Data Science Ethics

Concepts, Techniques and Cautionary Tales

DAVID MARTENS

OXFORD
UNIVERSITY PRESS

Contents

About the Author xiii

1. Introduction to Data Science Ethics 1
   - 1.1 The Rise of Data Science (Ethics) 1
   - 1.2 Why Care? 2
   - 1.3 Right and Wrong 4
   - 1.4 Data Science 7
   - 1.5 Data Science Ethics Equilibrium 10
   - 1.6 The FAT Flow Framework for Data Science Ethics 13
   - 1.7 Summary 29

2. Ethical Data Gathering 31
   - 2.1 Privacy as a Human Right 32
   - 2.2 Regulations 36
   - 2.3 Privacy Mechanisms 45
   - 2.4 Cautionary Tales: Backdoors and Messaging Encryption 62
   - 2.5 Bias 70
   - 2.6 Cautionary Tales: Bumps, Gorillas, and Resumes 75
   - 2.7 Human Experimentation 77
   - 2.8 Cautionary Tales: Dating, Happiness, and Ads 81
   - 2.9 Summary 82

3. Ethical Data Preprocessing 85
   - 3.1 Defining and Measuring Privacy 86
   - 3.2 Cautionary Tales: Re-identification 92
   - 3.3 Defining and Selecting Variables 102
   - 3.4 Cautionary Tale: Pregnancy and Face Recognition 104
   - 3.5 Fair Relabelling 111
   - 3.6 Cautionary Tale: Biased Language 115
   - 3.7 Summary 117

4. Ethical Modelling 121
   - 4.1 Privacy-Preserving Data Mining 122
   - 4.2 Discrimination-Aware Modelling 136
   - 4.3 Cautionary Tale: Predicting Recidivism and Redlining 143
   - 4.4 Comprehensible Models and Explainable AI 148
   - 4.5 Cautionary Tale: Explaining Webpage Classifications 166

4.6 Including Ethical Preferences: Self-Driving Cars 168
4.7 Summary 170

5. Ethical Evaluation 173
5.1 Ethical Measurement 173
5.2 Ethical Interpretation of the Results 178
5.3 Ethical Reporting 182
5.4 Cautionary Tale of Diederik Stapel 188
5.5 Summary 189

6. Ethical Deployment 191
6.1 Access to the System 191
6.2 Different Treatments for Different Predictions 195
6.3 Cautionary Tales: Censoring Search and Face Recognition 198
6.4 Honesty and DeepFake 200
6.5 Governance 202
6.6 Unintended Consequences 205
6.7 Summary 211

7. Conclusion 213

Bibliography 217
Index 249