The functioning of SDSs is based on text and data mining (TDM). TDM uses techniques from natural language processing, machine learning, information retrieval, and knowledge management for the automated analysis of digital content (structured and unstructured data), in order to extract information, identify patterns, and discover new trends, insights or correlations. In the context of the European GDPR, TDM is defined as “any automated analytical technique aiming to analyse text and data in digital form to discover information, to identify patterns or trends, or to perform specific tasks or research”.

Over recent years, new providers have proposed services to automatically extract information from websites, such as fare comparison engines that enable passengers to compare prices and flight times, or smart disclosure systems that provide consumers with the possibility to compare different products and services. But is all this useful and legal? A case study will show that the terms and conditions (T&C) of a representative set of platforms are not compliant with the European legal framework, and the practice of TDM can endanger the rights of both consumers and data subjects.

Limitations to Text and Data Mining and Consumer Empowerment. Making the Case for a Right to “Machine Legibility”

Micklitz
March 30, 2018

The use of artificial intelligence (AI) tools and techniques, such as text and data mining (TDM), has been increasing exponentially in the last years. However, the European legal framework fails to provide a clear and firm regulatory framework covering TDM. The proposed exceptions are still not enough for embracing the potential of TDM and will exclude many useful applications. In the EU, the prohibition of TDM can dangerously undermine the legitimate activity of consumers, who have access to the precontractual information (T&C) and the information related to the processing (privacy settings), not only should information be legible to a human eye, but also to the tools that a platform uses for processing, machine learning, and natural language processing. This is why the automated analysis of digital content is an area where TDM can be indispensable, in order to extract information, identify patterns, and discover new trends, insights or correlations.

The updated version of the paper is expected to be published this spring in "International Review of Intellectual Property and Competition Law."