Misinformation and Disinformation

Victoria L. Rubin

Misinformation and Disinformation

Detecting Fakes with the Eye and AI

Springer

Victoria L. Rubin
Western University
London, ON, Canada

ISBN 978-3-030-95655-4 ISBN 978-3-030-95656-1 (eBook)
https://doi.org/10.1007/978-3-030-95656-1

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

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Introduction

Abstract How do we detect, deter, and prevent the spread of mis- and disinformation with the human eye and AI? How does theory inform the practice, and how do the evidence-based research and best practices in lie-catching and truth-seeking professions—inform AI? The book looks into well-established human practices such as the routines and processes used in detective work, journalism, and scientific inquiry, and how they contribute toward innovative AI solutions. The book explains the principles, inner workings, and recent evolution of five types of state-of-the-art AI technologies suitable for curtailing the spread of mis- and disinformation: automated deception detectors, clickbait detectors, satirical fake detectors, rumor debunkers, and computational fact-checking tools.

Keywords Misinformation; Disinformation; Fake news; False news; Falsehoods; Infodemic; Satire; Clickbait; Information manipulation; Online deception; Library and Information Science; LIS; Journalism; Psychology; Communication; Social Sciences; Infodemiology; Natural Language Processing; NLP; Artificial Intelligence; AI; Machine Learning; ML; Algorithms; Automated Deception Detection; News verification; Automated fact-checking; Satire detection; Clickbait identification

Have you ever been duped? Do you tend to believe what you see or hear? Do you naturally trust others? If so, like many of us, you may be an easy target for sinister pranks or serious fraud. The world is full of phone scammers, email spammers, and other con-artists who work hard to influence what we think, what choices we make, and what we ultimately do. The classic definition of deception is a message intentionally transmitted to bolster a false impression, idea, or belief in someone else's mind (Buller & Burgoon, 1996). Why do we fall for such messages? Being a mark for an online con-artist is an unenviable position but avoiding it by staying on top of the latest online tricks and false claims is not easy. So, how can we get to the truth and avoid being fooled? What shapes do manipulative influences take? How do we guard ourselves against unwanted deceptive attempts? With so much of our daily

lives spent online, we are learning how to be more vigilant, how to fight our unconscious biases, and how to be more critical of online sources. Are there any assistive technologies that we can rely on to alleviate these burdens? Or, is the battle ultimately in our own minds?

This book shares the results of my 15-year quest to understand deceptive behaviors: the shapes and forms of deceptive messages, how they are expressed in language, and how they can be spotted using both the human eye and AI (Artificial Intelligence). I focus on how theory informs practice and how the evidence-based research and best practices from multiple professions (e.g., journalists, detectives, social and computer scientists) can help us to systematically separate lies from truths, mistaken beliefs from facts, and myths from reality. How do we amplify and complement human intelligence with artificial ones? What have the leading research and development (R&D) labs achieved thus far toward the goal of scalable AI-solutions to curtail the spread of mis- and disinformation? How successful have their efforts been? Is AI mimicking the procedures and know how of the experts, or does it require entirely new systematic approaches?

My plunge into researching the capabilities of AI deception detection systems automatically forced me to scrutinize our innately human behaviors. I consider the very essence of human intelligence, by contrasting it with AI, and make observations about our manipulative use of language, as well as human ineptitude in spotting lies. While the human mind is the ultimate built-in detector, it requires preparation, awareness, and practice to achieve better thinking. As Hemingway once said to a reporter from *The Atlantic*, "every man should have a built-in automatic crap detector operating inside him. It also should have a manual drill and a crank handle in case the machine breaks down" (Manning, 1965). I review what may cause such breakdowns: the circumstances that make us susceptible to being fooled and manipulated. I also speculate about how to extricate ourselves from the powers of persuasive propaganda.

Driving Forces Behind This Book

What drove me to write this book is my desire to endorse the value of education and show its concrete application in the fight against the infodemic. Without any distinct political or partisan agenda, I promote rational thinking, well-informed decision-making, and deeper self-awareness. Of the two modes within which we process information, known in social and cognitive psychology as the dual processing model (Chaiken & Ledgerwood, 2007), one mode is to think in a quick, associative, relatively automatic, and superficial way. The other way is an effortful, reflective, systematic mode of thinking. Which mode we are in influences our outcomes such as the conclusions, judgments, and attitudes we form and commit to memory. I warn against the dangers of superficial thinking and emphasize the need to be critical of the information we encounter in digital environments.

My main claim in this book is that in the fight against the infodemic, some technological assistance is inevitable and likely to come from AI-enabled applications. In other words, our human intelligence can be, at least in part, enhanced with an artificial one. We need systematic analyses that can reliably and accurately sift through big data that comes at us in large volumes, with velocity, and in a variety of formats. Since opinions, claims, thoughts, news, requests, promises, and so forth are often expressed in words (or textual format), we are talking in part about linguistic behaviors.

I have always been passionate about languages. I was born, raised, and educated in a Russian-speaking part of Ukraine. My working language for the last 25 years has been English. I have spent over half of my life in North America; progressed from one graduate degree to another in the States. By now, I have been a professor and the director and principal investigator of the Language and Information Technology Research Lab (LIT.RL) at Western University in Canada for about 15 years.

As a multilingual, I converse in up to six different languages including French, Spanish, and Japanese. I also occasionally get by in a few Slavic languages due to their similarities to my native Russian and Ukrainian. Eager to discover new cultures and use my linguistic gifts, I have relocated from country to country, and traveled extensively in Europe, Mexico, and Japan. How we use languages in the context of our everyday social lives has been my perpetual curiosity.

What specifically fascinates me is how we use language under challenging circumstances, especially when what we think and what we say does not exactly match. During my travels, I kept running into differences in attitudes about lying across languages and cultures. For instance, Japanese speakers puzzled me with their apparent inability to refuse face-to-face requests. It would force them at times to invent fake excuses or be vague, instead of simply saying "No" (as described earlier in Rubin (2014).) Spanish speakers in Mexico would rather give me wrong driving directions than admit that they simply did not know the route. Many improvized freely with their knowledge of local autostradas.

Lying and deception may be distinctly cultural, yet universal, in the sense of their relevance to our human condition, and our ability to share ideas through language. Most cultures deeply frown upon serious prevarications but seem to excuse white lies such as lying about surprise birthday parties. Some cultures disagree on whether to disclose the terminal diagnosis to fatally sick patients (see, for example, Blum (2007)). Universally, people lie to avoid hurting their loved ones, to get themselves out of predicaments, and for the sake of self-preservation, image-management, or other personal gains. The more I studied deceptive behaviors, the more I was confronted with variations in deception's forms, formats, motives, and justifications. I looked to put the pieces of the truth-deception puzzle together.

Intended Readers

If you are interested in issues concerning truth and deception, digital fake news, information disorder, mis- or disinformation, and the use of AI to curtail the infodemic, this book is for you. If you have a persistent online presence and generate online content (as a blogger, "citizen journalist," or other website creator), or if you simply consume large volumes of social media daily, you may be looking for this book as a primer on deception research in the humanities and social sciences, with an additional technological digest of "know-hows." You can build on these foundations for educational campaigns, policy making, or for your personal educational, financial, or political advantage.

Anyone that deals with large amounts of information, professionally or otherwise, probably realizes that AI can help solve the problem of the infodemic, but they may lack the mathematical or computer science background needed to understand how this can be done. I am a computational linguist and information scientist by training. Over the past 15 years, I have taught future information professionals (librarians, archivists, record managers, metadata specialists, etc.) in our library and information science (LIS) programs. In my experience, more LIS students are comfortable reading in the humanities and social sciences than in the more technical computational disciplines. Some graduate students are certainly more tech-savvy than others, but most share a curiosity for innovative technological solutions. They often wonder what promise intelligent systems hold in solving the problem of the mis- and disinformation. How can AI be combined with instruction and training that information professionals offer at their workplaces? Even if we gloss over some technicalities, we can still discuss the inner workings of the AI systems. This conversation can reveal what otherwise may seem like magic tricks to those who do not read specialized AI literature. Considering the pros and cons of AI systems can stimulate intelligent conversations about their capabilities, the principles that guide their research and development, and the rationale behind the adoption of these systems.

If you are a programmer or an expert in information retrieval or other areas within computer science broadly, you can still benefit from the overview of the human side of language computing. I explore psychological, philosophical, and communication models that account for human linguistic behaviors: the human factor behind the data-driven world of AI computing. For more technical readers who are well versed in AI and natural language processing (NLP), my overview may feel a bit lightweight, but it will lead you to the primary literature for more technical details.

To bridge the disciplinary boundaries, I combine psychological, philosophical, and linguistic insights into the nature of truth and deception, trust and credibility, cognitive biases, and logical fallacies. I then translate these insights into practical terms by drawing on the professional practices of expert lie detectors and truth seekers. I explain how their expert routines can be automated and augmented by AI methods.

Information professionals and technology users more broadly will find this book useful because it accumulates multiple perspectives on mis- and disinformation, otherwise scattered across diverse professional journals, into one volume. A few books about "fake news" prioritize the political and historical perspective (e.g., Bennett & Livingston, 2020). Others are filled with real-life examples of fraudulent news reports to advise practicing journalists and reporters (e.g., Silverman, 2021). Books from a psychological viewpoint focus on the human susceptibility to misinformation. Media literacy infographics instruct you on how to fight fake news (IFLA, 2021) and are popularly available in thin report brochures (e.g., Cooke, 2018), which were produced in the wake of the 2016 U.S. Presidential Election, and more recently, the COVID-19 pandemic (e.g., Ostman, 2020). An occasional chapter in library and information literature explains AI countermeasures to digital attacks like phishing, spamming, and social bots, but their skepticism about an eventual AI solution is palpable (Dalkir & Katz, 2020). In my research and writing, I am cautiously optimistic about the success of technological solutions, especially when AI is applied in combination with heightened critical thinking. None of the books I have reviewed so far combine aspects of human and artificial intelligence as two necessary parts for the resolution of the infodemic problem as definitively as I have here. That is the main contribution of my book: it explains how these two intelligences fit together.

Book Contributions

This book translates descriptions of the AI detection of mis- and disinformation into digestible portions about the principles, processes, techniques, analyses, and other research considerations. One of the aims is to facilitate the dialog between tech developers and less tech-savvy digital media users. The secrecy of technological "know-hows" often keeps users in the dark, preventing us from understanding how algorithms make choices. There are, however, hundreds of research labs and institutions that publish findings and innovative solutions in scientific journals and conference proceedings. Barriers arise from the complexity of specialized jargon or the background knowledge required to access and understand AI literature (e.g., Shu & Liu, 2019).

This book is an overview and culmination of over 10 years of NLP R&D in my LiT.RL lab[1]—from the earliest features for discerning deception automatically to more successful prototypes of clickbait and satire detectors. My doctoral students and I have published extensively on the topic in specialized journals and conference proceedings, but I felt the need to share it with a wider professional audience across disciplinary boundaries. My doctoral students in the lab are ambitious and capable collaborators with a variety of skills. They assist me in data collection,

[1] See https://victoriarubin.fims.uwo.ca/research/.

management, analysis, programming, and testing. What precipitated this book are years of brain-storming sessions, active discussions in the lab, and our interviews with journalists, satirists, clickbait creators, and usability study participants. We have collected real-world samples of fake news from political campaigns, contrasted fabricated bluff and real news reports, and participated in computer science programming challenges to distinguish the two. It is now time to disseminate what we have learned about the infodemic of mis- and disinformation more widely, both the human factors and the automated ways of detection, since if neither users nor systems are able to filter out rubbish, the accumulation of inaccurate information threatens the very usefulness of information retrieval.

Book Structure at a Glance

Between the introductory and conclusive chapters, the book contains two parts of four chapters each. **Part I** of the book seeks to understand the nature of deception and why we fall for it, how the human mind conceives of the truth and how we distinguish it from deception. This section establishes the theoretical footing for our understanding of how we interact with new information. To do this, I use evidence-based research from interpersonal, social, and cognitive psychology, computer-mediated communication, and insights from library and information science, as well as discussions in philosophy.

Chapter 1 frames the problem of mis- and disinformation as the proliferation of deceptive, inaccurate, and misleading information in digital media and information technologies. Discussing divergent terms and their nuances, the chapter narrows the field down to several types of fakes that can be identified with the naked eye and AI. My infodemiological model, the (Rubin, 2019) Disinformation and Misinformation Triangle, serves as a starting point in which we recognize the three minimal interacting causal factors that fuel the infodemic—susceptible digital media users, virulent fakes, and toxic digital environments. Three corresponding interventions—automation, education, and regulation—are proposed to interrupt the interaction among these factors.

Chapter 2 surveys psychological studies on lying and what distinguishes various kinds of deception from truths. I explore the cognitive biases that predispose us to being manipulated into believing untruths. Chapter 3 taps into library and information science insights into trust and credibility, their perceived components, and their connection to other markers of high-quality information. Chapter 4 establishes that truth may be seen from different philosophical perspectives, and how your view of reality impacts how you establish facts and build up your knowledge.

Each chapter in **Part I** brings your attention to central chapter concepts and their interconnections. I invite you to ponder key questions before you proceed to read the research synthesis that addresses each question. For instance, consider what truth, facts, and reality actually mean to you, and how they relate to each other in

your understanding before you read about how philosophical worldviews guide us in establishing truths and presuming reality in Chap. 4.

Part II focuses on how the theoretical and empirical knowledge from **Part I** can be applied in practice. Both human and automated stepwise procedures can be put to good service by truthfully informing the public, using the best knowledge of experts in their fields. Alternatively, there are well-documented techniques for influencing the public mind, for commercial or political gains, regardless of what is known by the experts.

Chapter 5 discusses law enforcement, scientific inquiry, and investigative reporting as examples of well-established traditions for truth-seeking. Each applied field has its own systematic ways of collecting strong, supportive evidence and conducting thorough inquiries to reach valid conclusions. Ideally, the best practices of such inquiries lead to establishing facts and reliable knowledge. When experts are not well-trained, diligent, or honest, mistakes and missteps may happen despite these established systems, as exemplified by cases of wrongful convictions, scientific dishonesty, and journalistic fraud.

By contrast, many practices in advertising, public relations, and marketing, canvassed in Chap. 6, have the intent to persuade and manipulate the public opinion from the outset. Advertising techniques used in marketing campaigns and political propaganda frequently engage in truth-bending. Their persuasion mastery often exploits human biases and logical fallacies. Manipulative techniques need to be recognized first before we can resist their powers. I describe activities to identify biases, logical fallacies, manipulative advertising techniques, and other propaganda tricks.

Chapter 7 culminates this book with a thorough review of the AI systems that can help our human eyes identify and call out fakes of various kinds for the benefit of the public good. I explain, in plain language, the principles behind automated deception detectors, rumor debunkers, satire and clickbait detectors, and automated fact-checkers, both how they work and where they tend to fail.

In the **Conclusion**, I offer suggestions on how to incorporate the lessons from each chapter of the book into media and information literacy education to help curtail the infodemic. Chapter 8 aggregated the key arguments and claims about the use of automated ways of detecting various online fakes and puts forward ten recommendations for educational, AI-based, and regulatory interventions, which are articulated summatively, as a package of countermeasures to control the infodemic. Reverting to rigorous systematic thinking is the way to inoculate the public against the mindless and unquestioning consumption of manipulated content.

Some of the AI solutions described in this book have user interfaces and can be installed on personal computing devices. I encourage proactive citizens, librarians, and information professionals to download and experiment with such sample systems. Collaborations across disciplines are now needed to assess how effective AI solutions are when compared to traditional instruction in classrooms, in libraries, or at home. Now that AI can help to identify and label fakes, the public should be more widely informed of the pros and cons when considering its adoption. Further research should evaluate how to best present AI predictions to digital media users to

build confidence in their use. Decisions about which AIs to adopt and how to interact with them have societal level impacts, so these choices should not be left solely to giant tech companies. This is especially true when deciding how to incorporate automation with another society-wide measure—the legislative regulation of digital environments to disrupt mis- or disinformation cycles. Our society should then be able to make better-informed decisions about important matters such as our health care, laws, public policies, finances, and voting preferences. While we debate and decide on how to best regulate toxic social media platforms, information professionals and educators can direct their attention to cultivating more discerning minds and encouraging more tech-savvy digital media use. The book can help you make sense of what AI does and how, possibly, spur conversations on various flavors of fakes and how to identify them with an eye and AI.

References

Bennett, W. L., & Livingston, S. (2020). *Disinformation age: Politics, technology, and disruptive communication in the United States*. Cambridge University Press.

Blum, S. D. (2007). *Lies that bind: Chinese truth, other truths*. Rowman & Littlefield Publishers. Retrieved from http://ebookcentral.proquest.com/lib/west/detail.action?docID=1351099

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory*, *6*(3), 203–242.

Chaiken, S., & Ledgerwood, C. (2007). Dual process theories. In *Encyclopedia of social psychology*. SAGE.

Cooke, N. A. (2018). *Fake news and alternative facts: Information literacy in a post-truth era*. ALA Editions.

Dalkir, K., & Katz, R. (Eds.). (2020). *Navigating fake news, alternative facts, and misinformation in a post-truth World*. IGI Global. https://doi.org/10.4018/978-1-7998-2543-2

IFLA. (2021, February 18). How to spot fake news. *The International Federation of Library Associations and Institutions*. Retrieved from https://www.ifla.org/publications/node/11174

Manning, R. (1965, August). Hemingway in Cuba. *The Atlantic*. Retrieved from http://www.theatlantic.com/magazine/archive/1965/08/hemingway-in-cuba/399059/

Ostman, S. (2020, March 20). *Fighting fake news in the pandemic: Using your library's digital reach to thwart misinformation*. American Libraries Magazine. Retrieved from https://americanlibrariesmagazine.org/blogs/the-scoop/covid-19-fighting-fake-news-pandemic/

Rubin, V. L. (2014). Pragmatic and cultural considerations for deception detection in Asian languages. Guest editorial commentary. *TALIP Perspectives in the Journal of the ACM Transactions on Asian Language Information Processing (TALIP)*, *13*(2), 1–8. https://doi.org/10.1145/2605292

Rubin, V. L. (2019). Disinformation and misinformation triangle: A conceptual model for "fake news" epidemic, causal factors and interventions. *Journal of Documentation*, *ahead-of-print*. https://doi.org/10.1108/JD-12-2018-0209

Shu, K., & Liu, H. (2019). Detecting fake news on social media. *Synthesis Lectures on Data Mining and Knowledge Discovery*, *11*(3), 1–129. https://doi.org/10.2200/S00926ED1V01Y201906DMK018

Silverman, C. (Ed.) (2021). *Verification handbook for disinformation and media manipulation (Online)*. DigitalJournalism.com. Retrieved from https://datajournalism.com/read/handbook/verification-3

Contents

Part I Human Nature of Deception and Perceptions of Truth

1 The Problem of Misinformation and Disinformation Online 3
1.1 Introduction 4
1.1.1 Infodemic 4
1.1.2 Mis- and Disinformation 7
1.2 Framework for Causes and Interventions 8
1.3 Immediate Digital Literacy Steps 11
1.4 Terminology in Mis- and Disinformation Research 12
1.4.1 Digital Deception 13
1.4.2 "Fake News" 13
1.4.3 Information Disorder and Other Problematic Content 14
1.5 Computational Fields 16
1.5.1 AI, Computational Linguistics, and Natural Language Processing 17
1.5.2 Shift from Factuality to Subjectivity Analysis 18
1.6 Varieties of Fakes Identifiable with AI 20
1.6.1 Serious Journalistic Fraud, or Falsehoods 20
1.6.2 Rumors or Hoaxes 21
1.6.3 Satirical Fakes 22
1.6.4 Deepfakes 23
1.6.5 Resulting AI Technologies 23
1.7 Conclusions 23
References 24

2 Psychology of Misinformation and Language of Deceit 29
2.1 Introduction: Deception Is Uncooperative Communication 30
2.2 Motivations to Deceive and Disinform 30
2.2.1 Deceptive Content 32
2.2.2 Serious and Everyday Lies 32
2.2.3 Incentives to Disinform 33

2.3 Many Faces of Deception . . . 36
2.3.1 Definitions of Deception . . . 36
2.3.2 Deception Is Pervasive . . . 37
2.3.3 Deception Varieties and Typologies . . . 37
2.4 Why Mis-/Disinformation Is Effective . . . 40
2.4.1 External Circumstances and Human Propensities . . . 41
2.4.2 Cognitive Biases and Individual Traits . . . 42
2.5 Language of Deceit and Cues to Its Detection . . . 44
2.5.1 Human Inability to Detect Deception . . . 44
2.5.2 Cues to Lie Detection by Mode of Perception . . . 45
2.5.3 Verbal Cues . . . 46
2.5.4 Incentives to Deceive . . . 47
2.6 Putting Psychological Research to Practice: Interventions . . . 47
2.6.1 Fact-Checking and Content Moderation . . . 47
2.6.2 Inoculating . . . 49
2.6.3 Labeling . . . 50
2.6.4 Literacy Education . . . 53
2.7 Summary . . . 54
References . . . 55

3 Credibility Assessment Models and Trust Indicators in Social Sciences . . . 61
3.1 Introduction . . . 62
3.2 Credibility in Information Retrieval . . . 63
3.3 Credibility Indicators in Social Sciences . . . 64
3.4 Trust Modeling . . . 70
3.5 Credibility Indicators in Journalistic Fact-Checking . . . 73
3.6 Seals of Approval in Libraries and Health Informatics . . . 75
3.7 Composite Reputation Ratings and Popularity Scores . . . 77
3.8 Trust and Distrust Markers in Language . . . 79
3.9 Contemporary Solutions to Credibility Assessments and Trust Ratings of Online Content . . . 80
3.9.1 Digital Realities of Social Technologies in the 2020s . . . 80
3.9.2 "Nutritional" Information Facts for Online Content as Explainable AI . . . 81
3.10 Conclusions . . . 89
References . . . 91

4 Philosophies of Truth . . . 95
4.1 Introduction . . . 96
4.2 Artificial Intelligence and Applied Ethics (Morality) . . . 98
4.2.1 Moral Virtues . . . 101
4.3 Theories of Truth . . . 104
4.3.1 What Is Truth? . . . 104
4.3.2 The Correspondence Theory . . . 105
4.3.3 The Coherence Theory . . . 108

4.3.4 Pragmatist Theories . . . 109
4.3.5 Epistemic (Anti-Realist) Theories . . . 109
4.3.6 Deflationary (or Deflationist) Theories . . . 110
4.4 Practical Applications of Philosophies of Truth . . . 111
4.4.1 Ordinary Life Decisions . . . 111
4.4.2 On Truth, Reality, Fact, and Knowledge . . . 112
4.4.3 Contemporary Debates of "Post-Truth" and "Truth Decay" . . . 113
4.5 Application in AI and NLP . . . 114
4.6 Conclusions . . . 117
References . . . 118

Part II Applied Professional Practices and Artificial Intelligence

5 Investigation in Law Enforcement, Journalism, and Sciences . . . 123
5.1 Introduction . . . 124
5.2 Lie Detection in Law Enforcement and Justice System . . . 126
5.2.1 Polygraph (Lie Detector) . . . 126
5.2.2 Verbal Content Analyses . . . 128
5.2.3 Statement Validity Assessment (SVA) and the Undeutsch Hypothesis . . . 129
5.2.4 Content-Based Criteria Analysis (CBCA) . . . 130
5.2.5 Reality Monitoring (RM) . . . 133
5.2.6 Integration of CBCA and RM Approaches . . . 134
5.2.7 Suitability for Automation . . . 134
5.2.8 A Word of Caution . . . 135
5.3 Investigative Reporting and Fact-Checking in Journalism . . . 136
5.3.1 Journalistic Goals and Aspirations in Pursuit of Facts . . . 136
5.3.2 Skeptical Way of Knowing in the Best Traditions of Investigative Reporting . . . 139
5.3.3 Fact-Checking within Investigative Journalism and as a Newly Emerged Genre . . . 140
5.3.4 Newsworthiness, Sensationalism, and "Yellow Journalism" . . . 142
5.3.5 News Literacy for all . . . 144
5.3.6 Implications for AI . . . 145
5.3.7 Lessons Learned from Journalistic Practices . . . 147
5.4 Scientific Method of Inquiry . . . 147
5.4.1 Scientific Process . . . 148
5.4.2 Scientific Principles . . . 149
5.4.3 Falsifiability . . . 149
5.4.4 Reproducibility (Transparency) and Replicability (Generalizability) . . . 150
5.4.5 Systematicity . . . 151

5.4.6 Lessons Learned from the Scientific Method 151
5.5 Conclusions 152
References 153

6 Manipulation in Marketing, Advertising, Propaganda, and Public Relations 157
6.1 Introduction 158
6.2 Marketing and Advertising 160
6.2.1 Marketing Communications as a Promotional Tool 160
6.2.2 Digital Advertising Formats 161
6.2.3 Ad Revenue Model 164
6.2.4 Advertising Paradigms: Salesmanship in Print, Seduction, and Others 164
6.3 Persuasion in Advertising 166
6.3.1 Classic Modes of Persuasion 166
6.3.2 Advertising Strategies 167
6.3.3 Codes, Standards, and Regulation in Advertising Industry 168
6.3.4 Puffery 169
6.3.5 False Claims and Other Acts of Misleading Advertising 170
6.3.6 Regulatory Actions and Antitrust Investigations 172
6.3.7 Negative Attitude toward Ads, Distrust, and Ad Avoidance 173
6.4 Covert Advertising in Many Guises 174
6.4.1 Irresistibility of Native Advertising 174
6.4.2 Native Ad Varieties: Advertorials, Sponsored Content, Branded Links, and Others 176
6.4.3 Overdue Measures: Regulation, Clearer Labeling, and Awareness of the Masquerade 178
6.5 Virality 180
6.5.1 Elusive Properties of Virality 180
6.5.2 Compelling Word-of-Mouth (WOM) 181
6.5.3 Emotional Vs. Informative Appeals in Advertising 181
6.5.4 Viral Marketing as a Measure of Success in Digital Advertising 182
6.6 Mass Propaganda 182
6.6.1 Consistent and Enduring Efforts to Influence Minds 182
6.6.2 Propagandist Techniques 183
6.6.3 Public Relations 185
6.6.4 Conspiracy Theory Peddling 186
6.7 Implications for Artificial Intelligence and Counterintelligence 187
6.7.1 Virality, Popularity, and Rumor Propagation in AI-Research 187
6.7.2 Ad-Blockers Wars 188

6.7.3 Google's AdWords & AdSense 189
6.7.4 Facebook's Multiplatform Targeted Advertising 190
6.7.5 Bot Wars: Automation of Advertising, Customer Services, and Social Engagement 191
6.8 Broader Regulations and Legislation as Countermeasures 192
6.8.1 Avoiding Governmental Overreach in Freedom of Speech and Citizens' Liberties. 192
6.8.2 Regulatory and Legislative Measures Targeting Social Media Platforms. 195
6.8.3 Regulatory and Legislative Measures Targeting Offenders 195
6.8.4 Regulatory and Legislative Measures Targeting Government Capacity 196
6.8.5 Regulatory and Legislative Measures Targeting Citizens, Civil Society, and Media Organizations 197
6.8.6 Cutting the Finances of "the Disinformation for Hire" Industry 198
6.9 Conclusions 200
References. 200

7 Artificially Intelligent Solutions: Detection, Debunking, and Fact-Checking 207
7.1 Introduction 208
7.1.1 AI Basic Terms and Concepts Demystified 209
7.1.2 AI Goals in Solving the Mis- and Disinformation Problem 216
7.1.3 Chapter Roadmap: AI Detection by Type of Fake 218
7.2 Automated Deception Detectors 218
7.2.1 Linguistic Feature-Based Approaches to Identifying Falsehoods 218
7.3 Automated Clickbait Detectors 228
7.3.1 Clickbait Defined 228
7.3.2 Clickbait Predictive Features and Clickbait Detection Performance Measures. 228
7.3.3 Sample Clickbait Detector User Interface 232
7.4 Automated Satire Detectors 235
7.4.1 Satirical Fakes Defined 235
7.4.2 Methodologies for Computing Satire. 236
7.4.3 Predictive Feature Specifications. 237
7.4.4 Methodology for Binary Classification (Satirical Fakes Vs. Legitimate News) 238
7.4.5 Neural Network Methodologies for Computing Satire 240
7.4.6 Sample Satire Detection UIs and Suggested Improvements. 242

7.5 Automated Rumor Debunkers . . . 246
7.5.1 Rumor Detection Stages or Sub-Tasks . . . 246
7.5.2 Rumor Features . . . 248
7.5.3 Other Methodologies for Rumor Debunking . . . 248
7.6 Automated Fact-Checkers . . . 252
7.6.1 Commercial Web-Based and Mobile Applications . . . 256
7.7 Conclusions . . . 257
References . . . 258

8 Conclusions: Lessons for Infodemic Control and Future of Digital Verification . . . 265
8.1 The Main Premise of the Book . . . 266
8.1.1 Socio-Technological Solutions to Mis- and Disinformation . . . 266
8.1.2 Infodemiological Framework for Causes and Interventions . . . 268
8.2 Concrete Recommendations for Educational Interventions . . . 269
8.2.1 Teaching to Verify your Sources . . . 269
8.2.2 Authoritative Fact-Checking and Debunking . . . 270
8.2.3 Relying on the Scientific Consensus . . . 270
8.2.4 Relying on Professional Verification Experts and Distrusting Manipulators . . . 270
8.2.5 Inoculating and Preemptive Messaging . . . 271
8.2.6 Identifying and Filling Niches in Public Awareness . . . 271
8.2.7 Self-Directed Coaching and Literacy Resources . . . 272
8.2.8 Making Sense of Credibility Indicators, Warning and Labeling Systems . . . 272
8.2.9 The Broader Argument for Informed Rationality and Systematic Training . . . 273
8.3 Automated Interventions . . . 273
8.3.1 AI-Based Tools for Combatting Mis- and Disinformation . . . 273
8.3.2 Claim 1: AI Is Inherently Tied to Human Practices . . . 274
8.3.3 Claim 2: Innovative AI Solutions Are Feasible and in Existence but Are Imperfect . . . 274
8.3.4 Claim 3: Explainable, Accountable, and Ethical AI as the Future of Content Verification . . . 276
8.4 Regulatory and Legislative Interventions . . . 276
8.5 Policy-Making Recommendations for Governments and Institutions . . . 277
8.6 Summary . . . 278
References . . . 279

Index . . . 281

List of Abbreviations

ACL	Association for Computational Linguistics
AFP	Agence France-Presse
AFT	(Wikipedia's) Article Feedback Tool
AI	Artificial Intelligence
AIDA	Awareness-Interest-Desire-Action (in advertising)
AOI	Area of Interest
API	American Press Institute
BBB	Better Business Bureau
BILD	Building Industry and Land Development Association (Toronto, Canada)
3 Cs	Times of "Conflict, Crisis, and Catastrophe" (in research about rumors)
CBC	Canadian Broadcasting Corporation (Canada)
CBCA	Content-Based Criteria Analysis
CIHR	Canadian Institutes of Health Research
CMC	Computer-Mediated Communication
CNE	Cámara Nacional Electoral (National Court on Elections, Argentina)
CTR	Click-through rate
DAGMAR	Defining Advertising Goals for Measured Advertising Results
DRIP	Differentiate, Reinforce, Inform, and Persuade (a marketing model)
EMA	European Medicines Agency (for the scientific evaluation, supervision, and safety monitoring of medicines in the EU)
ERIC	Education Resources Information Center (U.S.)
EU	European Union
FTC	Federal Trade Commission (U.S.)
GIGO	Garbage In, Garbage Out
HCI	Human–Computer Interaction
HTML	HyperText Markup Language
IAB	Interactive Advertising Bureau (Europe, U.S.)
IAMAI	Internet and Mobile Association of India
ICT	Information and Communications Technology
IFCN	Poynter's International Fact-Checking Network

IFLA	International Federation of Library Associations and Institutions
IMC	Integrated Marketing Communications
IPA	Institute for Propaganda Analysis (U.S.)
IR	Information Retrieval
JSTOR	Journal Storage (a database for Internet access of scholarly journals)
LIS	Library and Information Science
LiT.RL	Language and Information Technology Research Lab (Western University, Canada)
LOP	Location of Presence
MAIN	Modality, Agency, Interactivity, and Navigability (a model)
MCQ	Memory Characteristic Questionnaire
MIS	Management Information Systems
ML	Machine Learning
MMS	Multimedia Message Service
mRNA	Messenger Ribonucleic Acid (vaccine technology)
NACOLE	National Association for Civilian Oversight of Law Enforcement (U.S.)
NLP	Natural Language Processing
NSERC	Natural Sciences and Engineering Research Council of Canada
NSF	National Science Foundation (U.S.)
OBA	Online Behavioral Advertising
PKM	Persuasion Knowledge Model
PPC	Pay-Per-Click (paid search marketing)
PR	Public Relations
R&D	Research and Development
RM	Reality Monitoring
RIRO	Rubbish In, Rubbish Out
SEM	Search Engine Marketing
SEO	Search Engine Optimization
SERPs	Search Engine Results Pages
SMS	Short Message Service
SSHRC	Social Sciences and Humanities Research Council of Canada
SVA	Statement Validity Assessment
UI	User Interface
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNICEF	United Nations International Children's Emergency Fund
URAC	Utilization Review Accreditation Commission
URL	Uniform Resource Locator
VDPV	Vaccine-Derived Poliovirus
WHO	World Health Organization
WOM	Word-of-Mouth
WOT	Web of Trust (an online reputation and Internet safety service)