenda: data ethics and critical data studies*. University of Westminster Press, p 154

- Rios A, Joshi R, Shin H (2020) Quantifying 60 years of gender bias in biomedical research with word embeddings. In: Proceedings of the 19th SIGBioMed workshop on biomedical language processing, pp 1–13
- Samberg R, Vollmer T, Althaus S, Bamman D, Benson S, Butler B, Worthey G et al (2021) Building legal literacies for text data mining
- Shah N (2011) Ethical issues in biomedical research and publication. *J Conserv Dent JCD* 14(3):205
- Taylor, H. A. (2019). An Overview of Ethics and Public Health Data Collection. *The Oxford Handbook of Public Health Ethics*, 315.
- Zhang Y, Chen Q, Yang Z, Lin H, Lu Z (2019) BioWordVec, improving biomedical word embeddings with subword information and MeSH. *Scientific Data* 6(1):52
- Zhao J, Wang T, Yatskar M, Cotterell R, Ordonez V, Chang KW (2019) Gender bias in contextualized word embeddings. arXiv preprint [arXiv:1904.03310](https://arxiv.org/abs/1904.03310)