**Using Various Lenses to Understand Big Data Ethics**

David Broom

Northcentral University

TIM\_8530: Big Data Development

Dr. Joseph Issa

August 10, 2023

**Using Various Lenses to Understand Big Data Ethics**

**Introduction**

Big data has many ethical and practical issues that must be addressed.  The use of big data can affect the public, government, and environment.  The agency of individuals and their free will can be compromised with big data agendas and purposes.  There are many aspects and elements to consider on how big data can make an impact like transparency, privacy, consent, regulation, and economic evaluation.

A philosophical and practical approach can be used to evaluate the impact of most big data ethical aspects.  Utility, moral duties, honesty, integrity, and responsibility are some of the key philosophical components to big data ethics.  An individual’s rights and how big data affects their privacy, and security is an important consideration in big data ethics.  Practical strategies of big data ethics involve police and procedures  that make responsible handling of big data.

**Privacy and Consent**

             Privacy can be described as an individual’s right to control the collection and use of that individual’s data.  A big concern in big data regulation is an individual’s right to control their personal information.  Big data technologies allow for a vast collection of information and privacy concerns are becoming a real concern for organizations and individuals.  It is important to have safeguards needed to protect the privacy of individuals.

It is important in the context of big data, to obtain informed consent.  There should be some form of effective communication with an individual about any shared data and what purpose and uses collected data are serving.  Individuals should be made aware of their abilities to and right concerning control of their personal data.

Two other types of consent are called granular and revocable consent.  Granular consent offers individuals choices about how their personal data is shared.  This type of consent is more empowering and provides more control over how their data is shared.  Revocable consent gives individuals the ability to withdraw any permissions to share or collect data.  With revocable consent individuals have more control of their information in the data life cycle.

An imperative ethical aspect of big data is the balance between the use of big data and the privacy concerns that come with big data.  A breach of trust and potential harm can be the results of poor consent procedures.  Another issue that can occur with big data is legal consequences for organizations because of privacy breaches.

**Data Security**

             Data breaches and encryption are two important aspects to big data ethics.  Data security can be assessed by the ability to protect confidentiality, integrity, and availability of large amounts of data.  As big data grows in size and complexity so does the need for robust security measures.  Access controls, encryption, secure data storage, and some of the elements of security to help protect the users of big data.

With access controls only authorized users of a data set will be able to view, modify and interact with the data set.   Access controls prevent unauthorized data breaches.  Encryption converts data into a code so that it can’t be accessed by unauthorized users.  Both data at rest and data in transit should be encrypted in a big data environment (Budke & Ferguson, 2017).

Continuously monitoring data helps keep data secure and will find any unauthorized users.  It would be important to keep regular security logs and perform audits of system access.   Big data is secured across many nodes and servers, and it would be very important for all locations to be secure. Each node and server must be configured and monitored to protect against A diagram of data ethics

Description automatically generatedvulnerabilities.

Figure An example of a big data ethics framework.

https://ethicalboardroom.com/information-governance-achieving-data-ethics-privacy-and-trust/

             Some of the challenges to big data security are scale and complexity, data governance, regulatory compliance, and a dynamic threat landscape.  User awareness and training is a critical part of a security strategy for employees and users involved with big data.  The landscape of cybersecurity is constantly changing and new threats and vulnerabilities are reasons to stay up to date with common security practices.

**Bias and Fairness**

Discriminatory and unfair outcomes from data driven processes can be referred to as the bias and fairness of big data ethics.  Identifying and mitigating any biases that are present in algorithms or data used for decision making can help make more equitable systems with just outcomes.  Several different biases classified as algorithmic bias, data bias, migration strategies and fairness metrics are used to categorize various areas in big data where unfairness can be identified.

Algorithms for big data analysis can pick up bias from the data that is involved in the training.  These biases can be discriminatory and produce unfair results, particularly for data sets that are reflecting a historical bias.  A data bias can occur because of historical and societal factors.  An example would be that a hiring data set would contain mostly males  and the algorithm trained off that set would give more favor to males.

Using fairness metrics and frameworks to quantify and assess how the decisions from algorithms are treating users can result in less disparity of different demographic groups.  To identify and address bias migration strategies can be used to promote fairness.  Some migration strategies include re-sampling, re-weighing, and fairness aware algorithms which can ensure certain groups are not unfairly disadvantaged.

Some of the main changes to big data bias and fairness are data quality, intersecting biases, tradeoffs, and transparency.  It’s important to be transparent on how algorithms work and make decisions.  Achieving better fairness can consider tradeoffs with other desirable objectives like accuracy.  It is possible for multiple biases to intersect and build complexity to the identification of fairness concerns.

**Transparency, Explainability, and Exploitability**

The clarity and comprehension of the processes, algorithms, and decision-making systems that are involved in big data analysis are referred to as the transparency and explainability of big data ethics.  These aspects of transparency and explainability should be understandable to stakeholders, individuals, regulators, and the public.  There are two types of transparency which are model transparency and algorithmic transparency.

The transparency of data driven algorithms of machine learning involves making the inner workings more accessible and understandable.  The features and variables that are weighted in making predictions should be transparent.  Understanding how an algorithm operates including specific computations that occur during analysis is referred to as algorithmic transparency.  Providing clear and interpretable explanations of outputs and decisions made by data driven systems is called explainability.

A diagram of a diagram

Description automatically generated

Figure A diagram of big data ethics concepts.

https://research.aimultiple.com/data-collection-ethics/

Documenting and keeping records of data processing and analysis is also under the context of transparency.  Transparency ensures relevant parties can review, audit and understand any processes involved in big data.  Trust and accountability are two elements that improve with transparency.  Transparency will help build trust with the public and stakeholders.  Also transparency can hold organizations and individuals accountable for the decisions made on analysis (O’Mathúna & Iphofen, 2022).

Exploitability occurs when data is accessed, used, or shared in ways that violate privacy, consent and legal standards.  Data misuse, security vulnerabilities, and third party access are some characteristics of exportability in big data.  It is important to respect privacy and consent which can help minimize exploitability in big data.  Reducing the level of exploitability can prevent harm and negative consequences which result from the unauthorized use of data.

**Ownership and Control**

A diagram of a model framework

Description automatically generated             The rights and authority over data that is stored, collected, and processed is referred to the ownership and control of big data.  Ownership and control involves a clear framework for who owns the data, who can access the data, and how the data can be used.  To elaborate and label aspects of ownership and control it is important to include data ownership, user control, access permissions, and data licensing and agreements.

Figure A layout of Big Data Governance

https://www.researchgate.net/figure/Strategy-to-Govern-Open-Data-Ownership\_fig1\_319018708.

The legal rights of individuals or organizations who claim possession of data that is generated and collected are referred to as data ownership.  Individual agency, ability and access to modify or delete their own data makes reference to the term user control.  It is important for organizations to establish protocols for recovering and granting access to certain data sets.  Access permissions along with data licensing can recognize and uphold individual rights and avoid exploitation of individuals.

Some challenges and considerations for data ownership and control include data sharing and collaboration, legal and regulatory frameworks, and emerging technologies.  It is difficult to balance the need to share data for research, innovation and public good with ownership of that data.  It is important to establish clear agreements and protocol for data sharing.  Due to different jurisdictions having different laws complexities in data ownership will arise.  Organizations established in different regions can have a difficult time navigating through opposing and conflicting laws and rules.

Emerging technologies like block chain decentralized data ownership is being explored for solutions to problems with data ownership.  Understanding these new technologies is not easy and implementation can be difficult as well.  Establishing a good framework for ownership and control is essential for respecting individuals rights, preventing data exploitation, and making responsible data practices.

**Regulation and Compliance**

A diagram of data governance

Description automatically generatedTwo crucial aspects of big data ethics are regulation and compliance.  Adhering to legal frameworks, industry standards and ethical guidelines can greatly help protect the privacy of individuals and help maintain trust.  Regulation and compliance can also promote transparency and ensure responsible data handling and usage.  The government and industry bodies will set regulatory frameworks and guidelines to govern the collection, storage, processing, and sharing of data.

Figure Data Governance And Security Diagram

https://www.imperva.com/learn/data-security/data-governance/

An important example of a big data regulation is the General Data Protection Regulation (GDPR) set forth by the European Union.  This regulatory framework was set forth in 2018 and provides a comprehensive set of regulations that will help protect the rights and privacy of individuals.  The GDPR requires organizations to obtain explicit consent and provide transparency about data usage.  The regulatory legislation requires measures of security implementation (Giacalone et al., 2022).

Adhering to legal and regulatory requirements made by relevant authorities is referred to as compliance.  Compliance implements processes and systems inside of an organization to make sure that the data practices are following established rules.  The Health Insurance Portability and Accountability Act (HIPAA) is an instance of regulation made by the United States that pertains to health care data.  The purpose of HIPAA is to protect the privacy and security of healthcare data (*Health Insurance Portability and Accountability Act of 1996 (Hipaa) | Cdc*, 2022).

The balance between regulation and compliance is fundamental to responsible data handling.    The legal foundation in regulatory frameworks along with compliance ensures that laws are followed.  Ethical considerations can guide decision making beyond just following state law requirements. Using ethics, compliance, and regulation together will allow responsible use of big data in many different industries and contexts.

**Cost-Benefit Analysis**

Cost-Benefit analysis (CBA) is a decision-making tool that involves costs associated and potential benefits with a certain course of action. Applying cost-benefit analysis to big data helps an organization look at whether an investment in big data is justified.  The cost of collecting, storing, and analyzing large volumes of data is studied, evaluated, and weighed with returns of the big data infrastructure investment.  Ethical questions of CBA and its impact on the great society and public arise such as if the benefits are distributed equally.  It could be possible that big data could compound inequalities of societal wealth (McAfee & Brynjolfsson, 2012).

A diagram of a project cost analysis

Description automatically generatedCBA and big data can bring up the issue of utility and many other related questions. CBA may be coaxing and enticing society to view data as a valuable resource.  Does the value produced from big data vindicate the potential risks it can pose? Some practical elaborations of CBA include infrastructure and technology costs, data quality and governance, compliance and security and the competitive advantage.

Figure An example of one type of CBA framework.

https://www.researchgate.net/figure/Three-Layer-Model-for-Cost-Benefit-Analysis\_fig1\_262066123

The costs that come with big data and technology are associated with processing and storage capabilities.  It is important to consider the cost of all tools required to make an effective analysis as well.  Data quality initiatives of CBA help ensure data accuracy, completeness, and consistency which will improve the reliability of insights found from big data.  Compliance and security is a critical element of big data CBA which has organizations consider expenses related to regulatory adherence, cyber security measures, and possible fines for non-compliance.

**Conclusion**

Big data ethics topics address issues that affect individuals’ rights and privacy.  Also organizations that use big data gain to benefit from regulation compliance.  Big data ethics  is complex and as technology changes new questions and concerns arise.  Big data ethics were examined there various aspects like, cost-benefit analysis, transparency, access controls and ownership and control.

**References**

Budke, C. A., & Ferguson, J. A. (2017). Data integration and e-commerce threats challenging providers. *Missouri Medicine*, *114*(6), 419–423. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6139979/

Giacalone, M., Sinitò, D. C., Calciano, M. V., & Santarcangelo, V. (2022). A novel big data approach for record and represent compliance in the covid-19 era. *Big Data Research*, *27*, 100290. https://doi.org/10.1016/j.bdr.2021.100290

*Health insurance portability and accountability act of 1996 (Hipaa) | cdc*. (2022, June 28). https://www.cdc.gov/phlp/publications/topic/hipaa.html

McAfee, A., & Brynjolfsson, E. (2012, October 1). Big data: The management revolution. *Harvard Business Review*. https://hbr.org/2012/10/big-data-the-management-revolution

O’Mathúna, D., & Iphofen, R. (Eds.). (2022). *Ethical issues in covert, security and surveillance research* (First edition). Emerald Publishing Limited.