Reconsidering the data protection framework for use of publicly available personal data

1.1. Introduction

In early 2020, reports[1] of a new facial recognition tool from a private organization, that was trained using publicly available images from social networking sites started surfacing. In 2015 a dataset was created using photos of individuals that had the Creative Commons license for training facial recognition algorithms that could be downloaded off the internet[2]. The common thread through these incidents is that entities used personal data disclosed by individuals in one context in a different one without notifying such use to the individuals who disclosed it. From the perspective of protecting the privacy of the individual, the primary question is whether an individual's expectation of privacy diminishes once they have disclosed data in a specific context even if further use of such data is in a different context. Since, there seems to be a general understanding that the data in public sphere has diminished protections related to its further use, it is necessary to determine the characteristics of a digital platform that render any personal data posted on it as part of the public sphere. These characteristics are important in determining if an individual loses the rights associated with their personal data once it is considered to be in the public sphere.

With the rapid change in digital technologies and introduction of new products associated with such technologies there seems to be an ever-growing consensus that the impact of such products on privacy in public needs to be examined. For example, Google Glasses' clandestine recording capabilities lead to organizations creating "Glass free zones" to protect the privacy of their clients in what can be termed as traditionally publicly accessible spaces[3]. The impact of the ubiquitous surveillance of individuals in public places as a result of advanced facial recognition systems have been documented to such an extent that many have called for a ban of use of such systems[4]. Several countries[5] have rules regarding usage of drones in public spaces which include not photographing an individual without their

[1] Kashmir Hill, The Secretive Company that might end privacy as we know it, *The New York Times,* 18 January 2020 https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html accessed 3 October 2021

[2] Kashmir Hill and Aaron Krolik, How Photos of Your Kids Are Powering Surveillance Technology, *The New York Times,* 11 October 2019, https://www.nytimes.com/interactive/2019/10/11/technology/flickr-facial-recognition.html accessed 3 October 2021

[3] Olya Kudina and Melis Bas, '"The end of privacy as we know it": Reconsidering public space in the age of Google Glass', in Bryce Clayton Newell, Tjerk Timan and Bert- Jaap Koops(eds), *Surveillance, Privacy and Public Space* (Routledge 2018)

[4] Tambiama Madiega and Hendrik Mildebrath, 'Regulating facial recognition in EU' (European Parliamentary Research Service, 16 September 2021) <https://epthinktank.eu/2021/09/16/regulating-facial-recognition-in-the-eu/ >accessed 25 December 2021
'An Act To Increase Privacy and Security by Regulating the Use of Facial Surveillance Systems by Departments, Public Employees and Public Officials', State of Maine
Gregory Barber, ' San Franscico bans agency use of facial recognition tech' (Wired, 14 May 2019) < https://www.wired.com/story/san-francisco-bans-use-facial-recognition-tech/> accessed 25 December 2021

[5] Freedom from Drone Surveillance Act 725 ILCS 1679 (Illinois); Criminal Code Section 934.50 (Florida); Drone Rules, 2021 (India); Drones (UAS), European Union Aviation Safety Agency

consent[6]. These examples relate to privacy invasive measures using digital technologies that lead to constant surveillance of individuals in physical public spaces. However, such constant surveillance also extends to digital spaces such as social media platforms.

Facial recognition algorithms have been developed by scraping millions of images on social media that have been considered publicly available[7]. EU data protection authorities[8] have disapproved of the invasive data collection practices of credit rating agencies and data brokers dealing with direct marketing data where personal data of individuals was being compiled from their social media profiles. Hence, it is essential that the importance of privacy in physical public spaces is extended to digital public spaces as well. The interconnected and interoperable nature of digital platforms combined with the ability to create vastly detailed sensitive profiles[9] of individuals through data aggregation techniques makes defining what is public for the digital space rather complicated.

There are multiple legal methods to safeguard the privacy of the individual. EU's General Data Protection Regulation(Regulation) has been considered as a model standard of data protection legislations and encompasses the internationally accepted data protection principles in its provisions. The Regulation provides for setting up of independent data protection authorities with wide investigative and enforcement powers resulting in a body of guidance documents and enforcement actions specifying the details regarding implementation of the provisions of the Regulation. Despite the absence of a federal data protection legislation, the judiciary in the United States of America and the Federal Trade Commission have set certain precedents with respect to protecting the privacy of the individuals under their Fourth Amendment provisions. Many of the incidents of use of publicly available personal data have been litigated in both the US[10] and the EU[11] providing for a baseline understanding of the protection currently being offered to publicly available personal data. Both these countries take different approaches to protecting the personal data

[6] 'Rules on recreational use of drones', Government of Netherlands <https://www.government.nl/topics/drone/rules-pertaining-to-recreational-use-of-drones> last accessed 26th December 2021

[7] n(1); n(2)

[8] Information Commissioner's Office, "Investigation into data protection compliance in the direct marketing data broker industry" (*Information Commissioner's Office*, October 2020), < https://ico.org.uk/media/action-weve-taken/2618470/investigation-into-data-protection-compliance-in-the-direct-marketing-data-broking-sector.pdf> last accessed 26 December 2021;
European Data Protection Supervisor, " Opinion 11/2021 on the Proposal for a Directive on consumer credits ", (*European Data Protection Supervisor*, 26 August 2021)< https://edps.europa.eu/system/files/2021-08/opinion_consumercredit-final_en.pdf> last accessed 27 December 2021

[9] Sara Geoghegan and Dana Khabbaz, ' Reproductive Privacy in the Age of Surveillance Capitalism' (*Electronic Privacy Information Centre,* 7 July 2022) <https://epic.org/reproductive-privacy-in-the-age-of-surveillance-capitalism/> accessed 1 August 2022
Kristin Cohen, ' Location, health, and other sensitive information: FTC committed to fully enforcing the law against illegal use and sharing of highly sensitive data' (*Federal Trade Commission*, 11 July 2022) <https://www.ftc.gov/business-guidance/blog/2022/07/location-health-other-sensitive-information-ftc-committed-fully-enforcing-law-against-illegal-use> accessed 1 August 2022

[10] ACLU v Clearview https://www.aclu.org/cases/aclu-v-clearview-ai

[11] Ian Carlos Campbell, 'Clearview AI hit with sweeping legal complaints over controversial face scraping in Europe', (*Verge,* 27 March 2021) https://www.theverge.com/2021/5/27/22455446/clearview-ai-legal-privacy-complaint-privacy-international-facial-recognition-eu

of individuals. This thesis will examine the two dominant methods narrowed down based on their adoption in legislative instruments and jurisprudence of courts i.e. privacy as control and privacy as social norms.

Privacy as control: Is the notice and consent model the answer?

In the case of digital spaces, it can be argued that once an individual discloses their personal data on a specific platform, they lose control over further disclosure of such data and by extension they lose control over protecting their personal data. The notice and consent model has been considered as a method to enable an individual's control over their personal data. According to this model, the entity processing personal data is required to provide the individual with details of the processing operations to the individual in clear and concise manner. Based on the information provided, the individual has the choice to consent to the processing operation and provide the relevant personal data. EU's GDPR is an example of such a model.

EU relies on the GDPR among other legislations[12] to protect the informational privacy of the individuals. When using personal data in a context that is different from its initial disclosure[13] and that is not directly obtained from the data subject[14], entities are required to provide details of the processing operations to the data subject. Theoretically, this does give the individual the ability to control the use of their personal data even after the initial disclosure as they are required to be notified of additional details of the processing operations. However, it might not have a noticeable impact as the data to be processed has been already disclosed i.e., it is not possible for the individual to know for certain that their personal data is being processed by a non-related entity till the time the data protection authorities fine them for non-compliance with the notice obligations. If the individual does receive a notice from an entity processing their personal data that has already been disclosed, the pitfalls of the notice and consent model[15] continue to exist. Since the personal data in these cases exists in the open, the harms arising from non-consensual use and improper notices for complex data processing operations is graver than when the data has to be directly obtained from the individual.

[12] 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; Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data

[13] Article 14(4)

[14] Article 14

[15] Mikella Hurley & Julius Adebayo, 'Credit Scoring In The Era Of Big Data', (2017) 18 Yale J.L. & Tech https://digitalcommons.law.yale.edu/yjolt/vol18/iss1/5 accessed 3 October 2021 ; Ananny, Mike and Kate Crawford, 'Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability' (2018) 20 New Media & Society 973 – 989 < https://journals.sagepub.com/doi/10.1177/1461444816676645>

Privacy as social norms: Is reliance on social norms as the basis for determining the expectation of privacy of the individual the answer?

Social norms as a basis for determining privacy of the individuals has been referred[16] to in the fourth amendment jurisprudence of the US. Fourth amendment is related to search and surveillance capabilities of law enforcement authorities in the course of investigation of a criminal activity. Courts have examined the legality of searches and surveillance in the context of the reasonable expectation of privacy of the individual. This reasonable expectation of the individual is evaluated based on the social norms prevalent in relation to the activity in question.

For the purposes of this thesis, social norms refer to the norms and practices that have been accepted by the society at large. The problem with relying on social norms as the primary justification for recognizing privacy rights are twofold: social norms don't necessarily have to be ethically and morally right[17] . The rate at which digital technologies develop an action cannot be postponed for the social norm regarding their acceptability to stabilise.[18] The recent calls for banning facial recognition systems are the perfect example. They were seen as an acceptable measure to ensure public security. However, the harms resulting from biased input data, false positives etc. have called into question their earlier acceptability.

Neither of these two interpretations of privacy can be successfully applied to grant sufficient protection to publicly available personal data. The contextual integrity framework rejects the notion of privacy as control and instead defines privacy in terms of expectations regarding the appropriate flow of information.

Privacy as Contextual integrity

Helen Nissenbaum, in her theory on contextual integrity(CI)[19] explains the importance of context in relation to privacy and information of the individual. She suggests that every area of life is governed by the norms of appropriate flow of information which are mostly set by the relevant social norms. By doing this, the theory moves past the strict dichotomy between private and public sphere and highlights that each context is governed by a set of its own social norms. A dating application, for instance, can have myriad data types ranging from location to sexual orientation to health-related data (details of sexually transmitted diseases, allergies etc.) The framework suggests that all these different data types on one single platform will have different context specific norms because the privacy constraints that

[16] California v Greenwood 486 U.S. 35; Rakas v. Illinois, 439 U.S. 128; Georgia v. Randolph, 547 U.S. 103 ; City of Ontario v. Quon, 560 U.S. 746

[17] Matther J Tokson, Ari Ezra Waldman, "Social Norms in Fourth Amendment Law" (2021). Michigan Law Review, Vol. 120, https://ssrn.com/abstract=3767261 accessed 11 December 2021

[18] Bert Jaap Koops, "Privacy Spaces" (2018). 121 West Virginia Law Review
611 https://ssrn.com/abstract=3157169 accessed 11 December 2021

[19] Helen Nissenbaum, "Privacy as contextual integrity." (2004) 79 Washington Law Review
https://crypto.stanford.edu/portia/papers/RevnissenbaumDTP31.pdf accessed 3 October 2021

people hold over information is related to the characteristics of the background situation[20]. By focusing on the norms surrounding a specific context of information gathering and disclosure of data, the framework points towards the fluid dimensions of a digital public space. It is this regard for context of disclosure and respect for the norms of appropriate flow of information that need to be better translated into data protection legislation to safeguard rights of individuals related to publicly available personal data.

US and EU approach data protection in different ways. The lack of a federal data protection legislation in the former results in requiring a subjective assessment of what amounts to an expectation of privacy that is accepted by the society. The latter places heavy reliance on an individual's ability to comprehend complex documents on details of the processing operation to safeguard their right to personal data protection. An in depth analysis of the safeguards offered in the US and safeguards offered in the EU have to be compared to the position advocated by the contextual integrity framework to propose additional safeguards for publicly available personal data in the digital space.

The objective of this thesis is to examine the existing safeguards offered to publicly available personal data in the US and EU and compare them to the contextual integrity framework. The thesis will then analyse methods to incorporate the moral and political terms of the contextual integrity framework to the data protection frameworks in the jurisdictions identified.

1.2. Methodology

The thesis will rely on comparative doctrinal legal research. The primary focus of the doctrinal research will be of isolating the legal protections offered to publicly available personal data in the US and EU. To arrive at a conclusive list of the extent of protections offered, the fourth amendment jurisprudence related to identifying the reasonable expectations of privacy of the individual from the US and the notice and consent model relied on by the GDPR and the associated Article 29 working party and EDPB guidelines from the EU will be analysed. An analysis of case law related to web scraping in both jurisdictions will be undertaken as it is one of the processes through which publicly available personal data is collected.

The adequacy of these protections will be evaluated against the CI framework to examine if norms surrounding contextual disclosure have been adequately translated into the legal frameworks. Theories that highlight the importance of the right to privacy of the individual will be referred to, to contextualize the need for legal reform of protections offered to publicly available personal data on digital platforms. By examining the baseline protections offered by each of the legal frameworks, measures to incorporate the norms of information flow of the CI framework will be suggested.

[20] Helen Nissenbaum, *Privacy in Context: Technology, Policy and the Integrity of social life(*2009, Stanford University Press) 331

EU and US have been identified as in scope for research as both these countries represent different approaches to privacy protection i.e., privacy as control and as social norms. It is acknowledged that both countries have other legislations and policies that might have provisions dealing with the subject matter. However, the scope has been narrowed down to fourth amendment jurisprudence from the US and the GDPR in the EU as these largely reflect the attitudes of the regulator and legislator towards protections offered to PPD.

2. Publicly available personal data in the United States and European Union

For US, this section will identify the meaning of publicly available personal data through caselaw related to the fourth amendment of the Constitution by focusing on the analysis of the reasonable expectation of privacy test by courts. For the EU, this section will isolate the meaning of publicly available personal data by relying on the provisions of the GDPR and the guidance issued by the European Data Protection Board.

2.1. Publicly available personal data in the United States

The fourth amendment protects US citizens from unlawful search and seizure by law enforcement authorities[21]. A lawful search requires a search warrant which is authorised based on probable cause. The courts while examining the legality of a search, in the absence of a search warrant, in fourth amendment cases deliberate on two issues[22]: was there a search and was the search reasonable? It is in determining whether the search was reasonable that reasonable expectation of privacy of an individual plays a role. The cases that involve determination of what amounts to a reasonable expectation of privacy do discuss the accessibility of the information as one of the criteria for such an expectation. Even though the fourth amendment is relied on specifically in the context of law enforcement searches, the courts acceptance of reasonable expectation of privacy in such cases provides us with an understanding of what is deemed to be an acceptable expectation of privacy within US legal system. This acceptable level of privacy needs to be analysed to determine if it is efficient in providing the necessary protections to personal data that is publicly available even outside the context of law enforcement.

2.1.1 Reasonable expectation of privacy test

The concurring opinion of Justice Harlan in Katz [23] set forth the two strands of the reasonable expectation of privacy test: Has the person exhibited actual expectation of privacy? Is the society prepared to recognize this expectation as reasonable? The cases that relied on this test consider the accessibility of the information concerned in determining the reasonableness of the expectation of the individual to maintain their privacy.

[21] U.S. Const. amend. IV

[22] Andrew D. Selbst, 'Contextual Expectations of Privacy' (2013). 35 Cardozo Law Review 643 (2013), <https://ssrn.com/abstract=2093594 > last accessed 20 November 2021

[23] Id.

Accessibility of information as a precursor to validate reasonable expectation of privacy

The Supreme Court in *Maryland*[24] ruled that the numbers dialed from a telephone call were not subject to the protections offered by fourth amendment as the number dialed is exposed to the phone company so that the phone company can complete the call. Similarly, the court in *Miller*[25] held that the drawer of a check couldn't have a reasonable expectation of privacy in the information contained on the check as that information was disclosed to many third parties as a result of the banking system. When this reasoning is extracted to fit into the personal data protection sphere, it would mean that disclosure in one context gives the receiver of that information free reign to use it in different contexts.

The Court in *Ciraolo*[26] held that observation of a private property that is shielded from public street but clearly visible to an aircraft flying in public air space is not in violation of fourth amendment in the absence of a warrant. However, when law enforcement used a heat sensor to detect marijuana cultivation inside a closed garage the court held that the use of a device that wasn't in general public use is unreasonable and in violation of the reasonable expectation of privacy of the individual[27]. In *Ciraolo*, the private field was in plain view for aircrafts flying above whereas in *Kyllo* the information was detected only due to the use of the heat sensor and was not in plain view.

Based on the jurisprudence so far, the acceptability of the reasonable expectation of privacy seems to depend on the accessibility of the information concerned and the reasonability of the tools required to access said information. The use of an aircraft for aerial surveillance of an open field was justified while the use of heat sensor to map heat signatures within a private garage wasn't. The former was in plain view and the individual concerned didn't exhibit an actual expectation of privacy by putting a shed over the field for example. In the latter, it was a closed garage and that in itself exhibits an actual expectation of privacy and the use of heat sensor violates such an expectation since it wasn't foreseen by the individual.

In *Miller* and *Maryland,* we see that accessibility of information is provided a rather wide interpretation. In both cases, the fact that specific information was disclosed to banks and phone company systems, respectively, was seen as sufficient reason to not warrant any expectation of privacy over further use. This interpretation doesn't consider the importance of contextual disclosure of data. In both the cases above, it can be argued that by disclosing personal data to only specific group the individuals were exhibiting a reasonable expectation of privacy. It is also important to note that in both these cases the information had to be shared by individuals as part of the services that were being provided by the banks and phone companies.

[24] Smith v. Maryland, 442 U.S. 735
[25] United States v Miller, 425 U.S. 435 (1976)
[26] California v Ciraolo 476 U.S. 207
[27] Kyllo v. United States, 533 U.S. 27

With respect to personal data shared on social media, the court found that hypothetical accessibility of the individual's social media page meant that the said information was public[28]. Similarly, it has been argued[29] that there aren't any reasonable expectations of privacy in the context of SOCMINT collection as users are expected to know from the terms and conditions that the data disclosed may be shared with others.

Based on the caselaw, what is publicly available personal data is dependent on the accessibility of the information concerned to the third party and if extraordinary tools need to be in use to access that information. This essentially means that personal data once revealed can be considered to be publicly available personal data without regard for the expectations of the individual related to the context of disclosure. This also puts an unreasonable burden on the individual to be aware of the state-of-the-art technologies that entities could use to collect their personal data and then safeguard themselves against the same.

2.2 Publicly available personal data in the European Union

The GDPR is the primary data protection legislation in the EU. It permits the European Data Protection Board(EDPB) and the data protection authorities to publish guidance and recommendation that aid entities in achieving better compliance with its provisions. Reference to personal data that can be considered public can be found in these sources. The analysis in this section does not refer to personal data processing by law enforcement authorities as there is sufficient reference to publicly available personal data in the Regulation that applies to commercial data controllers and processors. GDPR does not define publicly available personal data. However, there are two references to personal data that could be considered public for the purposes of our analysis.

Article 14 requires a data controller to provide a notice containing the relevant details of the processing operation to the data subject in cases where the data is not directly obtained from them. There is no explicit reference to public in this provision however, the fact that data controllers are obliged to provide a notice to the data subject even in cases where they are not the primary source of the said information seems to indicate that the level of protection offered to this category is similar to the one that is directly obtained from the individual.

The second reference to data made public is with reference to lawful grounds of processing special categories of personal data. Article 9 prohibits processing of special categories of personal data unless the specified exceptions apply. One of the exception is for personal data that has been "manifestly made public by the data subject." The definition of this phrase has not been provided in the Regulation. However, the EDPB in its guidance[30] on targeting social

[28] Sandler v Calcagini 2008 U.S. Dist. LEXIS 54374

[29] J Bartlett and L Reynolds, 'The state of the art 2015: a literature review of social media intelligence capabilities for counter- terrorism' (*Demos*, September 2015) < https://www.demos.co.uk/wp-content/uploads/2015/09/State_of_the_Arts_2015.pdf> last accessed 2 August 2022

[30] European Data Protection Board, 'Guidelines 08/2020 on the targeting of social media users '(*European Data Protection Board,* 13 April 2021) < https://edpb.europa.eu/system/files/2021-

media users' data expands on what can be considered as manifestly made public for the purposes of the Regulation. Such a determination should include an analysis of the default private settings of the data subject, the nature of the social media platform, information provided to the data subject regarding the public nature of the information disclosed.

3. Existing safeguards for publicly available personal data in the US and EU

The previous section laid out the characteristics of personal data to be considered publicly available in the US and EU. In the US, the reasonable expectation of privacy standard used to determine legality of law enforcement actions under the fourth amendment was considered to conclude that personal data that can be accessible relatively easier without the use of complex tools to access such data will be considered public. In the EU, despite the lack of a definition for publicly available personal data in the GDPR, the guidance issued by the EDPB[31] also takes into consideration the accessibility of the personal data concerned to determine which special categories of data are "manifestly made public" for the purposes of article 9 of GDPR. It is important to note that regardless of a finding of publicly available personal data, GDPR does not exempt application of the provisions of the Regulation for such data. With a clear understanding of what is publicly available personal data and the resultant implications of such findings on the application of data protection legislations, this section will outline the safeguards that are applicable to the personal data concerned. This analysis will be based on the legislations identified in the previous chapter as well as the jurisprudence of the relevant courts. This analysis will be conducted based on the data collection strategy of web scraping as it is one of the most relied on method to collect publicly available personal data from digital platforms.

3.1 United States of America and web scraping

One of the most prominent cases on web scraping in recent times is that of *HiQ v. LinkedIn*[32]. HiQ is a data analytics company that scraped information that was disclosed by LinkedIn users on LinkedIn's platform without the authorization of the users as well as LinkedIn. The information collected was used to help employers in identifying employees at risk of being recruited by other firms. HiQ's action was challenged by LinkedIn under the Computer Fraud and Abuse Act (CFAA), which is a cybersecurity legislation that was enacted to curtail hacking. The ninth circuit, while recognizing the privacy interests of the users, held that HiQ's actions weren't in violation of the terms of the CFAA as the information that was obtained wasn't password protected by the users or the platform. While examining LinkedIn's assertion of their responsibility to respect the privacy of their users, the court argued that there isn't evidence to conclude that users maintained an expectation of privacy over the information they disclosed in their public profiles. It was concluded that in the off chance that some users retain an expectation of privacy the same cannot be significant enough to outweigh HiQ's business interest in accessing, analysing and communication information derived from public LinkedIn profiles. The conclusion that disclosure of information by the user as part of their public profile implies lack of further

04/edpb_guidelines_082020_on_the_targeting_of_social_media_users_en.pdf> last accessed 2 August 2022

[31] id

[32] HiQ Labs, Inc. v. LinkedIn Corp., 938 F.3d 985 (9th Cir. 2019)

expectation of privacy is in line with the conclusion of the courts in *Maryland, Miller* and other cases referred to in section 2.

In *United States v Chavez*[33], legality of the law enforcement's actions in searching the defendant's facebook account under the fourth amendment was examined. While evaluating the reasonableness of the expectation of privacy of the defendant, the court had to verify if they intentionally took steps to avoid access of their data to the public at large[34]. It was found that the defendant took active steps to exclude the public from accessing select content on their Facebook profile. The court likened such restriction to sealed packages and private telephone calls i.e., the individual expects privacy in non-public content. The court disregarded the government's argument that despite active steps to restrict the information disclosure to the public, the defendant still shared it with "hundreds" of Facebook friends. It was held that the defendant's legitimate expectation of privacy is protected by the fourth amendment.[35]

This case can be distinguished from that of *hiQ* wherein the court ruled excessively on the fact that the information that was scraped was part of the public profile of the individual. It is acknowledged that the court deciding the case in hiQ relied on the CFAA and not the reasonable expectation of privacy standard relied on by the court in *Chavez*. However, the objective of both the cases was the same i.e., to analyse the subjective expectations of privacy of the individuals in question. In the former we see that a user's public profile is devoid of any additional safeguards regarding further use. In the latter, we see that if there is an intentional action by the individual to restrict access to their data on a specific platform, it will be regarded as a legitimate expectation of privacy at least for cases under fourth amendment. Hence, it can be argued that safeguards against web scraping will be acknowledged by the court if the social media profile is at least partly restricted on the platform.

Due to the nature of privacy enforcement in the United States, it is difficult to come up with an exact list of safeguards offered to publicly available personal data. The only safeguard they seem to have for data that is partially restricted on social media is that law enforcement authorities will need a warrant to ensure legality of search under the fourth amendment. However, the legality of web scraping of such partially restricted content is unclear.

3.2. European Union (& UK) and web scraping

In stark contrast to the case law in the US, enforcement actions by the data protection authorities in the EU and UK against Clearview AI have provided for rich literature on the applicability of GDPR and the safeguards awarded to personal data of individuals in case of web scraping. Data protection authorities [36] have found Clearview's data collection method of web scraping to gather

[33] 423 F. Supp. 3d 194 (W.D.N.C. 2019)

[34] Para 202

[35] Para 205

[36] Information Commissioner's Office, 'ICO issues provision view to fine Clearview AI, (*Information Commissioner's Office*, 29 November 2021) < https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2021/11/ico-issues-provisional-view-to-fine-clearview-ai-inc-over-17-million/> last accessed 15 February 2022

publicly available information online to be unlawful and in violation of GDPR, specifically provisions related to lawful ground of processing personal data, data storage limitation and the information obligations[37]. They have also issued guidance[38] on web scraping and re-use of publicly available online data for direct marketing wherein they reiterate the importance of complying with the data protection requirements related to obtaining lawful consent, providing notice to data subjects, data minimization etc. Despite the reiteration of the applicability of data protection requirements to web scraping activity, the below sections will argue that complying with such requirements is either difficult or not sufficient to enable adequate safeguards for publicly available personal data.

For initiating any valid processing operation, the entity must identify the lawful grounds of processing[39] they intend to rely on. Based on the activity, in the current case of web scraping entities will need to examine if they intend to rely on the lawful basis of consent[40] or that of legitimate interest of the controller[41].

3.2.1 Lawful basis of consent

GDPR defines[42] consent to mean any "freely given, specific, informed and unambiguous indication of the data subject's wishes" to agree to the processing of personal data. The EDPB issued guidelines[43] that interpret these terms to aid entities in complying with relevant requirements related to consent. It has been argued[44] that that idea of informed consent is to ensure that the individual stays in control of their personal data. This opportunity to exercise control must be provided by the entities processing personal data by using clear and plain language to explain the processing operations to the data subject.[45] Consent will not be a valid legal ground in the absence of real choice i.e. if the data subject is compelled to provide consent due to any negative repercussions in the absence of such consent.[46] The EDPB acknowledges consent fatigue and verbosity of privacy policies and provides ways to address the same by

Hamburg Commissioner for data protection and freedom of information, 'Consultation prior to an order pursuant to article 58(2)(g) GDPR', (*NOYB,* 27 January 2021) https://noyb.eu/sites/default/files/2021-01/545_2020_Anhörung_CVAI_ENG_Redacted.PDF last accessed 15 February 2022

[37] Article 12, 13 and 14

[38] Hunton Andrews Kurth, 'CNIL publishes guidance on web scraping and reuse of publicly available online data for direct marketing' (*Hunton Privacy Blog,* 4 May 2020) < https://www.huntonprivacyblog.com/2020/05/04/cnil-publishes-guidance-on-web-scraping-and-re-use-of-publicly-available-online-data-for-direct-marketing/> (original source in French) last accessed 15 February 2022

[39] Article 6

[40] Article 6(1)(a)

[41] Article 6(1)(f)

[42] Article 4(11)

[43] European Data Protection Board, 'Guidelines 05/2020 on consent under Regulation 2016/679' (*European Data Protection Board,* 4 May 2020) <https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_en.pdf> last accessed 15 February 2022

[44] id

[45] id

[46] id

providing for options in the browser settings[47]. If we assume that privacy policies are as clear as they are expected to be, there are other concerns for scraping publicly available personal data.

A request for consent must be initiated by the entities processing personal data prior to commencing any processing operation. In the case of web scraping, this becomes tricky. Usually, entities engaged in web scraping are not the same entities who own the platform from where data is being scrapped. This can be examined through Clearview AI's scrapping activities from Facebook. Users share their data with Facebook meaning that for the purpose of that transaction Facebook is the data controller and the user can be expected to read the privacy policies of Facebook to understand the details on how their information is being processed. Clearview AI is not a data processor of Facebook i.e. Facebook hasn't outsourced a data processing operation of scraping to Clearview AI. Since it's not a data processor, Facebook is not obligated to include the web scraping activities of Clearview AI in their privacy policy. Clearview AI is merely a third party in this equation. It is unclear how a third party is expected to initiate a processing operation in this case as there isn't a reasonable method for them to provide a privacy notice to the users of Facebook prior to extracting the data. The other option is to extract the data and then provide notice to those users whose data has been extracted. This still leaves the question about ways to isolate the users whose data has been extracted and ways to contact and provide the users with the relevant information as Clearview AI have argued that they have no means of verifying the identity of the users whose information they collect[48]. Hence, even though consent of the individual is a very important safeguard with respect to using publicly available personal data it is unclear how that can be achieved in practice.

3.2.2 Lawful basis of legitimate interest of data controller or third party

Article 29 working party[49] has issued guidelines on conducting a balancing test between the interests and rights of the data subject and the legitimate interests of the data controller or third party. The most important requirement is that the legitimate interest be acceptable under the law. Web scraping is merely a data collection method. The validity of the legitimate interest will depend on the purpose of such data collection. For this thesis, it is assumed that the purpose of data collection is lawful. The part of the balancing test that is contentious when it comes to publicly available personal data is the risk assessment of the data subjects.

The Italian Data Protection Authority[50] analysed the possibility of Clearview AI relying on the legitimate interest ground to scrape publicly available personal data. It found that the legitimate interest of the company was to make profit. This interest when balanced with the rights and interests of the data subject wherein the data collected had the potential to detect different aspects of their private life and the magnitude of data collection veered towards protection of the

[47] id
[48]Injunction order against Clearview AI, https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9751362 > last accessed 20 February 2022
[49] Article 29 working party, 'Opinion 06/2014 on the notion of legitimate interests of the data controller under article 7 of Directive 95/46/EC', (*European Commission,* 9 April 2014) < https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf> last accessed 20 February 2022
[50] n(48)

data subjects thereby invalidating the applicability of the legitimate interest ground in this specific case.

The subsequent question will be to examine the applicability of this lawful ground for cases where the profit seeking motive is not as obvious as was the case with Clearview AI. According to the guidelines issued, the risk assessment will need to take into consideration if the data is publicly disclosed or if it is made accessible to many persons[51]. After this is an analysis of the harm that is caused or is likely to be caused to the data subject. In the case of web scraping, the harm that is expected to be caused will arise at a later stage i.e., after the data has been collected and moulded into the required format for the purposes of the processing operation. For example, if data is being scraped to train a facial recognition algorithm or to create a profile of individuals who are Left leaning or Right leaning, the resultant harm of that action cannot be traced back to an individual data subject. However, it can be argued that an individual's expectations of usage of personal data disclosed was violated at the stage of data collection. The harms caused to the expectations of the data subject have not been considered as primary cause of action in any data protection enforcement cases so far. This essentially means that web scrapers can potentially rely on the lawful ground of legitimate interest of the data controller or third party if their actions are not considered to be profit seeking.

3.2.3 Information obligations

Regardless of which lawful ground web scraper relies on, they will have to comply with the information obligations prescribed in GDPR. For collecting publicly available personal data, entities will have to rely on article 14 which provides details on the information to be provided to the data subject where personal data has not been directly obtained from them. This information includes the categories of personal data collected, information regarding data subject rights, the source of personal data etc. Theoretically this is a very strong obligation in favor of the data subject that ensures they are aware of the processing operations and their rights associated with such processing. However, there are two implementation issues with regards to publicly available personal data and web scrapers.

The first is the practical considerations of providing such information to the data subject which has been outlined in the section on consent (3.3.1). The second is the time frame that is available to the entity to provide this information. According to article 14(3)(a), the data controller is expected to provide the required information to the data subject within a reasonable period and at the latest within one month since the initiation of the processing operation. For a period of about a month the data subject is unaware about further use of their personal data. This assumes that the entity using the personal data provides the required information according to the law. Since the personal data that the entity requires has already been disclosed, the data subject has no practical way of verifying if a third party that is unrelated to the initial data controller has access to their information. This second problem arises only if the data controller for the purposes of web scraping has verified the identity of the data subject and has a method to provide the privacy notice.

[51] n(49)

4. Contextual integrity and safeguards in US and EU

The previous section laid out the safeguards available to publicly available personal data in the US and EU through the lens of web scraping as a data collection method. It was concluded that due to the nature of privacy enforcement in the US, there is a lack of extensive safeguards for personal data disclosed in a specific context apart from law enforcement authorities requiring a warrant to ensure legality of search under the fourth amendment. With respect to private entities collecting such data, there does not appear to be any safeguards for personal data partially restricted on a digital platform. In the EU, the information obligations and the legal grounds of processing personal data do offer a diminished level of protection on account of implementation challenges. This section explains the contextual integrity framework proposed by Helen Nissenbaum and applies it to web scraping activities to suggest that such operations are in violation of the framework. The existing safeguards in the US and EU are examined within the context of the norms identified in the framework to suggest next steps in the process of re-examining the safeguards for publicly available personal data. The gaps in the existing safeguards contributing to diminished protections for personal data disclosed in a specific context will be examined through the objective of safeguarding individuals' expectations associated with contextual disclosure proposed by the CI framework.

4.1.Contextual Integrity Framework

The framework for contextual integrity argues that individuals' expectations of privacy over their information are relatively more complex than the mere public and private nature of the information and that these expectations are dependent on the context of data collection[52]. In other terms, there exist multiple contexts all which are governed by distinctive rules that operate within each specific context. It rejects the argument that non sensitive information can be freely transmitted on the basis that such information is still tagged with the context in which it was collected[53]. The underlying principle of the framework is that all areas of life are governed by norms of information flow and the legitimacy of the norms depends on the context in which the action takes place[54]. The framework argues that there are two types of informational norms i.e., norms of appropriateness and norms of flow or distribution.[55]

The norms of appropriateness determine if it is appropriate to disclose information about an individual in a specific context.[56] For example, an individual who discloses their sexual orientation in a gay pride parade does not have to disclose the same information in their workplace because the context of disclosure have changed[57]. The norms of distribution determine if the distribution or flow of the information abides by the contextual norms of the

[52] n(20)

[53] n(19)

[54] id 300, Digital book edition

[55] id

[56] id

[57] Ferdinand Schoeman, Gossip and Privacy, in GOOD GOSSIP 72, 73 (Robert F. Goodman &Aaron Ben-Ze'ev eds., 1994)

information flow.[58] For example in the context of healthcare, the distribution of information that has been shared by the patient with the healthcare provider is limited by the confidentiality obligations of the healthcare provider.

Both norms take into consideration the types of information, the roles of the actors involved and the principles of transmission to determine if the action is in line with the rules of a given context i.e., an examination of the contexts, actors, attributes, and transmission principles[59]. The actors that are part of the norm are senders of information, recipients of information and information subjects[60]. The attributes are the data types or types of information that are being considered[61]. The norms examine if the disclosure and further transmission of these data types are appropriate in each context. For example, it is appropriate for a healthcare provider to require sensitive reproductive information, but it is not appropriate for a workplace to require similar information. The transmission principles are constraints of the flow of information from the actors within the transmission[62]. Confidentiality, consent, notice, reciprocity are some examples of the constraints that are usually imposed.

4.1.1 The framework in action

As mentioned in the previous chapter, web scraping is one of the most prominent data collection methods used to collect publicly available personal data off the internet. This section will examine the way the CI framework would respond to the introduction of web scraping of personal data from user profiles on Facebook and then analyse if the existing protections identified in US and EU consider this "decision heuristic "of the CI framework[63]. This example is like the web scraping activities of Clearview AI. The decision heuristic is as follows:

Information flows

There are three information flows: a user creates an account and by such creation provides Facebook their information, detailed information from their profile is provided to friends and others depending on the settings selected by the user, some information is made available to the search engines for indexing purposes.

Prevailing context

The prevailing context is the overall objective of the platform in question since the user has disclosed their personal information in furtherance of these objectives. This context is characterized by the roles, activities and internal values that operate within the social settings. Roles refers to the capacities in which the individual in the context acts i.e., the users of the social media platform. Activities refers to the practices and actions in which the roles engage i.e., users of the social media platform exchange personal information such as likes, dislikes, photos, interests etc. with a group of individuals pre-determined in their user

[58] n(19), 119
[59] n(20) 297
[60] id 298
[61] id 302
[62] id 306
[63] id 379

settings. Internal values are the objectives/goals/purposes around which a context is situated and are the defining features of the context. As an example, the author suggests that values in educational context would be transmitting knowledge and social values to the society's young.[64] According to Facebook, it is an online social networking service that helps the user "connect with friends, family and communities of people who share their interests"[65]

The interesting question here is determining if the business model of Facebook which is built on its data sharing activities with third party advertisers forms a part of the context of information sharing for the individual. This question needs to be answered in the background of the objective of the CI framework. The framework aims to assess the reasonable expectations of privacy of the individual based on the context in which the information is shared. Including the business interests of the company as part of this context effectively means that the reasonable expectations of privacy of the individual would be dependent on the business interests of the company and not the other way around. Such an interpretation will diminish the meaning of privacy. Hence, the business model of Facebook will not form a part of the context for the purposes of the decision heuristic.

Identification of information subjects, senders, and recipients:
Information subjects are the users of the site. Senders of information are users who make their profiles visible to other users. Recipients include social network, other users, and potential public if the user has provided access to a limited set of profile data to be visible via search engines. Web scraping will expand the scope of the recipients by making all the information available to the automated script that scrapes the data.

Transmission principles
These principles govern information sharing within a particular context. The transmission centres around the settings determined by the information subjects i.e., they get to decide if the information is shared with all their friends on the platform, with a specific list or with users of the platform that are not friends as well.

Entrenched information norms
The norms give the users the ability to control the information flows based on their preferences in the settings tabs and the privacy policy they consented to prior to creating the account in the first place. The web scraper will disrupt these norms as it tends to harvest all available profile data which includes information that was restricted by user preferences and the scraping wasn't something that the user had consented to in the first place. Prima facie assessment reveals a violation of the CI framework.

Nissenbaum recognised that strong reliance on entrenched informational norms can lead to conservatism wherein no new practice will have the potential to challenge the status quo even in cases where the new practice is beneficial to the individuals[66]. To address this, Nissenbaum

[64] id 131
[65] Meta, About section,< https://about.facebook.com/technologies/facebook-app/> last accessed 29 March 2022
[66] n(20) 334

suggested an examination of the overall moral superiority of the new practice by assessing the legitimacy of the new practice. This legitimacy can by established by analysing the moral and political factors[67] that are affected i.e., the potential threats to autonomy and freedom, effects on power structures, implications for justice, equality etc. In addition to these, she argues for an examination of the effectiveness of the new practice in achieving the relevant contextual values[68]. These two additional evaluations posed by Nissenbaum will be examined in the context of web scraping of user profiles from Facebook by a third party below.

Evaluation I

By bypassing the settings chosen by the user and Facebook (in cases where web scraping is contractually prohibited through the terms of use of the platform), one of the indirect harms caused is the loss of the user's control of their information flows on the platform i.e., information that was shared within the context of a social networking service were exposed to the public using the automated scraping algorithm. This action also results in the harm of thwarted expectations as the user's reasonable expectation of their choices in the general settings being adhered to has not been complied with. By scraping the information that the user intended to disclose to a specific community determined by their settings, the context of disclosure has been expanded to the general community. This expansion in audience can also lead to chilling effects on the part of the user as the inability to predict changes in the context of disclosure can nudge the user to self-censor.

Evaluation II

As mentioned earlier, in the context of social media platforms, the contextual values are with respect to the engagement of the users of the platform based on mutual or shared interests. Web scraping expands the audience of the information that has been uploaded by the user and processes the information for secondary purposes. These secondary purposes do not advance the contextual values of enabling engagement with other users on the platform.[69] Hence, the new practice does not have any connection with the objectives of the platforms and therefore isn't necessarily superior in terms of achieving the objectives of the context. Therefore, web scraping activity undertaken by unrelated third parties to account for the expectations of the user with respect to contextual disclosure on Facebook and by extension any digital platform.

4.2.Existing safeguards and the CI framework

As identified in the previous chapter, the safeguards for publicly available personal data in United States of America that deal with web scraping are non-existent as a result of which, it is not considered inappropriate to share information disclosed in context A to be further disclosed without user approval in context B. This indicates the lack of recognition of context

[67] id 364

[68] id 365

[69] Michael Zimmer, "Addressing Conceptual Gaps in Big Data Research Ethics: An Application of Contextual Integrity." Social Media + Society, (2018). https://doi.org/10.1177/2056305118768300 last accessed 25 April 2022

relative information norms thereby disregarding the norms of appropriateness and distribution of the CI framework.

In the EU, two essential safeguards for publicly available personal data were identified under the GDPR i.e., risk assessment prior to relying on the legitimate interests ground for processing personal data and information obligations. One of the criteria to conduct the balancing exercise between the legitimate interests of the data controller and the rights and interests of the data subject is if the data concerned was publicly disclosed or was made accessible to many persons.[70] However, the extent to which the balancing will result in favour of the data subject in case of contextual disclosure is unclear. A reference to the importance of the context of initial disclosure was provided by the European Data Protection Supervisor (EDPS) in relation to repurposing data from profiles through algorithms. The EDPS concluded by insisting on compliance with the purpose limitation principle to address such algorithms.[71] An accurate determination of the importance of contextual disclosure in the risk assessment can be arrived at only after further clarifications from the data protection authorities or the Member States concerned.

With respect to information obligations, an enforcement notice [72] penalizing Clearview AI for scraping billions of publicly available images of UK residents was published. Amongst many other provisions of the UK GDPR, it was held to be in violation of article 5(1)(a) for processing personal data unfairly i.e., without informing them and the processing did not fall within reasonable expectations of the individual. However, the question of how web scrapers should provide notice has not been addressed.

However even if the challenges associated with web scrapers determining an effective method to issue a privacy notice to individuals is addressed, compliance with the information obligations is difficult. The information obligations require the data controller to provide privacy notice to the data subjects with all relevant details of the processing operations in clear, concise, and simple language. The onus is on the data subject to determine the legitimacy of the operations and then proceed with either accepting or rejecting such processing. Based on the numerous studies[73] that have been conducted to examine the actions of the data subject, this reliance on a rational informed user is delusionary as privacy policies have remained unreadable. In cases where data subjects do read and understand the

[70] n(49)

[71] European Data Protection Supervisor, 'Opinion 03/2018 EDPS opinion on online manipulation and personal data'(*European Data Protection Supervisor,* 19 March 2018) https://edps.europa.eu/sites/default/files/publication/18-03-19_online_manipulation_en.pdf last accessed 28 March 2022

[72] Information Commissioner Office, 'Enforcement notice, Clearview AI', (*Information Commissioner's Office,* 18 May 2022) https://ico.org.uk/media/action-weve-taken/enforcement-notices/4020437/clearview-ai-inc-en-20220518.pdf last accessed 28 March 2022

[73] Brooke Auxier, Lee Rainie et al, 'Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information' (*Pew Research Centre,* 15 November 2019) < https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/> last accessed 28 March 2022
Brooke Auxier, Lee Rainie et al, ' American attitudes and experiences with privacy policies and laws' (*Pew Research Centre,* 15 November 2019) < https://www.pewresearch.org/internet/2019/11/15/americans-attitudes-and-experiences-with-privacy-policies-and-laws/> last accessed 28 March 2022

privacy policies, the power asymmetry between the data subject and the entity collecting the personal data can result in the lack of bargaining power on the part of the former to suggest changes to the same.[74]

The decision heuristic of the CI framework relies on an examination of the new practice in the background of the objectives and goals of the prevailing context to determine if there was a violation of contextual integrity. These objectives and goals of the prevailing context form an essential characteristic of understanding the expectations of the data subject with respect to the disclosure of information on a specific platform. Data collected via web scraping is seldom used to advance the objectives and goals of the prevailing context i.e., Clearview AI created an altogether new product based on the scraped data and did not advance the objectives of the platform from where the data was scraped. Neither the US nor the EU acknowledge this essential characteristic in their legislations or jurisprudence. This reliance on overall objective of the CI framework needs to be translated into legal frameworks of US and EU to examine the validity of the expectations of the data subject for further use of the data. Despite recognition of a diminished form of contextual disclosure in the EU through its information obligations and lawful grounds of processing, in practice these obligations fall short of providing adequate protections to the informational norms identified by the CI framework.

5. Recommendations for changes in the safeguards to publicly available personal data

So far, the safeguards provided to publicly available personal data have been examined through the lens of the data collection strategy used by web scrapers in the US and European EU. Section 4 explains the decision heuristic of the contextual integrity framework through the example of web scraping of Facebook profiles of individuals and concludes that web scraping as a processing activity fails to consider privacy of the individual in terms of contextual disclosure. With this background, this section shall examine the alternatives provided to address the challenges of the notice and consent model and examine if any of the alternatives suggested address the change in context and prescribe additional safeguards for fair processing of publicly available personal data.

5.1. Alternatives to the notice and consent model

One of the alternatives[75] suggested requires the entity processing personal data to conduct a harm assessment process in which they will be required to assess the legality of their data collection and processing practices. This is subsequently evaluated by an auditor who then makes further recommendations if any to mitigate prospective harms to the data subject

[74] Michiel Roen, "Beyond consent: improving data protection through consumer protection law" (2016) Internet Policy Review 5 (1). https://policyreview.info/articles/analysis/beyond-consent-improving-data-protection-through-consumer-protection-law. Last accessed 28 March 2022

[75] World economic forum (2020), 'Redesigning data privacy: Reimagining notice and consent for human technology interaction' <https://www3.weforum.org/docs/WEF_Redesigning_Data_Privacy_Report_2020.pdf> last accessed 25 April 2022

because of the practices of the companies. This audit report can be opened for further scrutiny of the public by publishing the report[76]. To ensure the principles of contextual integrity are followed, the harm assessment process can be expanded to consider the moral and political effects of the processing operation as required by Evaluation 1 of the decision heuristic of the CI framework i.e., the potential threats to autonomy and freedom, effects on power structures, implications for justice, equality etc[77]. The assessment can be expanded to include evaluation 2 of the CI framework i.e., the effectiveness of the processing operation in achieving the contextual values of the context in which the personal data was collected in the first instance.[78]

By requiring the process to be audited by a qualified individual as well as disclosing the report to the public, subject matter experts can evaluate the benefits and risks of the processing operation identified. However, the feasibility of this alternative depends on the entity processing the personal data disclosing details of their processing transparently. The CI framework doesn't provide for additional factors or standards to assess evaluation 1 and 2. Due to the subjective nature of "threats to autonomy and freedom, effects on power structures" it is possible that the entities assessing their practices against these standards arrive at different conclusions that may not technically be incorrect per se due to absence of additional factors to assess them against. This can render any such assessment a mere theoretical exercise. For this exercise to be effective, data protection authorities will need to develop guidance on interpreting the subjective criteria of evaluation 1 and 2. This will ensure that the assessment carried out by the entities processing data is uniform regardless of their business interests.

Another alternative, specifically, with respect to data sharing with third parties is the introduction of data trusts[79]. They are legal instruments that are appointed as stewards to manage the personal data on behalf of the beneficiary i.e., the individuals disclosing personal data[80]. By ensuring that the data trust is fiduciarily responsible to its beneficiaries, there is an assumption that the trust can meaningfully negotiate over the terms and conditions of data sharing pertaining to the data that is in their trust. There have been different approaches to operationalise the data trust mechanism in practice. A data trust can be created implicitly[81] whenever individuals share their personal data with the entity collecting data i.e., by uploading personal data on Facebook, Facebook can be said to be holding the personal data as a trust and is expected to operate with undivided loyalty to the beneficiaries which are the users who have disclosed the data. However, this approach was considered by a few[82] as being impractical as data collectors might have to be unduly loyal to their shareholders and not the users providing the data.

[76] id

[77] n(20) 364, Digital book edition

[78] id 365

[79] n(75)

[80] Sean Martin McDonald,' Reclaiming data trusts' (*Centre for International Governance Innovation*, 5 March 2019) https://www.cigionline.org/articles/reclaiming-data-trusts/?ref=https://githubhelp.com last accessed 25 April 2022

[81] Lilian Edwards, "The Problem with Privacy " (2004). International Review of Law Computers & Technology, Vol. 18, No. 3, pp. 263-294, https://ssrn.com/abstract=1857536 last accessed 25 April 2022

[82] Sylvie Delacroix and Neil Lawrence, "Bottom-Up Data Trusts: Disturbing the 'One Size Fits All' Approach to Data Governance "(October 12, 2018). Forthcoming in International Data Privacy Law https://ssrn.com/abstract=3265315 or http://dx.doi.org/10.2139/ssrn.3265315 last accessed 25 April 2022

The other approach[83] is to create a data trust as a bottom-up mechanism where in the data subjects pool their rights over their personal data within the legal framework of a Trust. These explicit trusts are required to be run by independent trustees that are bound by the terms and conditions and the governance structure of the trust. Such[84] an explicit trust can result in a market of data trusts with different governance structures which gives data subjects an option to decide which trust they would want to pool their personal data into. However, this will require more active role of data subjects. They will need to contemplate the potential risks of data sharing based on the terms and conditions of each data trust and then make an active choice. Albeit, the frequency of these decisions will be low, data subjects will still be required to undertake complicated balancing exercise that will again depend on their understanding of the Trust's operations.

The information fiduciary concept argues for establishing fiduciary obligations on digital companies owing to the vulnerability and dependence of users created by their business model of information capitalism[85]. These fiduciary obligations go beyond that of mere good faith and encompass duties of care, confidentiality, and loyalty towards their users. These obligations travel with the data[86] i.e., if entity A has a fiduciary responsibility towards its users, those fiduciary obligations travels to entity B with whom entity A has shared the relevant data. Entity A can share the relevant data only after receiving substantial guarantees from entity B about the fiduciary responsibility. However, applying this in the case of processing of publicly available personal data is tricky. In the Clearview AI case, the fiduciary responsibility of safeguarding the personal data of the individual in the context in which it was disclosed rested with the digital platforms on which individuals had voluntarily disclosed the data. Clearview AI did not seek authorization from the digital platform prior to scraping the data which means the fiduciary responsibility could not be contractually transferred.

In response to increase in big data analytics, there has been a growing chorus of arguments to shift policy attention towards regulating the actual uses of big data and less on the protections offered during the collection stage[87]. Scholars have argued[88] that principles such as data minimisation and purpose limitation are antithetical to the big data business models as they depend on collecting vast amounts of personal data for developing valuable new uses for personal data, uses which might not be apparent at the collection stage. International discussions[89] reiterate the need to shift the focus of regulatory intervention to the uses of big data owing to the pervasive nature of data collection that is difficult to monitor. These

[83] id
[84] id
[85] Jack Balkin, 'The fiduciary model of data privacy' (*Harvard Law Review,* 9 October 2020) https://harvardlawreview.org/2020/10/the-fiduciary-model-of-privacy/ last accessed 25 April 2022
[86] id
[87] Recommendation 1 in President's Council of Advisors on Science and Technology (PCAST),!Big Data and Privacy: A Technological Perspective, (North Charleston, S.C.:Create Space, 2014),49.
Craig Mundie, "Privacy Pragmatism: Focus on Data Use, Not Data Collection," (*Foreign Affairs*, 12 February 2022) https://www.foreignaffairs.com/articles/2014-02-12/privacypragmatism last accessed 25 April 2022
[88] Omer Tene and Jules Polonetsky, "Big Data for All: Privacy and User Control in the Age of Analytics," Northwestern Journal of Technology and Intellectual Property 11, no 5 (2013):259–60
[89] Fred H. Cate, Peter Cullen, and Victor Mayer Schönberger, "Data Protection Principles for the 21st Century: Revising the 1980 OECD Guidelines Books by Maurer Faculty. 23.

arguments opt for adopting a top-down approach by outlawing certain uses of personal data which can be expanded to include publicly available personal data. This does protect the individuals from the potential harms that arise in the end stage of processing such as discrimination, bias, inaccurate determinations etc in cases where the personal data has not been legitimately acquired but doesn't extend to cases in which it has been legally acquired.

However, the harms that arise at the collection stage i.e. loss of autonomy, lack of transparency compounded with power and information asymmetry between entity collecting data and individual providing the data, lack of control over contextual disclosure leading to potential chilling effects on speech are not addressed by use based regulations. There is an assumption that by correcting the harms that arise at the decision stage of the processing operation, individuals will not be negatively impacted by the unconstrained data collection. This assumption however is not based on any empirical evidence.[90] Even without empirical evidence, the assumption doesn't possess any merit. This can be explained with the case of Clearview AI, which scrapped publicly available images off the internet to develop a facial recognition algorithm that was subsequently sold to law enforcement officials. The use of the processing operation that will need to be regulated is developing a facial recognition algorithm. If, a legislation prohibiting such an action is enacted, the initial stage of collection i.e., the scrapping action is not regulated by such a legislation. The harms associated with such an action such as that of loss of autonomy, lack of transparency, chilling effects and others cannot be addressed by the regulation that bans facial recognition algorithms. However, an argument can be made that a ban on developing facial recognition algorithms would have disincentivised Clearview AI from scrapping the images in the first place. This argument may be true with respect to facial recognition algorithms, but due to the even increasing uses of personal data an entity deploying web scrapers can always sell the data they have scrapped to other entities whose processing operations have not been legally prohibited. Hence, the harms arising at the collection stage of publicly available personal will remain.

5.3. Way Forward

The objective of this thesis has been to prove that the traditional private/ public dichotomy that is translated into data protection legislations and jurisprudence does not bode well in the age of complex digital technologies. Neither of the alternatives to the notice and consent model mentioned above address the considerations of the decision heuristic of the contextual integrity framework with respect to publicly available personal data i.e., specifically the moral and political factors impacted due to a change in the context of processing and the effectiveness of the new processing operation in achieving the contextual values of the context of initial disclosure. However, a combination of these alternatives with few additional safeguards has the potential to take into consideration contextual disclosure.

[90] Joris Van Hoboken, "From collection to use in privacy regulation? A forward-looking comparison of European and us frameworks for personal data processing"(2016) Exploring the Boundaries of Big Data, 231 < http://www.jorisvanhoboken.nl/wp-content/uploads/2017/11/VanHoboken_Collection_and_Use_2016.pdf> last accessed 25 April 2022

A combination of use-based regulation with fiduciary obligations on the entities processing publicly available personal data is proposed. This proposal will be explained through the example of Clearview AI. Clearview AI scrapped images of individuals that was publicly available on digital platforms.[91] Facial metrics obtained from the database of these images was then used to develop a facial recognition algorithm that was subsequently sold to law enforcement agencies and private corporations[92].

The recommendation proposed suggests that the digital platforms have a fiduciary responsibility to ensure confidentiality of personal data uploaded on their platforms. This fiduciary responsibility can be exercised by ensuring that the user profiles are obscure by default[93] i.e., instead of relying on the users of the platform to change the privacy settings and prevent users external to the platform from having access to the data, the digital platforms should configure default settings that make it difficult (or impossible if technically feasible) for automated attempts to crawl through their platforms and scrape the data.

Clearview AI would be expected to conduct a harm assessment process that takes into consideration evaluation 1 and 2 of the decision heuristic of the contextual integrity framework i.e., the moral and political factors impacted by their processing operations and the effectiveness of their processing operation in achieving the contextual values of the context of disclosure. This assessment should be evaluated by an independent auditor affiliated with the authority who is responsible for safeguarding the privacy rights of the individual. In the event, that assessment is approved by the auditor, the final uses of the processing operation need to be examined.

A use-based regulation can be introduced that outlaws specific uses of publicly available personal data such as facial recognition algorithms, sentencing and predictive policing algorithms, credit scoring algorithms etc. The auditor, in cooperation with the data protection authority, can be tasked with the responsibility of evaluating the final uses of the processing operation that relies on publicly available personal data against the prohibited uses outlined in the use-based legislation of each jurisdiction. In case the use is not prohibited by the legislation, a request for authorization of scraping publicly available personal data can be sent to the digital platforms. The digital platforms can contractually pass on the fiduciary responsibility associated with the personal data to Clearview AI.

This proposal addresses the power imbalances between the individual and the obscure third parties by passing the responsibility of safeguarding their interests to the auditors affiliated with the regulatory authorities tasked with data protection compliance. It also addresses the conservatism of the contextual integrity framework by incorporating evaluation 1 and 2 of the decision heuristic framework in the harm assessment process, thereby not just safeguarding the

[91] n(1)

[92] Consent order of permanent and time limited injunctions against defendant Clearview AI , Case No. 2020 CH 04353 https://www.aclu.org/legal-document/consent-order-permanent-and-time-limited-injunctions-against-defendant-clearview-ai

[93] Woodrow Hartzog, "Facebook's failure to end 'Public by Default" (*Medium,* 2 November , 2018) https://medium.com/s/story/facebooks-failure-to-end-public-by-default-272340ec0c07

rights of contextual disclosure but ensuring efficient processing operations are not negatively impacted.[94]

5.4. Concluding remarks

With the advent of new technologies for collection of personal data, it is imperative that the legislative and judicial approach to safeguarding re-use of personal data disclosed on digital platforms be reconsidered. Existing approaches such as the notice and consent model of the EU and the accessibility rule to determine reasonable expectation of privacy of the individual in the US are underequipped to deal with the individual's expectations and attitudes surrounding contextual disclosure of their data. The first step in this shift is to stop relying on safeguards that solely depend on the "public" or "private" nature of personal data and start examining the informational norms surrounding the context of disclosure. By relying on the contextual integrity framework, introducing a legislative obligation of fiduciary responsibility on the data controllers combined with the drafting of use-based regulation as safeguards for processing publicly available personal data may be recommended. One topic of further research is an examination of the typology of privacy harms that originate in the digital public sphere. The literature on privacy harms needs to be compounded with empirical evidence on the expectations and attitudes of the individuals who have disclosed personal data in one context to suggest any further legislative changes.

[94] n(20) 334, Digital book edition