[Invitation To Comment] Public Consultation Paper: Ethical, Legal and Social Issues Arising from Big Data and Artificial Intelligence (AI) Use in Human Biomedical Research

a. Responsible Data Usage

1   Responsible data usage ensures that data is used in a fair and transparent manner without compromising data integrity. This protects individual privacy and control of personal data. It also reduces the risk of discrimination or injustice, or inaccurate research outcomes as a result of bias in AI algorithms and the data used.

Ethical Considerations:

   i.   *How should human biomedical research data be managed and used responsibly to avoid the risk of discrimination arising from participation or biases in biomedical research?*
  ii.   *How should researchers or developers ensure that data fed into algorithms is not biased and will not result in bias-driven outcomes when developing AI algorithms?*
 iii.   *What is the ethical significance of 'social licence' in ensuring responsible big data and AI use in biomedical research?*

b. Data Ownership, Custodianship, and Stewardship

2   Data ownership may be referred to as legal rights to have complete control over data elements. While data owners have responsibilities to ensure data security and accessibility, data custodianship is an assignment of responsibility by the data owner to the custodian to manage the use and access to the data through safe keeping of data (i.e., safe custody and storage). Data stewardship complements data custodianship and involves coordination with multiple parties to develop and implement policies for managing and sharing data with responsible third parties in the interests of biomedical research, as well as the training and educating of stakeholders about the importance of responsible data management.

Ethical Considerations:

   i.   *How much control/right/power do individuals have after contributing their data for biomedical research?*
  ii.   *What are the responsibilities of a data custodian?*
 iii.   *Should a person who had provided biological materials or data but did not participate in the subsequent processing or analysis have intellectual property rights in the data?*
  iv.   *How should large volumes of biomedical data shared across countries in international research collaborations be managed?*

c. Data Privacy, Accessibility and Security

3   Data privacy, accessibility and security frameworks will be necessary to ensure that individuals' personal data rights and interests are protected through robust technical security systems while facilitating access to the protected data by legitimate third parties.

Ethical Considerations:

i.   *How can providers/researchers ensure robust data security and proper access control while maintaining data privacy?*
ii.   *How can institutions or organisations managing data stored in multiple on-site servers or cloud repositories ensure appropriate data accessibility?*

d. Data Anonymisation, De- and Re-identification of Data

4   Anonymisation, de- and re-identification of data are tools used in biomedical research to enable data to be analysed while protecting the data contributors' privacy. As analyses involving big data and AI algorithms in biomedical research rely increasingly on large volumes of personal, health and medical data, conventional methods of de-identification and anonymisation may no longer be adequate in protecting privacy.

5   In Singapore, whole genomic data[1] is considered identifiable personal data if linked to identifiers or indirect identifiers, or on its own. It is not considered as personal data if the data is anonymised and protected against re-identification. In our local context of healthcare and biomedical sciences research where appropriate levels of data treatment[2] and managerial controls[3] are in place, whole genomic data that has been de-identified is considered as anonymised data.

Ethical Considerations:

i.   *Are current methods of anonymisation and de-identification still applicable when large volumes of personal, health and medical data are used in big data and AI research?*
ii.   *Should genetic data be considered exceptional and treated differently from other types of personal and health data?*

e. Revisiting Consent in the Arena of Big Data and AI

6   With advancements in technology, health information may be derived from novel methods such as consumer platforms, social media, wearables and sensors. Consent for mobile data collection is often carried out through the internet of things (IoT) or edge computing, which is also known as 'instant' data or real-time data generated by sensors or users where data can be processed more quickly and closer to where it is generated, allowing real-time interaction with users. This is unlike cloud computing which processes large datasets on centralised remote servers, making real-time interaction difficult. In such scenarios, taking

---

[1] Genomic data is the DNA data of organisms that reveal how differences in DNA affect human health and disease. Whole genomic data i.e., Whole Genome Sequence refers to the entire DNA of the genome of an organism and Whole Exome Sequence refers to the protein coding sequences from the genome.
[2] Examples of data treatment controls include removal of direct identifiers such as national identification numbers, names, and addresses, and generalisation of indirect identifiers such as age and demographic data.
[3] Examples of managerial controls include applying access controls such as physical, technical and/or administrative measures to restrict access to only authorised parties, designing approval processes and process controls that minimise risks of collusion, and prohibiting unauthorised re-identification of individuals.

informed consent where patients and individuals can understand clearly and sufficiently how their data will be used in biomedical research becomes challenging.

7   In cohort studies which typically involve longitudinal studies on a group of individuals to study the development of diseases over long periods of time, ensuring that the rights, interests, and welfare of research participants are protected often entails obtaining consent from participants for usage of data prior to their participation in research. However, obtaining consent for the use of real-world data such as electronic health records, insurance claims data, and data from health-monitoring devices can be more challenging, because the data is often not collected primarily for research purposes.

Ethical Considerations:

   i.   *What are the differences between consent taking for health and medical data that are collected via various sources and novel methods?*
   ii.   *What are the differences between consent for use of data between cohort studies and that of real-world data?*

f. Responsibility to the Public in Data-Sharing for Research

8   Responsible data sharing can lead to research that benefits individuals, communities, and society. The challenge lies in ensuring that data sharing for research is conducted ethically, equitably, and with proper respect for privacy.

Ethical Consideration:

   i.   *How can the benefits of biomedical research be shared with participants whose data is used?*

g. Ethical Considerations and Issues Specific to AI

9   Ethical principles that may apply to AI in biomedical research include transparency[4], explainability[5], and justifiability[6]. Other key considerations include reliability and safety, accountability, human agency, equitable access and model security to minimise potential harm to individuals and parties involved in research projects.

Ethical Considerations:

   i.   *What are the ethical considerations and responsibilities of biomedical researchers and AI developers in complying with best standards to ensure clinical safety of AI models?*
   ii.   *What are the ethical considerations in ensuring equitable access to AI technologies in research?*

---

[4] Transparency refers to the need to openly inform and communicate with various stakeholders, at all stages of the AI system's development and implementation, how the AI system is designed, developed, and applied.
[5] Explainability of AI systems refers to the interpretability of input, output, and behaviour of the AI model and how it contributes to the outcome of the prediction.
[6] An explainable AI model enables biomedical researchers, users, and patients to understand the AI models' predictions and how the outcomes were derived.