https://doi.org/10.6092/issn.1825-1927/18870 ISSN 1825-1927

Moderation of illegal content and social media scraping.

Privacy and data protection constraints in the processing of publicly available data by law enforcement authorities[*]

Flavia Giglio[†]

Abstract.

Social media scraping can be used in connection to a wide range of law enforcement tasks, and to investigate various forms of crime. At the same time, these practices create peculiar fundamental rights concerns, linked to the rights to privacy, data protection and freedom of expression. To complicate the matter, the recently adopted EU legal instruments on the moderation of illegal content create new grounds for potential social media monitoring activities by law enforcement authorities, possibly leading to an ehanced use of social media scraping techniques. In light of these legislative novelties, the paper aims to present potential fundamental rights criticalities of the use of social media scraping technologies in this context, when aimed to collect and process publicly available data.

Keywords:

Social media scraping, content moderation, law enforcement authorities, crime prevention

1 Introduction. Social media scraping as a form of Open Source Intelligence

Open Source Intelligence (OSINT) refers to actions aimed to gather and analyse data that is accessible to any individual or entity, within the remits of the fight against crimes. Considering open source data as the raw material that can be extracted by publicly available surces to collecte information on a given subject, OSINT represents the stage where such information is discriminated and distilled in order to respond to a question that guides the screening of such sources.[1] The benefits of OSINT for the activities of law enforcement authorities (LEAs) in the criminal justice domain has been recognised by the European Union (EU), while also sparking discussions about potential fundamental rights-related concerns on such practices. In this regard, it was observed that the use of OSINT should be put in place with sufficient guarantees to protect publicly available personal data, such as including settings that allow the end-users to adapt the possibility to access personal data according to the facts and circumstances of the use.[2] The practice to monitoring social media content in order to extract publicly available data from social networks is known as Social Media Intelligence (SOCMINT), at it falls under the broader umbrella of OSINT practices.

Potential privacy-related criticalities of the LEAs' activities aimed to repurpose publicly accessible data from the Internet are connected to the use of web crawling and web scraping technologies. Web crawlers are tools used to navigate the web in order to gather sources and index them, according to certain criteria. Web scraping, instead, refers to the act of extracting data from previously identified online sources.[3] Therefore, data scraping allows to retrieve unstructured data from the Internet. This data

[*] This publication was prepared in the context of the FERMI project, funded by the European Union's Horizon Europe research and innovation programme under grant agreement No 101073980.

[†] KU Leuven, Centre for IT & IP Law (CiTiP); ✉ flavia.giglio@kuleuven.be

[1] Steele, Robert David. "Open source intelligence." *Handbook of intelligence studies* 42.5 (2007): 129-147.

[2] Shere, Anjuli. "Reading the Investigators their Rights: A review of literature on the General Data Protection Regulation and open-source intelligence gathering and analysis." *The New Collection* 3, Oxford, 2020, 3-21.

[3] Vanden Broucke, Seppe, Baesens, Bart, "From web scraping to web crawling." *Practical Web Scraping for Data Science. Best practices and examples with Python*, Apress, Berkeley, 2018, 155-161.

Copyright © 2023 Flavia Giglio
This work is licensed under the Creative Commons BY License.
https://creativecommons.org/licenses/by/4.0/

can, in turn, be stored and analysed in a central database. In the context of SOCMINT, this process can be used for emotion and sentiment analysis purposes, for example in order to predict the outcome of an election based on what users share online. For this type of use, retrieving posts or comments that mention a specific candidate's name may not be necessary, as sentiment analysis algorithms can infer certain information about individuals' opinions even when the analysed content is not explicit.[4]

In the context of LEAs' activities, the results of an horizon scanning exercise conducted in 2022 on the use of OSINT by investigators of eight different countries leads to the conclusion that LEAs make use of the latter instrument of data scraping, as they analyse data available online. This is especially true in the context of social media analysis, considered as one of the most important capabilities available to the investigators.[5] The study also highlighted the ability of such instruments to return a massive amount of data, with little capacity to limit the volume of the retrieved information. This characteristic rises doubts about how to assess the proportionality of an investigation with regard to the access to data.[6]

When it comes to detecting a certain type of content online, social media scraping techniques that retrieve data from a certain source usually work along with natural language processing (NLP) libraries, as a means to organise the gathered data in a structured way. NLP algorithms allow to classify scraped content against a database of previously provided examples, normally sets of words that illegal or harmful content is likely to contain. As it will be further explored in the next sections of this paper, such algorithms suffer some fundamental limitations from a technological point of view. These limitation add up to those that characterised data scraping techniques in general.

The present paper aims to analyse the privacy and data protection constrains that should in in place in social media scraping practices put in place by LEAs, with a specific focus on social media monitoring and content moderation activities. It first highlights why social media content can be a source of publicly available data whose data protection guarantees should be kept in high consideration, due to the connection between the rights to privacy and data protection and the enjoyment of online freedom of expression. In turn, it analyses the existing EU content moderation landscape, with special focus on some of the existing provisions that exemplify the enhanced presence of LEAs, and public actors in general, in the context of content moderation. Afterwards, it highlights the fundamental rights concerns arising from the use of NLP algorithms to carry out content moderation and categorise publicly available data scraped from social media. Finally, it explores the legal status of publicly available data as defined in the EU legal framework, in the European Court of Human Rights (ECtHR) and in the EU Court of Justice (CJEU) case law. Using the principles extracted by this normative framework, potential social media scraping practices performed by LEAs are assessed in light of the rights to privacy and data protection.

The next section of the paper investigates the privacy and data protection concerns arising from the processing of publicly available data extracted by social media content. Due to the pivotal role played by social media in fostering freedom of expression, the protection of this category of publicly available data from unduly interferences of public authorities is particularly challenging as well as important in ensuring the enjoyment of fundamental rights.

2 The peculiar status of social media as a source of publicly available data

LEAs' practices involving social media scraping are a type of OSINT presenting its own peculiarities in relation to fundamental rights. Social media sources are by nature characterised by broad visibility and accessibility. When it comes to monitoring activities of LEAs, this characteristic creates an asymmetry between the visibility of online social life and police activities. This asymmetry can create concerns, especially with the view of extreme scenarios when the LEAs' monitoring activities become to overreaching that every subject is treated as a potential suspect, with a consequent "criminalisation of online spaces".[7] Moreover, in traditional social media platforms, expressions made through posts or comments can be easily associated with their authors. An example is Facebook, whose community guidelines requiring to use the users' legal name was subject to debate, due to

[4] Khder, Moaiad Ahmad, "Web scraping or Web crawling. State of Art, Techniques, Approaches and Application." *International Journal of Advances in Soft Computing & Its Applications* 13.3 (2021), 144-168.

[5] Bayerl, Petra Saskia, et al., "Future Challenges and Requirements for Open Source Intelligence in Law Enforcement Investigations: Results from Horizon Scanning Exercise." *European Law Enforcement Research Bulletin* 21 (2022), 21-38.

[6] *Idem*.

[7] Trottier, Daniel, "Open source intelligence, social media and law enforcement: Visions, constraints and critiques." *European Journal of Cultural Studies*, *18*.4-5 (2015), 530–547.

the alleged disregard for situations where anonymity can benefit to the expression of minorities and human rights activists.[8] Therefore, compared to traditional public spaces that could be monitored by LEAs, the expectation of anonymity of individuals expressing themselves on social networks is largely diminished. These privacy-related preoccupations can have a fallout on another fundamental right: freedom of expression.

In the era of social media, social networks amount to a new public forum of great importance for exchange of ideas and information, ultimately influencing the public debate.[9] Given this status of social networks as facilitators to freely express opinions, it has become crucial to consider the implications on freedom of expression inherent to any activity aimed to monitoring or analysing what is shared online.[10] This is especially in light of the fact that social media, and the Internet as a whole, have become a powerful instrument for political activism and social movements, facilitating the organisation of demonstrations and resistance acts everywhere in the world, including in less democratic countries. In this context, the creation of content on social media platforms, and the interactions deriving from it, give a powerful chance to enhance collaboration and communication.[11]

The disproportionate use of social media scraping technologies can result in a state of mass surveillance capable of having a chilling effect on freedom of expression. This risk is enhanced when considering that, in line with the OSINT logic, information about the content posted online can be combined with other data freely available in multiple Internet sources, returning a very comprehensive picture of individuals to LEAs. The dangers of aggregating data from multiple sources was underlined by the UK Information Commissioner' Office and other data protection authorities in a joint statement on data scraping and data protection, published in 2023. The statement pointed out that this aspect of data scraping could lead to individuals losing power over their personal data, as information that they decide to delete could still be processed and circulated by entities which scraped it. It also acknowledged the risk of "monitoring, profiling and surveilling individuals" among those to be considered when using data scraping techniques.[12]

Conversely, the expansion of platforms where to express opinions has profoundly changed the landscape of traditional media, affording an unprecedented freedom to individuals to share any type of content online. This enhanced freedom came with the emergence of distortions in the use of social media, such as the sharing of malicious content or the perpetration of harmful actions through online platforms. The misuse of social media can have severe effects to the detriment of persons or society. In such cases, the intervention of LEAs and the enactment of criminal laws protecting the public values at stake is called for.[13]

The search for a fair balance between the need of LEAs to monitor social media and the protection of fundamental rights is therefore particularly challenging.

To complicate this scenario, content moderation of online illegal content emerged as a relatively new area where LEAs can play a role, adding up to the grounds at least for social media crawling, and potentially for social media scraping. The next section of this paper analyses the existing EU framework on content moderation, and highlights how this framework can impact on the role of LEAs in content moderation practices.

3 The EU legislative landscape on content moderation and the role of LEAs

The growing attention of the policy debate on the dissemination of forms of illegal or harmful content is demonstrated by the recent legislative initiatives at the EU level to establish duties of care of digital platforms with regard to tackling said content. . The involvement of LEAs in tackling certain types of content can derive from the fact that their dissemination is a criminal offense *per*

[8] Sander, Barrie, "Freedom of expression in the age of online platforms: the promise and pitfalls of human rights-based approach to content moderation." *Fordham International Law Journal*, 43.4 (2020), 939-1006.

[9] Vladimir Kharitonov v. Russia, App no 10795/14 (ECtHR 23 June 2020), par. 33. Melike v. Turkey, App no 35786/19 (15 June 2021), par. 44. Magyar Helsinki Bizottság v. Hungary [GC] App no 18030/11 (ECtHR 8 November 2016), par. 168.

[10] Scott, Jeramie D. "Social media and government surveillance: The case for better privacy protections for our newest public space." *Journal of Business and Technology Law*, 12.2 (2017), 151-164.

[11] Sandoval-Almazan, Rodrigo, Gil-Garcia, G. Ramon, "Towards cyberactivism 2.0? Understanding the use of social media and other information technologies for political activism and social movements." *Government Information Quarterly*, 31.3 (2014), 365-378.

[12] UK Information Commissioner's Office and others, "Joint statement on data scraping and the protection of privacy", August 2023. https://ico.org.uk/media/about-the-ico/documents/4026232/joint-statement-data-scraping-202308.pdf

[13] Coe, Peter. "The social media paradox: an intersection with freedom of expression and the criminal law." *Information & Communications Technology Law* 24.1 (2015), 16-40.

se, or it is made otherwise illegal. This is the case for terrorist content, as defined in the Terrorist-Content Regulation (TERREG)[14] through a reference to the Counter-Terrorism Directive[15], or of the dissemination of child sexual abuse material online,[16] for which the EU is currently acting in order to enshrine a specific set of rules in a regulation.[17] Moreover, the concerns about the effects of disinformation have led the EU Member States to take legislative initiatives to tackle the phenomenon, which sometimes resulted in the criminalisation of certain forms of false information with a potential to cause societal harms.[18] Additionally, LEAs can be appointed with tasks related to the monitoring of content shared online when non-legislative initiatives provide for such measures in order to safeguard public order, or other relevant interests such as the integrity of election processes, that can be undermined by the spread of harmful content itself.[19] The provisions enshrined in the TERREG and DSA that are hereby analysed provide an example of how the EU legislator tried to address this necessity.

The TERREG, adopted in 2021, represented an important step in the direction of regulating the spread of a type of content that has been made uniformly illegal across the EU.[20] More recently, the Digital Services Act (DSA) was adopted with the aim to harmonise rules imposed on digital services providers concerning the moderation of illegal content.[21] While the TERREG only focused on terrorist content, defining it by reference to the definitions of terrorist offenses as delineated in the Counter-Terrorism Directive[22], the DSA defined "illegal content" as any content which has been made illegal either by EU law or by domestic laws of Member States.[23] In other words, the provisions enshrined in the legal instruments can apply differently across the EU, depending on which type of content has been made illegal in the national jurisdictions. However, it is noteworthy that the DSA shows a willingness of the EU legislator to tackle the worrying phenomenon of online disinformation. Various recitals of the DSA highlight that the instrument was conceived to address societal risks from dissemination campaigns.[24] Moreover, very large online platforms and very large search engines[25] are obliged to perform a risk assessment of a number of systemic risks for their platforms[26], in order to take adequate mitigating measures.[27] One of the systemic risks triggering these obligations concerns "any actual or foreseeable negative effects on civic discourse and electoral processes, and public security".[28] While this formulation is very broad, the definition of this risk resonates with the content of former EU policy documents addressing the matter of disinformation. In particular, the communication of the European Commission adopted in 2018 to define a common EU approach to disinformation identifies the hampering of democratic political and policy-making processes as one of the threats posed by this phenomenon.[29]

The next two subsections provide an in-depth analysis of how the TERREG and the DSA encourage LEAs to monitor social media spaces in order to tackle illegal content, but also with a broader view to prevent and control crimes.

[14] Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online, [2021] OJ C 110 (TERREG).

[15] Directive (EU) 2017/541 of the European Parliament and of the Council of 15 March 2017 on combating terrorism and replacing Council Framework Decision 2002/475/JHA and amending Council Decision 2005/671/JHA, [2017] OJ C 177 (Counter-Terrorism Directive).

[16] Directive 2011/93/EU of the European Parliament and ofthe Council of 13 December 2011 on combating the sexual abuse and sexual exploitation of children and child pornography, and replacing Council Framework Decision 2004/68/JHA, [2011] L 194/1.

[17] Proposal for a Regulation of the European Parliament and of the Council laying down rules to prevent and combat child sexual abuse, [2022] COM(22)209.

[18] Ó Fathaigh, Ronan, Natali Helberger, Naomi Appelman. "The perils of legally defining disinformation." *Internet policy review* 10.4 (2021), 2022-40.

[19] van Hoboken, Joris, Ronan Ó. Fathaigh. "Regulating Disinformation in Europe: Implications for Speech and Privacy." *UC Irvine Journal of International, Transnational and Comparative Law* 6 (2021), 9-36.

[20] TERREG (n. 13).

[21] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) [2022] OJ L 277 (DSA).

[22] TERREG (n. 13), art. 2(7).

[23] DSA (n. 18), art. 3(h).

[24] DSA (n. 18), recitals 2-9-69-83-84-88-95-104-106-108.

[25] DSA (n. 18), art. 33.

[26] DSA (n. 18), art. 34.

[27] DSA (n. 18), art. 35.

[28] DSA (n. 18), art. 34.1(c).

[29] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, *Tackling online disinformation: a European Approach* [2018] COM/2018/236.

3.1 The role of LEAs in identifying and removing illegal content

Both the TERREG and the DSA are examples of legal instruments that institutionalise the possibility for competent authorities to issue removal orders in their specific scope.[30] While the competent authorities empowered by the Member States to issue such orders do not have to be necessarily LEAs, at least the process of national designation pursuant to the TERREG shows how the Member States mostly selected police bodies to carry out this function.[31]

Moreover, the DSA established rules on the so-called "trusted flaggers". These entities are designated based on their particular expertise in the area of illegal content, their independency from online platforms' providers, and their diligence, accuracy and objectivity in carrying out their tasks.[32] Such entities have, like any other user of online platforms, the possibility to flag an allegedly illegal content to the providers of online platforms to have it removed, pursuant to the terms and conditions of the service or to a EU or domestic law making that content illegal.[33] However, contrary to other individuals entitled to resort to this notice and action mechanism, the trusted flaggers' notices must be given priority in the processing by online platforms. As confirmed by the recitals of the DSA, the status of trusted flaggers can be recognised, among others entities, also to national LEAs or Europol.[34]

The concept of "trusted flagger" is not new in the area of content moderation, as it resembles the functioning of the already existing Internet Referral Units (IRUs) in the area of terrorist content. The EU Internet Referral Unit was established within the European Counter-Terrorism Center of the EU Agency for Law Enforcement Cooperation (Europol) in 2015 with the aim to support EU Member States in their counter-terrorism efforts by tackling terrorist content.[35] Its establishment was functional to coordinate the activities of the national IRUs across the EU. The mechanism of referrals, which implies a notice made by LEAs based on the terms and conditions of online platforms, is referred to in the recitals of the TERREG, as a means alternative to removal orders that can be used to tackle terrorist content online.[36] This tool was criticised by the Fundamental Rights Agency (FRA) in its comments to an early proposal of the TERREG, which aimed to institutionalise it through a dedicated provision. The FRA expressed its concerns with regard to the fact that the proposal did not clearly delineate the circumstances where the referral mechanism was to be preferred to a removal order, with a risk of the former being used because of less stringent requirements – and diminished safeguards – at the national level.[37] While the provision was not included in the final text of the TERREG, the mechanism is still in place at both the EU and national level. Moreover, the competences of the EU IRU were recently enlarged beyond the scope of tackling terrorist content, to other types of dangerous information. In general, the EU IRU has shown a certain degree of flexibility in adapting its range of actions to different categories of content depending on national LEAs' needs.[38]

The IRUs have been identified as one of the areas where online platforms and public authorities' cooperation results in a co-production of security.[39] In fact, while LEAs are entitled to use referrals, the final assessment about a flagged content is left to the discretion of online platforms, which decide based on their terms and conditions. This aspect differentiates the IRUs from the trusted flaggers, as the latter are entitled, pursuant to the DSA, to flag content that is either against the terms and conditions of online platforms or against EU or Member States' law.[40] However, the characterisation of trusted flaggers significantly resembles that of the IRUs, and renewed the question of opportunity of allowing LEAs to act on this ground, instead of making use of the

[30] TERREG (n. 13), art. 3, DSA (n 18), art. 9.

[31] European Commission, List of national competent authority (authorities) and contact points, https://home-affairs.ec.europa.eu/policies/internal-security/counter-terrorism-and-radicalisation/prevention-radicalisation/terrorist-content-online/list-national-competent-authority-authorities-and-contact-points_en

[32] DSA (n. 18), art. 22.

[33] DSA (n. 18), art. 16.

[34] DSA (n. 18), recital 61.

[35] Council of the European Union, "Justice and Home Affairs Council, 12-13 March 2015", 2015, https://www.consilium.europa.eu/en/meetings/jha/2015/03/12-13/.

[36] TERREG (n. 13), recital 40.

[37] Opinion of the European Union Agency for Fundamental Rights, *Proposal for a Regulation on preventing the dissemination of terrorist content online and its fundamental rights implications*, 2019.

[38] Kilpatrick, Jane, Jones, Chris, "Empowering the police, removing protections: the new Europol Regulation", Statewatch, 2022. https://www.statewatch.org/media/3615/empowering-the-police-removing-protections-new-europol-regulation.pdf

[39] Bellanova, Rocco, De Goede, Marieke, "Co-Producing Security: Platform Content Moderation and European Security Integration." *JCMS: journal of common market studies* 60.5 (2022), 1316-1334.

[40] DSA (n. 18), art. 22.

removal order mechanisms. Due to transparency and rule of law concerns, the possibility enshrined in the DSA to designate LEAs and Europol as trusted flaggers has been criticised.[41]

Due to the importance of the dissemination of content on the Internet to exercise freedom of expression, initiatives trying to tackle illegal content are subject to a close scrutiny in the academic community. As noted in a report of 2023 about legislative initiatives taken across the world to combat disinformation, such initiatives can have a chilling effect on freedom of expression, due to the vagueness of the definitions of illegal content and the sometimes disproportionate sanctions associated with its spread.[42] Similar concerns were expressed in occasion of the adoption of the TERREG, due to the divergence of interpretations across the EU about what amounts to terrorist content.[43] The TERREG itself demonstrates the awareness of the EU legislator about the risk for content moderation practices to become an instrument to silence minorities or dissenting opinions in the political realm. In fact, one of its provisions specifies that material which represents an expression of polemic or controversial views in the course of public debate, shall not be considered to be terrorist content.[44] In this regard, it has also been noted how important it is to distinguish between terrorist content and content raising awareness about terrorism, which should not be banned. In general, a common difficulty in content moderation activities is that certain types of content, such as terrorist and extremist content, are more complicated to uniformly define compared to other, such as child sexual abuse material.[45]

In reason of its recent adoption, and the broad scope of the DSA with regard to what constitutes illegal content, its impact on social media monitoring practices by LEAs in the quality of trusted flaggers remains to be seen. However, the functioning of IRUs already provides an example of how LEAs are involved in patrolling social media. Given the amount of information available online, IRUs resort to OSINT capabilities to detect illegal content online.[46] Besides the huge number of pieces of content and online platforms analysed[47], the 2021 Consolidated Annual Report of Europol states that, once identified, the content is manually assessed by a Unit's expert to verify whether it has wrongfully been categorised as terrorist.[48] This activity adds up to that of the national IRUs, which provide Europol with the open source information they collected.[49] In turn, all the information obtained thanks to this activity is then included in a centralised database. In fact, one of the main interests of the EU IRU is not only to detect illegal content, but also to retrieve it, as it falls within the competence of Europol, providing a legal basis for such retention.[50]

Moreover, while the EU IRU was originally create to tackle terrorist content, its reach was broadened in relation to its cooperation with national IRUs. In 2021, its tasks included tackling content related to right-wing violent extremism, online terrorist propaganda and violent jihadist ideology, migrant smuggling during the "Belarus Crisis". Moreover, it also acted in the context of an Internet Archive platform aimed to strengthen public-private cooperation in content moderation.[51] The analysis of the EU IRU's work shows how the connection between social media activities and offline crimes can justify a policy-making approach enlarging the grounds for content monitoring. The potential for LEAs' monitoring activities to go beyond the mere detecting of illegal content is discussed in the next subsection.

[41] Komaitis, Konstantinos, Rodriguez, Katitza, Schmon Christoph, "Enforcement Overreach Could Turn Out To Be A Real Problem in the EU's Digital Services Act", *Electronic Frontiers Foundation*, 2022. https://www.eff.org/deeplinks/2022/02/enforcement-overreach-could-turn-out-be-real-problem-eus-digital-services-act

[42] Lim, Gabrielle, Bradshaw, Samantha, "Chilling Legislation: Tracking the Impact of"Fake News" Laws on Press Freedom Internationally." *Center for International Media Assistance*, 2023. https://www.cima.ned.org/wp-content/uploads/2023/06/CIMA-Chilling-Legislation_web_150ppi.pdf

[43] Berthélémy, Chloé, "EU Terrorist Content Online Regulation Could Curtail Freedom of Expression across Europe." *EDRi*, 2021. https://edri.org/our-work/eu-terrorist-content-online-regulation-could-curtail-freedom-of-expression-across-europe/

[44] TERREG (n. 13), art. 1.3.

[45] Chang, Brian. "From Internet Referral Units to International Agreements; Censorship of the Internet by the UK and EU." *Columbia Human Rights Law Review* 49 (2017), 114-212.

[46] Chang (n. 43), p. 135.

[47] Europol Consolidated Annual Activity Report, 2021. https://www.europol.europa.eu/cms/sites/default/files/documents/Consolidated%20Annual%20Activity%20Report%202021.PDF.

[48] Europol EU Internet Referral Unit Transparency Report, 2021. https://www.europol.europa.eu/cms/sites/default/files/documents/EU_IRU_Transparency_Report_2021.pdf-.

[49] Chang (n. 43), p. 135.

[50] Bellanova, De Goede (n. 37).

[51] Kilpatrick, Jane, Jones, Chris (n. 39).

3.2 The role of LEAs beyond the identification and removal of illegal content

While not all of the content that is made illegal under EU law leads to its dissemination being a criminal offense under EU or Member States' law, the connection between the spread of illegal content online and the risk of commission of certain forms of crimes was highlighted in the EU policy debate on content moderation. Prior to the adoption of the TERREG and the DSA, the Commission recognised, in a more general communication on the moderation of illegal content, that LEAs have a duty to prosecute crimes, and online platforms are responsible for preventing that their services are misused in order to commit them.[52] Therefore, the intersection between the dissemination of certain content and the area of criminal justice can result in a potential overlapping between LEAs' tasks and the monitoring of social media beyond the tackling of illegal content. In particular, LEAs can be involved in monitoring content published on social networks as a response to the strict connection between the dissemination of said content and the commission of criminal offenses in the offline world. In other words, social media monitoring activities can serve the purpose of act in a crime prevention perspective, by analysing social media content to detect and anticipate potential criminal offense.

The connection between the dissemination of certain content online and the possibility of online activities resulting in a broad range of threats is confirmed by the DSA In its recitals, the regulation specifies that the concept of illegal content should be defined in relation to existing rules in the offline environment. Thus, it should encompass any information that not only is made illegal by a specific law, but is also connected to illegal activities in general. As an example of online content that can signal criminal offenses, the DSA mentions the non-consensual sharing of private images, online staking, the sale of non-compliant or counterfeit products, copyrights infringements.[53]

The monitoring of social media provides LEAs with an affordable source of information to observe individual behaviours and social relations. Due to the sociological premise that the observation of individuals in social environments can help gain knowledge about future crimes, monitoring activities deriving from the necessity to moderate illegal content can result in further actions by LEAs.[54] The analysis and social media data can be use not only to take down harmful content, but to direct LEAs' efforts in their crime prevention activities. Therefore, data extracted for content moderation purposes can be also used in other, and broader, contexts. An example is the possibility of using social media data inferred by the content shared online to make behavioural predictions about individuals. AI-driven predictive models are versatile, as they can be used with any amount of publicly available personal data retreived from social media in order to make a broad range of predictions. In combination with certain assumptions on the way human beliefs and choices are formed, such predictions can result in conclusions on motivations and inclinations that are not always accurate. In the case of predictions of LEAs based on publicly available data, there is risk of social control and stigmatisation deriving from these practices, which can in turn harm the autonomy of individuals in expressing themselves online.[55]

In light of the possibility for social media data to be used for purposes that go beyond what is traditionally understood as content moderation, it is all the more important to carefully assess the impact of data scraping techniques on the protection of personal data that are publicly available. Such an evaluation should be carried out by considering the technological features of data scraping practices, and in particular, in the context of social media, the premises and pitfalls of NLP algorithms in detecting illegal content, or suspicious activities in general. The next section will explore the fundamental rights' concerns arising from these technological means.

4 The application of social media scraping technologies in the law enforcement domain. Natural Language Processing techniques to classify online content

As already observed, NLP technologies are a precious tool in order to detect and categorise certain types of online content. However, it has been noted that such technologies, while widely used, are not yet at a developing stage allowing their use without

[52] Communication from the Commission to the European Parliament, the council, the European Economic and Social committee and the Committee of the Regions, *Tackling Illegal Content Online Towards an enhanced responsibility of online platforms* [2017] COM/2017/0555.

[53] DSA (n. 18), recital 12.

[54] Susser, Daniel, "Predictive policing and the ethics of preemption." *The ethics of policing: New perspectives on law enforcement* (2021), 268-292.

[55] Ploug, Thomas. "The right not to be subjected to AI profiling based on publicly available data—privacy and the exceptionalism of ai profiling." *Philosophy & Technology* 36.1 (2023), 14.

any human supervision.[56]

First, NLP algorithms are not able to take into account contextual clues and circumstances that, for certain types of content, is essential in order to assess its nature.[57] This is, for example, the case for terrorist content, for which the importance of context in the identification is made explicit in the recitals of the TERREG[58], and disinformation, as falsehood of a piece of information is an highly contextual element.[59] This point is connected to the inherent difficulty in categorising certain types of content. In some cases, such as with regard to violent extremist content, defining what falls within this category can be challenging due to the lack of an uniformly accepted definition, as opposed to the case of other illegal content, such as child sexual abuse material, which is more straightforwardly identifiable.[60] Secondly, the margin of error of such technologies may result in wrongful labelling of content, due to the difficulties of identifying harmful content online by matching the content with a pre-defined set of words. In other words, false positive and false negatives are a risk to be considered.[61] Thirdly, as a recent report of the Fundamental Rights Agency outlined through an empirical research how algorithms used to identify hate speech are susceptible to be biased, and their use can lead to discriminatory effects for persons, based, for example, on their religious or political affiliation.[62]

Notwithstanding these limitations, the range of possible applications of NLP technologies in the context of LEAs' activities seems wide. A brief literature review of research projects investigating the potential of NLP for the criminal justice area reveals that this technology is proposed in the context of social media monitoring as a potential solution to investigate illicit COVID-19 product sales[63], human trafficking[64], religious and political extremism.[65] Moreover, NLP's potential uses for LEAs have also been identified in situations where the spread of illegal content itself represents the source of societal harm, as opposed to scenarios where it is connected to ulterior crimes perpetrated offline. This is the case for studies investigating the use of NLP and its limitations to tackle hate speech[66] and disinformation.[67]

The ongoing discussion about the use of data scraping technologies on social media by using NLP requires a legal analysis of the privacy and data protection implications of using such techniques to monitor online environments. The aim of this analysis is to draw more general conclusions about the possible impact of such technologies to fundamental rights and freedoms in the area of law enforcement. The next section explores the legal status of publicly available data collected from social media and their protection at the EU level.

5 The legal status of publicly available data collected from social media

The present section discusses the status of publicly available data from a privacy and data protection standpoint, then applying the outcomes of the analysis to social media scraping practices performed by LEAs. Firstly, it analyses what protection is granted to publicly available data pursuant to the EU data protection framework. Secondly, it analyses privacy and data protection guarantees afforded in relevant judgements of the ECtHR. Finally, it focuses on the CJEU case law, by extracting some useful principles to guide the lawful processing of publicly available data from the EU Court's decisions.

[56] Krotov, Vlad, Leiser Silva. "Legality and ethics of web scraping." *Twenty-fourth Americas Conference on Information Systems, New Orleans* (2018).

[57] Gorwa, Robert, Bonns, Reuben, Katzenback, Christian "Algorithmic content moderation: Technical and political challenges in the automation of platform governance." *Big Data & Society* 7.1 (2020), 2053951719897945.

[58] TERREG (n. 13), recital 11.

[59] Pielemeier, Jason. "Disentangling disinformation: What makes regulating disinformation so difficult?." *Utah Law Review* 4 (2020), 917-940.

[60] Chang (n. 43), p. 137.

[61] Gorwa, Binns, Katzenback (n. 53).

[62] Fundamental Rights Agency, *Bias in algorithms. Artificial Intelligence and discrimination*, 2022.

[63] Mackey, Tim Ken, et al. "Big data, natural language processing, and deep learning to detect and characterize illicit COVID-19 product sales: infoveillance study on Twitter and Instagram." *JMIR Public Health and Surveillance* 6.3 (2020), 360-376.

[64] Granizo, Sergio L., et al. "Detection of possible illicit messages using natural language processing and computer vision on twitter and linked websites." *IEEE Access* 8 (2020), 44534-44546.

[65] Torregrosa, Javier, et al. "A survey on extremism analysis using natural language processing: definitions, literature review, trends and challenges." *Journal of Ambient Intelligence and Humanized Computing* 14.8 (2023), 9869-9905.

[66] Poletto, Fabio, et al. "Resources and benchmark corpora for hate speech detection: a systematic review." *Language Resources and Evaluation* 55 (2021), 477-523.

[67] de Oliveira, Nicollas R., et al. "Identifying fake news on social networks based on natural language processing: trends and challenges." *Information* 12.38 (2021).

5.1 The EU legal framework

From a data protection point of view, the activities involving the scraping and classification of data from content shared on social media amount to processing of personal data when the detection of illegal content implies that such content can be associated to the civil identity of a user. In other words, such technologies make a certain data subject identifiable, thus falling within the scope of the EU data protection framework.[68] The processing attains in this case to personal data that are publicly available online. Personal data is considered as publicly available when its access is not limited to a specific group of persons, with different nuances about what should amount to such a free access. However, regardless of the divergencies over the definitory approach to publicly available data, the EU legal system extends data protection guarantees to publicly available data, even in the law enforcement context, as evident from the Europol legal framework.[69]

The 2016 Europol Regulation mentions data deriving from publicly available sources among those that the EU agency can process in the course of its activities.[70] Nontheless, the restrictions enshrined in the regulation on the possibility to share publicly available data with private parties[71], as well as the rights of erasure, rectification or access restriction to which data subjects are entitled with regard to this type of data[72], confirms that the public availability does not entail a release from any obligation to respect data protection principles.

Personal data that can be retrieved from content shared online represents a peculiar type of publicly available content, for two orders of reasons. First of all, the data mining on social media can allow to extract implicit and potentially useful information that regards sensitive data of the users, such as their political affiliation and religious beliefs. This aspect can be especially problematic in the context of profiling to be used in predictive policing activities.[73] Secondly, while other publicly available data is made available online by third parties, the content that is shared online is made available by the publishers themselves, triggering the application of a specific regulatory framework when LEAs process, in particular, sensitive data of the data subjects.

Under the Law Enforcement Directive (LED), governing the processing of personal data by competent authorities in relation to the prevention, detection or prosecution of crimes[74], special categories of personal data, including data pertaining to political or philosophical beliefs, should only be processed when strictly necessary and based on specific legal grounds, subject to appropriate safeguards for fundamental rights.[75] One of the grounds justifying the processing is the fact that sensitive data have been manifestly made public by the data subject.[76] From the letter of the law it can be drawn that, while the processing of publicly available sensitive data by LEAs is possible in this case, the processing still needs to comply with the substantial requirement of strict necessity, and the procedural requirements of providing appropriate safeguards to data subjects.

Both the Working Party 29 (WP 29) and the European Data Protection Board (EDPB) contributed to delineating the meaning of "manifestly public". While a clear intention of the data subject to make personal data available should be present to consider the processing lawful, the WP 29 provided some specifications with regard to publicly available data in social media. In this case, most of the individuals publishing content online do not expect their data to be accessible by police authorities. A combination of elements should play a role in the assessment about this expectation of privacy, such as the accessibility of the source where

[68] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119, art. 4(1).

[69] Gottschalk, Thilo. "The Data-Laundromat? Public-Private-Partnerships and Publicly Available Data in the Area of Law Enforcement." *European Data Protection Law Review* 6 (2020), 21-40.

[70] Regulation (EU) 2016/794 of the European Parliament and of the Council of 11 May 2016 on the European Union Agency for Law Enforcement Cooperation (Europol) and replacing and repealing Council Decisions 2009/371/JHA, 2009/934/JHA, 2009/935/JHA, 2009/936/JHA and 2009/968/JHA [2016] OJ L 135 (2016 Europol Regulation), art. 17.2.

[71] 2016 Europol Regulation (n. 66), art. 37.

[72] Regulation (EU) 2022/991 of the European Parliament and of the Council of 8 June 2022 amending Regulation (EU) 2016/794, as regards Europol's cooperation with private parties, the processing of personal data by Europol in support of criminal investigations, and Europol's role in research and innovation [2022] OJ L 169.

[73] Mitrou, Lilian, et al., "Social media profiling: A Panopticon or Omniopticon tool?." *Proceedings of the 6th Conference of the Surveillance Studies Network*, 2014.

[74] Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA [2016] OJ L 119 (LED).

[75] LED (n. 70), art. 10.

[76] LED (n. 70), art. 10(c).

the data is published, the visibility of the information and the fact that the subject made the data about herself or himself public, as opposed to third parties.[77]

The EDPB Guidelines on the use of facial recognition technologies shed further light on the possibility to scrape publicly available data online, adopting a restrictive interpretation of what can be considered as publicly available online.[78] While the guidelines focused on the risks associated with the processing of biometric data obtained from scraping publicly available images online, some concepts expressed by the EDPB can have an application to the case of data on political views inferred from posts and other activities performed on social media. The Guidelines stated that, in the case of social networks or online platforms, the privacy features chosen by the user are not sufficient to consider that personal data are manifestly made public, and that such data can be processed for identification purposes without consent. Moreover, the requirement of the processing being "strictly necessary" refers to conditions even stricter than the conditions of necessity as normally required under the LED. This means that the processing should be *indispensable*, and the LEAs should have a limited margin of appreciation in assessing the necessity. Any processing of general or systematic nature should be excluded in this context. In fact, only objective criteria to define whether the circumstances and conditions of a certain situation justify the processing of sensitive data. The strict necessity requirement could not be met in the case of measures entailing the population of police databases with data collected on a mass-scale and in an indiscriminate way from online sources.

The stringed stance taken by the EDPB about facial recognition technologies used on images scraped from social media derives from concerns about the use of these tools in public spaces. The EDPB underlined how a general provision allowing to use such technologies in public spaces where individuals have an expectation of anonymity can lead to a chilling effect with regard to rightful actions, such as joining an association or a demonstration. Moreover, according to the sensitivity of the data processed, such tools can be used to put pression not only on the general public in a way that may impair their ability to take part to public life, but also on key actors such as political opponents and journalists. Finally, it should be considered that the use of sensitive data to populate police databases is a type of processing which is prone to discriminatory effects based on gender, origins or political opinions.[79]

The combined interpretation of the LED provisions and the analysis of the WP29 and the EDPB leads to some first reflections about the data that can be scraped by analysing content shared on social media. First, with regard to the elements that allow to draw the conclusion that certain sensitive data are made manifestly available by data subjects, it is not enough to simply derive this intention of the data subject from the sharing of content on a publicly available source. In the case of social media content, the matter is complicated by the fact that, as observed above, automated means of data analysis can infer implicit information from content posted online, by categorising it according to pre-established criteria. A case-by-case assessment should be carried out to evaluate whether a classification based on data inferred by online content can automatically be considered as a manifestation of the intention to make such data publicly available. Without any doubts, the data subject manifestly chooses to publish content online, but the consequent classification based on implicit meanings of the content does not necessarily reflect the extent to what she or he accepted the possibility of having her or his sensitive information subject to public scrutiny.

Secondly, even when sensitive data is made manifestly available, the requirement of strict necessity is still to be met to comply with the LED provisions. The assessment on the strict necessity of processing sensitive data is linked to the intricacies of evaluating the value of social media data in the LEAs' activities. As noted above, the dissemination of a certain content can amount to a criminal offense *per se*, or it can signal a preparatory act to commit an offline criminal offense, or, lastly, it is indicative of an abstract threat to a public good that deserves legal protection, such as public order. The first case, where the dissemination of online content can be directly connected to an actual harm to individuals, shows a significant evidentiary value of the content itself and the data related to the content. For example, the dissemination of hate speech or defamatory content online can fall within this category.[80] To the opposite side of the spectrum, content criminalised as disinformation, or otherwise made illegal due to the potential disturbance of a general public good such as public order, require a more careful assessment on whether collecting and processing publicly available data from social media can be connected to an actual LEAs' need in relation to a specific harm. Terrorist content can be allocated in the middle of these two scenarios, as, while the connection to potential, yet serious danger of a criminal offense is clear in this type of content, the risk is that of an excessive anticipation of the protection offered by criminal

[77] Jasserand, Catherine. "Article 10. Processing of Special Categories of Personal Data." *LED Commentary, OUP (forthcoming Fall 2023)*, 2023.

[78] European Data Protection Board, *Guidelines 05/2022 on the use of facial recognition technology in the area of law enforcement*, 2022.

[79] *Idem*.

[80] Birritteri, Emanuele, "La disinformazione tra politica e diritto. Dimensione istituzionale, strategie preventive e dinamiche punitive.", *Diritto Penale Contemporaneo* 4 (2021), 304-334.

laws with respect to the possibility of the dissemination resulting in a threat to the safety of persons. The assessment on the causal link with criminal offenses, let alone a specific harm, can be challenging in the last two cases described.[81]

Further elements in the concrete scenarios should be considered when evaluating to what extent rights to privacy and data protection can be restricted. However, the case law of the European Court of Human Rights (ECtHR) and of the Court of Justice of the EU (CJEU) can provide some additional elements on how social media scraping technologies should be used, depending on the types of crimes to be tackled, and on the level of interference with fundamental rights. The case law of the two Courts is explored in the next two subsections.

5.2 5.2 The ECtHR case law

The ECtHR traditionally questioned the lawfulness of interferences with the rights to privacy and data protection with regard to data acquired by open sources by focusing on the storage and subsequent use of such data. As regards the mere searching and consultation of data on online platforms, the question is open on whether this may cause an infringement of fundamental rights.[82]

The ECtHR was very active in delineating the concept of an expectation of privacy in public spaces, in the context of interferences by LEAs aimed to combat crime. In Rotaru v. Romania, the court stated that "public information can fall within the scope of private life where it is systematically collected and stored in files held by the authorities".[83] The conclusion is notable, as it argued against the conviction of the government that the applicant had willingly engaged in political activities and published pamphlets concerning his political views, therefore waiving his right to anonymity.[84] The Court argued that the ability to establish social relationships fell within the scope of the right to a private life, as protected under Article 8 of the European Convention of Human Rights (ECHR). Moreover, it noticed that public information falls within the scope of private life, even more "where such information concerns a person's distant past".[85] It is evident from this decision how the systematic nature of the collection and storage of publicly available data is a key element in assessing whether an interference by public authorities is legitimate. This conclusion can be drawn by other cases, where the Court did not find an infringement of Article 8, due to the absence of a systematic collection and processing of data.[86]

The ECtHR took a further step in addressing the issue of the right to privacy with respect to publicly available data in the case Catt. v. UK. In the case, the police held in an "extremism database" personal data related to the participation of the applicant in a number of demonstrations, and his association with some political organisations whose protests tended to become violent.[87] The data were retained pursuant to the definition of "domestic extremist", which significantly varied in its interpretation among public authorities. This vagueness gave raise to concerns with regard to the ambiguous criteria leading to the collection and storage of data, with the risk of the legal basis being used as an *ad hoc* instrument to collect information of individuals.[88] However, as also noted in one of the opinions to the judgement, the ECtHR did not focus further on the quality of the law in question.[89] Instead, the Court acknowledged that the monitoring of the protests, and consequent collection of data of the applicant, were pursuing a legitimate aim, due to the tendency of such protests and groups therein to become violent, and the willingness of the applicant to take part to them in a public way.[90] Nevertheless, the illimited retention of the data, the lack of review on the database and the ineffective remedies provided to the data subject were a decisive element that brought the Court to declare an infringement of Article 8.[91] Moreover, the Court highlighted the sensitiveness of the data collected, as related to political activities of the applicant. The retention of such data could cause a chilling effect to the freedom of expression of individuals, thus playing a role

[81] Sabella, Pietro Maria. "Il fenomeno del cybercrime nello spazio giuridico contemporaneo. Prevenzione e repressione degli illeciti penali connessi all'utilizzo di internet per fini di terrorismo, tra esigenze di sicurezza e rispetto dei diritti fondamentali." *Informatica e diritto* 26.1-2 (2017), 139-176.

[82] Edwards, Lilian, Urquhart, Lachlan. "Privacy in Public Spaces: What Expectations of Privacy Do We Have in Social Media Intelligence?'(2016)." *International Journal of Law and Information Technology* 24.3 (2016), 279-310.

[83] Rotaru v Romania App no 28341/95 (ECtHR, 4 May 2000).

[84] *Idem*, par. 92.

[85] *Idem*, par. 43.

[86] Edwards, Urquhart (n. 78).

[87] Catt v UK App no 43514/15 (EctHR, 24 January 2019).

[88] *Idem*, par. 97.

[89] *Idem*, Concurring Opinion of Judge Koskelo joined by Judge Felici.

[90] *Idem*, par. 108.

[91] *Idem*, par. 100 and ff.

in the evaluation about the infringement.[92] This acknowledgement of the sensitive nature of data playing a role in the level of data protection should be read in light of the Court recognising, in a different case on mass retention of sensitive data, that "an individual's concern about the possible future use of private information retained by the authorities is legitimate and relevant to a determination of the issue of whether there has been an interference".[93]

More recently, in the case Glukhiv v. Russia, the ECtHR reiterated that an interference with the right to privacy is possible even when data collected and stored by public authorities only pertain to public activities.[94] The Court also recognised that individuals can have a reasonable expectation of privacy when engaging in certain activities, since they are not aware that such activities are recorded.[95] Therefore, a violation of the right to a private life was found. Such violation derived not only from the consideration of a possible chilling effect due to the fact that the data processed by public authorities led to the arrest of the applicant based on information collected on him engaging in a solo demonstration. The Court also noted that the intrusiveness in his private life was disproportionate compared to the offense committed by the subject (and the underlying objective of public interest pursued). In fact, the offense in question was administrative in nature, and was linked to the lack of prior approval of the solo demonstration by the competent authorities.[96]

These judgements should be read in light of the extensive jurisprudence of the ECtHR embracing a very broad interpretation of the right to freedom of expression. The Court went as far as to establishing a protection for ideas that may offend, shock or disturb the public[97] and for the dissemination of information whose truthfulness can be called into question[98]. Furthermore, the Court established that governments, due to their dominant position, retain a discretion in adopting criminal-law measures to preserve public order, but should at the same time exercise restraint in resorting to criminal sanctions to silence criticisms and political opponents.[99] The ECtHR also recognised that the necessity to limit freedom of expression in the context of the dissemination of false allegations aimed to undermine the ability of citizens to obtain accurate information in the context of elections. However, strong procedural safeguards should be in place in enacting such limitations.[100]

The analysis of the ECtHR case law allows to enrich the discussion about the rules that should guide social media scraping activities by LEAs. First and foremost, the importance of protecting sensitive data pertaining to political views of individuals is strongly remarked in the analysed judgements. In the case of content shared on social media, the concerns regarding freedom of expression should play a role in assessing the level of interference with fundamental rights. Indeed, such concerns could for example be linked to the preoccupation of data subjects about retained data could be used in the future. In the case of social media content explicitly or implicitly revealing political affiliations or other sensitive information, a valid concern could be identified in the possibility that a change of political flag in a government could lead to a very different treatment of individuals with certain political views. In this regard, the systematic collection of personal data could be problematic, especially in light indefinite retention periods. This element is important, as it not only creates grounds for potential misuse of the retained data depending on the political climate, but also diminishes the power of individuals to present themselves to the public as they see fit. This, in turn, can lead to a crystallisation of the profiles obtained by LEAs, in spite of the changing nature of individual characteristics related to their expressions in social life. Finally, the necessity proportionality of an interference with private life should also be evaluated taking into account the possible consequences of such an interference on the individuals. In this sense, the seriousness of the offenses that LEAs are trying to prevent should provide a standard with regard to the intrusiveness of social media scraping practices, and the amount and quality of personal data collected. A legislative intervention allowing to modulate the intrusiveness of monitoring activities according to the gravity of crimes that LEAs are attempting to prevent would allow to substantiate the requirements of necessity and proportionality by reference to objective elements that should guide the compliance with them.

[92] *Idem*, par. 112.

[93] Marper v UK App no 30562/04 (EctHR, 4 December 2008), par. 71.

[94] Glukhiv v Russia App no 11519/20 (EctHR, 4 July 2023).

[95] *Idem*, par. 66.

[96] *Idem*, par. 88

[97] Handyside v UK App no 5493/72 (EctHR, 7 December 1976).

[98] Salov v Ukraine App no 65518/01 (EctHR 6 September 2005)

[99] Incal v Turkey App no 41/1997/825/1031 (EctHR 9 June 1998).

[100] Brzeziński v Poland App no. 47542/07 (EctHR 25 July 2019).

5.3 5.3 The CJEU case law

While the CJEU did not address the interference with privacy and data protection when collecting publicly available data, it is useful to briefly summarise is findings as regards the lawfulness of data retention measures adopted by national authorities in the context of the fight against crime. While the decisions of Tele2 Swerige and Digital Rights Ireland focused in particular on the retention of traffic and location data, they generally confirmed that the mere retention of personal data can amount to an interference with data protection, due to its potential deterrent effect to freedom of expression.[101] Moreover, in Tele2 Swerige, the EU Court stated that the legal bases allowing for retention measures should be precise enough to direct such measures to individuals that have a link, at least an indirect one, with the objective pursued by the measure, such as fighting a certain crime or safeguarding public security – in particular, the court mentioned at least an indirect link with potential criminal proceedings.[102] In the case La Quadrature du Net, the Court addressed the question on the lawfulness of preventive retention measures of data pertaining, among others, to the civil identity of individuals. The CJEU found that, based on the sensitivity of information that such data can reveal when read in combination with other data, such as those on the IP addresses, only the fight against serious forms of crime, such as terrorist offenses, can justify such a severe interference with fundamental rights.[103] Contrary to the need to adopt meansures against a genuinely present and foreseeable threat to national security, the fight against crime cannot justify an illimited and indiscriminate retention of such data. Retention measures should therefore be limited in geographical terms or to a certain period of time, pursuant to objective and non-discriminatory criteria.[104] The real-time collection of data and automated analysis of data collected in police database must also respond to the same objective of combating a particularly serious crime, and both the measures should be carried out only when strictly necessary and only if surrounded by appropriate procedural guarantees.[105] The stance of the CJEU on data retention opted for a nuanced approach to bulk collection of personal data, which, it has been argued, is in line with similar judgements of the ECtHR. In fact, both the Courts focused their assessments on the level of infringement with private life, and its proportionality to the seriousness of the threat to be tackled with the measures in question.[106]

The CJEU case law provides some principles to come to final remarks about the use of social media data by LEAs. While the Court does not encompass an exhaustive list of criminal offenses to be categorised as serious, it recognised that the seriousness of crimes should play a decisive role in the assessment of necessity and proportionality. This point is crucial with regard to social media monitoring practices, as the sharing of content online could justify a systematic and indiscriminate processing of personal data only where a) at least an indirect link with a criminal offense can be inferred from the collection of data, and b) the criminal offense in question is serious enough to justify the interference with data protection. In the area of illegal content, it is important to consider each type of content differently, depending on the actual harm that can be caused by its dissemination. This assessment should be based on strong factual indicators that the spread of content is connected to the commission of a serious offense in the offline world, or to a direct harm to individuals online. Conversely, the prevention of minor offenses, or an excessively tenuous causal link between the content and a negative impact on society, should call for a lesser intrusion into private life. This consideration should be informed by the fact that, for example, the spread of false content leading to civil unrest often derives not by the dissemination of a single content, but from the systematic re-sharing of the same content over and over. While in this case it can be useful to monitor online environments to prevent exhalations and violent behaviours, the final result of disturbance of public order cannot be linked to a single piece of content online. Different reasonings can inform the stage of investigations, as finding a culprit of an actual offense is the underlying objective of social media monitoring in this scenario. On the other hand, in the case of crime prevention the causal link between online content and criminal behaviours offline is by nature weaker, but should not be forced, as not to result in a departure from any connection between the processing of personal data and an actual suspicion of criminality.[107]

[101] Joined Cases C-293/12 and C-594/12 Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources and Others (C-293/12) and Kärntner Landesregierung and Others (C-594/12) ECLI:EU:C:2014:238; C-203/15 Tele2 Sverige AB v Post –ochtelestyrelsen and Secretary of State for the Home Department v Tom Watson and Others ECLI:EU:C:2017:214 (Tele2 Sverige).

[102] Tele2 Sverige (n. 97), par. 198.

[103] Joined Cases C 511/18, C 512/18 and C 520/18 La Quadrature du Net and Others v Premier Ministre and Others ECLI :EU :C :2020 :791 (La Quadrature du Net), par. 156.

[104] La Quadrature du Net (n. 99), par. 168.

[105] La Quadrature du Net (n. 99), par. 172 and ff.

[106] Tzanou, Maria, Spyridoula Karyda. "Privacy international and quadrature du Net: One step forward two steps back in the data retention saga?" *European Public Law* 28.1 (2022), 123-154.

[107] Quattrocolo, Serena. "Hacking by Law-Enforcement: Investigating with the Help of Computational Models and AI Methods." *Artificial Intelligence,*

Ultimately, the CJEU case law provides a guide when assessing how to comply with the requirement of strict proportionality, as established under the LED for the processing of sensitive data.

6 Conclusions

A project report published by authors of the Cardiff School of Journalism, Media and Cultural Studies in 2015 outlined the results of a study conducted on the use of social media by the UK police in the context of radicalisation and extremism. From the interviews conducted with UK LEAs, the authors deducted that social media monitoring is used to analyse online activities prior to public events. The outcomes of the analyses are in turn used to adjust pre-emptive and real-time tactics in response to potential community tensions. This use was defined as "situational awareness", and included searches on social media according to key words in order to conduct risk assessments on such events. Sentiment analysis and geo-localisation where instead found not to be so common purposes in social media monitoring. The study outlined a need for more transparency in the use of social media by LEAs, including about the legal bases allowing for retention of data collected online, based on previous assessments on what may constitute a threat.[108]

Social media scraping and consequent retention of data by LEAs can undoubtedly have an impact on the exercise of freedom of expression. This type of practices can produce a chilling effect on individuals where not carried out transparently, and especially when a systematic monitoring activity leads to the processing, including retention, of sensitive data, such as those on political affiliations or religious beliefs. Moreover, it has been observed how, pursuant to the EU legislation and ECtHR and CJEU case law, publicly available data that are manifestly made available by individuals do not fall, for this, outside of the scope of data protection.

The rise of politically motivated crimes across Europe, however, shows that the collection of sensitive data from social media can be based on legitimate objectives of public interests.[109] The issue of striking a balance between the protection of publicly available data collected from social media content and the need to prevent or fight such crimes is not susceptible to be solved with easy solutions.

However, two considerations can be made to spark further discussions about this matter. First of all, it would be of the outmost important that social media scraping technologies are used according to clear and precise rules about when such use is lawful. The requirement of the quality of the law as enshrined in Article 52 of the EU Charter of Fundamental rights should be interpreted in this case as requiring a clarification about the circumstances and conditions justifying the processing of social media data, so to avoid arbitrary decisions in the context of law enforcement. Such a clarification should take into account the need to consider the seriousness of the criminal offense that LEAs are trying to tackle while monitoring online environments. As interferences with fundamental rights can only be acceptable when necessary and proportionate to the objective of public interest pursued by public authorities, social media scraping practices should be put in place in a way that allows to adapt their intrusiveness to the gravity of the threats at stake. While this conclusion stems from the supranational principles guiding the legislative action in this realm, it should also be noted that the legislative efforts in applying the necessity and proportionality principles in this realm should be aware of the pivotal role played by the technologies used to implement the laws. In this sense, while the legislators should clarify when social media scraping is allowed in the context of LEAs' activities, it is also important to adapt the use of technological means in a way that allows such principles to be concretely respected. Depending on the goals pursued by social media monitoring activities, and the seriousness underlying seriousness of the threats to be tackled, such technologies should be designed in order to allow for less intrusive forms of personal data processing when such a processing is not strictly needed for the purposes in question. A careful assessment on how these technologies should be shaped in order to respect privacy and data protection principles is therefore of the outmost importance, and should be the object of further, interdisciplinary academic efforts.

Strictly connected to the former point is the need to define which publicly available data can be scraped from social media, and to outline criteria on their analysis and retention based on the seriousness of the potential criminal offenses and on the level of suspicion raised by the content published online, or related online activities. These rules are especially needed in light of the

Computational Modelling and Criminal Proceedings: A Framework for A European Legal Discussion, Springer, 2020, 37-71.

[108] Dencik, Lina, et al. "Managing 'threats': uses of social media for policing domestic extremism and disorder in the UK." *Media Democracy Fund, Ford Foundation and Open Society Foundations*, 2015. https://orca.cardiff.ac.uk/id/eprint/85618/1/Managing-Threats-Project-Report.pdf.

[109] Glöckner, Paul, Stuchtey, Tim, "Turning sensitive data into knowledge. The need for a common understanding of politically motivated crimes in Europe", *FERMI – Fake News Risk Mitigator*, 2023. https://fighting-fake-news.eu/articles.

provisions of the TERREG and the DSA, and in particular the existence of the IRUs, and the institutionalised status of trusted flaggers. It has already been noted how the IRUs' competences have been expanded over the years. In the case of the trusted flaggers, the underlying assumption in the DSA that what is illegal offline should also be illegal online leaves open the question of how much data should actually be analysed or retained from publicly available sources, depending on the abovementioned criteria. This is in light of the fact that both the IRUs and the LEAs acting as trusted flaggers can potentially act pursuant to the terms and conditions of online platforms, and not actual legal basis, when issuing their removal orders, and can do it without a judicial or otherwise independent scrutiny being necessary, pursuant to the provisions regulating them. This interaction between public authorities and the privately enforced framework established by online platforms results in legal uncertainty, that can have a negative impact not only on the rights to privacy and data protection, but also to freedom of expression.

More precise rules about social media monitoring, to be based on the seriousness of crimes and able to narrow down the categories of publicly available data to be processed and the types of processing to be carried out in different scenarios, would ultimately lead to enhanced legal certainty. In the context of social media scraping activities, content moderation, and LEAs' involvement in monitoring online environments, this could avoiding, or at least limit, detrimental effects to the enjoyment of fundamental rights.

Conclusively, the current legislative evolutions on content moderation should be accompanied by a reflection on the possible effects on the criminal justice systems, and legislators should take accountability to avoid that the legitimate objective to tackle illegal or harmful content online degenerates in practices of mass surveillance that, in the era of social media, are easier to put in place than ever.

7 Bibliography

P. S. Bayerl, et al.,*Future Challenges and Requirements for Open Source Intelligence in Law Enforcement Investigations: Results from Horizon Scanning Exercise,* in "European Law Enforcement Research Bulletin" 21 (2022).

R. Bellanova, M. De Goede, *Co-Producing Security: Platform Content Moderation and European Security Integration*, "JCMS: journal of common market studies", 60.5 (2022).

C. Berthélémy, *EU Terrorist Content Online Regulation Could Curtail Freedom of Expression across Europe*, "EDRi*"*, 2021.

E. Birritteri, *La disinformazione tra politica e diritto. Dimensione istituzionale, strategie preventive e dinamiche punitive,* in "Diritto Penale Contemporaneo", 4 (2021).

B. Chang, *From Internet Referral Units to International Agreements; Censorship of the Internet by the UK and EU*, "Columbia Human Rights Law Review", 49 (2017).

P. Coe, *The social media paradox: an intersection with freedom of expression and the criminal law*, in "Information & Communications Technology Law", 24.1 (2015).

L. Dencik, et al., *Managing 'threats': uses of social media for policing domestic extremism and disorder in the UK,* Media Democracy Fund, Ford Foundation and Open Society Foundations, 2015.

N. R. de Oliveira, et al. *Identifying fake news on social networks based on natural language processing: trends and challenges*, in "Information"*,* 12.38 (2021).

L. Edwards, L. Urquhart, *Privacy in Public Spaces: What Expectations of Privacy Do We Have in Social Media Intelligence?* In "International Journal of Law and Information Technology", 24.3 (2016).

P. Glöckner, T. Stuchtey *Turning sensitive data into knowledge. The need for a common understanding of politically motivated crimes in Europe*, FERMI – Fake News Risk Mitigator, 2023.

R. Gorwa, R. Bonns, C. Katzenback, *Algorithmic content moderation: Technical and political challenges in the automation of platform governance,* in Big Data & Society, 7.1 (2020).

T. Gottschalk, *The Data-Laundromat? Public-Private-Partnerships and Publicly Available Data in the Area of Law Enforcement*, in "European Data Protection Law Review", 6 (2020).

S. L. Granizo, et al., *Detection of possible illicit messages using natural language processing and computer vision on twitter and linked websites*, in "IEEE Access", 8 (2020).

C. Jasserand, *Article 10. Processing of Special Categories of Personal Data*, in *LED Commentary, OUP (forthcoming Fall 2023).*

M. A. Khder, *Web scraping or Web crawling. State of Art, Techniques, Approaches and Application,* in "International Journal of Advances in Soft Computing & Its Applications", 13.3 (2021).

J. Kilpatrick, C. Jones, *Empowering the police, removing protections: the new Europol Regulation*, *Statewatch*, 2022.

K. Komaitis, et al., *Enforcement Overreach Could Turn Out To Be A Real Problem in the EU's Digital Services Act*, Electronic Frontiers Foundation, 2022.

V. Krotov, S. Leiser, *Legality and ethics of web scraping*, *Twenty-fourth Americas Conference on Information Systems, New Orleans* (2018).

G. Lim, S. Bradshaw, *Chilling Legislation: Tracking the Impact of "Fake News" Laws on Press Freedom Internationally*, Center for International Media Assistance, 2023.

T. K. Mackey, et al., *Big data, natural language processing, and deep learning to detect and characterize illicit COVID-19 product sales: infoveillance study on Twitter and Instagram*, in "JMIR Public Health and Surveillance", 6.3 (2020).

L. Mitrou, et al., *Social media profiling: A Panopticon or Omniopticon tool? Proceedings of the 6th Conference of the Surveillance Studies Network*, 2014.

R. Ó Fathaigh, et al., *The perils of legally defining disinformation,* in "Internet policy review", 10.4 (2021): 2022-40.

J. Pielemeier, *Disentangling disinformation: What makes regulating disinformation so difficult?* In "Utah Law Review", 4 (2020).

T. Ploug, *The right not to be subjected to AI profiling based on publicly available data—privacy and the exceptionalism of ai profiling*, in "Philosophy & Technology", 36.1 (2023).

F. Poletto, et al., *Resources and benchmark corpora for hate speech detection: a systematic review*, in "Language Resources and Evaluation"55 (2021).

S. Quattrocolo, *Hacking by Law-Enforcement: Investigating with the Help of Computational Models and AI Methods,* in *Artificial Intelligence, Computational Modelling and Criminal Proceedings: A Framework for A European Legal Discussion*, Springer, 2020.

P. M. Sabella, *Il fenomeno del cybercrime nello spazio giuridico contemporaneo. Prevenzione e repressione degli illeciti penali connessi all'utilizzo di internet per fini di terrorismo, tra esigenze di sicurezza e rispetto dei diritti fondamentali,* in "Informatica e diritto", 26.1-2 (2017).

B. Sander, *Freedom of expression in the age of online platforms: the promise and pitfalls of human rights-based approach to content moderation,* in "Fordham International Law Journal", 43.4 (2020).

R. Sandoval-Almazan, G. R. Gil-Garcia, *Towards cyberactivism 2.0? Understanding the use of social media and other information technologies for political activism and social movements,* in "Government Information Quarterly", 31.3 (2014).

J. D. Scott, *Social media and government surveillance: The case for better privacy protections for our newest public space*, in "Journal of Business and Technology Law", 12.2 (2017).

A. Shere, *Reading the Investigators their Rights: A review of literature on the General Data Protection Regulation and open-source intelligence gathering and analysis,* in "The New Collection", 3, Oxford, 2020.

R. D. Steele, *Open source intelligence*, *Handbook of intelligence studies,* 42.5, London, Routledge, 2007.

D. Susser, *Predictive policing and the ethics of pre-emption*, in B. Jones (Ed.) *The ethics of policing: New perspectives on law enforcement,* NYU Press, 2021.

J. Torregrosa, et al., *A survey on extremism analysis using natural language processing: definitions, literature review, trends and challenges*, in "Journal of Ambient Intelligence and Humanized Computing", 14.8 (2023).

D. Trottier, *Open source intelligence, social media and law enforcement: Visions, constraints and critiques,* in "European Journal of Cultural Studies", *18*.4-5 (2015).

M. Tzanou, and K. Spyridoula K., *Privacy international and quadrature du Net: One step forward two steps back in the data retention saga?* In "European Public Law", 28.1 (2022).

S. Vanden Broucke, B. Baesens, *From web scraping to web crawling,* in *Practical Web Scraping for Data Science. Best practices and examples with Python*, Apress, Berkeley, 2018.

J. van Hoboken, and R. Ó. Fathaigh, *Regulating Disinformation in Europe: Implications for Speech and Privacy,* in "UC Irvine Journal of International, Transnational and Comparative Law", 6 (2021).