


# Chapter 7

## Ethical and Societal Implications of Large Language Models: Can We Trust Machines With Human Language?

Wasswa Shafik

 <https://orcid.org/0000-0002-9320-3186>

*Dig Connectivity Research Laboratory (DCRLab), Uganda & School of Digital Science, Universiti Brunei Darussalam, Brunei*

### ABSTRACT

*Large language models (LLMs) are powerful AI tools with broad applications, but they also pose significant ethical and societal challenges. These include risks of inherent bias in training data, leading to biased outputs that may amplify stereotypes or unfair treatment. LLMs raise concerns regarding data privacy, as they may inadvertently use sensitive or personally identifiable information. Additionally, the widespread use of LLMs for generating disinformation poses threats to public trust and information integrity. Addressing these issues requires transparent model development, responsible data use, and strong ethical guidelines to mitigate risks while maximizing the societal benefits of LLMs.*

### INTRODUCTION

Large language models<sup>1</sup> (LLM) have become increasingly prominent, from state-of-the-art performance in commercial technology to their status as objects of general interest and concern. Companies use millions of dollars to sophisticate and

DOI: 10.4018/979-8-3693-8387-2.ch007

parade their models as both indicators of technological sophistication and guides for personal improvement (Pijeira-Díaz et al., 2024). At the same time, discussions surrounding their detrimental impacts on society feature prominently in news stories, blog posts, and journals. Misinformation, disinformation, adversarial, and dangerous uses of artificial intelligence<sup>2</sup> (AI), as well as the treatment of AI ethics as the primary strategy in mitigating these impacts, are becoming regular conversations among scholars, practitioners, and industry stakeholders (Jiang et al., 2024).

Although the research in ethical and data-driven concerns is growing, none of the existing literature has systematically mapped the potential harms and benefits of these models, the actors involved, and the plausible consequences of societal uptake. In this chapter, we discuss a range of concerns that industry, technologists, policymakers, and society, more generally, must confront when discussing and using language models at scale (Khan et al., 2024). There is an increasing need to take seriously the challenges that modern language models pose, in addition to the growing public concern for the societal impacts. Thus, we focus the scope of this paper on language models. We demonstrate the risks and benefits of modern language models and map out logistic, distributive, communal, epistemic, and power asymmetries at play. We contend that modern language models are both a danger to society and its stakeholders as well as a force for potential benefits (Smetana et al., 2024).

The use of large language models has significantly broadened the scope of modern natural language processing, becoming a dominant approach over the last few years. In essence, these models are solving tasks by predicting the probability of the words they generate, making use of past words in the sequence. As such, they tend to “learn” patterns in language, ranging from common spelling patterns to abstract notions over millions of tokens. In short, the text generation capabilities of these models have the potential to be disruptive in the AI community and, by extension, in society (Zietsch et al., 2024). As language models continue to advance, questions about the ethics of developing and using systems with the potential to influence society become both more important and more immediate.

But while research on language models, in general, has rapidly accelerated, research on their implications and ethical considerations similarly lags. Given the state of AI and the need to continue working out what it means to algorithmically manipulate language, quantify truth, strip privacy from communication, blur work with play, and shift how people connect with technology, thinking about the consequences of releasing a large language model is an essential next step. The following paper thus seeks to answer two fundamental but thorny questions. First, what new ethical, social, political, and practical risks or challenges arise from the use of large language models? Second, who should be accountable for those risks or challenges? These questions weave together an array of issues that hide beneath the headline,

30 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/chapter/ethical-and-societal-implications-of-large-language-models/390568](http://www.igi-global.com/chapter/ethical-and-societal-implications-of-large-language-models/390568)

## Related Content

---

### Applying Geospatial Data to Choose the Optimal Route During the Road Design Stage

Ehyaudin Musema Ahmed, Mohammed Nuru Mohammed and Mohammad Gouse Galety (2024). *Geospatial Application Development Using Python Programming* (pp. 209-273).

[www.irma-international.org/chapter/applying-geospatial-data-to-choose-the-optimal-route-during-the-road-design-stage/347439](http://www.irma-international.org/chapter/applying-geospatial-data-to-choose-the-optimal-route-during-the-road-design-stage/347439)

### Theoretical Foundations of Large Language Models

Yashodeep Bharat Deshmukhand Abhishek Mukhopadhyay (2026). *Theory, Practice, and Future Direction of Large Language Models* (pp. 33-74).

[www.irma-international.org/chapter/theoretical-foundations-of-large-language-models/390561](http://www.irma-international.org/chapter/theoretical-foundations-of-large-language-models/390561)

### Topological Insights Into Weather Dynamics in the Indian Context: An Application of Clustering and Mapper Algorithm

Azarudheen S., Daphne Julia Menezes and Vivan Clements (2024). *Ethics, Machine Learning, and Python in Geospatial Analysis* (pp. 279-303).

[www.irma-international.org/chapter/topological-insights-into-weather-dynamics-in-the-indian-context/345913](http://www.irma-international.org/chapter/topological-insights-into-weather-dynamics-in-the-indian-context/345913)

### Innovating Reality: The Pioneering Applications of Generative AI

Rabi Shankar Panda, Anjana Mishra and Abhishek Mohanty (2024). *The Pioneering Applications of Generative AI* (pp. 85-105).

[www.irma-international.org/chapter/innovating-reality/350779](http://www.irma-international.org/chapter/innovating-reality/350779)

### Geospatial Data Storage and Management

Islom Mamatov, Mohammad Gouse Galety, Ruslan Alimov, A. V. Sriharsha, Fanar Fareed Hanna Rofoo and Gurram Sunitha (2024). *Ethics, Machine Learning, and Python in Geospatial Analysis* (pp. 150-167).

[www.irma-international.org/chapter/geospatial-data-storage-and-management/345908](http://www.irma-international.org/chapter/geospatial-data-storage-and-management/345908)