


# Chapter 10

## Data Ethics and Privacy

**Thanaporn Phattanaviroj**

*Asia University, Taiwan*

**Massoud Moslehpour**

 <https://orcid.org/0000-0001-8808-2407>

*California State University, USA*

**Ankita Manohar Walawalkar**

*Asia University, Taiwan*

### ABSTRACT

*The publicity about information does not appear to decrease nor do the disgraces. Confidentiality openings in the collection, use, and distribution of data have affected all the major technology users, be it Facebook, Google, and AI go beyond the business world with administrations, cities, and educational and health organizations. However, with the speedy growth of social media and advanced technology such as mobile phone apps, various investors gather and use great quantities of data, ignoring ethics and privacy. This chapter explores and discusses the ethical considerations surrounding data collection and use, with a specific focus on privacy concerns related to Large Language Model (LLM) training data. The research further discovers topics such as anonymization, data rights, and consensus, aiming to highlight the importance of ethical practices in conducting data. Additionally, the inclusion of case studies on data misuse and privacy breaches serves to provide real-world examples, emphasizing the need for vigilance and responsible approaches in the realm of data collection and utilization.*

DOI: 10.4018/979-8-3693-3860-5.ch010

Copyright © 2024, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

## INTRODUCTION

The integration of data into different aspects of social life, including politics, social work, law enforcement, commercial models, public health, and editorial services, is becoming ever more pervasive (Quach et al., 2022). In recent years, the debate has expanded to include issues like algorithmic bias, where data-driven algorithms may perpetuate or even exacerbate existing social inequalities. High-profile cases in areas like predictive policing and credit scoring have demonstrated the need for ethical frameworks to ensure fairness, accountability, and transparency in the use of data-driven technologies (Agarwal et al., 2022). Data offers vast occasions to progress isolated and community life, as well as our situation (reflect the development of smart cities or the difficulties caused by carbon emissions). Inappropriately, such occasions are also coupled with noteworthy ethical and privacy challenges (Stahl et al., 2023). The general usage of progressively more information often individual, if not complex (big data) and the increasing dependence on algorithms to investigate them to shape choices and to make results (including mechanism learning, artificial intelligence, and robotics), as well as the regular decrease of human contribution or even mistake over many instinctive processes, pose pressing matters of equality, accountability, and respect of human rights, amongst others.

While information was often defined in terms of measure or even the 6Vs (i.e., volume, variety, velocity, veracity, virtue, and value) (W. Chen & Quan-Haase, 2018), the term has gained new meaning and is no longer controlled by either size or numerically collected info. Slightly, data involve all organized and formless information collected, stored, related, and evaluated either online or offline (Pascucci et al., 2023).

Consequently, researchers from several corrections are progressively attracted to exploring the ethics of data and the privacy inferences for both workers and manufacturers (Wiltshire & Alvanides, 2022). The study of data ethics necessitates new identifications as the complications range from information gathering, curation, and informed agreement to analytics (Mittelstadt & Floridi, 2016). The data have not completely resolved matters such as impartiality, correctness, accurateness, and inclusivity; quite differing, they have introduced new prejudices, objectivities, and procedures of domination (Rodrigo-Ginés et al., 2024). More does not permanently mean better, available does not continuously mean ethical, and accessibility does not always mean effective.

A large language model (LLM) is the linguistic typical with enormous restrictions that experience pretraining responsibilities (e.g., concealed linguistic model and autoregressive forecast) to recognize and procedure humanoid language, by modeling the contextualized writing semantics and possibilities from large amounts of insert data. For instance, LLMs have demonstrated to improve code safety

31 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/chapter/data-ethics-and-privacy/354400](http://www.igi-global.com/chapter/data-ethics-and-privacy/354400)

## Related Content

---

### Rational Unified Process and Unified Modeling Language - A GOMS Analysis

Keng Siau (2001). *Unified Modeling Language: Systems Analysis, Design and Development Issues* (pp. 107-116).

[www.irma-international.org/chapter/rational-unified-process-unified-modeling/30574](http://www.irma-international.org/chapter/rational-unified-process-unified-modeling/30574)

### A Generic Framework for Defining Domain-Specific Models

Arnor Solberg, Jon Oldevik and Audun Jensvoll (2003). *UML and the Unified Process* (pp. 23-38).

[www.irma-international.org/chapter/generic-framework-defining-domain-specific/30535](http://www.irma-international.org/chapter/generic-framework-defining-domain-specific/30535)

### Software Evolution with XVCL

Weishan Zhang, Stan Jarzabek, Hongyu Zhang, Neil Loughran and Awais Rashid (2005). *Software Evolution with UML and XML* (pp. 152-189).

[www.irma-international.org/chapter/software-evolution-xvcl/29613](http://www.irma-international.org/chapter/software-evolution-xvcl/29613)

### Forward Engineering and UML: From UML Static Models to Eiffel Code

Liliana Favre, Liliana Martinez and Claudia Pereira (2003). *UML and the Unified Process* (pp. 199-216).

[www.irma-international.org/chapter/forward-engineering-uml/30542](http://www.irma-international.org/chapter/forward-engineering-uml/30542)

### Unraveling the Inner Workings of Massive Language Models: Architecture, Training, and Linguistic Capacities

C. V. Suresh Babu, C. S. Akkash Anniyappa and Dharma Sastha B. (2024). *Challenges in Large Language Model Development and AI Ethics* (pp. 239-279).

[www.irma-international.org/chapter/unraveling-the-inner-workings-of-massive-language-models/354398](http://www.irma-international.org/chapter/unraveling-the-inner-workings-of-massive-language-models/354398)