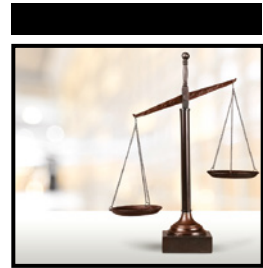


GRANT FLEMING
PETER BRUCE



RESPONSIBLE DATA SCIENCE

Responsible Data Science



Responsible Data Science

**Transparency and Fairness
in Algorithms**

Grant Fleming
Peter Bruce

WILEY

Copyright © 2021 by John Wiley & Sons, Inc., Indianapolis, Indiana

Published simultaneously in Canada

ISBN: 978-1-119-74175-6

ISBN: 978-1-119-74177-0 (ebk)

ISBN: 978-1-119-74164-0 (ebk)

Manufactured in the United States of America

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at www.wiley.com/go/permissions.

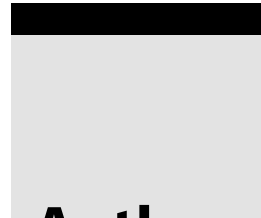
Limit of Liability/Disclaimer of Warranty: The publisher and the author make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation warranties of fitness for a particular purpose. No warranty may be created or extended by sales or promotional materials. The advice and strategies contained herein may not be suitable for every situation. This work is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional services. If professional assistance is required, the services of a competent professional person should be sought. Neither the publisher nor the author shall be liable for damages arising herefrom. The fact that an organization or Web site is referred to in this work as a citation and/or a potential source of further information does not mean that the author or the publisher endorses the information the organization or website may provide or recommendations it may make. Further, readers should be aware that Internet websites listed in this work may have changed or disappeared between when this work was written and when it is read.

For general information on our other products and services please contact our Customer Care Department within the United States at (877) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at booksupport.wiley.com. For more information about Wiley products, visit www.wiley.com.

Library of Congress Control Number: 2021933659

Trademarks: Wiley and the Wiley logo are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates, in the United States and other countries, and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc. is not associated with any product or vendor mentioned in this book.



About the Authors

GRANT FLEMING is a data scientist at Elder Research, Inc. His professional focus is on machine learning for social science applications, model interpretability, civic technology, and building software tools for reproducible data science.

PETER BRUCE is the Chief Learning Officer at Elder Research, Inc., author of several best-selling texts on data science, and Founder of the Institute for Statistics Education at Statistics.com, an Elder Research Company.



About the Technical Editor

ROBERT DE GRAAF is a data scientist and statistician from Melbourne, Australia. He is the author of *Managing Your Data Science Projects* and coauthor of *SQL Cookbook, 2nd edition*. He is husband to Clare and father to Maya and Leda, and enjoys playing guitar and learning new languages.



Acknowledgments

First and foremost, we acknowledge the support of Elder Research, Inc., and of John Elder (chairman) and Gerhard Pilcher (CEO) in particular. We have benefited greatly from the technical and philosophical conversations we have shared with our colleagues. Elder Research has been most generous in permitting us to pursue this project. At the same time, this book has not been reviewed or edited by the company, and we, the authors, bear sole responsibility for all opinions, errors, and omissions.

We thank, especially, our coauthors on select chapters. Will Goodrum lent his expertise to the legal issues explored in Chapter 3, “The Way AI Goes Wrong, and the Legal Implications,” and Sam Ballerini helped enormously in writing the code for the facial recognition CNN example in Chapter 8, “Auditing for Neural Networks.”

Robert de Graaf served as technical editor, raising important points and contributing in many places to a better book. This book certainly would have been incomplete without his input.

Our editorial team at Wiley has been most supportive throughout the process. Jim Minatel, associate publisher, embraced our vision from the beginning. Our editor, Jan Lynn, kept us on track and patiently shepherded the various pieces of the project to all come together. Saravanan Dakshinamurthy handled the production side of things, Louise Watson did the copy-editing, and Pete Gaughan managed the process behind the scenes.

We would like to thank Matthew Dwinnell and Amy Zhang for their tips on working with HTML and CSS, as well as our co-workers Brittany Pugh and Chris Lee for their advice and feedback. Grant would like to thank his professors and mentors, including Edward Munn Sanchez, Lite Nartey, Edward R. Carr, Gregory Magai Patterson, Jennifer Bess, Mark Schaffer, Patrick Jessee,

and Brandie Wagner, for their endless support of his efforts and spirit. Peter's appreciation goes to Galit Shmueli, his coauthor on other book projects, with whom he has had lively conversations on ethical issues surrounding the practice of data science.

Finally, we express our appreciation to our students at `Statistics.com` who have made constructive and useful comments on the material presented here.



Contents at a Glance

Introduction		xix
Part I	Motivation for Responsible Data Science and Background Knowledge	1
Chapter 1	Responsible Data Science	3
Chapter 2	Background: Modeling and the Black-Box Algorithm	27
Chapter 3	The Ways AI Goes Wrong, and the Legal Implications	49
Part II	The Responsible Data Science Process	71
Chapter 4	The Responsible Data Science Framework	73
Chapter 5	Model Interpretability: The What and the Why	99
Part III	RDS in Practice	135
Chapter 6	Beginning a Responsible Data Science Project	137
Chapter 7	Auditing a Responsible Data Science Project	173
Chapter 8	Auditing for Neural Networks	225
Chapter 9	Conclusion	265
Index		273



Contents

Introduction	xix
Part I Motivation for Responsible Data Science and Background Knowledge	1
Chapter 1 Responsible Data Science	3
The Optum Disaster	4
Jekyll and Hyde	5
Eugenics	7
Galton, Pearson, and Fisher	7
Ties between Eugenics and Statistics	7
Ethical Problems in Data Science Today	9
Predictive Models	10
From Explaining to Predicting	10
Predictive Modeling	11
Setting the Stage for Ethical Issues to Arise	12
Classic Statistical Models	12
Black-Box Methods	14
Important Concepts in Predictive Modeling	19
Feature Selection	19
Model-Centric vs. Data-Centric Models	20
Holdout Sample and Cross-Validation	20
Overfitting	21
Unsupervised Learning	22
The Ethical Challenge of Black Boxes	23
Two Opposing Forces	24
Pressure for More Powerful AI	24
Public Resistance and Anxiety	24
Summary	25

Chapter 2	Background: Modeling and the Black-Box Algorithm	27
	Assessing Model Performance	27
	Predicting Class Membership	28
	The Rare Class Problem	28
	Lift and Gains	28
	Area Under the Curve	29
	AUC vs. Lift (Gains)	31
	Predicting Numeric Values	32
	Goodness-of-Fit	32
	Holdout Sets and Cross-Validation	33
	Optimization and Loss Functions	34
	Intrinsically Interpretable Models vs. Black-Box Models	35
	Ethical Challenges with Interpretable Models	38
	Black-Box Models	39
	Ensembles	39
	Nearest Neighbors	41
	Clustering	41
	Association Rules	42
	Collaborative Filters	42
	Artificial Neural Nets and Deep Neural Nets	43
	Problems with Black-Box Predictive Models	45
	Problems with Unsupervised Algorithms	47
	Summary	48
Chapter 3	The Ways AI Goes Wrong, and the Legal Implications	49
	AI and Intentional Consequences by Design	50
	Deepfakes	50
	Supporting State Surveillance and Suppression	51
	Behavioral Manipulation	52
	Automated Testing to Fine-Tune Targeting	53
	AI and Unintended Consequences	55
	Healthcare	56
	Finance	57
	Law Enforcement	58
	Technology	60
	The Legal and Regulatory Landscape around AI	61
	Ignorance Is No Defense: AI in the Context of Existing Law and Policy	63
	A Finger in the Dam: Data Rights, Data Privacy, and Consumer Protection Regulations	64
	Trends in Emerging Law and Policy Related to AI	66
	Summary	69
Part II	The Responsible Data Science Process	71
Chapter 4	The Responsible Data Science Framework	73
	Why We Keep Building Harmful AI	74
	Misguided Need for Cutting-Edge Models	74
	Excessive Focus on Predictive Performance	74

	Ease of Access and the Curse of Simplicity	76
	The Common Cause	76
	The Face Thieves	78
	An Anatomy of Modeling Harms	79
	The World: Context Matters for Modeling	80
	The Data: Representation Is Everything	83
	The Model: Garbage In, Danger Out	85
	Model Interpretability: Human Understanding for Superhuman Models	86
	Efforts Toward a More Responsible Data Science	89
	Principles Are the Focus	90
	Nonmaleficence	90
	Fairness	90
	Transparency	91
	Accountability	91
	Privacy	92
	Bridging the Gap Between Principles and Practice with the Responsible Data Science (RDS) Framework	92
	Justification	94
	Compilation	94
	Preparation	95
	Modeling	96
	Auditing	96
	Summary	97
Chapter 5	Model Interpretability: The What and the Why	99
	The Sexist Résumé Screener	99
	The Necessity of Model Interpretability	101
	Connections Between Predictive Performance and Interpretability	103
	Uniting (High) Model Performance and Model Interpretability	105
	Categories of Interpretability Methods	107
	Global Methods	107
	Local Methods	113
	Real-World Successes of Interpretability Methods	113
	Facilitating Debugging and Audit	114
	Leveraging the Improved Performance of Black-Box Models	116
	Acquiring New Knowledge	116
	Addressing Critiques of Interpretability Methods	117
	Explanations Generated by Interpretability Methods Are Not Robust	118
	Explanations Generated by Interpretability Methods Are Low Fidelity	120
	The Forking Paths of Model Interpretability	121
	The Four-Measure Baseline	122

	Building Our Own Credit Scoring Model	124
	Using Train-Test Splits	125
	Feature Selection and Feature Engineering	125
	Baseline Models	127
	The Importance of Making Your Code Work for Everyone	129
	Execution Variability	129
	Addressing Execution Variability with Functionalized Code	130
	Stochastic Variability	130
	Addressing Stochastic Variability via Resampling	130
	Summary	133
Part III	RDS in Practice	135
Chapter 6	Beginning a Responsible Data Science Project	137
	How the Responsible Data Science Framework Addresses the Common Cause	138
	Datasets Used	140
	Regression Datasets—Communities and Crime	140
	Classification Datasets—COMPAS	140
	Common Elements Across Our Analyses	141
	Project Structure and Documentation	141
	Project Structure for the Responsible Data Science Framework: Everything in Its Place	142
	Documentation: The Responsible Thing to Do	145
	Beginning a Responsible Data Science Project	151
	Communities and Crime (Regression)	151
	Justification	151
	Compilation	154
	Identifying Protected Classes	157
	Preparation—Data Splitting and Feature Engineering	159
	Datasheets	161
	COMPAS (Classification)	164
	Justification	164
	Compilation	166
	Identifying Protected Classes	168
	Preparation	169
	Summary	172
Chapter 7	Auditing a Responsible Data Science Project	173
	Fairness and Data Science in Practice	175
	The Many Different Conceptions of Fairness	175
	Different Forms of Fairness Are Trade-Offs with Each Other	177
	Quantifying Predictive Fairness Within a Data Science Project	179

Mitigating Bias to Improve Fairness	185
Preprocessing	185
In-processing	186
Postprocessing	186
Classification Example: COMPAS	187
Pework: Code Practices, Modeling, and Auditing	187
Justification, Compilation, and Preparation Review	189
Modeling	191
Auditing	200
Per-Group Metrics: Overall	200
Per-Group Metrics: Error	202
Fairness Metrics	204
Interpreting Our Models: Why Are They Unfair?	207
Analysis for Different Groups	209
Bias Mitigation	214
Preprocessing: Oversampling	214
Postprocessing: Optimizing Thresholds	
Automatically	218
Postprocessing: Optimizing Thresholds Manually	219
Summary	223
Chapter 8 Auditing for Neural Networks	225
Why Neural Networks Merit Their Own Chapter	227
Neural Networks Vary Greatly in Structure	227
Neural Networks Treat Features Differently	229
Neural Networks Repeat Themselves	231
A More Impenetrable Black Box	232
Baseline Methods	233
Representation Methods	233
Distillation Methods	234
Intrinsic Methods	235
Beginning a Responsible Neural Network Project	236
Justification	236
Moving Forward	239
Compilation	239
Tracking Experiments	241
Preparation	244
Modeling	245
Auditing	247
Per-Group Metrics: Overall	247
Per-Group Metrics: Unusual Definitions	
of “False Positive”	248
Fairness Metrics	249
Interpreting Our Models: Why Are They Unfair?	252
Bias Mitigation	253
Wrap-Up	255
Auditing Neural Networks for Natural	
Language Processing	258

	Identifying and Addressing Sources of Bias in NLP	258
	The Real World	259
	Data	260
	Models	261
	Model Interpretability	262
	Summary	262
Chapter 9	Conclusion	265
	How Can We Do Better?	267
	The Responsible Data Science Framework	267
	Doing Better As Managers	269
	Doing Better As Practitioners	270
	A Better Future If We Can Keep It	271
Index		273



Introduction

In this book, we will review some of the harmful ways artificial intelligence has been used and provide a framework to facilitate the responsible practice of data science. While we will touch upon mitigating legal risks, in this book we will focus primarily on the modeling process itself, especially on how factors overlooked by current modeling practices lead to unintended harms once the model is deployed in a real-world context.

Three core themes will be developed through this book:

- Any AI algorithm can have a harmful, dark side: once they are applied in the real world, AI algorithms can cause any number of harms. An algorithm designed to help police catch murderers can later be appropriated by totalitarian states to persecute dissidents; an algorithm that expands the availability of financial credit for the vast majority of people may nonetheless intensify bias against minorities.
- The dark sides of AI algorithms are created or deepened by current modeling approaches. By focusing only on technical considerations like maximizing predictive performance, data scientists ignore the potential for their model to aggravate biases against certain groups, generate harmful predictions, or otherwise be used by other groups in the future for malicious purposes.
- New modeling approaches are needed if we want to use AI more responsibly. If data scientists and their users are going to continue to use AI algorithms to make consequential decisions, then they ought to do so with consideration for a broader range of technical and societal factors than are normally considered.

New U.S. diplomats in training used to be told “not to give unintentional offense.” Our primary goal for this book is to tell you a variant of this: that there are a number of specific actionable steps that you, the reader, can begin taking to reduce the risk of causing unintentional harm with your models.

In particular, this book focuses on how to make models more transparent, interpretable, and fair. It will present illustrations and snippets of code in a way that a technically literate manager or executive can understand, without necessarily knowing any programming language.

What This Book Covers

Chapter 1, “Responsible Data Science,” provides historical background for the ethical concerns in statistics and an introduction to basic modeling methods. In Chapter 2, “Background: Modeling and the Black-Box Algorithm,” we define various types of predictive models and briefly discuss the concepts of model transparency and model interpretability. Chapter 3, “The Ways AI Goes Wrong, and the Legal Implications,” reviews the landscape of the types of ethics and fairness issues encountered in the practice of data science (e.g., legal constraints, privacy and data ownership concerns, and algorithms “gone bad”) and finishes by distinguishing interpretable models from black-box models. In Chapter 4, “The Responsible Data Science (RDS) Framework,” we discuss the desired characteristics of a Responsible Data Science framework, summarize the attempts by other groups at creating one, and combine the lessons learned from these other groups with those presented in the book up until this point to construct our own framework, the aptly named the Responsible Data Science (RDS) framework. Chapter 5, “Model Interpretability: The What and the Why,” prepares the reader for implementing the RDS framework in later chapters by doing a deeper dive into model interpretability and how it can be achieved for black-box models. We begin setting up a responsible data science project within our framework and performing initial checks on two datasets in Chapter 6, “Beginning a Responsible Data Science Project.” In Chapters 7, “Auditing a Responsible Data Science Project,” and Chapter 8, “Auditing for Neural Networks,” we delve into case studies in auditing conventional machine learning models and deep neural networks for failure scenarios, fairness, and interpretability. Finally, we conclude the book in Chapter 9, “Conclusion,” with a look to the future and a call to action.

Who Will Benefit Most from This Book

Much has been written elsewhere about the legal issues relevant to AI; thus, our primary audience is not corporate general counsels. Instead, this book is intended for the following two groups:

- Data-literate managers and executives
- Business-literate data scientists and analysts

Although the focus placed on responsibility in data science is relatively new, many people have been trained in the myriad wonderful things that AI can accomplish. They have also read in the news about the ethical lapses in some AI projects. These lapses are not surprising, because relatively few data scientists are trained in how to adequately understand and control their AI while maintaining high predictive performance in models. Hence, we aim this book at data science managers and executives and at data science practitioners.

Practitioners will learn of the ways in which their models, intended to provide benefits, can at the same time cause harm. They will learn how to leverage fairness metrics, interpretability methods, and other interventions to their model or dataset to audit those models, identifying and mitigating possible issues prior to deployment or result delivery. Through worked examples, the book guides users in structuring their models to have a greater consideration for ethical impacts, while assuring that best practices are followed and model performance is optimized. This is a key differentiator for our book, as most responsible AI frameworks do not provide specific technical recommendations for fulfilling the principles that they lay out.

Managers of data science teams, and managers with any responsibilities in the analytics realm, can use this book to stay alert for the ways in which analytical models can run afoul of ethical practices, and even the law. More importantly, they will learn the language and concepts to engage their analytics teams in the solutions and mitigation steps that we propose. While some code and technical discussion is provided, following it in detail is by no means needed. The overall presentation in the book is at a level that provides managers who are at least somewhat familiar with analytics the ability and tools to instill responsible best practices for data science in their organizations.

Finally, a word to individual data scientists. You may think that your project has no implications in the ethical realm. The real-world context for deployment may seem innocuous, the modeling task may seem harmless, and the content of this book may not seem relevant to your project. Though the ideas and techniques presented in this book are primarily discussed in the context of

ethically fraught models, they are still useful as the basis for best practices in other modeling contexts. After all, there is a great degree of overlap between traditional best practices for modeling and best practices for responsible data science. Doing data science more responsibly, in the manner that we lay out in this book, improves understanding of the relationships between a model and its real-world deployment context, improves transparency and accountability through better guidelines for documentation, and reduces the risk of unanticipated biases creeping into models by providing workflows for model auditing. Plus, who knows when that innocuous-sounding project may later turn out to have a dark side?

Looking Ahead in This Book

The responsible practice of data science covers a lot of ground in different dimensions.

- **Formal legal and regulatory requirements:** Clearly, any company or individual developing or implementing data science solutions will want to stay on the right side of the law. The most famous attempt to regulate AI is the GDPR; it runs over 80 pages and is quite detailed. It was developed to meet the demands of a specific point in time, but there is no guarantee that it will be a useful guide in the future. Things change rapidly in the field of AI, and the GDPR is like a boulder placed in the path of a stream—sooner or later, the stream will find ways around the obstacle. There are already a number of publications on this topic, and our audience is not the corporate general counsel but rather the manager and the data science practitioner. So, while this book will touch on key laws in this area, such as the GDPR, it will not do so in great depth.
- **Bad actors:** In many cases, the pernicious use of AI is neither inadvertent nor the result of lack of understanding—it is intentional. Deep learning has been put to malicious use by cyber hacks who can digest and analyze multilayered defense mechanisms to determine quickly where weaknesses lie. When those who are responsible for data science development and implementation have malevolent intentions, a lecture on responsibility and a course on ethics will not have much impact. This book will note countermeasures that can have some effect, but dealing with bad actors, like dealing with regulators, is not the primary focus of this book.
- **AI out of control:** In many cases, those deploying AI are responsible parties, obeying the law, and yet their AI has in some sense “escaped their full control” after deployment. Perhaps it has morphed into something that was not initially intended, or perhaps it has triggered effects and

reactions that were unanticipated. Maybe not all decision-makers in the organization that designed the AI, or affiliated stakeholders throughout the project, fully understood or appreciated from the beginning all of the ways that their AI project would operate in a real-world context. The disconnect between the goals of the model and the realities of the real-world context might make it so that even a perfectly accurate model can cause a great deal of harm. This overarching issue is the main focus of the book: how executives, managers, and practitioners can follow best practices in ethical data science—in particular, how they can better understand, explain, and gain control over their AI implementations.

Special Features

DEFINITION Throughout the book, we'll explain the meanings of terms that may be new or nonstandard.

NOTE Inline boxes are used to expand further on some aspect of the topic without interrupting the flow of the narrative.

Small general discussions that deserve special emphasis or have relevance beyond the immediately surrounding content are called out in general sidebar notes.

Code Repository

Code referred to in the text of each of the chapters, plus updates and expanded code for generating additional results, can be found in the repositories at www.wiley.com/go/responsibledata-science and github.com/Gfleming/responsibledata-science. Unless otherwise noted in the text, the code to reproduce the results within each of the chapters can be found by navigating to the appropriately named chapter subfolders at either of the links (e.g., the code for Chapter 6 can be found in the `responsibledata-science/ch6` subfolder.) The README file within the head of the code repository folder provides instructions for setting up your software environment, and the README files within each of the chapter subfolders provide additional information about the code for that chapter.

Part

I

Motivation for Responsible Data Science and Background Knowledge

In This Part

Chapter 1: Responsible Data Science

Chapter 2: Background: Modeling and the Black-Box Algorithm

Chapter 3: The Ways AI Goes Wrong, and the Legal Implications

Responsible Data Science

Data science is an interdisciplinary field that combines elements of statistics, computer science, and information technology to generate useful insights from the increasingly large datasets that are generated in the normal course of business. Data science helps organizations capture value from their data, reducing costs and increasing profits, and also enables completely new types of endeavors, such as powerful information search and self-driving cars. Sometimes, data science projects can go awry, when the predictions made by statistical and machine learning algorithms turn to be not just wrong, but biased and unfair in ways that cause harm. History has shown that the dual good and evil nature of statistical methods is not new, but rather a characteristic that was present from nearly the moment that they were conceived. However, by adjusting and supplementing statistical and machine learning methods and concepts, we can diagnose and reduce the harm that they may otherwise cause.

In popular and technical writing, these issues are often captured by the general term “ethical data science.” We use that term here, but we also use the more general phrase “responsible data science.” Ethics can refer in some usages to narrow “rules of the road” that pertain to a particular profession, such as real estate or accounting. Our goal here is broader than that: presenting a framework for the practice of data science that is ethical, but not in a narrow sense: it is *responsible*.

The Optum Disaster

In 2001, the healthcare company Optum launched Impact-Pro, a predictive modeling tool. Impact-Pro was an early success for predictive analytics (predating the term *data science*), and a decade later, Steven Wickstrom, an Optum VP, touted its use cases. For healthcare providers, it could “support steerage to appropriate programs” and “identify members [patients] with gaps in care, complications, and comorbidities.” Optum termed these *care opportunities* in one document (i.e., opportunities for more revenue), but they are also of interest to those concerned with cost management: the correct early intervention in a health problem can cost significantly less than more drastic action later. For insurers, information on health risks for specific groups and individuals could be used to set premiums more accurately than is possible using traditional underwriting criteria.

DEFINITION DATA SCIENCE We use the term *data science* broadly to cover the process of understanding and defining a problem, gathering and preparing data, using statistical methods to answer questions, fitting models and assessing them, and deploying models in an organizational setting. We consider artificial intelligence (AI) to be part of data science, and we also consider the “science” component of data science to be important.

In 2019, though, a research team found that the tool was fundamentally flawed. For one important group—African Americans—the tool consistently underpredicted need for healthcare. The reason? The tool was essentially built to predict future spending on healthcare, and prior spending was a key predictor for that goal. And prior spending is a function not just of need, but also of ability to pay for and gain access to healthcare. Relative to other ethnic groups in the United States, African Americans have been (and continue to be) less insured, are less able to access healthcare, and possess fewer financial resources for covering healthcare expenses. In Optum’s data, therefore, African Americans had less prior spending and, hence, less predicted future need. As a result, African Americans were less targeted for preventive intervention and necessary follow-up healthcare than were other people with similar health profiles. Neither the model nor the data provided to it were able to account for the unanticipated and overlooked societal inequities lurking beneath.

Optum was blindsided. The company thought it had built a tool that was a winner on all fronts: improving health outcomes by being smarter about required follow-up care, and managing costs better in the bargain. Instead, it found itself the focus of widespread bad publicity and was pilloried for creating a product that exacerbated racial bias and widened the healthcare gap faced by African Americans. New York state regulators opened an investigation, and the controversy continued into 2020. At the time of writing, Optum continues to market Impact Pro.