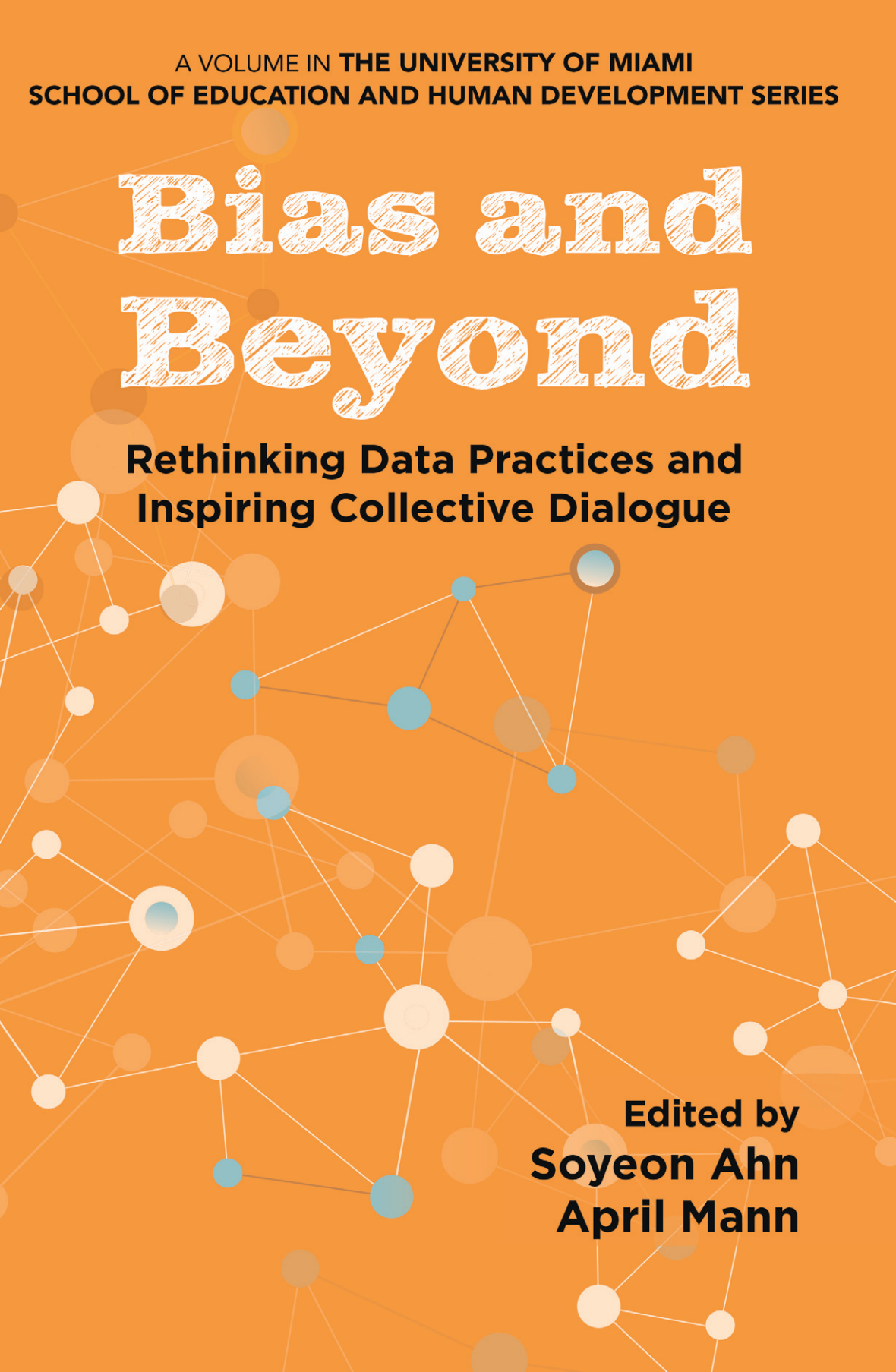
A VOLUME IN **THE UNIVERSITY OF MIAMI SCHOOL OF EDUCATION AND HUMAN DEVELOPMENT SERIES**

Bias and Beyond

Rethinking Data Practices and Inspiring Collective Dialogue



Edited by
Soyeon Ahn
April Mann

Bias and Beyond

The University of Miami School of Education and Human Development Series

Series Editor: Marie Guerda Nicolas, University of Miami, USA

This series aims to focus on the study, promotion and integration of educational, psychological, and physical well-being in multicultural communities. Books in the series will target the production of knowledge and the preparation of the next generation of leaders, researchers, and agents of change and well-being in education and the community.

OTHER TITLES IN THIS SERIES

Real Talk
Multiliteracies in English as an Additional Language Classrooms
Expanding Literacy Practices Across Multiple Modes and Languages for Multilingual Students

Bias and Beyond

Rethinking Data Practices and Inspiring Collective Dialogue

Edited by

Soyeon Ahn
University of Miami, USA

And

April Mann
University of Miami, USA

emerald PUBLISHING

United Kingdom – North America – Japan
India – Malaysia – China

Emerald Publishing Limited, Floor 5, Northspring, 21-23 Wellington Street, Leeds LS1 4DL

First edition 2026

Editorial matter and selection © 2026 Soyeon Ahn and April Mann.
Individual chapters © 2026 The authors.
Published under exclusive licence by Emerald Publishing Limited.

Cover photo: iStock and Nobi_Prizue

Reprints and permissions service
Contact: www.copyright.com

No part of this book may be reproduced, stored in a retrieval system, transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without either the prior written permission of the publisher or a licence permitting restricted copying issued in the UK by The Copyright Licensing Agency and in the USA by The Copyright Clearance Center. Any opinions expressed in the chapters are those of the authors. Whilst Emerald makes every effort to ensure the quality and accuracy of its content, Emerald makes no representation implied or otherwise, as to the chapters' suitability and application and disclaims any warranties, express or implied, to their use.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-80686-666-3 (Print hardback)
ISBN: 978-1-80686-668-7 (Print paperback)
ISBN: 978-1-80686-665-6 (Online)
ISBN: 978-1-80686-667-0 (Epub)

Typeset by TNQ Tech
Cover design by TNQ Tech

CONTENTS

About the Editors vii

About the Contributors ix

Preface xvii

1 Rethinking Data Life Cycle Practices: The Value of Applying Reflexivity Methods to Current Data Practices 1
EllenMarie McPhillip

2 What Are We Even Measuring?: The Role of Survey Questions in Demographic Data Collection 13
Joelle A. Dorsett, Alexis R. Franklin, Regine Thermy Jean Baptiste, Marisol L. Meyer and Guerda Nicolas

3 Using Course Evaluations to Assess Diversity, Equity, and Inclusion: A Case Study on Measuring Difficult Concepts With Problematic Instruments 31
April Mann, Cameron Riopelle and Adina Sanchez-Garcia

4 Garbage In, Garbage Out!: A Commentary on Bias in Standardized Assessments 51
Murat Kasli, Celia Rubio and Cengiz Zopluoglu

5 From a Different View: Addressing Bias in Qualitative Research ... 65
Jessica Wendorf Muhamad, Kallia O. Wright and Tyler R. Harrison

6 Current Practices of Meta-Analytic and Systematic Reviews in Education Toward Unbiased, Inclusive, and Representative Research Evidence 89
Soyeon Ahn, Ryan C. Schooley, Beck Graefe and Ching-Hua Chuan

7 Adult College (Re)Enrollment Advice and Information Networks: A Descriptive Analysis 109
Willis A. Jones

8 Systemic Racial Disparities in Large Data on School District Finance and Housing: Three Empirical Illustrations 123
Bruce D. Baker and Wendy Castillo

9 Considering Quality Indicators of Mixed Methods Research 155
Wendy Cavendish and Ignacio Barrenenchea

ABOUT THE EDITORS

Soyeon Ahn, PhD is a Professor in the Research, Measurement, and Evaluation Program at the University of Miami's School of Education and Human Development. Her research centers on identifying and mitigating bias across the data lifecycle, with a focus on advancing equity, justice, and inclusion in education and social science.

She investigates how measurement tools, statistical models, and algorithmic systems can perpetuate inequities—particularly in culturally diverse and underserved populations—and develops equitable analytic frameworks and methods. As founder of the Responsible Educational Data Analytics Hub (REAHub), she leads interdisciplinary initiatives to promote ethical, community-informed, and justice-centered data practices.

Dr Ahn has authored over 70 peer-reviewed publications and serves on editorial boards for journals focused on quantitative methods and equity. Her recent work includes a co-edited volume on bias and inclusion in the data lifecycle and research on AI-assisted meta-analysis. She has developed academic programs that prepare students to use data for social impact and has mentored over 50 doctoral students. Across her research, teaching, and leadership, Dr Ahn champions the responsible use of data to drive meaningful and inclusive change.

April Mann, EdD, is the Director of the Writing Center and a senior lecturer in the Department of Writing Studies at the University of Miami. She facilitates a Faculty Working Group for Teaching with AI that brings faculty together to discuss their innovative work with AI in the classroom and a Summer Writing Institute that helps faculty enrich their curricula through writing. April is co-author of several publications, most recently "The linguistic leap: Understanding, evaluating, and integrating AI in language education" in the *Journal of Language Teaching* and "Crossing the bridge: Writing and research bridge programming for an intensive English program" in *PILH: Information Literacy and Writing Studies, Volume II – Upper Level and Graduate Studies*. Ed. Grace Veach. Purdue University Press, 2019. She is also

the author of "Structure and Accommodation: Autism and the Writing Center," in *Autism Spectrum Disorders in the College Composition Classroom: Making Writing Instruction more Accessible for all Students*. Eds. Walsh, Lynda and Gerstle, Val. Marquette University Press, 2011. In addition, she was also a contributor to and reader for *Composing Inquiry: Methods and Readings for Investigation and Writing*. Ed. Margaret Marshall, Prentice Hall, 2004.

ABOUT THE CONTRIBUTORS

Bruce D. Baker is Professor in the Department of Teaching and Learning at the University of Miami and Co-director and creator of the School Finance Indicators Database. His most recent books from Harvard Education Press include Educational Inequality and School Finance: Why Money Matters for America's Students, and School Finance and Education Equity: Lessons from Kansas. He is a recently inducted Fellow of the American Educational Research Association, and serves in advisory and research fellow roles for the National Education Policy Center, Learning Policy Institute and Brown's Promise. He has consulted on state school funding policy design with Vermont, New Hampshire, Kansas, Texas, Colorado, Wyoming, Washington, Virginia, Missouri, Delaware, Oregon, testified in numerous state and federal courts on school funding inequalities and his work on racial inequalities in school funding has been cited by the U.S. Supreme Court. He is presently collaborating with the World Bank on a global indicator system.

Regine Thermy Jean Baptiste, MSEd earned her master's degree in Mental Health Counseling with a concentration in Marriage and Family Therapy from the University of Miami. She completed her undergraduate studies at Université d'Etat D'Haïti, where she majored in Psychology at Faculté des Sciences Humaines (FASCH) and Public Administration at Institut National d'Administration, de Gestion et des Hautes Etudes Internationales (INAGHEI). She also earned a certificate in American Sign Language at the Institut Haitien de Langues de Signes (IHLS). Her research focuses on family systems, parenting, and attachment in the Black Community. She is interested in working with Haitian families, children, and communities to promote positive parenting, self-esteem, and mental health outcomes. She is also interested in promoting inclusion and mental health among the deaf community in Haiti.

Ignacio Barrenenchea holds a PhD in Educational and Psychological Studies from the University of Miami, an MA in International Education from The George Washington University, an LLM in International Legal Studies from New York University, and a Law degree from Universidad Torcuato Di Tella. His career bridges academic research, public policy, and international cooperation. He currently serves as Chief of Staff at Argentina's National Secretariat of Education, where he coordinates the implementation of key educational policies and partnerships with multilateral organizations. His research agenda focuses on teacher well-being from an ecological perspective, combining mixed methods and community psychology frameworks. He has published in leading academic journals such as the *Journal of Professional Capital and Community*, the *Journal of Educational Change*, and the *International Journal of Educational Development*.

Wendy Castillo is an Assistant Professor in Urban Education and Quantitative Methods. She is a co-author of "How to QuantCrit: Applying Critical Race Theory to Quantitative Data in Education." Previously she was also the Senior Director of Equity, Data & Impact at the National Urban League where she led their data strategy and program evaluation efforts. Dr Castillo has experience with quantitative research methods, including large-scale randomized controlled trials, meta-analysis, cost analysis, and measurement design and validation. Her research interests lay at the intersection of race and quantitative data and methods including QuantCrit and other racial equity frameworks to advance and improve the lives of students of color. Dr Castillo studied Education Policy and earned her PhD at the University of Pennsylvania. She also holds a MSEd in Educational and Social Change from the University of Miami and a BA in International Relations from Brown University.

Wendy Cavendish, PhD, is a Professor in the Wellness and Inclusive Services in Education Department at Rowan University. She has taught applied research methods and mixed methods research courses to graduate students for the last 10 years. Wendy's research has an interdisciplinary focus and most recently, her mixed methods research has focused on identifying the policies, institutional processes, and practices in schools and other social institutions that facilitate more equitable transition outcomes for multiply marginalized youth. Her previous research projects have centered student and teacher voices to inform transformative change in schools and communities.

Ching-Hua Chuan is an Assistant Professor of Interactive Media at the University of Miami. She received her PhD in computer science from the

University of Southern Viterbi School of Engineering, and her BS and MS degrees in electrical engineering from the National Taiwan University. Dr Chuan's research interests include artificial intelligence (AI), human-centered computing and music information retrieval, with an interdisciplinary focus on responsible AI and technologies for social good. She has published refereed articles in journals and conferences on human-machine communication, human-centered AI, augmented/mixed reality, and machine learning applications. She was the recipient of the Arthur Page Legacy Scholar in 2022 on developing a theory-based mobile app for cultivating sustainable environmental-friendly behaviors. She also received the best paper award from the Journal of Current Issues and Research in Advertising in 2022, and the best new investigator paper award at the Grace Hopper Celebration of Women in Computing in 2010.

Joelle A. Dorsett is a doctoral candidate in counseling psychology at the University of Miami. She earned her bachelor's degree in psychology, with a minor in Spanish, from the University of Florida. Joelle's research focuses on developing and implementing evidence-based, culturally responsive interventions that support the mental health of children, adolescents, and young adults with chronic medical conditions, particularly within Black communities. Her work examines how physical health challenges, cultural identity, and systemic inequities intersect to shape psychological well-being. She is particularly interested in promoting resilience and adaptive coping among youth and young adults navigating these overlapping experiences. Clinically, she has provided therapy to diverse clients across hospital, community, and university settings, addressing concerns related to chronic illness, caregiver-child relationships, disruptive behavior, self-esteem, and identity development. Joelle is committed to advancing equity-driven approaches that bridge research, clinical practice, and community engagement to promote holistic wellness.

Alexis R. Franklin (she/her) is a 6th-year PhD candidate in the School of Education and Human Development at the University of Miami. She earned her bachelor's degree in Psychology from Rhodes College in Memphis, Tennessee, in 2018. Alexis's research focuses on developing and evaluating culturally informed interventions that promote mental health wellness among marginalized populations. Her work seeks to bridge the gap between research and practice by centering cultural context and community strengths within evidence-based approaches. Beyond her research, she is dedicated to mentorship, advocacy, and community building efforts that promote the wellbeing of marginalized communities.

Beck Graefe, PhD, is an Assistant Professor in the Department of Clinical and School Psychology in the College of Psychology at Nova Southern University (NSU). He is an educator and researcher who emphasizes student centered and innovative approaches to teaching and mentorship, committed to promoting the development of analytical skills for application across applied contexts. His research interests include: bias in quantitative methods, intersectional quantitative methods, longitudinal modeling, meta-analysis, structural equation modeling, and scale development. He has experience on interdisciplinary research teams that focus on a range of topics: innovative pedagogies, health education, health and educational disparities, and youth mental health and substance use behaviors.

Tyler R. Harrison, PhD is Professor Emeritus in the Department of Communication Studies at the University of Miami. His work focuses on the design, implementation, and evaluation of communication interventions to improve individual, organizational, and societal health and wellbeing. His work has focused on clinical trial accrual, organ donation, cancer, and climate adaptation. He is co-editor of the award-winning book, *Organizations, Communication, and Health* and co-author of *From Numbers to Words: Reporting Statistical Results for the Social Sciences.* His research has appeared in such outlets as *Journal of Communication, Journal of Health Communication, Journal of Applied Communication Research, Journal of Contingencies and Crisis Management, and PLoS Climate.*

Willis A. Jones is Associate Dean for Academic Affairs at the Annette Caldwell Simmons School of Education and Human Development at Southern Methodist University, where he also serves as Associate Professor of Higher Education in the Department of Education Policy and Leadership. His research uses quantitative methods to examine how institutional behaviors, strategies, and policies shape outcomes in higher education.

Murat Kasli is a Psychometrician at the National Council of State Boards of Nursing (NCSBN), where he has contributed to high-stakes examination development and validation for over 3 years. He has provided psychometric support for a variety of testing programs including the NGN (Next Generation NCLEX®) and REx-PN examinations. He earned a PhD in Research, Measurement & Evaluation from the University of Miami in 2022. His research interests include statistical methods for aberrant behavior detection, item response theory, and machine learning applications in psychometric research, particularly for adaptive testing environments.

EllenMarie McPhillip serves as Associate Dean of Undergraduate Academic Services and Innovation at the University of Miami's School of Education

and Human Development. Since joining UM in 1999, she has developed transformative programs bridging education, entrepreneurship, and social impact. Her key initiatives include co-founding **HBCUSTART@Miami-Herbert** (2019), a weeklong residency for HBCU students interested in launching businesses or nonprofits, and founding **The Hyperion Council** (2002), enabling students to consult with entrepreneurs and nonprofits in at-risk communities across 11 countries in South America and the Caribbean. She also developed **BUS101/MGT100** (2006), a foundational course in the business school's curriculum. She launched the podcast **The BadAss Women's Network** (2019), exploring how "difficult" women navigate leadership challenges and professional loneliness. Previously, she worked at *The New York Times Co.*, *Newsday*, and *Scholastic Inc.* She holds degrees from McGill University (BA) and New York University (MA).

Marisol L. Meyer, PhD graduated with a BA in Psychology and Anthropology at Dartmouth College and received her PhD in Counseling Psychology at the University of Miami. Her research, clinical work, and teaching focus on how unique community strengths and evidence-informed/evidence-based practice can be coalesced to promote wellness. Marisol has coordinated and co-designed interventions and curricula addressing community-selected topics such as trauma-informed mental health support, emotion socialization, and ethnic racial identity development. These programs have reached community-leaders, teachers, healthcare professionals, caregivers, and youth. Her experiences engaging with community needs through research and intervention have given her extensive experience in program manualization and adaptation. Marisol prioritizes a community-focused dissemination approach, prioritizing the act of sharing research findings with the community-members with whom she proudly collaborates and the broader public.

Jessica Wendorf Muhamad, PhD, is an Associate Professor in the School of Communication at Florida State University and Director of the PEAKS Lab (Participatory Engagement, Action, and Knowledge for Social Change). Her research focuses on developing and evaluating participatory-based interventions that support positive attitude and behavior change, particularly in health and organizational settings. She studies how communication—especially narrative, dialogue, and collaborative engagement—shapes individual and collective responses to health challenges and crisis. Dr Wendorf Muhamad has conducted applied research in varied environments, including healthcare systems, social service organizations, and disruptive events. Her work bridges academic research with practical application across communication, health, and social service sectors, often in collaboration with professionals and local organizations. She employs innovative methods that emphasize collaboration, practical relevance, and responsiveness to

community needs. She has published in top journals in communication and health and works to translate research into effective, impactful interventions.

Guerda Nicolas is a licensed clinical psychologist originally from Haiti who earned her doctoral degree in clinical psychology from Boston University. She completed her predoctoral training at Columbia University Medical Center and her postdoctoral training at the New York State Psychiatric Institute/Columbia University, Department of Child Psychiatry. Before joining the University of Miami, she held faculty positions at Boston College and the College of Saint Elizabeth in New Jersey. A multicultural (Haitian American) and multilingual psychologist (Spanish, French, and Haitian Creole), Dr Nicolas's research focuses on developing culturally effective mental health interventions for people of color, particularly immigrant children, adolescents, and families. She also examines social support networks within Caribbean populations, especially among Haitians, as well as spirituality and mental health among Black communities. Her scholarship, publications, and leadership in professional organizations reflect her commitment to advancing culturally responsive approaches to mental health and community well-being.

Cameron Riopelle is the Director of Research Data & Open Scholarship and a Librarian Associate Professor at the University of Miami Libraries. He received his PhD in Sociology from the University of Illinois at Urbana-Champaign in 2017, his Master of Science in Statistics (concentration: Applied Statistics) from UIUC in 2015, and Master of Arts in Sociology from UIUC in 2012. Before coming to the University of Miami, he taught at UIUC, Illinois Wesleyan University, and Millikin University. His research interests include teaching pedagogy and research methods. He has co-authored articles in journals including Clinical Psychological Science, Families in Society, Journal of Geophysical Research, Child Abuse & Neglect, the American Journal of Emergency Medicine, the Journal of Abnormal Child Psychology, American Behavioral Scientist, and the Annals of the American Association of Geographers.

Celia Rubio is a current doctoral student in the Research, Measurement and Evaluation program in the Department of Education and Psychological Studies at the University of Miami. Celia completed her undergraduate studies at the University of Miami with a major in Psychology and a double minor in Biology and Art and Art History. Her research focuses on applications of Rasch measurement theory, with particular interest in multilevel Rasch models and Explanatory Item Response Modeling (EIRM). She is especially interested in how these methods can be used to examine

complex latent constructs and to model individual growth trajectories, including how such trajectories may vary as a function of group membership. Her broader goals involve improving measurement practices in the social sciences through the use of more nuanced and interpretable statistical modeling techniques.

Adina Sanchez-Garcia is the Director of Undergraduate Studies and Associate Chair in the Department of Writing Studies at the University of Miami. She holds an EdD in Applied Learning Sciences and an MA in English, both from the University of Miami. Her research focuses on assessment, curriculum design, online pedagogy, and writing-enriched curricula, with a particular interest in how generative AI informs each of these areas.

Ryan C. Schooley, PhD is a Staff Psychologist with the Center for Counseling and Psychological Services at Washington University in St. Louis. He completed his doctorate in counseling psychology at the University of Miami and has worked at the University of Utah Counseling Center, the University of Miami Counseling Center and Athletic Department, and a community clinic in South Florida. Ryan has published and presented scholarly work related to whiteness and racial socialization, racialized aspects of mental health, and efforts toward social change. He has also facilitated multiple intergroup dialogue courses concerning race, ethnicity, SES, and religious traditions. Prior to his doctoral work, Ryan graduated from the Community Development and Action master's program at Vanderbilt University. While at Vanderbilt, he worked with a community organizing initiative attempting to increase access to affordable housing in a rapidly gentrifying city. Ryan is interested in actionable research projects toward more liberatory therapeutic practice and social change.

Kallia O. Wright (PhD, Ohio University) is an Assistant Professor in the School of Communication at the University of Miami, teaching courses including intercultural communication, interracial communication, and communication theory. Her research focuses on investigating culture's impact on health communication. She has several publications in journals such as, *Communication Teacher*, *Journal of Applied Communication Research*, and *Social Science and Medicine*.

Cengiz Zopluoglu is an Associate Professor in the Department of Special Education and Clinical Sciences at the University of Oregon. His research interests are centered around item response theory, psychometrics, latent variable modeling, and integrating machine learning methods into educational measurement.

This page intentionally left blank

PREFACE

In recent decades, the volume and diversity of data generated daily have increased extraordinarily. To process this surge of data, computational models and algorithms have been developed, and these models and algorithms shape decision-making across sectors, including public health, education, public policy, and business. In addition, technological innovations have used this data to enhance the efficiency and convenience of our daily lives.

However, these advances also bring significant societal concerns. Data governance, privacy, the quality and application of collected data, as well as policies and practices derived from advanced data collection are among the emerging challenges. Many people outside of the data sphere may not be aware of the critical role data plays in decision making. This has led to widespread data illiteracy and a gap in knowledge, experience, and skill between those who create and control data and those most affected by its use. This gap is exacerbating social disparities among individuals and communities at local, regional, and global levels.

These challenges have inspired the creation of this book. In it, we both explore the multifaceted issues around data and issue a call to action. Throughout the chapters, we explore the use and misuse of data and search for tentative solutions that can function iteratively to engage with the changing data landscape. By critically engaging with the transformative power and implications of data, we hope to promote more inclusive and responsible approaches to its creation and its use.

This book not only invites reflection but also serves as a practical guide for action. The insights and recommendations presented here can empower readers to address challenges and issues within their own contexts. Whether you are a researcher seeking to refine methodologies, a policymaker aiming to develop more representative policies, or a partitioner seeking for more inclusive practices, this book provides perspectives and tools to make changes.

At the core of this book lies the concept of the data lifecycle – a structured yet iterative process which describes the way data is created and used to drive the decision-making process. The data lifecycle, although varied across different disciplines, typically encompasses the following four stages: (1) formulating questions/problems; (2) measuring, collecting, and manipulating data for analysis; (3) analyzing and communicating the data to inform policies and practice development; and (4) refining these policies and practices through iterative reprocessing of the previous steps.

Our book traverses throughout the data lifecycle, highlighting the importance of using diverse data types - quantitative, qualitative, numbers, strings, audio, and video - to deepen our understanding of global phenomena and tackle complex societal challenges. Therefore, our discussions here extend beyond scenarios involving exclusively quantitative data, which is frequently but mistakenly regarded as the only form of data. Additionally, we recognize that data within the lifecycle can be either structured or unstructured, and it may vary significantly in volume, from small to large.

While we are all, consciously or unconsciously, involved in the data lifecycle, we do not all have the opportunity to engage fully in this process. Often data collection, analysis, and decision-making are managed by a select few or groups, excluding individuals and even whole communities. This intentional and unintentional exclusion can lead to policies and practices that do not represent the needs of those most impacted. Such exclusion highlights the long-lasting barriers that can limit representation and equity for all individuals and communities in the data lifecycle, reinforcing existing biases.

Thus, another core principle of our book involves addressing the challenges of applying Diversity, Equity, and Inclusion (DEI) principles throughout the data lifecycle. By reconsidering how problems are defined and how data is collected, analyzed, interpreted, and used, it is possible to better ensure that all groups, including historically marginalized groups, are represented, and that data-driven decisions support less biased, more equitable and more sustainable outcomes.

Our book examines some of these issues and provides guidance on critically evaluating traditional data practices and understanding how to implement data practices that prioritize inclusion, representation, and ethical accountability. By empowering readers with the knowledge and skills to actively engage in the data lifecycle, the book highlights the potential of data when used responsibly, ethically, and inclusively.

This book is intended for anyone involved in the collection, analysis, or application of data—researchers, educators, policymakers, and practitioners. It also seeks to engage those whose lives are most impacted by policies and practices, encouraging them to become more knowledgeable and skilled and thus active participants in data life cycle. By empowering a

diverse audience with the knowledge and perspective to engage with data, this book attempts to bridge the gap between those who create and control data and those most affected by its use.

We aim for this book to serve not merely as an academic exploration but as an invitation to reflect on how data impacts our lives, communities, and broader society. The book prompts readers to consider who has access to data and its analysis for decision-making, thinking critically about how data can be leveraged for social good. This book hopes to provide a variety of perspectives on the challenges encountered during the data lifecycle, offering insights into how citizens can contribute to a datasphere that is more conscious and active. A more conscious use of data, we believe, will facilitate more equitable, less biased, and more inclusive data-driven decision-making.

Our book covers four phases of the data lifecycle, presented through nine distinctive yet interconnected chapters. Each offers unique perspectives, with authors sharing their own reflections on the challenges they have encountered, and the insights shaped by their individual positionalities, experiences, and expertise. To encourage active engagement and dialogue with the authors, we have intentionally chosen to preserve the authors' distinct voices and messages rather than standardizing the tone across chapters. We consider the multiplicity of voices in this book to better represent the complexity of the challenges encountered throughout the data lifecycle.

Chapter 1 sets the stage by offering an overview of the ethical, methodological, and societal challenges that emerged from the data revolution of 2010 to 2020. The author highlights the relationship between data collection and social inequities, stressing the need for critical reflection on biases and underlying assumptions. She calls for conscientious and deliberate engagement in the data lifecycle, particularly concerning data governance and tools, to reduce biases, enhance equity, and address ethical dilemmas, especially in advanced computation models and algorithms, such as AI.

The subsequent three chapters explore potential challenges that may occur during data collection and the subsequent measurement of phenomena. The chapters offer authors' insights and recommendations for addressing these challenges effectively. Chapter 2 critically reviews traditional survey methods used for collecting participants' social identities, focusing on measuring race, ethnicity, gender, and sexual orientation. The authors offer recommendations to address the systematic exclusion of marginalized populations and improve inclusivity. They aim to adopt more inclusive data practices and thereby achieve a more accurate representation of diverse identities in research and practices.

In chapter 3, the authors examine the way course evaluation data may provide unreliable measurements of diversity, equity, and inclusion (DEI). They present empirical evidence, based on an analysis of student response data for a DEI item, highlighting potential biases and misrepresentation.

They emphasize the importance of thoughtful and deliberate practices in designing survey items to ensure that data collection accurately captures participant views on complex issues, such as DEI, especially when the results are used for high-stakes decision making.

Chapter 4 builds on these themes, underscoring the importance of critically examining biases in test item(s) which are purportedly used to advance DEI. Drawing on case studies from multiple disciplines such as medicine, education, and law, the authors discuss how biased testing can lead to high-stakes negative consequences for individuals and communities. The authors advocate for the use of best practices in designing test items, including psychometrics and statistical methods such as Differential Item Functioning (DIF) analysis, panel reviews, stakeholders' involvement, and clear policies and practices, to address these biases and mitigate their impacts.

The remaining four chapters discuss how various data analytic approaches—qualitative, quantitative, and mixed methods—can identify and address potential biases in data to inform policies and practices. Using empirical data examples provided by the authors, discussion begins with qualitative methods in Chapter 5, transitions to statistical analyses in Chapter 6 through Chapter 7, and concludes with mixed methods in Chapter 8.

Chapter 5 examines the concept of bias in qualitative research, framing it as a potential influence that may distort findings while also recognizing subjectivity as a central element of methodology. The authors advocate for positioning subjectivity as an asset rather than a flaw, although this perspective might raise questions about rigor and validity. By acknowledging potential threats to rigor and offering strategies to accurately represent phenomena or individual experiences, the authors emphasize the importance of reflexivity and preparation to achieve a shared understanding between the researcher and participants.

Chapter 6 examines the extent to which research findings reflect diverse socio-cultural and ethnic groups, based on a thorough analysis of meta-analytic reviews published in a leading peer-reviewed educational journal. The authors highlight significant gaps in inclusivity and provision of unbiased evidence, particularly for historically marginalized populations. In response, they offer recommendations to improve the representativeness and generalizability of research findings, promoting more inclusive practices that better serve all individuals and communities, regardless of their backgrounds.

Building on concerns discussed in Chapter 6, Chapter 7 introduces the egocentric network analysis as a tool to explore the social networks and dynamics of people of color. Through empirical data analysis, the authors uncover critical differences across socio-demographic backgrounds in key dimensions of people's social networks. They offer practical recommendations on how social networks analysis can draw more equitable and representative policies and practices to better support people of color.

Chapter 8 expands on the recurring themes of systematic biases embedded in data and analysis throughout the book, using large educational data analysis, wherein racial discrimination in housing and school funding led to educational inequities. Using the QuanCrit framework, the authors illustrate how systematic racism is embedded within data, policies, and algorithms, echoing earlier discussions on bias in data collection and analysis. The chapter provides actionable insights for dismantling systematic disparities in education by incorporating racial measures in statistical modeling and policymaking.

The final chapter, Chapter 9, provides a comprehensive overview of mixed methods research, highlighting its capacity to address complex research questions and advance social equity. The authors particularly discuss core quality criteria to ensure rigor, suggest ways mixed methods can reduce biases and promote inclusivity, and conclude with recommendations for further development of this research approach.

Together, these nine chapters follow the path of the data lifecycle, offering diverse perspectives, critical insights, and practical recommendations to uncover biases and create more inclusive and less biased data-driven research and practices. The chapters are organized around the structure of the data lifecycle, but readers are encouraged to engage with the book in a way that best suits their interests and needs - whether by exploring specific chapters or following the sequence as intended. Each chapter reflects the unique expertise and lived experiences of the authors, acknowledging the limitations of any single book to address all tools and approaches available to conduct research across disciplines.

Our hope is that this book is a foray into the topics introduced here, aiming to open dialogue and encourage reflection rather than provide definite answers. Readers are invited to evaluate challenges and issues from their own experiences and expertise within their own contexts. Through increased public awareness and constructive dialogue, our aim is to inspire continuous, iterative improvement in how data is collected, analyzed, and applied. We look forward to this book serving as a foundation - an invitation to critically examine past practices, rethink current methodologies, and collaboratively shape a future where data-driven research and policy practices are more inclusive, less biased, and more representative of all individuals and communities.

This page intentionally left blank

CHAPTER 1

RETHINKING DATA LIFE CYCLE PRACTICES: THE VALUE OF APPLYING REFLEXIVITY METHODS TO CURRENT DATA PRACTICES

EllenMarie McPhillip
University of Miami, USA

ABSTRACT

The data revolution has demonstrated its potential to provide evidence that can improve decision making and transform lives. While its benefits have been consequential, researchers, ethicists, and popular and social policy advocates suggest that efforts to better manage the "datasphere" are necessary (Bergé et al., 2018). This chapter offers a brief overview of some current concerns and advocates for researchers to become more conscious about the potential for bias in their data management practices and to deploy more conscientious engagement with data governance policies and tools. In this exploration of the current landscape governing the data lifecycle, we suggest that efforts to better sensitize data producers and users to the unobserved, but influential, sociological forces embedded in the questions they ask, the methods they use, and the experiences that "count," may help avoid perpetuating misconceptions or cementing biases into accepted knowledge.

Bias and Beyond, pages 1–12
Copyright © 2026 EllenMarie McPhillip
Published under exclusive licence by Emerald Publishing Limited.
doi:10.1108/978-1-80686-665-620261001

Keywords: Big data movement; data life cycle; data equity; ethical paradoxes; data governance; knowledge production; reflexivity

The "Roaring Twenties" (1920–1929) and the "Swinging Sixties" (1960–1969) were decades of such profound political, socio-economic, and cultural shifts that they defined a generation, changed a nation, or transformed human experience. The technological sea change in the decade spanning from 2010 to 2020 prompted such similarly disruptive socioeconomic and cultural change that it may deserve its *own* distinctive appellation.

What sets that decade apart? During this period, data collection and production technologies captured, measured, stored, and leveraged private individuals' data—our interests, behaviors and activities—at a scale unprecedented, even unimaginable, in human history. No subject or metric was too trivial or personal to track and share. Forging beyond traditional metrics, such as life expectancies, we began to track what we ate, how many steps we took, or how many times we visited a particular retailer (and the route we took to drive there). This freer exchange of information has produced a "tsunami of data" (Vidal, 2017).

While the benefits of the data revolution have been widespread and consequential, such vast quantities of data and attendant sociological, cultural, and political changes have energized age-old, ethical questions and introduced new ones (D'Ignazio & Klein, 2023).

The promise of the digital economy prompted some researchers from science, technology, engineering, and math fields, the social sciences, and the humanities to suggest that emergent data technologies could make traditional scientific methods in research seem obsolete, or at least outdated (Anderson, 2008). In their view, big data technologies could "inherently" provide "meaningful and insightful knowledge about social, political, and economic processes" by revealing previously unseen or even unimagined patterns and relationships (Anderson, 2008; Kitchin, 2021). Harari (2016) asserted that *data-ism* has gone further still, beyond research, becoming a new religion. Harari argued in *Homo Deus* that data-ism is now the prevailing scientific worldview, unsaid, unconscious, but shared by the scientific establishment throughout the world, relegating humans to the role of mere algorithms for generating data (Harari, 2016). Although some might disagree with that characterization, such data-ism sets the context for the emergence of the "Big Data phenomena" and its transformative effects on many professions and businesses (Donatha, 2017; Lohr, 2013). Anticipating the direction of big data, Prensky (2009) suggested that such research would mean that

> scientists no longer have to make educated guesses, construct hypotheses and models, and test them with data-based experiments and examples. Instead,

> they can mine the complete set of data for patterns that reveal effects, producing scientific conclusions without further experimentation because they can rely on analysis of a complete, digitally stored data set. (Enhancing our ability to conduct deeper analysis section, para. 1)

On the one hand, big data has been transformative by empowering individuals interested in social change, giving them new ways to analyze and understand their experiences of systemic or institutional inequities. Previously unknown or veiled disparities within societies, especially for historically marginalized groups, were no longer so easily concealed. Governments, influential NGOs, mega-foundations, and research institutions shared their findings, using data to reveal resource gaps and inequities (Canali, 2016). In addition to its impact on scientific research and public policy, data gave businesses an extraordinary opportunity to stimulate and drive the growth of their enterprises (Forbes Technology Council, 2020). Organizations such as Alphabet (Google's parent company), Amazon, Apple, Meta (Facebook), X (previously Twitter), and Microsoft, invented technologies and applied strategies to become some of the largest repositories for collecting, storing, and analyzing individuals' personal data points. They capitalized upon these billions of personal data points and transformed them into *their* companies' unique and proprietary assets. As in the sciences, the conversion and leveraging of data insights into ever more valuable "products" and "byproducts" (Dammann, 2019) underpinned the big data phenomenon.

But today, many feel that the early unfettered optimism of the data revolution has been replaced by a collective wariness and anxiety. This *zeitgeist* stems from compelling evidence that both lay people and experts have also been producers and purveyors of data operating within a *laissez-faire* attitude that has created a vast web of flawed data practices (Marcus, 2018). The lack of consistent guardrails for managing data, combined with the speed at which these technologies are being adopted, has allowed discreditable "information" to be distributed to unsuspecting or unsophisticated audiences at lightning speed. Even for credible information, the current methods of collecting, protecting, and managing data have proven woefully inadequate in managing the enormous opportunities and associated ethical challenges (Pike, 2019). These data technologies have permeated the private and public sectors so rapidly that prevailing data management and governance practices were and are still overwhelmed. Businesses, in particular, seem to be operating with ad hoc data collection and storage practices, incomplete research and analyses, ineffective quality controls, and governance practices that have proven to be inadequate for guaranteeing data use or provenance (Deur, 2024). According to McKinsey's annual *Global Survey*, many managers give short shrift to data compliance as a "strategic objective" and view it as purview of the legal or compliance teams (Edquist et al., 2022; Marcus, 2018). As of

2021, few organizations (17%) had dedicated resources for data governance, and only 30% thought it was a top priority (Edquist et al., 2022).

However, as data usage has grown, the rules and expectations around it have skyrocketed, producing an array of ethical, legal, and financial challenges. Researchers, ethicists, and popular and social policy advocates suggest that if society is to take full advantage of the potential value of the data revolution, we must also *understand* the vulnerabilities of new data technologies and practices. From the problem formulation to the data collection, data manipulation, data analysis, and data reporting, the issues of security, privacy, and ownership suggest that we must learn to manage the data life cycle differently.

First, however, we take a step back to examine how data becomes information and information becomes knowledge.

HOW DATA BECOMES KNOWLEDGE

Since long before society became literate, the desire to explain a phenomenon, extract meaning from a set of events and actions, or understand the relationships between or among individuals and their world has been an essential force in the human experience (Campbell, 1949). Individuals relied on intuition, superstition, or mythology to answer their questions about the world around them. However, individuals' singular experiences did not contribute to the creation of general knowledge because what was *worth* knowing was managed through a matrix of social, religious, and political beliefs and relationships held by a privileged few. From the ancient world through to the Reformation (1517–1600, Europe), churches and kings (or priests and monarchs, or other religious and state leaders) throughout the world held "monopolies of knowledge" that served to reinforce their power and positions in society (Innis, 1952, pp. 14–15).

The first efforts to systematically challenge this knowledge matrix dates to the 17th-century, where empiricism, along with methodologies for collecting data, standardizing experimentation and systematizing documentation, would be considered "scientific" approaches and therefore reliably accurate for explaining relationships between and among phenomena. In communicating observations and interpretations to others, data could be converted from instinctual impressions to information and knowledge through scientific inquiry, experimentation, analysis, and application. An instructive figure in this transformation is John Graunt, who pioneered data collection and analysis methodologies. Using crude, self-created methods, Graunt sought to address a question that was as much a political flash point as a scientific inquiry: What was the current population of London?

By organizing and analyzing recorded births, christenings, and deaths in London, Graunt revealed meaningful migration patterns which provided

the first data-based estimate of the city's population (Conner, 2024). This work was both precedent-setting and prescient. Graunt's work became an early example of data's life cycle. The data points he identified had been a messy assemblage of numbers—found in an assortment of ledgers, Bibles, and morgues. He organized and made sense of this voluminous, disorganized mass of data points and, even more importantly, communicated his findings broadly. No one else had aggregated and connected this data as evidence of a story or cogent narrative, much less a rational explanation for spread of the plague or shifts in the city's population (Slack, 1985; Sutherland, 2005).

Although Graunt's analysis could be perceived as a mere chronicle of individual births and deaths, his work offers an early example of the powerful influence of a new segment of knowledge producers: men who were part of the class of intellectuals or socio-economically privileged elites who had the time and intellectual training to engage in scientific methodologies. It also elucidates the interconnected relationship of knowledge and power often concealed within it.

As Gramsci's theories would suggest, Graunt's status as an intellectual presupposed his function as a deputy of the ruling class, enforcing its rules and behaviors (Gramsci, 1971). As a haberdasher, Graunt was not an aristocrat, but he was a member of society in good standing, educated in the accepted mores and values of the day. His social position and education provided the platform, the tools, and the foundation from which he could exercise his intellectual curiosity. According to the political and social norms of his time, the questions he asked were plausible and worthy of engagement in large part because he was a man of reasonable social status. And, although his social and economic status would fall precipitously in his later life, his findings function as a Foucauldian example of how

> all knowledge is possible and takes place only within a vast network or system of power relationships that allow that knowledge to come to be, in order for statements accepted as "true" in any context to be uttered, and in order for what counts as knowledge to be generated in the first place. (Fruhling, n.d.)

Consequently, knowledge and power were and are inextricably bound in the creation of each other and interwoven in support of each other's privilege and power.

But Graunt also displayed profound individual agency. In formulating and asking the question, he accomplished three things: (1) He decided that this data, seemingly detritus or dross, could instead become a valuable source for answering important social questions; (2) He set up the organizing principle for how to think about what the data revealed; (3) He decided which data points were to be included or excluded and which insights would matter. Graunt's insider assumptions, the questions he

asked, and the processes he used demonstrate the activities that converted loose data points into information. The refinement and conversion of that information into acceptable, unassailable, "knowledge" to be disseminated to the masses required human agency (Liamputtong, 2010). In other words, Graunt's findings were seamlessly transformed into "bona fide knowledge" as long as he followed certain conventions (Greenley, 1995).

Since Graunt's time, in practice and theory, data scientists and managers have been trained to believe that the data life cycle should be guided by experts in that specific subject matter domain (Keller et al., 2020). This is logical, as domain knowledge provides context to define, evaluate, and interpret findings at each stage of the data life cycle (Keller et al., 2020) However, the foundational practice of privileging domain knowledge and methodologies also suggests the many ways data can be corrupted. From collection to analysis to communication, there are myriad opportunities for flawed evidence or biases to enter and influence every point of the data life cycle (Ho & Beyan, 2020). Additionally, the demand from historically marginalized voices to be represented in and recognized as sources of knowledge has raised questions about the potential biases represented within those domains and their traditional practices. Scholars such as Mary Lou Rasmussen have raised questions about the domains themselves, asking *how* we have come to know what we know, *who decides* what knowledge means, *and who* knows what needs to be done (2009, 2017). The failure to include or represent such voices exposes the power/knowledge dynamic that has traditionally benefited each society's privileged elites. Scholars such as D'Ignazio argue that "data is never this raw, truthful input, and it is never neutral. It is information that has been collected in certain ways by certain actors and institutions for certain reasons" (Corbyn, 2020). In fact, Lyon (2014) notes that "*how* data are generated and framed always has decisive effects on the outcomes of analysis" (Critique section, para. 6). This combination of critical challenges reminds us that the data life cycle is neither ahistorical nor free of embedded sociological assumptions. Therefore, domain-based data professionals may now be ill-equipped to detect the insidious influence or mitigate the ethical consequences of bias in their work product. Hence, domain expertise is necessary, but it is not sufficient.

Given the ubiquity of data and society's dependence on it, for data to be more accurate, it must become more representative and precise. Users and makers of data—which is now all of us—must develop new skill sets for detecting and decoding how and where flawed evidence or bias could be influencing our conclusions. The questions that are asked, the methods that are used, and the conclusions that are drawn can thus be part of a vicious circle or a virtuous circle of knowledge generation. Therefore, none of us, especially those who make data-based decisions, can abdicate our role

in shaping or communicating data (D'Ignazio & Klein, 2023; Hesse et al., 2015). However, without an understanding of how a person's life experiences and worldviews may affect a data point's meaning or value, we do little to mitigate the potential to generate inaccurate, harmful information or partial truths (Dodgson, 2019). It is therefore crucial to reimagine new ethical methodologies for gathering and deploying data insights. Critical frameworks can help guide responsible management of the data life cycle, and one framework that can be especially useful for improving data practices is reflexivity.

REFLEXIVITY

As suggested earlier, the fast-paced evolution of the data landscape is challenging every aspect of data management regimes. The principles and methods that have served the field for generations have also reflected and reinforced biases leading to misrepresentations or exclusions of particular groups or perspectives (Dastin, 2018) Consequently, as we move into the middle of the next decade of big data, more effective practices for detecting and decoding how and where flawed evidence or bias could be influencing findings or outcomes becomes essential. Reflexivity is a time-honored methodology often used in the social sciences because it facilitates

> a researcher's conscious and deliberate effort to be attuned to one's own reactions to respondents and to the way in which the research account is constructed, it helps identify and explicate potential or actual effect of personal, contextual, and circumstantial aspects on the process and findings of the study and maintain their awareness of themselves as part of the world they study. (Berger, 2015, p. 3)

With consistent prompts for critical reflection, data users can consistently assess the potential impact of their perspectives, assumptions, and biases throughout the data lifecycle's stages. By embedding such reflexive practices into data management regimes as core professional principals, rather than elective choices, knowledge producers can better (a) assure historical contextualization, (b) avoid introducing unwitting assumptions and cultural expectations into their data, (c) spotlight anomalies that could perpetuate bias or inequity in their final work product, and (d) mitigate the risk of accumulating flawed data or incautiously applying problematic data and data practices (Canali, 2016; Dodgson, 2019).

Concerns about untrustworthy data and data sources finding its way into the knowledge generation cycle have produced attempts to address historical biases (McCollough, 2000). However, those efforts did not mitigate the impact of the challenges that have led to some of ethical dilemmas

including issues surrounding individual privacy and security (Bormida, 2021). Consequently, deploying reflexivity is not simply an attempt to correct for bias. The goal is to prevent flawed or incomplete data, unverified or partial conclusions from entering unchecked into the data life cycle. Emphasizing the responsible use of data, critical thinking and context-awareness while applying data-driven approaches to social issues also requires the use of reliable methods that ferret out Ensuring that data are also more inclusive and multiculturally representative can improve the quality and integrity of knowledge derived from that data (Fassinger & Morrow, 2013). However, building and implementing new protocols such as reflexivity into data management practices has proven to be a daunting task.

These sophisticated technologies and data management practices are increasingly accessible to laypersons and experts alike, and milliseconds separate theory from real-time activation by official or casual decision makers. Additionally, the outsized influence of governments and private entities, particularly in business and community-based organizations, has disrupted and fragmented attempts to create a systematic, uniform set of principles that govern international and national data practices and ethical standards. The difficulties in development and implementation of new governance data practices are a function of what I call *ethical paradoxes in progress*. An ethical paradox in progress is a complex ethical situation or dilemma developing or unfolding in real-time. Society is still grappling with defining the terms, "ethics" in question, and calculating the meaning and consequences for ignorance or violation. We have not yet reached a consensus on the relevant principle(s), and therefore, neither understand the norms nor the standards for appropriate recourse in the face of uncertainty or unsavory action. Reflexivity can assist us here as well, promoting transparency in decision-making and data handling. It encourages data users to acknowledge any perspectives, values, and/or assumptions that might influence the interpretation of a data set. By integrating reflexivity into statistical methods, model choices, and data visualization techniques, organizations and researchers can work against reinforcing or introducing bias, power differentials, cognitive impairments, cultural differences, or situational pressures.

CHATGPT AND AI

While this essay has attempted to encourage decision-makers to be hyper-aware of intangible, influential forces and adopt a uniform set of principles that govern international and national data practices and ethical standards, the release of ChatGPT and other artificial intelligence (AI) tools have amplified the need to recalibrate our thinking about the use of data and these technologies. The language-based algorithms are based entirely