Chapter 1
A Comprehensive Review of Trustworthy, Ethical, and Explainable Computer Vision Advancements in Online Social Media

Fahim Anzum
https://orcid.org/0000-0002-1846-3957
University of Calgary, Canada

Ashratuz Zavin Asha
University of Calgary, Canada

Lily Dey
University of Calgary, Canada

Artemy Gavrilov
University of Calgary, Canada

Fariha Iffath
University of Calgary, Canada

Abu Quwsar Ohi
University of Calgary, Canada

Liam Pond
University of Calgary, Canada

Md. Shopon
University of Calgary, Canada

Marina L. Gavrilova
University of Calgary, Canada

ABSTRACT

Responsible, ethical, and trustworthy decision-making powered by the new generation of artificial intelligence (AI) and deep learning (DL) recently emerged as one of the key societal challenges. This chapter provides a comprehensive review of state-of-the-art methods that emerged very recently in the domain of trustworthiness, fairness, and authenticity of online social media. Furthermore, this chapter discusses open problems, provides examples of other application domains in the realm of computer vision and intelligent computing, recommends bias mitigation strategies, and provides insights on the future developments in this key research domain.

DOI: 10.4018/978-1-6684-8127-1.ch001

1. INTRODUCTION

Responsible, ethical, and trustworthy decision-making powered by the new generation of artificial intelligence (AI) and deep learning (DL) has recently emerged as a key societal challenge. Politicians, government officials, industry experts, and end users alike agree that in order for AI and DL to be used in commercial and public domains, safety mechanisms must be built to protect individuals. ("Directive on Automated Decision-Making," 2019). Recent theoretical, technological, and computational developments have enabled new-generation computing systems to process incredibly large amounts of visual information in a fraction of a second and to learn advanced data patterns in a human-like manner (Zhu et al., 2023). However, while advancements in information processing, data collection and information visualization have grown exponentially, research on the explainability and trustworthiness of visual systems has remained limited. In recent years, the mysterious nature of highly publicized commercial products, such as OpenAI's ChatGPT and Google's Gemini chatbots, has highlighted the need to understand the inner workings of such complicated systems. As a result, more than ever, people are turning to societal and regulatory bodies to solve difficult problems surrounding the ethics, trust, privacy, security, legal, and policy issues of these emerging AI technologies.

In online social media applications, the increasing influence of computer vision techniques underscores the critical importance of explainability and trust. Because these algorithms control the curation and dissemination of content that increasingly shapes public opinion, transparency is paramount. Understanding the inner workings of computer vision systems will not only protect individual privacy and improve user experiences, but also can help address potential training data biases and verify the authenticity of data.

Furthermore, explainable AI systems are inherently more trustworthy, meaning that users can better understand and have confidence in the decisions made by the AI, leading to increased acceptance and adoption of AI technologies. However, establishing trust in these AI-driven platforms requires a commitment to explainability, ensuring that users can understand and hold accountable the algorithms shaping their online interactions. This is why striking the balance between innovation and ethical transparency is key, creating a digital environment where computer vision can be used to the fullest extent without compromising trust in these systems. This chapter makes the following contributions:

1. Presenting a comprehensive overview of the existing research surrounding trustworthiness, bias, fairness, and explainability in online social media.