Preliminary and self-archivable version of: Morato, J., Diaz-Nafria, J.M., Sanchez-Cuadrado, S. (2024). Factors Affecting the Reliability of Information: The Case of ChatGPT. In: Guarda, T., Portela, F., Diaz-Nafria, J.M. (eds) Advanced Research in Technologies, Information, Innovation and Sustainability. ARTIIS 2023. Communications in Computer and Information Science, vol 1937. Springer, Cham. https://doi.org/10.1007/978-3-031-48930-3_12

Factors affecting the reliability of information: the case of ChatGPT

Abstract. The abundance of current information makes it necessary to select the highest quality documents. For this purpose, it is necessary to deepen the knowledge of information quality systems. The different dimensions of quality are analyzed, and different problems related to these dimensions are discussed. The paper groups these issues into different facets: primary information, its manipulation and interpretation, and the publication and dissemination of information. The impact of these interdependent facets on the production of untruthful information is discussed. Finally, ChatGPT is analyzed as a use case. It is shown how these problems and facets have an impact on the quality of the system and the mentions made by experts are analyzed. Different challenges that artificial intelligence systems face are concluded.

Keywords: Information quality, Information accuracy, trustworthiness, ChatGPT.

1 Introduction

The current overabundance of information has made it critical to select the highest quality sources of information. Information quality refers to the degree to which information meets specific criteria or standards, ensuring, amongst others, its accuracy, reliability, relevance, completeness, timeliness, and accessibility. In this work, we consider Information quality and data quality as interchangeable concepts, according to a relatively broad understanding in the literature, although the difference is also often argued, in which case, data is considered as an underlying supporter of information (Sequoiah-Grayson & Floridi, 2022; CIHI, 2017; Floridi, 2011; Díaz-Nafría et al, 2016). The reason to consider this correspondence—though not its equivalence—has been thoroughly discussed by one of the authors with the purpose to address information and meaning in a very wide sense, based on cross-disciplinary foundations (Díaz-Nafría & Zimmemann, 2013a, 2013b; Zimmermann & Díaz, 2012). Nevertheless, we will narrow here our focus to propositional semantic information, in a sense of semantics which is environmentally and pragmatically situated (op.cit.).

The academic literature provides various definitions and dimensions of *information quality*, highlighting its multifaceted nature. Redman (2001), states that the quality of information is the degree to which the data and the corresponding information accurately represent reality. In a similar way, Pipino et al. (2002) say that Information quality is the degree to which data correctly and consistently represents the real-world constructs it purports to represent. These senses are aligned to the classical theories of truth as both correspondence and coherence (Glanzberg, 2021), and can be conveniently formalised in terms of Floridi's Correctness Theory of Truth (Floridi, 2009, 2011). In this view, information as true-beared provides a proxy to a reality the