GRUR International, 69(6), 2020, 616–623

doi: 10.1093/grurint/ikaa047
Opinion

DAPHNE KELLER*

Facebook Filters, Fundamental Rights, and the CJEU's *Glawischnig-Piesczek* Ruling

The Court of Justice of the European Union's (CJEU) 2019 ruling in *Glawischnig-Piesczek v Facebook Ireland*** addresses courts' powers to issue injunctions requiring internet hosting platforms to proactively monitor content posted by their users. It answers important questions about limitations on such injunctions under the eCommerce Directive (Directive 2000/31/EC). But, as this Opinion explains, it leaves some much thornier questions unresolved.

Glawischnig-Piesczek holds that courts may, consistent with Art. 15 of the eCommerce Directive, require platforms to monitor for and remove specific content. Monitoring orders may not, however, require platforms to carry out an 'independent assessment' of the content. The ruling does not closely examine what kinds of injunctions or filtering technologies are permissible, nor does it explore fundamental rights considerations when courts are asked to order platforms to monitor their users. This Opinion lays out the case's technological, legal, and policy backdrop, and identifies important questions it leaves open for Member State courts. In particular, the Opinion suggests that *Glawischnig-Piesczek*'s limitation on 'independent assessment' will make it difficult for courts to devise injunctions that simultaneously follow the CJEU's guidance under the eCommerce Directive and meet the requirements of fundamental rights. It lists key fundamental rights considerations for future cases involving potential monitoring injunctions, including procedural considerations in cases affecting the rights of absent third parties.

I. Introduction

Some of the most hotly contested recent internet law and policy questions in the European Union (EU) have concerned platform monitoring requirements: laws or injunctions compelling hosting platforms like YouTube or Twitter to proactively search for or filter out illegal material that their users post.[1] A 2019 case decided by the Court of Justice of the European Union (CJEU), *Glawischnig-Piesczek v Facebook Ireland*, shed important new light on the subject.[2] It also left key questions unresolved. As this Opinion will discuss in more detail, the Court answered a *legislative* question, holding that certain monitoring injunctions are permissible under the eCommerce Directive – the law which, as implemented by Member States, has structured platforms' legal responsibility for user content in the EU for almost two decades.[3] But the CJEU did not discuss the closely related *fundamental rights* questions raised by monitoring requirements. As a result, the ruling leaves Member State courts to decide how rights under the EU Charter may limit or shape monitoring injunctions against internet platforms.

Glawischnig-Piesczek holds that while certain monitoring injunctions are permitted under the eCommerce Directive, those injunctions may not require platforms to carry out an 'independent assessment' of the content to be removed. Courts seemingly may not, under Art. 15, direct platforms to apply nuanced human judgment to correct the failings of automated filters. This greatly limits the available mechanisms to protect news reports, scholarship, and other lawful 'dolphins' caught in a filter's net. As a result, an injunction that meets this newly clarified interpretation of Art. 15 may be difficult to square with the EU Charter.

This Opinion proceeds in two parts. First, in the Background Section, I will lay out the *Glawischnig-Piesczek* case history as well as relevant considerations regarding fundamental rights, filtering technology, and legislative developments. This discussion will draw on my experience as Associate General Counsel to Google, a job that I held until transitioning to an academic role at Stanford in 2015. Second, in the Analysis Section, I will more closely examine the CJEU's reasoning under the eCommerce Directive, and explore the resulting fundamental rights questions. I will suggest that Member State courts should not issue monitoring injunctions without the factually grounded expectation that the results will

* Director of the Program on Platform Regulation at Stanford's Cyber Policy Center.

** See the text of the decision in this issue of GRUR International at DOI: 10.1093/grurint/ikaa056.

1 See, eg, Christina Angelopoulos, 'On Online Platforms and the Commission's New Proposal for a Directive on Copyright in the Digital Single Market' (2017) University of Cambridge <https://ssrn.com/abstract=2947800> accessed 5 February 2020; Sophie Stalla-Bourdillon and others, 'A Brief Exegesis of the Proposed Copyright Directive' (2017) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2875296> accessed 5 February 2020; Joris van Hoboken and others, 'Hosting intermediary services and illegal content online: An analysis of the scope of article 14 ECD in light of developments in the online service landscape' (2019) Institute for Information Law <https://www.ivir.nl/publicaties/download/hosting_intermediary_services.pdf>.

2 Case C-18/18 *Eva Glawischnig-Piesczek v Facebook Ireland Limited* EU:C:2019:821 ('*Glawischnig-Piesczek*').

3 Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market [2000] OJ L 178 (eCommerce Directive).

achieve a balanced outcome, protecting a plaintiff's or government's interests without disproportionately burdening the rights of third-party internet users. Achieving such balance will be very difficult, particularly given *Glawischnig-Piesczek*'s limitations on the use of human judgment to correct for mistakes made by automated filters. I will close with brief observations about procedural shortcomings in this and other intermediary liability cases, which can profoundly affect the rights of internet users, while providing them no opportunity to voice those interests before a court. Some of the analysis in this Opinion was presented at greater length in a Stanford Center for Internet and Society White Paper, 'Dolphins in the Net: Internet Content Filters and the Advocate General's *Glawischnig-Piesczek v. Facebook Ireland* Opinion'.[4]

II. Background

This Section briefly describes (1) the case, (2) concerns about filters and fundamental rights, (3) the state of filtering technology, and (4) recent and ongoing EU legislative developments.

1. The case

Eva Glawischnig-Piesczek, the former head of Austria's Green Party, asserted that she was defamed by a Facebook user post that called her a 'lousy traitor' (*miese Volksverräterin*), a 'corrupt oaf' (*korrupter Trampel*), and a member of a 'fascist party' (*Faschistenpartei*).[5] When Facebook did not remove the post, she initiated two proceedings in Austrian courts. In one, a criminal case, the court ultimately said that the post was not obviously unlawful.[6] In the second, an expedited civil proceeding, a different court said that the post was obviously unlawful, and that Facebook was therefore liable. While Facebook is generally immunized as a host under Austria's implementation of the eCommerce Directive, it forfeited that immunity in this case by failing to remove the post after being notified of its existence.

The Austrian second instance court affirmed Facebook's liability, and also held that it must proactively block future posts from making the same defamatory statements. The Austrian Supreme Court referred the case to the CJEU, asking whether orders to block 'identical' or 'equivalent' content were permissible under Art. 15 of the eCommerce Directive. That referral did not ask the CJEU about fundamental rights. (It did raise a second question not addressed here: whether Austrian courts could require global compliance with the order.) The Advocate General (AG) advised that such orders were permissible under certain circumstances.[7]

The CJEU's ruling will be examined in more detail in Section III.1 below. Broadly, it held that injunctions requiring platforms to proactively remove both identical and equivalent content are permitted by the eCommerce Directive. While Art. 15 of the Directive prohibits Member States from imposing 'general' monitoring obligations on platforms, the CJEU concluded that courts could nonetheless issue more specific injunctions to block particular content identified by the court. Those injunctions could not, however, require the platform to independently assess whether content violates the law. The Court did not address how such automated filters might work, or discuss fundamental rights considerations. The CJEU's next chance to consider the intersection of fundamental rights and platform monitoring requirements is expected to come when it reviews Poland's challenge to the recently enacted Copyright Directive.[8]

2. The fundamental rights backdrop

Fundamental rights issues have shaped the public policy discussion about internet monitoring and filtering practices for many years, and will be important to Member State courts as they seek to apply *Glawischnig-Piesczek*'s reasoning in future cases. The CJEU has noted in the past that requiring internet platforms to monitor users' communications burdens not only platforms themselves, but also their users' rights to (1) privacy and data protection, and (2) freedom of expression and information. Scholars have also identified concerns about (3) rights to a fair trial and effective remedy for people whose online expression and participation are 'adjudicated' as legal violations and terminated by platforms.[9] And an increasing body of research suggests filtering mandates also implicate internet users' rights to (4) equality and non-discrimination before the law. Two recent studies, for example, found that when automated content filters attempt to parse human language, they disproportionately silence lawful expression by members of minority or marginalized racial and linguistic groups.[10] Finally, a law or injunction that burdens these rights must (5) be formulated with sufficient precision to meet the 'provided for by law' requirement of EU Charter Art. 52. While this Opinion will focus on expression and information rights (and briefly touch on privacy and data protection in Section III.3), considerable work remains to be done exploring implications for all of these rights in the wake of *Glawischnig-Piesczek*.

The CJEU has, in the past, emphasized the risk that an automated filter 'might not distinguish adequately between unlawful content and lawful content, with the result that its introduction could lead to the blocking of

4 Daphne Keller, 'Dolphins in the Net: Internet Content Filters and the Advocate General's Glawischnig-Piesczek v. Facebook Ireland Opinion' (2019) Stanford Center for Internet and Society <https://cyberlaw.stanford.edu/files/Dolphins-in-the-Net-AG-Analysis.pdf>.

5 The full post, informally translated, reads: 'Lousy traitor. This corrupt oaf has never earned a single honest cent with real work in her whole life, but with our tax money is kissing the asses of these smuggled-in invaders to build them up into the most valuable of all. Let us finally prohibit this Green Fascist party.'

6 Higher Regional Court for Criminal Matters Vienna, 23 March 2018, 17 Bs 236/17f, 27.

7 Case C-18/18 *Eva Glawischnig-Piesczek v Facebook Ireland Limited* EU:C:2019:821, Opinion of AG Szpunar.

8 Andrew Liptak, 'Poland has filed a complaint against the European Union's copyright directive' (*The Verge*, 25 May 2019) <https://www.theverge.com/2019/5/25/18639963/poland-european-union-copyright-directivefiled-complaint-court-of-justice> accessed 6 February 2020.

9 Christina Angelopoulos and others, 'Study of Fundamental Rights Limitations for Online Enforcement through Self-Regulation' Institute for Information Law (IViR) (2015).

10 Mixed Messages?, 'The limits of automated social media content analysis' (*Center for Democracy and Technology (cdt)*, November 2017), <https://cdt.org/files/2017/11/Mixed-Messages-Paper.pdf>; Maarten Sap and others, 'The Risk of Racial Bias in Hate Speech Detection' (2019) University of Washington <https://homes.cs.washington.edu/~msap/pdfs/sap2019risk.pdf>.

lawful communications.'[11] In two copyright cases, it noted this concern about information and expression rights, and rejected filtering injunctions as incompatible with eCommerce Directive Art. 15.[12] The European Court of Human Rights (ECtHR) has reached a similar conclusion on grounds of fundamental rights alone. In *Magyar Tartalomszolgáltatók Egyesülete (MTE) v Hungary*, it ruled that Hungary violated Art. 10 of the European Convention by holding a hosting platform strictly liable for defamation posted by its users – a standard that would effectively require the platform to proactively monitor users' posts in order to avoid liability.[13] The analysis of the European Court of Human Rights looked beyond the risk of erroneous removals, identifying more systemic concerns that such legal obligations might lead platforms 'to close the commenting space altogether[.]'[14] That has proved prescient in recent years, with ever fewer small companies offering open forums for online expression – and major incumbents like Facebook or YouTube increasingly relying on Terms of Service to prohibit lawful speech and avoid controversy or risk in legal grey areas.

Human rights officials[15] and civil society organizations[16] have recently raised the alarm about filters and fundamental rights in response to the EU's draft Terrorist Content Regulation. As this Opinion went to press, the Regulation was in trilogue discussions to reconcile the Commission, Council, and Parliament legislative drafts – two of which would empower Member State authorities to order platforms to deploy filters.[17] One coalition of civil society organizations called the draft filtering mandate 'a gamble with European Internet users' rights to privacy and data protection, freedom of expression and information, and non-discrimination and equality before the law.'[18] Because filters cannot assess the context in which information appears, they noted, such a law would predictably lead platforms to suppress material that is unlawful in one context (such as an ISIS recruitment video) yet lawful when re-used in new ways (such as news reporting or counterspeech). The letter also pointed to the example of the German non-profit Syrian Archive, which lost over 100,000 videos documenting abuses in Syria from its YouTube channel – seemingly as the result of a filtering error.

Expert discussion of filters' errors or over-reach has often focused on the potential for human oversight to correct mistakes and protect lawful information. The Council of Europe, for example, has said that '[d]ue to the current limited ability of automated means to assess context,' platforms that use filters should 'ensure human review where appropriate.'[19] The EU Commission offered similar guidance in 2018, saying that if hosting providers choose to rely on 'automated means' to review content, they should provide 'effective and appropriate safeguards' such as human review to ensure that 'decisions to remove or disable access to content considered to be illegal content are accurate and well-founded.'[20] As will be discussed below, however, *Glawischnig-Piesczek* complicates the task of courts or lawmakers who may seek to require human judgment as an element of proactive monitoring regimes.

3. The technology

The CJEU's discussion in *Glawischnig-Piesczek* left open a very large question about what exactly the Austrian courts' injunction requires Facebook to do. It is clear, however, that the only way a platform of Facebook's size can proactively block specific content is by using content filters, or what the Court calls 'automated search tools and technologies.'[21] The exact specifications of these technologies will matter greatly for their impact on fundamental rights. For example, if Facebook had to automatically block every instance of the *text phrases* 'fascist party', 'lousy traitor', and 'corrupt oaf', it would almost certainly take down numerous lawful posts – news coverage, legitimate political commentary, or teasing between friends, for example. If Facebook's obligation instead is to block posts that combine these words with any *photo-*

11 Case C-360/10 *SABAM v Netlog NV* EU:C:2012:85 = GRUR Int 2012, 350, para 50; Case C-70/10 *Scarlet Extended SA v SABAM* EU:C:2011:771 = GRUR Int 2012, 153.

12 ibid.

13 *Magyar Tartalomszolgáltatók Egyesülete (MTE) v Hungary*, App No 22947/13 (ECtHR, 2 February 2016), para 82 (strict liability for platform 'allowing unfiltered comments' in defamation case violated Convention art 10); *compare Delfi AS v Estonia*, App No 64569/09, (ECtHR, 16 June 2015) (strict liability in hate speech case did not violate art 10); Daphne Keller, 'New Intermediary Liability Cases from the European Court of Human Rights: What Will They Mean in the Real World?' (*Center for Internet and Society*, 11 April 2016) <http://cyberlaw.stanford.edu/blog/2016/04/new-intermediary-liability-cases-european-court-human-rights-what-will-they-mean-real>.

14 *MTE* (n 13) para 86. CJEU precedent has not focused on the existence of open forums as a Charter art 11 concern, but has noted the threat to the commercial viability of forums, saying that a filtering mandate would 'result in a serious infringement of the freedom of the hosting service provider to conduct its business since it would require that hosting service provider to install a complicated, costly, permanent computer system at its own expense'. *Netlog* (n 11) para 46.

15 David Kaye, Joseph Cannataci and Fionnuala Ní Aoláin, 'Mandates of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression; the Special Rapporteur on the right to privacy and the Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism' (7 December 2018) <https://spcommreports.ohchr.org/TMResultsBase/DownLoadPublicCommunicationFile?gId=24234>. See also Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, '2018 thematic report to the Human Rights Council on content regulation' (*United Nations Human Rights Office of the High Commissioner*, 2018) <https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/ContentRegulation.aspx>.

16 Access Now and others, 'Letter to Members of European Parliament' (*Center for Democracy and Technology (cdt)*, 4 February 2019) <https://cdt.org/files/2019/02/Civil-Society-Letter-to-European-Parliament-on-Terrorism-Database.pdf>; Article 19 and others, 'Joint letter on European Commission regulation on online terrorist content' (*Article 19*, 6 December 2018) <https://www.article19.org/resources/joint-letter-on-european-commission-regulation-on-online-terrorist-content/>; WITNESS and others, 'Letter to the Committee on Civil Liberties, Justice and Home Affairs' (28 January 2019) <https://drive.google.com/file/d/1WTgl5hjJ_cAE1U0OjqaQ9AucU6HNlhoi/view>.

17 cf Commission, 'Proposal for a Regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online' COM(2018) 640, art 6; Council, 'Proposal for a Regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online - general approach' 2018/0331(COD), art 6; Parliament, 'European Parliament legislative resolution of 17 April 2019 on the proposal for a regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online' COM(2018)0640 – C8-0405/2018 – 2018/0331(COD), art 6.

18 Access Now (n 16).

19 Recommendation CM/Rec(2018)2 of 7 March 2018 of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries [2018].

20 Commission, 'Recommendation on measures to effectively tackle illegal content online', C(2018) 1177.

21 ibid para 46.

graphic image of the plaintiff (as the Austrian court seemingly intended[22]), an additional and separate set of issues about data protection would arise. To comply, Facebook would seemingly have to carry out biometric facial recognition scans affecting its other users, as well as non-Facebook users depicted in photos uploaded to the platform. Other filter variations, such as blocking 'shares' of the entire original post defaming Ms. Glawischnig-Piesczek, can be imagined, but were not identified or discussed in the case to date.[23]

Courts considering the likely consequences of monitoring orders will need reliable information about real-world filtering technologies. The most commonly deployed filters today are designed to find *duplicates* of known, specific content such as images, audio, or videos. Many major platforms rely on duplicate-detection filters like PhotoDNA[24] to find child sexual abuse images or videos, and to find violent extremist images or videos.[25] Duplicate-detection filters are not perfect, but the most sophisticated ones can often find near-duplicates, like images that have been cropped.[26] Duplicate-detection filters for written text are technically simpler and have existed for decades – their basic function is familiar to anyone who has used the 'find' function in a browser or a text editor like Microsoft Word. As tools for restricting human communication, text filters are notoriously error-prone, because specific words or phrases can so easily be unlawful in one situation but innocuous in another. Many comic examples of this problem are documented in Wikipedia's entry for the 'Scunthorpe problem.'[27] There are very serious examples, too. Numerous victims of racially-based harassment have gone online to describe their experiences, for example, only to be penalized or locked out of social media for repeating the specific slurs and racial epithets used by their attackers.[28]

Facebook's role as the 'anti-filtering' voice in *Glawischnig-Piesczek* is a complicated one, since in many other contexts – particularly concerning videos or images that support violent extremism – the company has been a major proponent of automated filtering tools. Most visibly, Facebook CEO Mark Zuckerberg has testified optimistically about future machine learning or artificial intelligence-based content moderation.[29] The company also helped create and promote a widely-used (and widely-criticized) tool for detecting duplicates of known extremist images and videos.[30] It currently reports a high rate of proactive, machine-based detection for the content it classes as 'terrorist propaganda.'[31]

The exact capabilities of content-detection tools used by Facebook – and other major platforms, including Google, which reported spending $100 million on YouTube's filtering system – are not publicly known.[32] Some portion of the platforms' proactive measures presumably consist of conventional duplicate detection. Other mechanisms likely include threat profiling based not on *content* but on uploaders' *behavior* (using spam-fighting tools to flag suspicious patterns of contacts, followers, or posting locations, for example).[33] Beyond these relatively longstanding means of identifying prohibited content, Facebook is known to be working on artificial intelligence tools to analyze novel material. Experts warn, however, that while such tools are getting better at challenges like distinguishing broccoli from marijuana, they are very far from parsing the nuances of human communication.[34] Review by humans remains critical for discerning the message conveyed by both new material and old material re-used in new contexts.

4. The evolving legislative landscape

Lawmakers in Brussels have debated filtering proposals twice in recent legislative processes. They are likely to do so again in the pending Digital Services Act (DSA) – a major legislative initiative expected to update the eCommerce Directive. The Copyright Directive, which passed in a series of hotly contested votes in 2019, created a de facto filtering mandate by requiring platforms to 'prevent further uploads' of specific works.[35] That Directive provides that

22 For the second instance Austrian court, 'the reference to "identically worded items of information" was to publications of *photographs of the applicant* with the same accompanying text'. (AG Op para 56, italics altered).

23 See Keller (n 4) 19-22 (discussing possible filter designs and their issues).

24 'How does PhotoDNA technology work?' (*Microsoft*) <https://www.microsoft.com/en-us/photodna> accessed 6 February 2020.

25 Olivia Solom, 'Facebook, Twitter, Google and Microsoft team up to tackle extremist content' *The Guardian* (London, 6 December 2016) <https://www.theguardian.com/technology/2016/dec/05/facebook-twitter-google-microsoft-terrorist-extremist-content> accessed 6 February 2020.

26 See generally, Evan Engstrom and Nick Feamster, 'The Limits of Filtering: A Look at the Functionality and Shortcomings of Content Detection Tools' (*Engine*, March 2017) <www.engine.is/the-limits-of-filtering>.

27 'Scunthorpe Problem' (*Wikipedia*) <https://en.wikipedia.org/wiki/Scunthorpe_problem> accessed 6 February 2020.

28 Tracy Jan and Elizabeth Dwoskin, 'A white man called her kids the n-word. Facebook stopped her from sharing it' *The Washington Post* (Washington, D.C., 31 July 2017) <https://www.washingtonpost.com/business/economy/for-facebook-erasing-hate-speech-proves-a-daunting-challenge/2017/07/31/922d9bc6-6e3b-11e7-9c15-177740635e83_story.html> accessed 6 February 2020.

29 Sydney Li and Jamie Williams, 'Despite What Zuckerberg's Testimony May Imply, AI Cannot Save Us' (*Electronic Frontier Foundation*, 11 April 2018) <https://www.eff.org/deeplinks/2018/04/despite-what-zuckerbergs-testimony-may-imply-ai-cannot-save-us> accessed 6 February 2020.

30 Solom (n 25).

31 'Community Standards Enforcement' (*Facebook*) <https://transparency.facebook.com/community-standards-enforcement#terrorist-propaganda> accessed 6 February 2020.

32 Paul Sawers, 'YouTube: We've invested $100 million in Content ID and paid over $3 billion to rightsholders' (*VentureBeat*, 7 November 2018) <https://venturebeat.com/2018/11/07/youtube-weve-invested-100-million-in-content-id-and-paid-over-3-billion-to-rightsholders/> accessed 6 February 2020.

33 cf Natasha Lomas, 'Twitter claims tech wins in quashing terror tweets' (*TechCrunch*, 19 September 2017) <https://techcrunch.com/2017/09/19/twitter-claims-tech-wins-in-quashing-terror-tweets/> accessed 6 February 2020.

34 Cade Metz and Mike Isaac, 'Facebook's A.I. Whiz Now Faces the Task of Cleaning It Up. Sometimes That Brings Him to Tears.' *The New York Times* (New York, 17 May 2019) <https://www.nytimes.com/2019/05/17/technology/facebook-ai-schroepfer.html> accessed 6 February 2020; James Vincent, 'Using AI to screen live video of terrorism is "very far from being solved," says Facebook AI chief' (*The Verge*, 20 May 2019) <https://www.theverge.com/2019/5/20/18632260/facebook-ai-spot-terrorist-content-live-stream-far-from-solved-yann-lecun> accessed 6 February 2020; Emma Llansó and others, 'Artificial Intelligence, Content Moderation, and Freedom of Expression' (Transatlantic Working Group, February 2020), <https://www.ivir.nl/publicaties/download/AI-Llanso-Van-Hoboken-Feb-2020.pdf>.

35 Council Directive 2019/790 of 17 April 2019 [2019] O.J. (L 130), art. 17.4; Michelle Kaminsky, 'EU's Copyright Directive Passes Despite Widespread Protests – But It's Not Law Yet' (*Forbes*, 26 March 2019) <https://www.forbes.com/sites/michellekaminsky/2019/03/26/eus-copyright-directive-passesdespite-widespread-protestsbut-its-not-law-yet/#24b0d6902493> accessed 6 February 2020.

'decisions to disable access to or remove uploaded content shall be subject to human review.'[36] The second recent filtering proposal, in the Terrorist Content Regulation, was, at the time this essay was published, in trilogue proceedings to reconcile the Commission and Council drafts (which include filtering mandates) with the Parliament draft (which does not).[37] All three drafts say that hosts using 'automated tools' to assess user content 'shall provide effective and appropriate safeguards' against improper removals, consisting 'in particular, of human oversight and verifications' of filters' decisions – though only in the Parliament draft is such human review clearly mandatory.[38] Both the Copyright Directive and the Terrorist Content Regulation are ambiguous about the new mandates' relationship with eCommerce Directive Art. 15.[39]

III. Analysis

Glawischnig-Piesczek leaves a series of unresolved issues for both courts and lawmakers about platform content moderation and filtering. It clears the way for courts to order some use of filters under the eCommerce Directive. And it leaves open the possibility for platforms to *voluntarily* employ human reviewers to carry out 'independent assessment' of those filters' output. But courts cannot *require* such assessment, because of the Court's interpretation of Art. 15. That will make it harder for courts themselves to devise orders that comport with fundamental rights.

In this Section I will discuss (1) the Court's reasoning about Art. 15, (2) the resulting fundamental rights questions for filtering injunctions, and (3) procedural concerns about *Glawischnig-Piesczek* and other cases in which litigation between private parties affects the rights and interests of the public and absent third-party internet users.

1. The Court's analysis of Art. 15

The Court in *Glawischnig-Piesczek* interprets Art. 15 to permit monitoring orders for content specified by a court, including content 'equivalent' to the post initially assessed and deemed illegal.[40] Such orders, however, must be capable of being carried out without platforms providing 'independent assessment' of the filtered content's meaning or legality.[41] As I will discuss in this Subsection, the Court's reasoning in support of this conclusion departs significantly from prior cases, and raises important questions about courts' leeway to mandate outcomes without examining the mechanisms platforms will use to achieve them – or the results for fundamental rights.

Article 15 of the eCommerce Directive says that Member States cannot impose on hosts any 'general obligation to monitor the information which they transmit or store[.]' Other Articles, including the core immunity for hosts in Art. 14, indicate that courts can order platforms to 'prevent' legal violations, however, and Recital 47 states that Art. 15 does not preclude 'monitoring obligations in a specific case.' In *L'Oréal v eBay*, the CJEU discussed the boundaries between prohibited 'general' monitoring and permissible 'specific' injunctions.[42] '[T]he measures required of the online service provider,' it explained, 'cannot consist in an active monitoring of *all* the data of *each* of its customers.'[43] By contrast, it said, courts could potentially issue more specific orders for a host to terminate a particular user's account, or make that user easier to identify.[44] Similarly, in *Tommy Hilfiger v Delta Center A.S.*, the CJEU held that courts could not require 'general and permanent oversight' over *all* customers, but could require measures aimed at 'avoiding new infringements of the same nature by the *same*' customers.[45]

In *Glawischnig-Piesczek*, the Court moves away from the standards set forth in these earlier cases, without citing or discussing them. The injunction it approves would seemingly require Facebook to monitor every post by every customer. Instead of defining prohibited 'general' monitoring as *monitoring that affects every user*, the Court effectively defines it as *monitoring for content that was not specified in advance by a court*. Permissible monitoring orders for identical and equivalent content, it holds, can cover all users – not just the person who originally 'requested the storage of that information.'[46]

This interpretation of the line between general and specific monitoring drives the Court's analysis of which content injunctions may cover. Under Art. 15, it concludes, platforms can be compelled to monitor for specific content 'which was examined and assessed by a court … which, following its assessment, declared it to be illegal.' (para. 35) This can include both 'identical' and 'equivalent' content, so long as the court's order defines the content precisely enough to avoid requiring independent judgment by the platform. As the CJEU explains,

> '[I]t is important that the equivalent information … contains specific elements which are properly identified in the injunction, such as the name of the person concerned by the infringement determined previously, the circumstances in which that infringement was determined and equivalent content to that which was declared to be illegal. Differences in the wording of that equivalent content, compared with the content which was declared to be illegal, must not, in any event, be such as to require the host provider concerned to carry out an independent assessment of that content.'[47]

This standard and the requirements of Art. 15 can be met, the Court continues, so long as any monitoring is 'limited to information containing the elements specified in the injunction, and its defamatory content of an

36 Directive 2019/790, art 17.9.
37 See n 18.
38 cf art 9.2 in the Commission, Council, and Parliament drafts (n 17).
39 The Commission Draft of the Terrorist Content Regulation, for example, states both that its requirements 'should not, in principle, lead to the imposition of a general obligation to monitor' in contravention of art 15, and that those requirements 'could derogate from the approach established in art. 15(1)'. Recital 19.
40 *Glawischnig-Piesczek* para 45; see also *SABAM* (n 11) para 34 (court cannot require intermediary 'to actively monitor all the data of each of its customers'); *Scarlet Extended* (n 11) para 36 (same).
41 *Glawischnig-Piesczek* para 45.
42 Case C-324/09 *L'Oréal SA v. eBay* EU:C:2011:474 = GRUR Int 2011, 839.
43 ibid para 139 (emphasis added).
44 ibid para 141-142.
45 Case C-494/15 *Tommy Hilfiger Licensing LLC and others v Delta Center A.S.* EU:C:2016:528 = GRUR Int 2016, 925, para 34 (interpreting the *L'Oreal* standard in a case involving a physical marketplace under art 11 of Directive 2004/48) (emphasis added).
46 *Glawischnig-Pieszcek* para 37.
47 ibid para 45.

equivalent nature does not require the host provider to carry out an independent assessment, *since the latter has recourse to automated search tools and technologies*.' (paras. 46-47, emphasis added.) This definition largely collapses the difference between 'equivalent' and 'identical' content, since both must be identified in the injunction with sufficient specificity to allow 'automated search tools and technologies' to reliably carry out a court's order.

How much human review of filters' work can a court require without violating *Glawischnig-Piesczek*'s rule against 'independent assessment'? At a minimum, there must be some room for platform employees to build and test the 'automated search tools and technologies' the CJEU envisions. Once the filter is in operation, perhaps platform employees can review its output for purely technical errors, when a filter flags or blocks the wrong content entirely, without crossing the line into providing 'independent assessment.' But in the cases that need human judgment the most – parodies, news reports, and other material flagged by filters for re-using the same content in a new way – courts cannot require platforms to provide that judgment.

Platforms implementing filtering orders under *Glawischnig-Piesczek* can still *choose* to provide human evaluation. Well-resourced platforms like Facebook or Google may do that. (One reason to do so would be to avoid potentially triggering duties of 'algorithmic explanation' for purely automated decision-making under Art. 22 of the General Data Protection Regulation.) Other platforms, concerned about the expense and risk of such review, will presumably take down anything flagged by a filter – just as many, today, honor any takedown demand without examining its validity.[48] Even platforms that choose to have employees check the decisions of filters, however, are incentivized to remove legal 'gray area' content in a filter-and-human-review system. The act of review can itself cause platforms to, in the AG's words, 'lose the status of intermediary service provider and the immunity that goes with it' – which provides a strong reason to err on the side of caution and take down content flagged by filters, whether or not it is covered by the injunction.[49] This legal incentive compounds the pre-existing practical risk of human review providing what Ben Wagner has called a 'rubber-stamping mechanism in an otherwise completely automated decision-making system.'[50]

By specifying an *outcome* (identical and equivalent content must come down) without examining the *means* by which platforms can achieve it, *Glawischnig-Piesczek* creates interesting tensions with other cases. One is the ECtHR's *MTE* ruling, which rejected the imposition of strict defamation liability on a platform. The Hungarian ruling reviewed in *MTE*, like the Austrian order in *Glawischnig-Piesczek*, did not spell out what the defendant hosting platform had to do. It only specified an outcome: to avoid liability, the platform must keep defamatory content from appearing. The ECtHR held that this created such 'foreseeable negative consequences' for the rights of third parties posting on the platform that it violated Art. 10 of the European Convention.[51] *MTE* differed from *Glawischnig-Piesczek* in that it involved potential liability for *any* defamatory content, not just content identified by a court. But it was also similar, in that the unintended harms to internet users' rights arose from a facially lawful injunction, requiring removal of unlawful information without examining the practical consequences for lawful information and expression.

The CJEU considered open-ended injunctions and the mechanics of blocking internet content in another case, *Telekabel Wien v Constantin Film*. There, the Court approved an injunction that required internet access providers to block particular websites, but did 'not specify the measures which that access provider must take'.[52] To avoid conflict with fundamental rights, the court held, such an injunction must include mechanisms to protect both the platform's rights (by allowing it to defend its implementation as reasonable in court) and those of internet users.

Access providers themselves, the *Telekabel* Court said, must 'ensure compliance with the fundamental right of internet users to freedom of information' by adopting measures that 'bring an end' to the specified infringement 'without thereby affecting internet users who are using the provider's services in order to lawfully access information.'[53] To protect against errors, 'the national procedural rules must provide a possibility for internet users to assert their rights before the court once the implementing measures taken by the internet service provider are known.'[54] The *Glawischnig-Piesczek* court did not examine this precedent, but some points of contrast between the risk to fundamental rights in the two cases seem clear. First, the risk of over-blocking appears considerably lower in *Telekabel Wien*, since it involved blocking only a specific website already examined by the court.[55] Second, users affected by an access provider's blocking errors in the *Telekabel* situation would presumably have a meaningful opportunity to learn of the problem when they could not access a site, making them better able (if, in reality, not particularly likely) to protect their rights in court.[56] In any case, the two rulings imply different degrees of involvement for internet companies themselves in protecting users' rights. While *Telekabel Wien* charged the platform with 'ensur[ing] compliance with the fundamental right of internet users', *Glawischnig-Piesczek* tells us that courts may not require platforms to carry out a key

48 Jennifer Urban and others, 'Notice and Takedown in Everyday Practice' UC Berkeley School of Law (2017) 41 <https://ssrn.com/abstract=2755628> accessed 5 February 2020.

49 AG Opinion (n 7) para. 36; see discussion in Keller (n 4) 31-35.

50 Ben Wagner, 'Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision-Making Systems' Policy & Internet 2019, 104-122 <https://doi.org/10.1002/poi3.198>.

51 *MTE* (n 13) para 86. Compare *CDT v Pappert* 337 F Supp 2d 606 (ED Penn 2004) (U.S. case striking down law requiring ISPs to block child pornography sites as insufficiently narrowly tailored where ISPs' compliance would foreseeably be overbroad and block lawful sites).

52 C-314/12 *UPC Telekabel Wien GmbH v Constantin Film Verleih GmbH and Wega Filmproduktionsgesellschaft mbH* EU:C:2014:192 = GRUR Int 2014, 469, para 64.

53 ibid paras 54-56.

54 ibid para 57.

55 It ordered the access provider to block that site by 'domain name and current IP ("Internet Protocol") address and any other IP address of that site' known to the provider. Para 12.

56 This is particularly true if Member State courts require access providers to notify users who attempt to access those pages, as the UK court did in *Cartier Int'l AG v British Sky Broadcasting* [2014] EWHC 3354 (Ch), paras 262-265.

step to protect those rights – assessing the content that is being blocked.[57]

2. Fundamental rights questions in post-*Glawischnig-Piesczek* filtering decisions

Courts issuing orders under *Glawischnig-Piesczek*'s newly clarified Art. 15 guidelines will face difficult questions about the orders' proportionality and compliance with fundamental rights. The limits on human review, in particular, will increase the difficulty of devising orders that, as the CJEU put it in *Telekabel Wien*, 'prevent the fundamental rights recognised by EU law from precluding the adoption of an injunction.'[58]

a) Basic fundamental rights questions about filters

Some high-level fundamental rights questions can be expected to arise in any case involving filters. For example, before issuing a filtering injunction, must a court first determine that the prohibited image, text, or other content will foreseeably violate the law in *every* new context where it might be re-used? If some portion of re-uses will be legal, or if a filter will foreseeably make mistakes, how does that affect fundamental rights assessment? When a court balances the filter's benefits for a claimant like Ms. Glawischnig-Piesczek against the rights of unknown future people affected by the errors, which rights must it consider?

Balancing a particular filter's benefits and harms requires an understanding of its real-world operations and consequences, including changing consequences over time as the filter continues to operate. In assessing the benefits side of the equation, filters designed to protect economic interests in intellectual property may lend themselves to the simplest assessment, but even their value is disputed.[59] In a case involving reputation and dignity rights, like *Glawischnig-Piesczek*, a filter's actual benefit for the plaintiff will likely depend on its rate of false positives and false negatives – and the resulting risk of critical 'Streisand effect' attention to the plaintiff. For filters intended to serve more complex goals, questions about benefits are even more fraught. It is not known whether filters intended to counter violent extremism, for example, are ultimately effective in reducing real-world violence. Experts have noted that over-zealous or poorly designed filtering and removal efforts could ultimately backfire – driving potential terrorist recruits into echo chambers in darker corners of the internet; silencing moderate voices in communities vulnerable to radicalization; and fueling mistrust and anger within those communities.[60]

The potential downsides of filters are more widely discussed in the fundamental rights literature, but still poorly understood, largely owing to the lack of transparency about platforms' existing filtering efforts. On that side of the fundamental rights balance are the interests of all internet users affected by filters – who, in Facebook's case, number in the billions.[61] As discussed in Section II.2, the affected rights can include privacy and data protection for all users, regardless of the errors produced by filters. For users affected by wrongful removals, they include rights to free expression and information, fair trial and effective remedy, and equality and non-discrimination under law. Such burdens are hard to quantify, but are unavoidably significant given the sheer scale of platform operations. For Facebook, processing over half a million comments every minute,[62] even an error rate of 0.01% would lead to the wrong outcome 50 times each minute, or 72,000 times each day.

b) Correcting for filters' mistakes

The *Glawischnig-Pieszcek* analysis of Art. 15, which precludes requiring 'independent assessment' of content, introduces another question: what filters might be reliable enough to adequately protect users' fundamental rights without human supervision and judgment? In the simplest case, for an order covering content that has *no* lawful uses in other contexts – as is the case for child sexual abuse material in many countries – this would be purely a question of technical accuracy in identifying duplicates. For content capable of lawful new uses, though, relying on automated content moderation without the addition of human judgment poses substantially greater challenges. The words or images at issue in *Glawischnig-Pieszcek*, which are potentially of interest to the public, journalists, scholars, and legal practitioners, seem particularly likely to reappear in new, legal contexts – this Opinion is an example. Extremist content raises similar problems, since it can reappear in news coverage, scholarly analysis, or counterspeech by opponents of violence. It is hard to imagine a filtering system that can protect such information and expression without relying on additional human judgment.

Even if courts could require platforms to carry out human review, it is unclear how well such review would correct for filters' mistakes. As discussed above, humans may merely rubber-stamp decisions produced by filters – and have incentives to do so in order to avoid legal risk. Indeed, researchers have identified high rates of over-removal and error even in purely human-operated notice and takedown systems.[63] Other means of correcting errors might include notifying affected users and allowing them to challenge removals using platform-operated appeal or 'counter-notice' systems. The efficacy of those systems, though, is questionable – and at best they provide a remedy for speakers, but not for users unknowingly deprived of access to information. In any case, court orders requiring platform-operated appeals might still require too much 'independent assessment' to be permitted under *Glawischnig-Pieszcek*. If requiring platform judgment is not an option, users may be left with only hard-to-

57 *Telekabel Wien* (n 52) paras 54-56.
58 *Telekabel Wien* (n 52) para 57.
59 See generally Martin Husovec, 'The Promises of Algorithmic Copyright Enforcement: Takedown or Staydown? Which Is Superior? And Why?' 42 COLUM. J.L. & ARTS 53 (2018).
60 See Daphne Keller, 'Internet Platforms: Observations on Speech, Danger, and Money' (*Hoover Institution*, 2018) 20-26 <https://www.hoover.org/sites/default/files/research/docs/keller_webreadypdf_final.pdf>.
61 'Company Information' (*Facebook*) <https://newsroom.fb.com/company-info/> accessed 6 February 2020.
62 ibid.
63 See studies cited in Daphne Keller, 'Empirical Evidence of "Over-Removal" by Internet Companies under Intermediary Liability Laws' (*Center for Internet and Society*, 2015) <http://cyberlaw.stanford.edu/blog/2015/10/empirical-evidence-over-removal-internet-companies-under-intermediary-liability-laws>.

exercise corrective measures that depend on judgment from someone other than the platform, like the judicial review described in *Telekabel Wien*.[64]

3. Litigation process problems

Glawischnig-Piesczek illustrates a problem that is common in intermediary liability litigation: no party before the court truly shares or represents the interests of internet users who will be affected by the case's outcome.[65] As the above discussion of filters illustrates, this is not just a matter of protecting the individual user whose post gave rise to the case. By shaping platforms' future actions, rulings can affect people throughout the internet's information ecosystem. One basic problem is that neither the plaintiff nor the platform has proper incentives to explain the shortcomings of technologies like filters. Platforms that depend on algorithmic content analysis for core economic functions like ad targeting, or that are banking on ambitious future AI-driven business models, are ill-suited to educate courts about potential failures of those very technologies.

Another problem is that while platform and user interests are sometimes aligned for questions about expression and information rights, they are less likely to be aligned for other rights including privacy and data protection. Platforms that already face data protection claims and investigations for their automated content processing are unlikely to argue, in a case like *Glawischnig-Piesczek*, that automated content filtering orders conflict with users' data protection rights.[66] Indeed, no party in *Glawischnig-Piesczek* seems to have briefed the courts about data protection concerns, despite CJEU precedent on the point.[67] Even the AG's Opinion – which, unlike the CJEU ruling, discusses fundamental rights – does not consider data protection arguments. Nor do any courts or briefs to date seem to have assessed the serious additional data protection issues raised by the Austrian court's initial order for Facebook to monitor for *images* of Glawischnig-Piesczek – which seemingly would require biometric scans of innumerable other people's pictures.

This lack of representation for user interests is not unique to *Glawischnig-Piesczek*. It is a structural problem in intermediary liability cases generally. That makes the difficulty of intervening at the CJEU particularly problematic in these cases, and makes active judicial concern for fundamental rights essential as Member State courts consider requests for monitoring injunctions in the wake of *Glawischnig-Piesczek*.

IV. Conclusion

The CJEU's ruling in *Glawischnig-Piesczek* appears to broadly support internet monitoring injunctions. But it leaves major and unresolved questions for courts seeking to reconcile such injunctions with fundamental rights. Devising orders or filtering technologies that fulfill the CJEU's specifications under Art. 15 while simultaneously respecting EU Charter requirements will be, at the very least, challenging.

ACKNOWLEDGEMENTS

I am grateful for feedback, review, and insights shared by Joan Barata, Jennifer Daskal, Othon Flores Juarez, Joris van Hoboken, Martin Husovec, Kate Klonick, Joe McNamee, Graham Smith, and Ben Wagner for this piece and its predecessors. This Opinion was largely completed during my time at the Stanford Center for Internet and Society (CIS), with additional support from the Knight Foundation. CIS funding information is available at <http://cyberlaw.stanford.edu/about-us>.

64 More creatively, platforms themselves might be required to route gray-area cases to courts, regulators, or even plaintiffs for evaluation. The 'rubber-stamping' risk would be strong if plaintiffs provided the human judgment to supplement decisions produced by filters. But at least they – unlike platforms – could opt against removal in legal gray area cases without incurring legal risk to themselves.

65 Five government interveners – Austria, Latvia, Portugal, Finland, and the EU Commission – did submit confidential filings. Government briefs, however, may not be the best sources for insight into evolving technologies or advocacy for expression and information rights in a case about vulgar criticism of a politician.

66 See, eg, Russell Brandom, 'Facebook and Google hit with $8.8 billion in lawsuits on day one of GDPR' (*The Verge*, 25 May 2018) <https://www.theverge.com/2018/5/25/17393766/facebook-google-gdpr-lawsuit-max-schrems-europe> accessed 6 February 2020.

67 *SABAM* (n 11). Also potentially relevant are cases like C-293/12 and C-594-12 *Digital Rights Ireland Ltd. v Ireland* EU:C:2014:238 (rejecting data retention law that required electronic communications service providers to retain data about communications made by all of their subscribers).