SPRINGER BRIEFS IN APPLIED SCIENCES AND TECHNOLOGY · COMPUTATIONAL INTELLIGENCE

Samiksha Shukla
Jossy P. George
Kapil Tiwari
Joseph Varghese Kureethara

Data Ethics and Challenges

Springer

SpringerBriefs in Applied Sciences and Technology

SpringerBriefs in Computational Intelligence

Series Editor

Janusz Kacprzyk, Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

SpringerBriefs in Computational Intelligence are a series of slim high-quality publications encompassing the entire spectrum of Computational Intelligence. Featuring compact volumes of 50 to 125 pages (approximately 20,000-45,000 words), Briefs are shorter than a conventional book but longer than a journal article. Thus Briefs serve as timely, concise tools for students, researchers, and professionals.

More information about this subseries at https://link.springer.com/bookseries/10618

Samiksha Shukla · Jossy P. George · Kapil Tiwari ·
Joseph Varghese Kureethara

Data Ethics and Challenges

Springer

Samiksha Shukla
Department of Data Science
Christ University
Bengaluru, Karnataka, India

Kapil Tiwari
Christ University
Bengaluru, Karnataka, India

Jossy P. George
Christ University
Bengaluru, Karnataka, India

Joseph Varghese Kureethara
Centre for Research
Christ University
Bengaluru, Karnataka, India

ISSN 2191-530X ISSN 2191-5318 (electronic)
SpringerBriefs in Applied Sciences and Technology
ISSN 2625-3704 ISSN 2625-3712 (electronic)
SpringerBriefs in Computational Intelligence
ISBN 978-981-19-0751-7 ISBN 978-981-19-0752-4 (eBook)
https://doi.org/10.1007/978-981-19-0752-4

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

Data is the essential element of contemporary society. This generation has evolved into a data-informed society. Advancements in technology in various forms catapult this evolution. Organizations utilize data for better decision-making, customer-centric facilities, and target marketing. Data repurposing is another emerging field, mainly used in the medical/pharmaceutical domain. With the emergence of data at a faster rate, the following quote has become appropriate over time, "**Data is the new Soil, Let us cultivate it.**" It is a buzzword and a fizzy word these days. The internet is becoming a data facilitator not only for researchers but also for home-makers, kids, and businesses. It is becoming a critical part of our lives in a growing data-driven society. The Internet of Things has changed the world forever. Research, Academia, Government and Private sectors, Medical institutes, Cultural heritage, meteorological institutions, etc., generate, utilize and share a huge amount of data. This massive use of data comes with many challenges for individuals, institutions, societies, and states.

Considering the widespread relevance of the data today, it is easy to understand how precious it is. Data-driven decision-making provides an ease to individuals, but there is a need to maintain privacy, confidentiality, and security of the data. Ethical data handling practices play a crucial role in a data-informed society.

Data pertaining to a living being is primarily a property of that being. Hence, proper permission is required for anyone else to use that private property. Some vital questions are: Has a living being got absolute authority over its data? Can a living being use its data? What is the right of others over the data of a person? Has the state got any authority over the data of a living being?

Living beings can generally be classified into plants, animals, and human beings. Although it is not a proper biological taxonomy, this classification is based on the activities and the nature of domestication of living beings. Data related to plants and animals have to be collected and used concerning the state's laws. However, data pertaining to human beings have to be treated respecting the person from whom the data is collected and the state laws that govern it. In various cases, the dependent and independent human beings have to be treated separately as far as data collection and data use are concerned.

Religions and social settings have played essential roles in determining the foundations of ethics. The foundation of ethics laid upon the 'good' and 'bad,' and 'right' and 'wrong.' The perspectives are absolute as well as contextual. Ethics has to be thought about as evolutionary. It is because several concerns in ethics have been replaced and reworked in the various stages of history and context. The theories such as teleological, axiological, deontological, formalistic, naturalistic, intuitionistic, non-cognitive, etc., are to be considered in any formal discussion on ethics.

In close association with the data comes the intellectual property of living organisms. One question is important: Should intelligence be restricted to natural intelligence? How much authority does one person have over one's intellect? How much does one person owe to the state and society for one's intelligence? When considering the technical term 'Intellectual Property,' it primarily refers to an individual's 'creative outputs.' Has an individual got absolute authority over one's creative outputs? International and national regulations and directives on intellectual property and intellectual property rights from multiple perspectives have to be discussed.

What to Expect from This Book

This book gives a thorough and systematic introduction to Data, Data Sources, Dimensions of Data, Privacy, and Security Challenges associated with Data, Ethics, Laws, IPR Copyright, and Technology Law. This book will help students, scholars, and practitioners to understand the challenges while dealing with data and its ethical and legal aspects. The book focuses on emerging issues while working with the Data.

Who This Book is For

This book is primarily aimed at the general public, who need to know about multiple aspects of dealing with the data. Particularly, this book comes as a ready-to-refer short handbook for the people involved in the data-related industry, viz., data scientists. As an emerging field of higher education, data science programmes can use this book as a reference book.

Road Map

- Chapter 1 covers the concept of Data, its type, and sources. It also introduces the various facets of the data, the science of data, data ownership, and the FAIR data principles (Findable, Accessible, Interoperable, and Reusable). The importance of use/abuse and overuse of the data. In the current era of the information world of faster data harvesting, data theft and misuse of the data is a critical concern.

- Chapter 2 of this book underlines what data privacy entails and its relevance and importance to individuals in a growing digitized world. The right to privacy and the ethics concerning data collection from individuals is discussed nuanced, tracing constitutional recognitions and governmental rulings of an individual's privacy as a fundamental right. The chapter also presents the legal disputes and cases concerning data privacy as instances to support the information furnished. The chapter thus comprehensively explores the intricacy of data privacy and its evolution.
- Chapter 3 focuses on the security aspect associated with the data. This chapter presents various types of security breaches and mechanisms to control them. The Data Security and Information Technology Act is also covered in this chapter.
- Chapter 4 looks at the ethical aspect associated with data and data science. The concept of data ownership is presented in this chapter. The importance of Informed Consent and Institutional Review Board (IRB) along with repurposing and research ethics are also covered.
- Chapter 5 presents the concept of Intellectual Property Rights (IPR) and Copyright Law.

Bengaluru, India

Samiksha Shukla
Jossy P. George
Kapil Tiwari
Joseph Varghese Kureethara

Contents

1 Data and Its Dimensions 1
1.1 Types of Data 1
1.2 Theory of Information 4
1.3 Sources of the Data 5
1.4 Facets of Data 9
1.5 The Science of Data 12
1.6 Data Ownership 12
1.7 Value of Data 13
1.8 Use/Misuse/Overuse of Data 13
1.9 Data and Information Technology Act 14
1.10 Case Study 15
References 16

2 Data Privacy 17
2.1 Data Privacy: Aspects, Challenges, and Definitions 18
2.2 Privacy and Confidentiality 24
2.3 Privacy as a Philosophy 25
2.4 Data Privacy Law in the United States, European Union and India 29
2.5 Operations and Concerns of Privacy 33
2.6 Models of Privacy 33
2.7 Information Loss 34
2.8 Infringement of Privacy 35
2.9 Case Studies 35
2.10 Conclusion 38
References 39

3 Data Security 41
3.1 Understanding Data Security 42
3.2 The Importance of Securing Data 47
3.3 Data Security Solutions 51
3.4 Governance and Compliance 53

3.5 Data Security Audits and Frameworks 55
3.6 Case Study 57
References 59

4 Data Ethics 61
4.1 Types of Ethics 64
4.2 Individual View on Ethics 68
4.3 Data Science Ethics 69
4.4 Data Ownership 69
4.5 Informed Consent 70
4.6 Institutional Review Board 70
4.7 Data Repurposing 71
4.8 Cases of Ethical Violations When Dealing with Data 72
References 72

5 Intellectual Property Right - Copyright 73
5.1 Copyright 74
5.2 Plagiarism 77
5.3 Copyright Infringement and Plagiarism 78
5.4 Permissible Limits of Copying 78
5.5 Originality and Creativity 79
5.6 Research and Publication Ethics 79
5.7 Case Studies 80
Reference 81