# Book Review

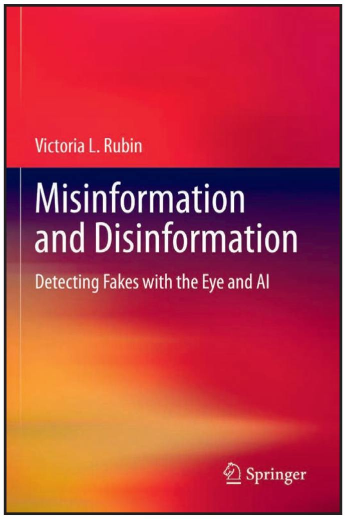
## Misinformation and Disinformation—Detecting Fakes with the Eye and AI

**Victoria L. Rubin**
Springer International Publishing, 2022
ISBN 978-3-030-95655-4, $69.99 USD, Hardcover: 289 pages

**Claire Laudy**
Thales
Palaiseau, France
claire.laudy@thalesgroup.com

## INTRODUCTION

“How do we detect, deter and prevent the spread of mis- and disinformation with the human eye and AI?” This is the question Victoria L. Rubin tries to answer in this book. She provides a large, detailed, and complete overview of the question, divided into two parts. Part 1 of the book focuses on the human interaction with information in order to understand the nature of deception and how the human mind conceives it and falls for it. In Part 2, Rubin explores how the theoretical knowledge described in Part 1 can be applied to develop automated artificial intelligence (AI)–based fake detection systems.


Victoria L. Rubin
Misinformation and Disinformation
Detecting Fakes with the Eye and AI
Springer

Rubin is multilingual, passionate about languages, and especially fascinated how language is used under challenging circumstances. As such, she has been particularly interested in studying how “lying and deception may be distinctly cultural, yet universal in the sense of their relevance to human condition”. Through this book, she offers an overview of over 10 years of her studies of natural language processing (NLP) in the LiT.RL Lab [1]. The book combines aspects of previous publications, adds important details, and puts previous work into perspective, offering a framework for future research and development on fakes and deception understanding and detection.

The book is intended for a broad spectrum of readers. The author makes sure to present the different theories in a simple manner. It is thus intended for people dealing with large amounts of information and online presence who are looking for a primer on deception research. It is also intended for programmers and information retrieval experts whose aim is to develop fake detection systems.

## PART I: HUMAN NATURE OF DECEPTION AND PERCEPTION OF TRUTH

### CHAPTER 1. THE PROBLEM OF MISINFORMATION AND DISINFORMATION ONLINE

Chapter 1 opens with useful definitions of the concepts of infodemic and infodemiology, as well as the distinction between mis- and disinformation. This chapter gives the reader the concepts necessary to understand mis- and disinformation and enjoy the following chapters. To define the concepts of infodemic and infodemiology, the author translates the triangle model for classical disease from epidemiology to the context of digital communication, identifying the three factors that allow the spread of diseases: compromised hosts, virulent pathogens, and conductive environments. This infodemiological model identifies the three interacting causal factors responsible for the spread of mis- and disinformation: automation, education, and regulation.

Here, the author sets the scene for the major importance of the subject she studies by reporting about the status of society about infodemic and mis- and disinformation. It appears there is a consensus among various organizations on the importance of the problem and the urge to tackle it. She states that assistance from AI to detect deception and fakes is inevitable and explains that the way we “accumulate knowledge from the past and the newest technological advancements can be combined […] to bring the current online infodemic under control”. The author also makes an inventory of the different types of fakes that can be identified with AI. She concludes that even with AI-based solutions, there will still be a need for a human in the loop in the detection and management of fakes.

### CHAPTER 2. PSYCHOLOGY OF MISINFORMATION AND LANGUAGE OF DECEIT

This chapter presents deception as an uncooperative communicative behavior. It describes the motivations for deception and disinformation. The different types of deceptions are studied, and a useful alignment is proposed for the varieties of deception described in various taxonomies in an earlier work in the field of psychology and communication research [2].

The author then enumerates the reasons deception is effective for us as humans. For instance, studies showed that the more information is repeated, the more it seems true. Furthermore, the less cognitive effort that is needed to understand the information, the more fluent (i.e., easy to integrate as one’s own) this information is, and finally the more true it looks [2].

The chapter describes several cues to detect deception in natural language that may also be used within automatic systems. Finally, existing responses to mis- and disinformation are listed, such as fact-checking, educating, inoculating (i.e., preventing by making aware of the phenomenon, as for vaccines in medicine), or labeling the information.

### CHAPTER 3. CREDIBILITY ASSESSMENT MODELS AND TRUST INDICATORS IN SOCIAL SCIENCES

This chapter proposes a wide survey and a deep analysis of the literature about credibility and trust. The credibility of a message depends on the characteristics of its source, contents, and medium of delivery. The author introduces many works that aim to describe these characteristics. Initiatives are presented, for instance, the common credibility assessment terminology definitions [3].

People tend to naturally trust others and the messages they receive. The author describes trust and distrust markers in language. These markers can be used in fake detection systems. Following that, she proposes ideas to develop AI systems may assist people in discriminating online information.

### CHAPTER 4. PHILOSOPHIES OF TRUTH

This chapter discusses various philosophical perspective about truth. The author summarizes the different philosophies of truth and untangles the key concepts of truth, reality, facts, and knowledge that are often confused one for another. This perspective leads us to think about what exactly the automated AI-based fake detection systems should look for. Facts may be wanted more than truth for users of these systems.

> ***"The credibility of a message depends on the characteristics of its source, contents, and medium of delivery".***

## PART 2: APPLIED PROFESSIONAL PRACTICES AND ARTIFICIAL INTELLIGENCE

### CHAPTER 5. INVESTIGATION IN LAW ENFORCEMENT, JOURNALISM, AND SCIENCES

In this chapter, the author investigates the best practices of three expert domains: law enforcement, scientific inquiry, and investigative journalism. Her aim is to find insights that would be useful for informing automated systems on the detection of deception and supporting the process of fact-checking. From law enforcement experts, she suggests that established tools and checklists for statement validity analysis used by detectives during police interrogations are potential guides to develop automated lie detection systems. The five Ws (who, what, when, where, and why) of journalism form a useful framework to examine the credibility of some online stories. She also emphasizes the use of rational observation and systematic questions, linked to the scientific method.

As a conclusion to this chapter, the author states that what unites experts in these three domains is their inquisitive critical mindset. She then makes the point that even if we develop automated AI-based fake detection systems, humans should still validate their results: "Technology advises and assists us but never replaces human judgement in determining what is truthful and what is disinformative".

### CHAPTER 6. MANIPULATION IN MARKETING, ADVERTISING, PROPAGANDA, AND PUBLIC RELATIONS

The chapter describes the different means used to propagate mis- and disinformation. It also describes AI techniques that mimic these means of propagation. Examples are taken from the marketing, advertising, and public relation domains, with specific and concrete examples from each. The author further explores what makes us—as humans—particularly vulnerable to viral conspiracy theories and how human biases are being exploited to propagate mis- and disinformation.

### CHAPTER 7. ARTIFICIALLY INTELLIGENT SOLUTIONS: DETECTION, DEBUNKING, AND FACT-CHECKING

This is the longest chapter in the book (58 pages). In this chapter, the author presents an overview of five large families of AI-enabled applications that may support humans in detecting and managing fakes. AI-based systems aim to accomplish three different tasks: assist in the detection itself, alert, and filter fake information. The author emphasizes the importance of having a good human–machine interface. Humans should remain in the loop and use AI only to assist them.

Five application families are reviewed: deception detectors, click-bait detectors, satire detectors, rumor debunkers, and fact-checkers. For each of these families of applications, the author offers detailed examples, together with some technical details of how they work. She keeps technical vocabulary to a minimum, in "favor [of] explaining step-wise procedures in principles, and wherever possible, offer some examples".

### CHAPTER 8. CONCLUSIONS: LESSONS FOR INFODEMIC CONTROL AND FUTURE OF DIGITAL VERIFICATION

This chapter concludes the book, giving arguments and claims about the use of automated ways of detecting online fakes. The author then gives recommendations for educational, AI-based, and regulatory interventions.

## SUMMARY

The book gives an overview on many historical, technological, and psychological aspects of the subject. The author explores many AI approaches, all based on statistical machine learning solutions for detecting fakes.

The author emphasizes that deceptions and fakes form a diverse set. The existing—and future—automated solutions for fake detection should thus focus on a limited objective, in order to be both relevant and efficient. She presents an interesting alignment of taxonomies of deception varieties. This should be widely used by researchers and solution providers as a pivotal model that would enable the rigorous description of the specific fakes they aim at detecting.

The book presents existing work mostly using statistical NLP. It is thus oriented toward using pragmatics and content to detect fakes. It would be interesting to expand the analysis toward meeting semantic and knowledge-based solutions, using a semantic description of human motivations and processes, which are well detailed by the author. This would enable understanding and explaining AI-based fake detection.

Reading the book is inspiring. For instance, a direct idea that comes to mind after reading Chapter 3 is to extend the work to provide a synthesis of all the reviewed models into a single complete one. Existing ontologies of trust could be used to align the different models, for instance. This work of synthesis is nicely done for the alignment regarding deception taxonomies proposed in Chapter 2.

The reading also raises the possibility of attempting to build semantic models to inform AI (i.e., choose which model from those presented in the book to use to inform machine learning approaches on a specific task). With the broad literature review and the analysis of existing works on the perception and detection of fakes by humans that the author provides, this semantic modelling step may be envisioned.

Finally, I recommend reading this inspiring book to members of the International Society of Information Fusion community who would want to have a different perspective on the subject of detecting fakes. The ideas and explanations given by Rubin may help our community find innovative ways of managing mis- and disinformation, using our usual solutions with enriched background and understanding of the way mis- and disinformation are produced and spread.

## REFERENCES

1. Victoria Rubin, prof., research lab director, writer, speaker, http://victoriarubin.fims.uwo.ca/research/.
2. Shane, T. The psychology of misinformation: why we're vulnerable, June 2020. [Online] https://firstdraftnews.org/articles/the-psychology-of-misinformation-why-were-vulnerable/.
3. Zhang, A. X., Ranganathan, A., Metz, S. E., Appling, S., Moon Sehat, C., Gilmore, N., et al. A structured response to misinformation: defining and annotating credibility indicators in news articles. In *Companion Proceedings of the Web Conference 2018*. Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2018, 603–612, doi:10.1145/3184558.3188731

**Claire Laudy** is a senior research engineer, at the research center of Thales, where she has worked for 22 years. She focuses on knowledge representation and modeling, (semantic networks, ontologies, conceptual graphs), semantic fusion and graph algorithms for high-level information management. She was involved in numerous research projects dealing with the new usage soft information (e.g., social media posts and citizen science) within fusion-based systems. Among others, she was involved in the promotion of the use of soft information towards operational experts used to work with sensor data only, such as marine biologists and Police and Safety organizations. Claire first worked as an engineer in the domain of Human-machine interaction and then obtained her Ph.D. from Sorbone Université in 2010 in the field of semantic information fusion.

Since 2007, Claire has been involved in the international research community on information fusion, through her presence at FUSION conferences. She published several book chapters on high level and soft information fusion. She has been a member of the ISIF Working Group on Evaluation Techniques of Uncertainty Representation (ETUR) since 2017 and was recently elected member of the Board of Directors of the International Society for Information Fusion (ISIF).

ISIF
Perspectives
On Information Fusion
June 2023
Volume 6
Number 1
GOING FOR GOLD: FINDING LOST GOLD CITIES IN ECUADOR
FUSION PROCESSES AND SITUATION CONTROL
25 YEARS OF ISIF: KEY DEVELOPMENTS, MEMORIES AND TRIBUTES
Publication of the INTERNATIONAL SOCIETY OF INFORMATION FUSION
ISIF 25th Anniversary