

# OECD Report on Artificial Intelligence, Data Scraping and the Challenges for Intellectual Property

abreuadvogados.com



On February 9, 2025, the Organization for Economic Cooperation and Development (OECD) published a report entitled “Intellectual Property Issues in Artificial Intelligence Trained on Scraped Data”, which addresses the implications of artificial intelligence (AI) in the field of intellectual property (IP).

This report focuses in particular on data extraction methods, namely data scraping, which have been widely used to train AI systems. The use of this type of method has raised legal challenges and issues, in particular but not limited to copyright, trademark, trade secret and database protection.

### **Advances in AI and the challenges they pose**

Artificial intelligence depends on the use of large volumes of data to train its models, and the quality of this data is fundamental to the effectiveness of the systems. This data can come from specific sources, such as licensed databases, or be collected through techniques that have been widely used, such as data scraping, which allows information to be obtained on a massive scale.

Although *Scraping* allows for a wider and more diverse collection of data, by covering all available data, it increases the risk of violating intellectual property rights (albeit unintentionally, since scraping can include protected works that are illicitly available). Much of the data collected comes from copyrighted works belonging to creators such as photographers, writers and artists.

The OECD report points out that, despite the existence of these rights, techniques such as *Scraping* make it difficult for authors to exercise them. In fact, the lack of transparency of many systems is highlighted as one of the main obstacles to the owners becoming aware of the infringement and, consequently, exercising their rights.

The report also mentions that, in general, intellectual property legislation, which often predates the development of these new techniques and technologies, is often unprepared to face the challenges posed by scraping. As a result, effective enforcement of IP rules becomes difficult, increasing the risk of infringement, even if unintentional, of exclusive rights.

### **The Role of Intellectual Property in Innovation and the Risks of Data Scraping**

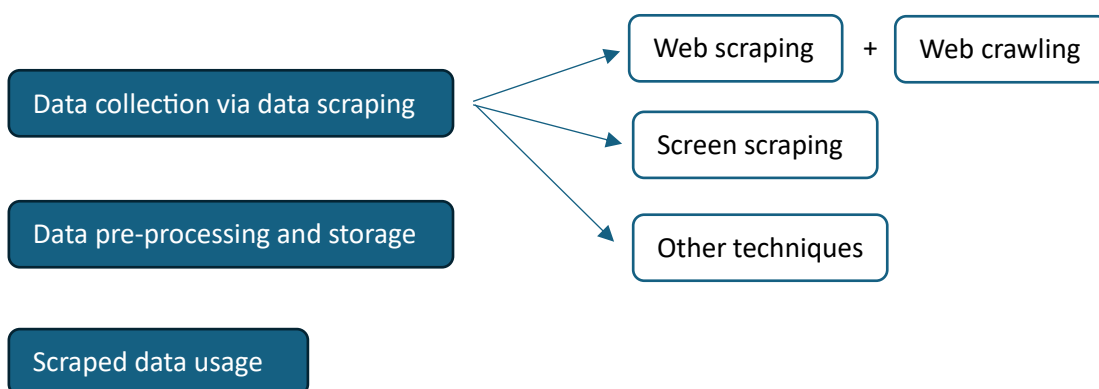
Given the crucial role that IP plays in protecting creativity and innovation, the report makes clear the urgency of updating legislation in this new digital age. However, it should be noted that the balance between technological development and the safeguarding of IP rights will be decisive in ensuring responsible and sustainable innovation.

### **Definition and Importance of Technical Concepts**

Given that one of the problems identified by the report is precisely the absence, on the one hand, and the disparity between different jurisdictions, on the other, of key concepts in the context of AI and IP, the report looks at them and proposes these definitions, including the definition of data scraping.

With regard to the definition of data scraping, the report proposes a working definition, which can be supplemented with other definitions focused on specific techniques and/or activities used in data collection. Thus, in this report, data scraping refers to *“the automatic extraction of AI training data from the web, online databases and from other sources, using automated software tools or scripts”*.

The report also provides definitions of the different techniques used in data scraping. These are:



These terms are often used interchangeably but, as the report highlights, it is essential to establish a common terminology to promote clarity and consistency, especially in legal terms.

### **Proposals for a Responsible Approach**

To respond to the challenges posed by data scraping and its impacts on IP, the report proposes several policy measures - underlining those already proposed in the EASD Recommendation (OECD, Recommendation of the Council on Enhancing Access to and Sharing of Data, 2021). Among them are:

- **Data Scraping code of conduct**  
Develop flexible and voluntary measures that take into account the different legal and regulatory approaches between jurisdictions, including the adoption of a cross-border “code of conduct” for data extraction and support for the development of technical tools and standardized contractual terms;

- **Standard contractual terms**  
Encourage the development of technical tools that protect intellectual property rights, allow control of access to data by rights holders and support licensing mechanisms, including privacy-enhancing technologies;
- **Technical tools**  
Implement awareness-raising initiatives to inform stakeholders of their rights and responsibilities, educating AI users in responsible use;
- **Data scraping awareness**  
Establish harmonized definitions and terminology for data mining activities, promoting consistency and shared understanding among stakeholders;

## **Conclusion**

The OECD report highlights the emerging legal challenges at the intersection between the evolution of AI systems and the protection of IP rights, especially with regard to data extraction techniques such as data scraping.

This study follows the OECD's principles for the development of AI, which advocate defending innovation and promoting the reliable use of this technology, respecting IP rights and promoting a more cohesive and transparent legal environment.



**Thinking about tomorrow? Let's talk today.**

**António Andrade** – Partner  
[antonio.andrade@abreuadvogados.com](mailto:antonio.andrade@abreuadvogados.com)

**Manuel Durães Rocha** – Partner  
[manuel.rocha@abreuadvogados.com](mailto:manuel.rocha@abreuadvogados.com)

**Ricardo Henriques** – Partner  
[ricardo.henriques@abreuadvogados.com](mailto:ricardo.henriques@abreuadvogados.com)

**Margarida Castillo Silva** – Associate Lawyer  
[margarida.c.silva@abreuadvogados.com](mailto:margarida.c.silva@abreuadvogados.com)

**Catarina Rocha Rodrigues** – Trainee Lawyer  
[catarina.r.rodrigues@abreuadvogados.com](mailto:catarina.r.rodrigues@abreuadvogados.com)