Stanford Law School | Stanford Center for Internet and Society

Dolphins in the Net: Internet Content Filters and the Advocate General's *Glawischnig-Piesczek v. Facebook Ireland* Opinion

Daphne Keller
Director of Intermediary Liability

2019

559 Nathan Abbot Way Stanford, CA 94305 cyberlaw.stanford.edu

Dolphins in the Net:
Internet Content Filters and the Advocate General's *Glawischnig-Piesczek v. Facebook Ireland* Opinion

Daphne Keller[1]
Stanford Center for Internet and Society
September 4, 2019

Table of Contents

[1] Director of Intermediary Liability, Stanford Center for Internet and Society (Stanford CIS); former Associate General Counsel to Google Inc. Stanford CIS's funding and other information is at http://cyberlaw.stanford.edu/about-us.

I. Introduction

One of the most important Internet law cases in recent years, *Glawischnig-Piesczek v. Facebook Ireland*, is currently pending before the Court of Justice of the EU (CJEU). The case, which concerns a Facebook user's vulgar comments about an Austrian politician, has received surprisingly little attention. The Advocate General (AG) issued his influential Opinion in June, and his recommendation for the Court's final ruling is troubling.[2] It would open the door to court orders requiring platforms to automatically detect and filter out particular content posted by users. It would also let national courts order platforms to globally remove users' posts, even in countries where the posts would be legally protected expression or information. The AG's analysis has serious implications for the EU's Intermediary Liability legal framework as well as currently-pending legislative changes. We should hope the Court does not adopt it.

As the European Commission pointed out at the hearing, the Court's Judgment will have consequences far beyond Facebook or this specific case. It is likely to shape the behavior of both large and small Internet platforms for years to come. By doing so, it will indirectly but seriously affect Internet users' rights, including rights to privacy and freedom of expression and information. The Court's conclusions will also likely influence the choices EU lawmakers make as they revise the EU's primary Intermediary Liability law, the eCommerce Directive.

The dispute in this case started when a Facebook user posted a news article about Eva Glawischnig-Piesczek, who then headed Austria's Green Party. The user's comments next to the article criticized the party's policy on refugees, and called its leader a "lousy traitor" (*miese Volksverräterin*), a "corrupt oaf" (*korrupter*

[2] Opinion of Advocate General Szpunar, Case C-18/18, *Eva Glawischnig-Piesczek v. Facebook Ireland Limited*, (June 4, 2019) http://curia.europa.eu/juris/document/document.jsf?text=&docid=214686&pageIndex=0&doclang=en&mode=req&dir=&occ=first&part=1&cid=7255506.

Trampel), and a member of a "fascist party" (*Faschistenpartei*).[3] Vulgar terms like that in a political context would be protected expression in many countries. But Glawischnig-Piesczek told Facebook they constituted defamation in Austria. Facebook took the position that the comments were not clearly unlawful. Therefore, it argued, under Austrian law, a court rather than a private company should decide whether the post should come down—and whether Glawischnig-Piesczek's reputational rights outweighed the user's expression rights and other users' rights to access information.

Glawischnig-Piesczek initiated two proceedings. In one, a criminal case, the court ultimately said that the post was not clearly unlawful. In the second, a civil case, a different court said that the post *was* obviously unlawful, and that Facebook was therefore liable for failing to remove it after receiving notice. The civil case is the one on appeal here. In an expedited preliminary injunction proceeding, the first instance court ordered Facebook to proactively monitor and block "identical" and "equivalent" posts in the future. On appeal, a higher court upheld a monitoring requirement for "identical" posts. Both parties then appealed to the Austrian Supreme Court. The appeal raised both the filtering issue and an issue not addressed in the lower courts: whether Facebook must remove the post globally, even in countries where it is legal.[4] The Austrian Supreme Court referred both the question about proactive monitoring and the question about global removal to the CJEU. In his Opinion, the AG advises the Court to rule that both are permissible in certain circumstances.

In this White Paper, I identify problems with the AG's recommendations, focusing in particular on the issue of filters. Many of the problems stem from the rushed process by which the case reached the CJEU, and the limited voices and

[3] The full post, informally translated, read "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."

[4] The Austrian appellate court analyzed whether Austrian law applied to the case, but not what geographic scope the injunction should have.

perspectives represented in the litigation. I will review this procedural shortcoming in Section II. The limited briefing on important aspects of the filtering issue seems to have affected the AG's conclusions about both fundamental rights and the eCommerce Directive. In Section III, I will analyze those legal conclusions, and suggest that they are inconsistent with the requirements of the EU Charter and the Court's precedent. In particular, they do not adequately address the problem of "dolphins in the net" – lawful user communications that are accidentally blocked by filters. Finally, in Section IV, I will very briefly discuss the issue of global content removal.

The AG in *Glawischnig-Piesczek*, Maciej Szpunar, is a capable jurist who has worked on important related cases. That includes two pending disputes about Google's implementation of the "Right to Be Forgotten," one of which raises a question analogous to the one in this case about global enforcement.[5] For the filtering questions in the Facebook case, unfortunately, his analysis of core issues is opaque. He says courts can compel Facebook to filter "identical" information across the entire platform, as well as "equivalent" information posted by the same user. But it is not clear what information he considers "identical" or "equivalent." Nor is it apparent what portion of the material blocked by the proposed filter would actually violate the law, or what corrective measures would be available if a filter suppressed the wrong expression and information.

That imprecision will make it very hard for the Court to assess the proposed filtering injunction's consequences for fundamental rights. Neither the AG nor the Court has the information needed to meaningfully weigh the injunction's benefits for the plaintiff against the burdens it will create for Facebook's other users. That includes potential harm to those users' data protection rights and

[5] Opinion of AG Szpunar, Case C-507/17, *Google v. Commission nationale de l'informatique et des libertés (CNIL)*, (January 10, 2019) ("Google Op.") http://curia.europa.eu/juris/document/document.jsf?text=&docid=209688&pageIndex=0&doclang=EN&mode=req&dir=&occ=first&part=1&cid=9734778. Groups representing journalists and other affected entities were able to intervene before the lower court in the Google case, giving the AG and Court the benefit of additional input in that case.

expression and information rights, as the Court has recognized in the past. Their rights to equality and non-discrimination are also implicated: recent research suggests that filtering errors disproportionately affect lawful speech by members of racial and linguistic minority groups.[6] Overall, Facebook's users number in the billions, so mistakes affecting even a tiny percent will still be very consequential.[7]

The AG and Court are not alone in lacking key information. Technologists and civil society groups engaged in the EU's ongoing policy debate about filters have charged that the public overall has inadequate information about their real capabilities and error rates. A 2019 letter from two dozen civil society organizations to the EU Parliament said that even some of the most widely deployed filters are "untested and poorly understood technologies," with great potential to harm "democratic values and individual human rights."[8]

Still, the public debate has generated some relevant information, little of which seems to have been surfaced to the court in this case. Civil society groups, technologists, and human rights bodies have all raised concerns about states relying on privately operated software to restrict expression and information.[9] European legal experts have carried out detailed analyses of hosts' immunities and filtering proposals under the eCommerce Directive and fundamental rights

---

[6] *See* Sap et al (2019) and Center for Democracy and Technology (2017), below note 11.
[7] Facebook reports 2.41 billion monthly active users. Company Information, https://newsroom.fb.com/company-info/.
[8] *Letter to Members of European Parliament* (February 8, 2019) https://cdt.org/files/2019/02/Civil-Society-Letter-to-European-Parliament-on-Terrorism-Database.pdf. The letter continued,

> The European public is being asked to rely on claims by platforms or vendors about the efficacy of [a widely used filtering] Database and similar tools—or else to assume that any current problems will be solved by hypothetical future technologies or untested, post-removal appeal mechanisms. Such optimistic assumptions cannot be justified given the serious problems researchers have found with the few filtering tools available for independent review. Requiring all platforms to use black-box tools like the Database would be 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. That gamble is neither necessary nor proportionate as an exercise of state power.

[9] *See* notes 10 through 15.

law.[10] This information is commonplace in discussions in Brussels, but, because of limits on who gets to participate in the case, does not seem to be before the Court.

A poorly considered ruling in this case could have major unintended consequences for the fundamental rights of Internet users around the world. To avoid those, the Court should take care, in its ultimate ruling, to articulate very clearly what questions it is resolving—and what questions it is not. It should not rely on untested claims or assumptions about filtering technology, or resolve legal questions that were not squarely raised and adequately briefed. It should, I believe, reject the Austrian court's injunction. If it does implicitly or explicitly endorse filtering in some circumstances, it should be very clear about how filtering injunctions can be reconciled with users' fundamental rights. At a minimum, injunctions should not issue without very strong showings that filters provide the best means to protect a plaintiff's or government's interests, and will not generate mistakes and false positives that disproportionately burden the rights of other Internet users. To the extent that the Court's conclusions require further fact-finding or legal analysis, its Judgment should make that clear.

Clarity in the Court's Judgment will be essential to the Austrian courts as this case continues. So will the Court's guidance about fundamental rights and the eCommerce Directive. EU policymakers, too, will look to the Court's Judgment in order to understand the rights and interests at stake as they revise the EU's platform liability laws.

---

[10] *See, e.g.*, Christina Angelopoulos, *On Online Platforms and the Commission's New Proposal for a Directive on Copyright in the Digital Single Market*, (2017), https://ssrn.com/abstract=2947800; Sophie Stalla-Bourdillon, Eleonora Rosati, Karmen Turk, Christina Angelopoulos, Aleksandra Kuczerawy, Miquel Peguera and Martin Husovec, *A Brief Exegesis of the Proposed Copyright Directive*, (2017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2875296; Joris van Hoboken, João Pedro Quintais, Joost Poort, and Nico van Eijk, *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), https://www.ivir.nl/publicaties/download/hosting_intermediary_services.pdf.

II. Process and Representation Issues

The procedure in this case has effectively deprived both the AG and Court of crucial information and analysis. Most importantly, they are not hearing from some of the people who will be most affected by the ruling, and are not hearing information and arguments that are familiar to participants in the public policy debate. That's troubling both for the likely quality of the resulting Judgment, and for its legitimacy in the eyes of European policy stakeholders.

Three factors in particular contribute to this problem. First, the case arose from an expedited preliminary injunction proceeding, so the factual and legal record before the Court is very slight. The Austrian courts never even addressed the complex questions around global takedown, for example. On filtering issues, the appeals court's discussion focused almost entirely on defamation law and harm to the plaintiff. It did not analyze the efficacy of filters in redressing that harm, or the unintended consequences the filters might have for other Facebook users.

Second, the advocacy in this case is imbalanced. As is common in Intermediary Liability cases, the people most concerned about expression and information rights are not parties, so the courts never hear from them. The plaintiff is the person who was harmed by an online post, but the defendant is not the post's author. Instead, the defendant is Facebook—a technology company, with interests that inevitably may not align with those of its users or other people in the Internet's information ecosystem.

Both of those problems might be addressed by intervention and briefing from civil society organizations or other expert groups. But those organizations did not recognize the case's importance in time to petition national courts to intervene in the expedited Austrian proceeding, and were not able to weigh in once the case arrived at the CJEU. That lack of input is the third, and perhaps most serious, factor distorting the picture presented to the Court and the AG. The only perspectives they see are those of the plaintiff, Facebook, and five government

interveners (Austria, Latvia, Portugal, Finland, and the EU Commission). The government briefs are confidential, so we don't know what they said. But it seems unlikely that they provided insight into evolving technology, which is at the heart of this case. And while the briefs from Member States likely represent the public interest as they see it, government lawyers may not be the best advocates for expression and information rights in a case about vulgar criticism of a politician.

The filtering debate as presented in the case looks very different from the one that participants in Brussels and throughout the EU would recognize. Many technologists,[11] academics,[12] and civil society organizations like EDRi,[13] Article 19,[14] and Access Now[15] have argued that filters pose major and poorly understood threats to fundamental rights. Human rights bodies, too, have raised serious concerns.[16] Facebook, in that larger public debate, is somewhere in the middle. It supports filters in some circumstances, and has come under sustained criticism for doing so. (To be clear, I have never heard of Facebook representatives endorsing a filter like the one proposed in this case, which would search users' textual posts for defamation—indeed, I have heard them express great skepticism

[11] *See, e.g.*, Evan Engstrom and Nick Feamster, *The Limits of Filtering: A Look at the Functionality and Shortcomings of Content Detection Tools*, (2017), www.engine .is/the-limits-of-filtering; Center for Democracy and Technology, *Mixed Messages?* (2017), https://cdt.org/files/2017/11/Mixed-Messages-Paper.pdf; Maarten Sap *et al*, *The Risk of Racial Bias in Hate Speech Detection*, (2019), https://homes.cs.washington.edu/~msap/pdfs/sap2019risk.pdf.
[12] See, e.g., Angelopolous, *supra* note 10.
[13] European Digital Rights, *Press Release: Censorship machine takes over EU's internet*, (March 26, 2019) https://edri.org/censorship-machine-takes-over-eu-internet/.
[14] Article 19, *Facebook congressional testimony: "AI tools" are not the panacea*, (April 13, 2018) https://www.article19.org/resources/facebook-congressional-testimony-ai-tools-not-panacea/.
[15] Denis Nolasco and Peter Micek, *Access Now responds to Special Rapporteur Kaye on "Content Regulation in the Digital Age"*, AccessNow (January 11, 2018) https://www.accessnow.org/access-now-responds-special-rapporteur-kaye-content-regulation-digital-age/.
[16] 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*, (December 7, 2018), https://spcommreports.ohchr.org/TMResultsBase/DownLoadPublicCommunicationFile?gId=24234; Council of Europe, *Recommendation CM/Rec(2018)2 of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries*, Committee of Ministers (March 7, 2018), https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=0900001680790e14.

about filtering text generally.) Facebook is a major proponent and backer of a controversial filter[17] for violent extremist videos and images, and its founder has extolled[18] the potential for AI-based content moderation generally.

In the circumscribed world of this case, though, Facebook is the sole filtering skeptic. All the other briefs apparently maintain that using software to automatically block defamation of a public figure while still respecting users' fundamental rights "must be possible."[19] If public interest organizations and other interested groups had submitted briefs in the case, the Court would have heard strong criticisms of that idea. In this essay, I will lay out many of those criticisms as they relate to expression and information rights. But there are other relevant areas where the interests of platforms and those of users affected by filters diverge. The lack of adequate briefing on those is a problem, too. By way of illustration, these include:

Privacy and Data Protection: The Court has clearly said in the past that using filters to automatically scan and assess users' every communication can burden their data protection rights. This issue barely comes up in the AG's Opinion, though. Had civil society groups intervened, that issue—and deeply intertwined questions about the monitoring that Facebook already uses to target ads—would almost certainly have been fleshed out. Privacy experts could have weighed in, for example, on the relevance of users' rights, under the GDPR, in relation to the kinds of automated decision-making carried out by filters. They could have discussed CJEU cases like *Digital Rights Ireland*, which rejected a law requiring electronic communications service providers to retain data about

[17] Olivia Solom, *Facebook, Twitter, Google and Microsoft team up to tackle extremist content*, The Guardian (December 6, 2016) https://www.theguardian.com/technology/2016/dec/05/facebook-twitter-google-microsoft-terrorist-extremist-content.
[18] Sydney Li and Jamie Williams, *Despite What Zuckerberg's Testimony May Imply, AI Cannot Save Us*, Electronic Frontier Foundation (April 11, 2018) https://www.eff.org/deeplinks/2018/04/despite-what-zuckerbergs-testimony-may-imply-ai-cannot-save-us.
[19] Par 55.

communications made by all of their subscribers.[20] They would also likely have had a lot to say about the Austrian courts' novel requirement that Facebook monitor its users' posts in search of *any* image of Glawischnig-Piesczek. Presumably that would require Facebook to use facial recognition technology. As the referring court and the AG frame the case, the facial recognition issue probably isn't in scope—which is fortunate, since the issue is massively complex and the Court has seemingly received no relevant briefs or analysis. The question of how filtering users' communications generally affects their privacy, though, *is* germane to the case, and should very much be part of the Court's analysis. The lack of briefing on it is disturbing.

Competition: If the Court opens the door to filtering orders in this case, it will inevitably cause other platforms to adjust their legal risk assessments, behavior, and product design—even if the ruling doesn't technically apply to them. That's consequential for Facebook's competitors. For giant companies like Google (where I worked until 2015), those consequences may be tolerable—Google has already spent at least $100 million[21] on filtering technology, and spends another $100 million[22] annually on content review. Smaller companies can't do that, though. Both immediate compliance costs and long-term uncertainty about what measures courts might require have major consequences for smaller companies, making both technologists and investors hesitant to try competing with today's incumbent platforms in the first place.[23] This concern was apparently discussed by the Commission in oral arguments, which is good, and illustrates the value of

---

[20] C-293/12 and C-594-12, *Digital Rights Ireland Ltd. v. Ireland*, (2014).
[21] Paul Sawers, *YouTube: We've invested $100 million in Content ID and paid over $3 billion to rightsholders*, VentureBeat (November 7, 2018) https://venturebeat.com/2018/11/07/youtube-weve-invested-100-million-in-content-id-and-paid-over-3-billion-to-rightsholders/.
[22] David Shepardson, *Google spends hundreds of millions of dollars on content review: letter*, Reuters (May 2, 2019) https://www.reuters.com/article/us-alphabet-google-youtube/google-spends-hundreds-of-millions-of-dollars-on-content-review-letter-idUSKCN1S81OK.
[23] Oxera, *The Economic Impact of Safe Harbours on Internet Intermediary Start-Ups*, Feb. 2015, https://www.oxera.com/wp-content/uploads/2018/07/The-economic-impact-of-safe-harbours-on-Internet-intermediary-start-ups.pdf.pdf; Booz & Co., *The Impact of U.S. Copyright Regulations on Early Stage Investment: A Quantitative Study*, http://www.strategyand.pwc.com/media/uploads/Strategyand-Impact-US-Internet-Copyright-Regulations-EarlyStage-Investment.pdf.

interventions to articulate points that neither plaintiffs nor defendants have a strong interest in raising.

Freedom of Expression and Information: Finally, public interest advocates could have helped elucidate freedom of expression and information issues. Facebook briefed the relevant law on this issue extensively and well. But the interests of the company's users are ultimately distinct from Facebook's own, and warrant different representation. Platforms' and users' interests do not align when platforms face choices between protecting users' expression and protecting themselves. Removing lawful speech, or adopting flawed enforcement tools that will inevitably do so, can be the safest and most cost-effective choice for platforms. It can help them avoid liability, stave off regulation, please advertisers, or appease influential critics. Many civil society groups charge that this is precisely what drove platforms to adopt the poorly understood filters that many use today.

The Internet is an ecosystem. Cases involving major platforms often have consequences far beyond the companies themselves. Courts should have the opportunity to hear from the individuals, journalists, political advocates, and others who will be affected by consequential decisions like this one. In the absence of those voices, it is unsurprising that the AG is sanguine about monitoring injunctions, and willing to assume that technology can block bad information without also blocking good information. The Court should not repeat that mistake.

III. Making Facebook Monitor and Filter Users' Posts

The AG concludes that Austrian courts can order Facebook to filter all of its users' posts for "identical" expression, and monitor the original user for "equivalent" expression. In this Section, I will briefly describe the relevant EU policy debate on

monitoring and filtering, which tees up many issues central to this case. I will then walk through some concerns with the AG's Opinion, beginning with issues of fundamental rights. The fundamental rights impact of the proposed filtering injunction is hard to assess, because it is not clear exactly what content the Austrian courts and AG believe should be filtered, what errors can be expected, and whether those errors can reasonably be remedied. Finally, I will review the AG's discussion of the eCommerce Directive – which is difficult to follow, but appears to say that hosts immunized under the Directive can be compelled to use software-based filters, yet risk serious liability if employees review the filters' removal decisions. That is a perverse result from a fundamental rights perspective, and conflicts with both the Court's precedent and recommendations of human rights bodies – as well as practices that many platforms have already adopted at the urging of EU lawmakers.

A. Policy Backdrop

EU policymakers have now invested several years in a major political debate about whether, when, and how platforms can be compelled to filter their users' online expression. Lawmakers in Brussels labored for years to reach political resolution to this question, even in the comparatively straightforward area of copyright. The Copyright Directive, which was enacted in a series of hotly contested votes in 2018 and 2019, ultimately included a filtering requirement. That outcome prompted street protests and newspaper blackouts, as well as a judicial challenge arguing that the mandate violates Internet users' fundamental rights.[24] As opponents pointed out during the political process, filters can't understand context.[25] That means if text, images, or videos violate the law in one

[24] Michelle Kaminsky, *EU's Copyright Directive Passes Despite Widespread Protests -- But It's Not Law Yet*, Forbes (March 26, 2019) https://www.forbes.com/sites/michellekaminsky/2019/03/26/eus-copyright-directive-passes-despite-widespread-protestsbut-its-not-law-yet/#24b0d6902493; Andrew Liptak, *Poland has filed a complaint against the European Union's copyright directive*, The Verge (May 25, 2019). https://www.theverge.com/2019/5/25/18639963/poland-european-union-copyright-directive-filed-complaint-court-of-justice.

[25] Julia Reda, *When filters fail: These cases show we can't trust algorithms to clean up the internet*, Julia Reda (September 28, 2017) https://juliareda.eu/2017/09/when-filters-fail/.

situation, filters will likely also block the same material in lawful uses like parody, journalism, or scholarship. Famous examples include a lecture by Harvard Law Professor Lawrence Lessig, which was automatically removed from YouTube because it used music clips to illustrate a legal point.[26]

The EU's political branches are now wrangling with a second monitoring proposal, in the draft Terrorist Content Regulation.[27] The Commission and Council versions of the Regulation required hosts to use upload filters to prevent content from reappearing.[28] They also encouraged platforms to carry out "human review" or "human oversight," meaning that platform employees should check filters' work and correct errors. Civil society organizations[29] and three UN human rights rapporteurs[30] responded to this new proposed filtering mandate with alarm. They pointed out that filters would not be able to tell when extremist material such as an ISIS recruiting video was reused in contexts such as news reporting, academic research, or counter-radicalization messaging. As a result, valuable and lawful expression would likely be blocked. Many critics questioned whether human review of filters' automated decisions would be sufficient to offset these harms, noting examples like YouTube taking down over 100,000 videos maintained by the German non-profit Syrian Archive as evidence of war crimes and human rights violations.[31] The most recent draft of the Terrorist Content Regulation, from the EU Parliament, eliminated the filtering requirement.[32] But many observers worry that the Parliament's new rapporteur

---

[26] *See* Reda *supra* note 25.
[27] EU Parliament, *Tackling the dissemination of terrorist contest online*, Legislative Resolution (April 17, 2019) https://www.europarl.europa.eu/doceo/document/TA-8-2019-0421_EN.pdf.
[28] EU Commission (September 12, 2018) https://ec.europa.eu/commission/sites/beta-political/files/soteu2018-preventing-terrorist-content-online-regulation-640_en.pdf Art. 6; EU Council (December 3, 2018) Art. 6, https://www.parlament.gv.at/PAKT/EU/XXVI/EU/04/57/EU_45743/imfname_10862334.pdf.
[29] Article 19 *et al*, *Joint letter on European Commission regulation on online terrorist content*, (December 6, 2019) https://www.article19.org/resources/joint-letter-on-european-commission-regulation-on-online-terrorist-content/.
[30] *Supra* note 16.
[31] Kate O'Flaherty, *YouTube keeps deleting evidence of Syrian chemical weapon* attacks, Wired (June 26, 2018) https://www.wired.co.uk/article/chemical-weapons-in-syria-youtube-algorithm-delete-video.
[32] *Supra* note 27.

for the trilogue process, conservative Polish MEP Patryk Jaki, will agree to reinstate the filtering mandate in the law's final version.[33] If so, that law will likely also, like the Copyright Directive, be challenged before the CJEU as a fundamental rights violation.

European political discussion has not focused on the kind of filter at issue in this case: one that is designed to restrict criticism of a public figure, and works by blocking text—rather than, as is more common, images or video. To my knowledge, no legislative body to date has seriously proposed this. (A possible exception is Singapore's controversial "fake news" law.[34]) Both defamation and speech about public figures are widely considered too context-dependent for even judges to assess easily, and thus particularly ill-suited for enforcement by automated tools.

A new and more general political debate about filtering is expected in 2019 and 2020, as EU lawmakers consider a proposed Digital Services Act. That legislative effort is expected to lead to changes in the core law at issue in this case, the eCommerce Directive—which, as implemented in Member States' laws, has defined platforms' legal obligations for almost two decades.

Alongside the EU's political debate is a major discussion of filtering from human rights experts. Representatives of both the UN and regional human rights systems around the world have raised grave concerns about relying on private companies and automated tools to police online expression and information.[35] The Council of Europe's Committee of Ministers, in a 2018 Recommendation, said

---

[33] Patryk Jaki, European Union, European Conservatives and Reformists Group Member, http://www.europarl.europa.eu/meps/en/197516/PATRYK_JAKI/assistants?.
[34] *Singapore fake news law a 'disaster' for freedom of speech, says rights group*, The Guardian (May 9, 2019), https://www.theguardian.com/world/2019/may/09/singapore-fake-news-law-a-disaster-for-freedom-of-speech-says-rights-group.
[35] 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*, (2018), https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/ContentRegulation.aspx.

> Due to the current limited ability of automated means to assess context, intermediaries should carefully assess the human rights impact of automated content management, and should ensure human review where appropriate. They should take into account the risk of an over-restrictive or too lenient approach resulting from inexact algorithmic systems, and the effect these algorithms may have on the services that they provide for public debate. Restrictions of access to identical content should not prevent the legitimate use of such content in other contexts.[36]

EU lawmakers, to their great credit, have been attentive to the evolving fundamental rights guidance on Intermediary Liability and filtering. The EU Commission's 2018 Recommendation on tackling illegal content online, for example, says 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."[37] This attention to the guard-rails created by fundamental rights will be essential as lawmakers revise the eCommerce Directive in the coming years. The Court's ruling in this case may be one of their most important inputs.

B. Legal Analysis

The AG's enumerated conclusions about filtering are:

1. The court can order Facebook to filter every post by every one of its users, in order to block "identically worded" content.
2. The court can order Facebook to filter "equivalent" content, but only from the account of the user who put up the original post.
3. Once a court determines that specific content is defamatory, Facebook must take down equivalent content if it is notified about it. (This part strikes me as relatively uncontroversial.)

---

[36] *Supra* note 16.
[37] European Commission, *Commission Recommendation on measures to effectively tackle illegal content online*, (March 1, 2018) https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=50095.

To arrive at these, he relies on some interim conclusions. Importantly, he concludes that the Court *can* order Facebook to use automated, software-based filters, but that it *cannot* require "active non-automatic filtering."[38] This seems to mean that the Court could not order Facebook to have its employees carry out human review of content flagged by filters. In other words, courts can order Facebook to monitor its users' posts—but only in the way that experts and human rights bodies have warned is likeliest to lead to damaging mistakes. As I will discuss in Subsection 1, that approach is hard to reconcile with fundamental rights guarantees. But appropriate fundamental rights analysis in this case is nearly impossible, given the lack of clarity about what is to be filtered, what errors are likely, and whether those errors can be corrected.

In Subsection 2 I will discuss the eCommerce Directive, and the AG's troubling implication that platforms using "active non-automatic" content review in any situation may forfeit immunity under Directive Article 14. Assuming that this refers to human review of user content, his reasoning suggests that existing filtering efforts, including the ones that both small and large platforms adopted at the urging of EU political bodies, expose the platforms to massive legal risk—not only in this case, but for any future claims in areas such as copyright, trademark, or defamation. This assessment of human review is also problematic for platforms' ordinary notice and takedown operations. If the simple act of looking at content immediately creates liability for platforms, they have strong reasons to avoid reviewing or moderating user content at all—or, if they do moderate, to err heavily on the side of taking content down. These conclusions are not required by the eCommerce Directive, and they place an unnecessary and disproportionate burden on Internet users' rights. The Court should not accept them.

[38] Par. 61.

1. Fundamental Rights

Both the CJEU[39] and the European Court of Human Rights (ECtHR)[40] have in the past disapproved rulings that would have effectively required platforms to proactively monitor users' expression. The CJEU specifically rejected an injunction requiring a social media platform to filter users' communications in a copyright case involving the Belgian collecting society SABAM. Both courts identified serious fundamental rights concerns with laws that would require platforms to monitor their users. One concern is that inspecting users' communications can invade their privacy or data protection rights. This privacy issue is little explored in the case law, and not addressed in the AG's Opinion. Another concern is that filters can impinge on users' rights to receive and impart information, whether by blocking lawful expression and information[41] or preventing platforms from hosting open forums for discussion in the first place.[42]

Recent public discussion about filters has largely focused on their errors and resulting removal of legal expression and information. But the concern about preserving open forums for online discussion has become increasingly pressing, as Internet users have consolidated onto a small handful of powerful private platforms like Facebook, YouTube, and Twitter. These major platforms, seeking

---

[39] Case C-360/10, *SABAM v. Netlog NV*, (2012); Case C-70/10, *Scarlet Extended SA v. SABAM*, (2011).

[40] *Magyar Tartalomszolgáltatók Egyesülete (MTE) v. Hungary*, App. No. 22947/13, Eur. Ct. H.R. 135 (2016), Par 82 (strict liability for platform "allowing unfiltered comments" in defamation case violated Convention Article 10); *compare Delfi AS v. Estonia*, App. No. 64569/09, Eur. Ct. H.R. 586 (2015) (strict liability in hate speech case did not violate Article 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 (April 11, 2016) http://cyberlaw.stanford.edu/blog/2016/04/new-intermediary-liability-cases-european-court-human-rights-what-will-they-mean-real.

[41] *Netlog* Par. 50 (filtering injunction "could potentially undermine freedom of information, since that system might not distinguish adequately between unlawful content and lawful content, with the result that its introduction could lead to the blocking of lawful communications").

[42] *MTE* Par. 61 (defendant "provided forum for the exercise of expression rights, enabling the public to impart information and ideas"), Par. 86 (strict liability would threaten "the comment environment of an Internet portal, for example by impelling it to close the commenting space altogether"). CJEU precedent has not focused on the existence of open forums as a Charter Article 11 concern, but has noted the threat to open forums' commercial viability, 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* Par. 46.

to avoid controversy or liability in legal grey areas, have prohibited content ranging from art to images of breastfeeding to photos of indigenous Amazonian women in traditional garb.[43] Many experts fear that the remaining smaller platforms that offer more open forums for expression may not remain economically viable if the law requires expensive filters—or if operators fear that a court might at any moment impose that requirement.

The fundamental rights case law on filtering doesn't provide definitive guidance for assessing any of these risks. But it does stand for the proposition that filters' real-world consequences matter. Courts and lawmakers can't just assume that filtering technology is perfect or adequate to meet the Charter's and Convention's requirements. They must factor in filters' real capabilities or shortcomings in order to assess how well filters serve legitimate government aims, what burdens they place on other fundamental rights, and whether that burden is proportionate or necessary.

The AG's Opinion makes this assessment hard. He recommends a filtering injunction without defining or analyzing three very basic things. First, what specific content will Facebook be ordered to filter? Second, how effective will the filters be in distinguishing legal from illegal content? Third, can Facebook or other future platforms affected by the ruling try to correct filters' errors using mechanisms like human review? These questions are essential to assess how

---

[43] Par Perrine Signoret, *Facebook: la justice se penche sur la censure de*, Le Monde (February 1, 2018) https://www.lemonde.fr/pixels/article/2018/02/01/censure-de-l-origine-du-monde-sur-facebook-une-attaque-contre-la-democratie_5250611_4408996.html; Jessica Reed and Becky Gardiner, *The beautiful breastfeeding images Facebook is missing out on*, The Guardian (February 23, 2012) https://www.theguardian.com/commentisfree/2012/feb/23/breastfeeding-images-facebook-missing; Waqas, *Brazil will sue Facebook for blocking picture of indigenous woman*, HackRead (April 20, 2015) https://www.hackread.com/facebook-blocking-brazil-indigenous-picture/; *see generally* Daphne Keller, *The EU's Terrorist Content Regulation: Expanding the Rule of Platform Terms of Service and Exporting Expression Restrictions from the EU's Most Conservative Member States*, Center for Internet and Society (March 25, 2019) http://cyberlaw.stanford.edu/blog/2019/03/eus-terrorist-content-regulation-expanding-rule-platform-terms-service-and-exporting.

many "dolphins" will be caught in filters' nets—in other words, how often Internet users around the world will be prevented from receiving or imparting lawful information. Neither the AG nor the Court should attempt to resolve the case without clearer answers.

a. What "Identical" or "Equivalent" Content Is to Be Filtered?

The AG says that Facebook can be ordered to monitor all of its users in order to block "identical" content, and also monitor the original user's account for "equivalent" content. Filtering "identical" content, he suggests, is simple. But it is hard to tell what content he considers to be "identical." That imprecision is a problem. Without knowing what content a filter is supposed to detect, it is hard to predict its likely accuracy, or how much lawful information and expression it might block by accident.

i. "Identical" Content

The filter for "identical" content that the AG contemplates seems to be one or more of the following:

Facebook must block every user from sharing copies of the original post using Facebook's "Share" button. This is an odd issue to litigate, because taking down the user's post should already, as a matter of Facebook's basic architecture, have eliminated any "shared" copies. You don't need a case like this to compel that outcome. Apparently, the Austrian government raised the "share" function in its brief, though, so it may be in play.[44]

Facebook must block every user from posting the text that the Austrian court ruled defamatory ("'lousy traitor of the people' and/or a 'corrupt oaf' and/or a member of a 'fascist party'"). As best I can discern, this is what the AG means by

[44] Par. 56.

"identical" content.[45] If that's right, the Opinion raises major concerns for Facebook users' information and expression rights. A filter blocking these bare phrases would prevent friends from calling one another oafs or traitors in jest, and prevent historians or journalists from writing about actual fascists. It would also block news coverage of this case—or academic and legal analysis, including the AG's own Opinion. These kinds of problems with text filters are nothing new. Wikipedia's entry on the Scunthorpe Problem documents notorious and often comic examples of text filtering failures going back to the 1990s.[46] There are very serious examples, too. Numerous victims of racially-based harassment have gone online to bear witness to their experiences, only to be penalized or locked out of social media for repeating words used by their attackers.[47]

Facebook must block every user from posting the text that the Austrian court ruled defamatory ("'lousy traitor of the people' and/or a 'corrupt oaf' and/or a member of a 'fascist party'") coupled with any photograph of the plaintiff. This seems to have been what the Austrian appellate court had in mind when it said Facebook must filter "identical" content.[48] From the perspective of defamation law, enjoining posts with images of the plaintiff makes some sense, since she would not have a right to stop people using those words about anyone but herself.

---

[45] In theory the "identical content" could be the Austrian user's entire written post. *See supra* note 3 for text. The AG does not include that text in his Opinion, though. He writes that he takes "'identically worded items of information' to mean … precise manual reproductions of the information which [the Austrian court] has characterised as illegal" (par. 56) and notes that the injunction below covered allegations "that the applicant was a 'lousy traitor of the people' and/or a 'corrupt oaf' and/or a member of a 'fascist party'." (Par. 14).

[46] Wikipedia, https://en.wikipedia.org/wiki/Scunthorpe_problem.

[47] Tracy Jan and Elizabeth Dwoskin, *A white man called her kids the n-word. Facebook stopped her from sharing it*, The Washington Post (July 31, 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. A recent U.S. case strangely merged this too-common fact pattern with the one at issue in *Glawischnig-Piesczek*. A politician's office posted video documenting protestors' crude allegations and suggestion of violence against him – then complained when Twitter removed the video for violating its policy against threats. Marc Rod, *Twitter reverses course, unlocks Mitch McConnell campaign account and leaves video that violated threats policy*, https://www.cnbc.com/2019/08/09/mitch-mcconnell-campaign-twitter-account-is-unlocked.html.

[48] 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". (Par. 56, italics altered).

But the Supreme Court referral asks only about "identically worded items of information," so the question to the CJEU appears to be about text.[49] The AG seems to interpret it that way: his Opinion doesn't discuss photo filtering or the distinct concerns it would raise.

From the perspective of fundamental rights, requiring a company like Facebook to employ facial-recognition-based filters would be radical. At a time when many privacy advocates want platforms to *stop* rolling out pervasive biometric identification, an injunction covering any photo of the plaintiff would make the technology mandatory—requiring Facebook to run facial scans on people who have nothing to do with this case.[50] Facebook's billions of users may (or may not) have consented to that. The other people depicted in their photos almost certainly haven't. If that's what this case were about, we'd expect the referral, briefing, and AG Opinion to discuss the depicted people's privacy and data protection rights.

More basically, an order to block content based on plaintiff's image or identity would raise fundamental rights issues similar to those the Court considered in *Google Spain* and is assessing in the newer "Right to Be Forgotten" cases now. While the doctrinal basis of data protection and defamation claims differ, they raise similar tensions between a claimant's rights to reputation, privacy, or data protection on the one hand, and other Internet users' rights to seek and impart information on the other. As I discuss in my more detailed article about Data Protection and Intermediary Liability, few courts have considered how the precedent assessing this balance of rights for search engines might apply to the very different situation of social media hosts like Facebook.[51] We do know, however, that the Article 29 Working Party's guidelines under *Google Spain* called for search engines to evaluate each individual item of online information

---

[49] Par. 56.
[50] Mark Scott and Naomi O'Leary, *Facebook's privacy push stumbles over EU rollout of facial recognition technology*, Politico (April 6, 2018) https://www.politico.eu/article/facebook-facial-recognition-privacy-data-protection-cambridge-analytica-mark-zuckerberg/.
[51] Daphne Keller, *The Right Tools: Europe's Intermediary Liability Laws and the EU 2016 General Data Protection Regulation*, 33 Berkeley Tech. L.J. 287 (2018) at 322-327.

separately before de-indexing, and particularly emphasized the complexity of this review in cases involving public figures.[52]

ii. "Equivalent" Content

It is also unclear what "equivalent" content Facebook can be required to monitor on the original user's account. The AG suggests that text is "equivalent" if it "scarcely diverges" from the original, containing perhaps a "typographical error" or "slightly altered syntax."[53] Filters like that might exist for uses like plagiarism detection, though I'm unaware of any literature on their efficacy and in practice they would likely depend heavily on human review. In any case, though, the AG notes that the Austrian court might mean something else.[54] This ambiguity makes the fundamental rights consequences of this part of the injunction particularly hard to assess.

Relying on filters to discern the meaning of written human communication, or even its sentiment, is a dubious proposition. One recent study of "natural language processing" filters found errors in one out of every four to five takedown decisions, and noted that the errors increased when speakers used slang, sarcasm, or languages that tech company employees didn't speak.[55] Another found that automated detection tools disproportionately mis-identified social media posts in African American English as "toxic."[56]

The Court should not uphold the unprecedented filtering injunction in this case without clearly understanding and describing what that injunction would require Facebook to do. Without clarity about what information, exactly, is to be filtered, it is all but impossible to assess the injunction's impact on fundamental rights.

---

[52] Article 29 Data Protection Working Party, *Guidelines on the Implementation of the Court of Justice of the European Union Judgment on "Google Spain and Inc. v. Agencia Española de Protección de Datos (AEPD) and Mario Costeja González" C-131/12*, (November 26, 2014) https://perma.cc/6LPR-TFRL.
[53] Par. 67.
[54] *Id.*
[55] Center for Democracy and Technology, *supra* note 11.
[56] Sap *et al*, *supra* note 11.

b. How Much of the Filtered Content Will Be Illegal?

The AG appears to posit that as long as filters accurately flag identical content, nearly all of that content will be illegal. In general, he says, "repetitions of an infringement actually characterised as illegal… should be characterised in the same way."[57] That factual proposition is, of course, central to the fundamental rights question in this case. As described above, it is very much contested. Basic questions about how often filters accurately identify duplicate content, and how often that content violates the law, have fueled years of argument in political, legal, and human rights circles.

Purely as a matter of formal logic, the AG's assumption would make sense if "identical" content were truly illegal in *every* context. In that case, filters' inability to assess context wouldn't matter. The court would only need to know whether filters can accurately identify duplicates as a technical matter.

In principle, the Austrian Supreme Court could have framed the question this way: "*assuming* that every identical copy of this information violates Austrian law, is the filtering injunction permissible?" But that's not what it did. Its question states only that the initial copy is illegal, and asks the CJEU to decide whether Facebook can be ordered to filter "other identically worded items of information."[58] Nothing in its referral asserts that every copy is illegal.

Some content really is, at least in some jurisdictions, illegal in every possible context. But the only examples I am aware of (outside of countries like China) are very extreme and harmful content, specifically child sexual abuse imagery

[57] Par. 65. A filter's effectiveness in identifying unlawful content depends on two things: (1) its technical accuracy in identifying duplicates, and (2) its ability to assess whether a duplicate is illegal in a new context. If this case were about detecting duplicate images or videos, or about using facial recognition technology to detect new pictures of the plaintiff, there would be important factual questions about (1), the filters' purely technical accuracy. Those questions are also important if filters are to block "equivalent" content of any sort, including text. For "identical" copies of words in text files (as opposed to words that appear in image files, etc.) though, it is probably safe to assume filters will be technically accurate. In that case the key question is about (2), legality when the words are used in a new context.

[58] Par. 22.

(CSAI). Even for CSAI, a ban on all possible copies, with no exceptions, can go awry. A former law of this sort in the U.S., for example, impeded efforts to report CSAI to law enforcement.

Because CSAI generally has no contextually-legal uses, automated filters for CSAI image and video files (but not text) on Internet platforms are widely accepted. Many companies rely on matching tools like PhotoDNA, which use digital hashes or fingerprints to find duplicates of videos and images that have already been identified as CSAI.[59] Some companies depend on these filters entirely, skipping the expense and emotional toll of having employees review deeply disturbing material. Others carry out human review, saying that the filters alone are not accurate enough.

Is the Austrian Facebook user's post about Glawischnig-Piesczek so completely illegal that, like CSAI, it violates the law in every possible context? It is hard to imagine that there is no lawful use, in journalism, research, or parody, for crudely critical words about a powerful politician.[60] The case itself, of course, makes the exact words still more important. Knowing precisely what made the user's post illegal is essential for journalists and legal researchers—and for publisher and platform lawyers assessing similar content in the future. If there are arguments under Austrian law or the EU Charter for making this information universally unlawful, the AG does not examine them.

Another ground for concluding that filters raise no fundamental rights concerns might be if a competent authority had assessed filters' expected errors and resulting burden on lawful information and expression, and determined the burden was proportionate and acceptable as a matter of law. That has not happened here. A judicial conclusion like that would require courts to have much

[59] Microsoft, *How does PhotoDNA technology work?*, https://www.microsoft.com/en-us/photodna.
[60] Compare *MTE* Par. 77 (rejecting monitoring requirement and noting that the terms at issue "albeit belonging to a low register of style, are common in communication on many Internet portals -- a consideration that reduces the impact that can be attributed to those expressions").

more information—starting with knowing what specific content is to be filtered. If the AG's analysis or the Court's future ruling in this case turns on an assumption that all or nearly all identical content is illegal, that assumption should be stipulated and labeled very clearly. If this is not a temporary assumption but a final legal conclusion, its factual basis should be clear and the supporting analysis under Austrian defamation law and the EU Charter should be rigorous.

A final possible argument might be that inaccurate filters are acceptable specifically on *Facebook*, because its users can always use other means to seek and impart information. The AG, to his credit, says nothing of the sort. An assumption like that would be dangerously out of touch with ordinary people's reliance on social media platforms for basic communication, as well as news and information on matters of public concern.[61] That reliance, and Facebook's resulting status as a de facto information gatekeeper, creates its own set of problems. But they are problems that make protecting Facebook users' rights *more* important, not less. A signal from the AG or Court that fundamental rights are less protected on Facebook than on other media or communication channels would raise major red flags for human rights organizations around the world, especially in countries with less rights-respecting regimes. In the EU, a ruling implying that users have weakened rights when using Facebook would be particularly relevant for users in France, Germany, and Poland who have claimed that Facebook's own practices burden their free expression rights.[62]

---

[61] Pew Research Center, *Majorities in most European countries get news from social media*, (May 8, 2018) https://www.journalism.org/2018/05/14/many-western-europeans-get-news-via-social-media-but-in-some-countries-substantial-minorities-do-not-pay-attention-to-the-source/pj_2018-05-14_western-europe_5-01/.

[62] Sarah Cascone, *After an 8-Year Legal Battle, Facebook Ends Its Dispute With a French School Teacher Who Posted Courbet's 'Origin of the World'*, (Aug. 5 2019), https://news.artnet.com/art-world/facebook-courbet-lawsuit-ends-1616752; David Meyer, *Court tells Facebook: Stop deleting 'offensive' comment*, The German View (April 13, 2018) https://www.zdnet.com/article/court-tells-facebook-stop-deleting-offensive-comment/; Panoptykon Foundation, *First court decision in SIN vs Facebook: the internet giant must not restrict the organisation's activities in its services*, (July 2, 2019) https://en.panoptykon.org/articles/first-court-decision-sin-vs-facebook.

The Court should not assume that filters' errors will be harmless. Nor should it assume, without close analysis, that those harms to Facebook's billions of users are proportionate or legally acceptable. A ruling that is vague on this issue would support disturbing incursions into the rights of ordinary people on the Internet, and provide little guidance to Member State courts or to lawmakers in the coming years.

c. What Happens if Filters Take Down Legal Content?

If filters do take down the wrong expression or information, the usual next question in human rights discussions is whether those errors can be caught and corrected. There are a number of conventional measures platforms could take to weed out errors—though their effectiveness is debated. The AG's interpretation of the eCommerce Directive, though, seems to deter him from endorsing measures that require platform employees to review content. His proposed protections for expression and information rights instead depend on direct involvement by courts.

Human review is the most commonly proposed correction for filters' mistakes, and is widely employed by platforms today. Facebook says its employees generally review content flagged by filters as duplicates of violent extremist images or videos, for example.[63] Many critics counter that such review may be, in Austrian Professor Ben Wagner's words, a "rubber-stamping mechanism in an otherwise completely automated decision-making system."[64] This concern about the limited efficacy of human review is reinforced by examples like YouTube's removal of Syrian Archive videos, as well as research showing high rates of over-removal in notice and takedown systems where human review is the norm.[65]

---

[63] Facebook Newsroom, *Hard Questions: What Are We Doing to Stay Ahead of Terrorists?*, (November 8, 2018) https://newsroom.fb.com/news/2018/11/staying-ahead-of-terrorists/.
[64] Ben Wagner, *Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision-Making Systems*. Policy & Internet, 11: 104-122 (2019) https://doi.org/10.1002/poi3.198.
[65] *See* studies cited in Daphne Keller, *Empirical Evidence of "Over-Removal" by Internet Companies under Intermediary Liability Laws*, Center for Internet and Society (October 12,

Platforms could also provide notice to the affected users when their posts are removed. Germany's NetzDG law, for example, requires this. Civil society groups around the world have called for notice to users in this situation, and said that users must be able to appeal takedown decisions to platforms' own moderation teams as a bare minimum for protecting fundamental rights. Human rights literature, too, supports this approach.[66] But the little we know about existing mechanisms of this sort is not encouraging, either. Platform transparency reports suggest that the rate of appeals from users accused of violating the law is usually under 1%—far below the documented rates of false accusations or takedowns.[67] Letting users appeal when their own posts are removed also does little to help *readers* who want to access information. It is a remedy for speakers only, not for the numerous ordinary people, reporters, and civil society organizations that rely on platforms like Facebook as important sources of information.[68]

The AG does not engage with corrective measures, like human review and counter-notice, that depend on hands-on platform engagement with user content. His only proposed correction for errors is to let users appeal takedowns in court, following a similar requirement in a case about ISPs blocking websites.[69] Making people go to court before they can access lawful information is a high bar, though. And with Facebook content removals, unlike ISP website blocks, it is hard to see how users seeking information will even know what is missing.

2015) http://cyberlaw.stanford.edu/blog/2015/10/empirical-evidence-over-removal-internet-companies-under-intermediary-liability-laws.

[66] *See* Manila Principles on Intermediary Liability, https://www.manilaprinciples.org; David Kaye (Special Rapporteur), Human Rights Council, Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression at 6, U.N. Doc. A/HRC/32/38 (May 11, 2016), http://ap.ohchr.org/documents/dpage_e.aspx?si=A/HRC/32/38 (Manila Principles "establish baseline protection for intermediaries in accordance with freedom of expression standards").

[67] Daphne Keller, *Counter-Notice Does Not Fix Over-Removal of Online Speech*, Center for Internet and Society (October 5, 2017) http://cyberlaw.stanford.edu/blog/2017/10/counter-notice-does-not-fix-over-removal-online-speech.

[68] Groups including Amnesty International and Reporters Sans Frontières recently noted Facebook's particular importance as an information source, in a letter raising concerns about filtering proposals. WITNESS et al, *Letter of January 28, 2019*, https://drive.google.com/file/d/1WTgl5hjJ_cAE1U0OjqaQ9AucU6HNlhoi/view.

[69] Par. 65, citing C-314/12, *UPC Telekabel Wien*, (2014).

The AG also says harm to fundamental rights can be avoided by ensuring that monitoring orders are limited in duration.[70] That seems doubtful. If Facebook has to build the filtering tool he recommends, it's hard to see why they would ever dismantle it. This particular order may be, as he notes, interlocutory. But the plaintiff could presumably sue again if Facebook lets these posts reappear. The best way for Facebook to avoid that is to keep filtering, even if the injunction expires. Once both the new filtering tool and legal precedent exist, other plaintiffs and governments around the world will surely demand that Facebook use it for their own purposes.

The Court should not disregard the best-known tools for correcting filters' errors. At the same time, it should not assume that human review or user appeals are truly effective—or that smaller platforms could even afford to pay employees to carry them out. Ultimately, the open questions about those measures' effectiveness are part and parcel of this case's core inquiry: whether and how filtering injunctions can be reconciled with fundamental rights in the first place.

2. The eCommerce Directive

The doctrinal questions referred by the Austrian court turn on Articles 14 and 15 of the EU's core Intermediary Liability law, the eCommerce Directive. Article 14 immunizes hosting providers from liability for users' content unless they know about that content or are "aware of facts or circumstances from which [it] is apparent." Article 15 says that Member States cannot impose on hosts any "general obligation to monitor the information which they transmit or store[.]"

The AG's discussion of these Articles is hard to parse. Many of his concerns seem driven not by the Articles' plain language, but by what he calls a "reading of Article 14(3) in conjunction with Article 15(1)[.]"[71] As he seems to see it, the two

[70] Par. 60.
[71] Par. 40.

Articles have a structural relationship, with each defining and limiting the other's scope. Under Article 15, courts can only order hosts to do things that the hosts could have done anyway without losing immunity under Article 14. By the same token, if Article 15 precludes a court from mandating an operation, a platform that carries out the same operation voluntarily may risk losing immunity under Article 14. European legal experts I've talked to all think the AG's analysis means slightly different things. Some of those things are very troubling—not only for this case, but for the EU's overall Intermediary Liability legal framework and the protection of fundamental rights.

The Court's analysis can and should be much simpler, tracking the Court's prior interpretations of the eCommerce Directive. As I discuss below, I think that precedent—in conjunction with fundamental rights considerations—should lead the Court to reject to Austrian court's filtering injunction as a "general obligation to monitor," in violation of Article 15. Whatever the Court rules about Article 15, it should take care not to needlessly unsettle longstanding interpretations of Article 14. Specifically, it should avoid the implication that platforms' efforts to weed out unlawful content put them at risk under the Directive, making them automatically and categorically "aware of facts or circumstances" that strip them of immunity.

a. Can Courts Order Platforms to Filter All User Content under Article 15?

The simplest question raised under the Directive, and the only one most observers expected to be at issue in the case, is whether the Austrian injunction violates Article 15's prohibition on "general" monitoring. The exact scope of that prohibition has always been unclear, because the Directive also says courts can order hosts to take proactive measures to "terminate or prevent" infringements.[72] The closest the Court has come to defining the distinction between prohibited

[72] Art. 14.

"general monitoring" and permissible specific measures was in the 2011 *L'Oréal v. eBay* case.[73]

In *L'Oréal*, the Court said that monitoring affecting an entire service is "general" and thus prohibited. "[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."[74] By contrast, the Court offered two examples of narrower injunctions that might be permitted: requiring a host to terminate a *particular user's* account, or requiring it to make *that user* easier to identify.[75] The Court reiterated a similar rule in *Tommy Hilfiger v. Delta Center A.S.* Using the legal standards applicable to intermediaries, the Court held that injunctions may not compel a marketplace operator to "exercise general and permanent oversight over its customers," but may require measures that "contribute to avoiding new infringements *of the same nature by the same market traders* from taking place."[76]

Following *L'Oréal*, the injunction to filter "identical" content would appear to violate Article 15, because it requires Facebook to monitor "all the data of each of its customers." The AG nonetheless recommends that the Court uphold the injunction, based in part on Directive Recital 47, which says that monitoring for a "specific case" is not "general."[77] The Recital doesn't add much clarity, though, because it does not define what constitutes a "specific case." The AG takes it to mean a specific item of content. Following this reasoning, Article 15 would seemingly allow courts to order hosts to monitor for any number of specific items, but not to monitor for illegality generally. But a "specific case" could also mean a particular dispute or wrongdoing by a particular user—which would be consistent with both *L'Oréal*'s and *Tommy Hilfiger*'s examples of permissible

[73] C-324/09.
[74] Par. 139.
[75] Par. 141-142 (emphasis added).
[76] Case C-494/15 (2016) Par. 34 (emphasis added) (interpreting the *L'Oreal* standard in a case under Article 11 of Directive 2004/48).
[77] Par. 59-60.

injunctions, all of which focused on individuals. Interpreting "specific case" this way would also, as *L'Oréal* requires, avoid making a host monitor "all the data of each of its customers."

The AG also concludes that courts can order hosts to use "software tools" but not "active non-automated filtering."[78] This part of his analysis may derive from the relationship he seems to see between Articles 14 and 15: Courts can order automated filtering under Article 15, because such filters are "passive" enough to be permitted under Article 14. Reading the Directive to permit automated filtering but not human error-correction puts it in real conflict with human and fundamental rights guidance, as discussed above. That interpretation of Article 15 is also in tension with its plain language, since pervasive, automated monitoring would seem to be "general," in the *L'Oréal* sense of applying broadly to all users or content at once, while more targeted investigations might not be.

The Court can avoid these doctrinal snarls by focusing on the Directive's language and the Court's own precedent in cases like *L'Oréal*. Both that precedent and fundamental rights considerations counsel rejecting the Austrian court's filtering injunction.

b. Does Filtering Cause Hosts to Lose Immunity under Article 14?

In his analysis of Article 14 "in conjunction with" Article 15, the AG wrestles for several paragraphs with the idea that any host engaged in general monitoring "might well lose the status of intermediary service provider and the immunity that goes with it."[79] Ultimately, he concludes that operating the specific filter proposed in this case would not strip hosts of immunity. But he implies that this is a close call—that under different facts, hosts might indeed lose immunity. This will likely come as an unpleasant surprise to the many platforms that already voluntarily filter content including Child Sexual Abuse Imagery or violent

[78] Par. 61.
[79] Par. 36.

extremism, and to the political bodies that urged them to do so. The European Commission, for example, assured platforms in 2018 that taking "voluntary proactive measures does not automatically lead to the hosting service provider concerned losing the benefit of the liability exemption provided for in Article 14."[80]

CJEU precedent supports the Commission's conclusion about Article 14. Hosts do not lose immunity merely because they rely on automated content detection and management tools. As AG Jaaskinen said in *L'Oréal*, it would be "surreal" to deny hosts immunity based on any use of technical measures to "intervene[] and guide[]" content.[81] The Court in that case held that eBay generally qualified for Article 14 protections. Only when the platform got too involved in optimizing and promoting particular listings, the Court said, did it risk losing immunity.[82] In *Google France*, similarly, the Court said Google was immune for hosted ad content, even though Google's algorithms automatically ranked the content and decided whether to display it. The Court specifically rejected the idea that these automated operations made Google insufficiently "passive" or stripped it of immunity.[83]

Platforms' reliance on automated filters to block online expression may, as discussed above, raise concerns about fundamental rights. But it does not eliminate their eCommerce Directive immunities. The Court should steer clear of any analysis that implies otherwise.

---

[80] *Supra* note 37.
[81] Opinion of Advocate General Jääskinen, Case C-324/09, *L'Oréal SA v. eBay*, (December 9, 2010) (Par. 146) http://curia.europa.eu/juris/document/document.jsf?text=&docid=83750&pageIndex=0&doclang=EN&mode=lst&dir=&occ=first&part=1&cid=2516654
[82] *L'Oreal* par. 116-17.
[83] *Google France SARL and Google Inc. v Louis Vuitton Malletier SA* (C-236/08) (2019) Par. 114-117.

c. Does Human Review Cause Hosts to Lose Immunity under Article 14?

Finally, and most troublingly, the AG's Opinion may imply that platforms lose immunity when employees review content flagged by filters. The Court should not endorse such reasoning, which would undermine not only the filter-plus-human-review mechanisms currently being employed by some platforms and debated by the EU's political branches, but also ordinary notice and takedown operations under the eCommerce Directive.

The implication that human review might sacrifice immunities arises from the relationship the AG describes between Articles 14 and 15. An Article 15 injunction, he says, "cannot have the consequence" of making a host "no longer neutral" for purposes of Article 14.[84] Injunctions can require hosts to use "software tools," but cannot require them to use "active non-automated filtering"—seemingly meaning human review.[85] Read together, these two statements may suggest that "non-automated filtering" or human review is both outside a court's power to order under Article 15, and outside the scope of an immunized platform's operations under Article 14.

If human review made platforms too "active" and forfeited the protections of Article 14, platforms that already employ human moderators to check filters' work would have a problem. So would the policymakers and human rights experts who urged them to do so.[86] The untold number of European platforms that carry out ordinary notice and takedown would also face serious new difficulties. Platform employees routinely review and assess user content in response to notices alleging illegality in order to *maintain* immunity under the Directive. It would make little sense if those same actions automatically caused platforms to *lose* immunity.

---

[84] Par. 40.
[85] Par. 61. It is possible that the AG intended this recommendation to be grounded in concern about the expense or difficulty of filtering.
[86] *See* EU Commission, *supra* note 28; EU Council, *supra* note 28; EU Parliament, *supra* note 27; Council of Europe, *supra* note 16; EU human rights rapporteurs, *supra* note 16.

A rule making platforms automatically liable for anything an employee looked at would create powerfully imbalanced incentives. Platforms would have reason to either avoid moderating user content at all, or else remove anything that might create legal risk. That incentive to readily honor all notices would hurt platforms' users. Legal notices to platforms raise false or dubious legal claims frequently—anywhere from 5%[87] to 70% of the time[88] according to academic studies. Abusive takedown demands come from governments suppressing critical journalism, scientists trying to hide errors in their work, religious organizations targeting dissenters, and businesses attempting to undermine competitors.[89] Interpreting the eCommerce Directive to strip immunity every time a platform receives a notice—and thus effectively encouraging the platforms to honor all notices—would be at odds with both fundamental rights and Directive Recital 46, which says platform removal operations must "be undertaken in the observance of the principle of freedom of expression."

Of course, in many *particular* cases, a platform may lose immunity upon reviewing content. That happens, under Article 14, when the unlawfulness is "apparent," so the platform knows or should know that the content is illegal. In this case, for example, the post that Facebook was notified about and reviewed was, as Austrian courts later determined, obviously and recognizably illegal. By failing to take it down, Facebook became liable. But that doesn't mean platforms are liable every time they review content. CJEU precedent tells us that merely knowing that content exists, and even that someone has claimed it is illegal, does not by itself make platforms liable. A notice does not "automatically preclude the exemption from liability provided for in Article 14," for example, if it is

---

[87] Sharon Bar-Ziv and Niva Elkin-Koren, *Behind the Scenes of Online Copyright Enforcement: Empirical Evidence on Notice & Takedown* (July 15, 2018). Connecticut Law Review, Vol. 50, 2017 https://ssrn.com/abstract=3214214.

[88] Daphne Keller, *DMCA Classic, DMCA Turbo: Major new empirical research on notice and takedown operations*, Center for Internet and Society (April 20, 2016) http://cyberlaw.stanford.edu/blog/2016/04/dmca-classic-dmca-turbo-major-new-empirical-research-notice-and-takedown-operations.

[89] *See* sources cited at Keller 2018, *supra* note 51 at fn 24, 32, 33.

"insufficiently precise or inadequately substantiated."[90] Some Member State courts have said that platforms don't have culpable knowledge or lose immunity when faced with disputed facts[91] or difficult questions of law[92] that require court resolution.

If Facebook were to operate a filter to find duplicates of the Austrian user's post in this case, "notices" from the filter would presumably function somewhat like notices from claimants.[93] Facebook could assume the filter was always right, or it could have its employees investigate. The AG's analysis seems to indicate that Facebook should do the former—that it should not have employees investigate or apply human judgment, because doing so would take away the company's Article 14 immunities. That conclusion is inconsistent with fundamental rights, and is not required by CJEU precedent. The Court should not adopt it.

IV. Allowing Austria to Order Global Content Takedowns

The jurisdiction question—whether Austria can order global content takedowns—in a sense has the opposite problem from the monitoring question. The facts are pretty easy to grasp. But the law is a mess.[94]

Unlike the filtering issue, the political branches don't want to touch this one. That's a shame, because the global takedown issue raises policy questions of the

[90] *L'Oreal* Par. 122.
[91] England and Wales High Court Decision (2011) http://www.bailii.org/ew/cases/EWHC/QB/2011/3031.html.
[92] Barcelona appellate court judgment 76/2013 of 13 February 2013. Spanish law had previously required court orders for all takedowns, but the Supreme Court limited that standard in the *Internautas* case, Sentencia Núm. 914/2006, Tribunal Supremo (Sala 1ª de lo Civil) (2009), http://www.uaipit.com/files/jurisprudencias/1329312151_stc914.2006.pdf.
[93] *See generally L'Oreal* Par. 122 (discussing obligations when platform finds content of its "own initiative").
[94] Around the world, judicial decisions are often messy because courts conflate doctrinally distinct issues like jurisdiction to adjudicate, private international law or choice of law, comity, and scope of remedies. Even in this case, the Austrian appeals court flagged the question whether an injunction should be restricted to Austria, but then discussed only the separate, simpler question of whether to apply Austrian law.

sort courts typically rely on other branches of government to answer. Laws that let courts in one country reach across borders to take down expression protected in another, or laws that lead tech companies to erect digital borders, have consequences for everything from foreign relations to competition and trade. Governments would probably resolve these issues better if the relevant ministries and expert bodies had to sit down and hammer out solutions. Instead, courts around the world have been left to sort things out case-by-case. That's a recipe for uncoordinated, piecemeal outcomes, with priorities determined by whatever question happens to come before a court first.

In this case, the only question the AG and Court technically have to answer is whether the eCommerce Directive prevents Austria from ordering global takedowns. The AG says that it does not, which seems right to me. He also says that questions about takedowns outside the EU can't be answered by looking to intra-EU jurisdiction rules under sources like the Brussels Convention.

The AG thinks EU legislators *have* answered some other relevant questions about the substantive law though. He says that while defamation law varies between EU Member States, "the applicable material rules are harmonized" across the EU under data protection laws like the ones in Google's pending "Right to Be Forgotten" case.[95] On this point, I think he is—for better or for worse—incorrect. The GDPR made data protection law the same throughout the EU for many things. But it did not eliminate national legal differences for "Right to Be Forgotten" claims.[96] Like the 1995 Data Protection Directive before it, the GDPR

[95] Par. 79. This issue is more immediately relevant for Google's case than for Facebook's. But it matters for defamation plaintiffs, since they can often reformulate their claims under data protection law – and may do so if that confers an advantage in obtaining global removal. *See* Ashley Hurst, *Data Privacy and Intermediary Liability: Striking a balance between privacy, reputation, innovation and freedom of expression, Part 1*, International Forum for Responsible Media Blog (May 14, 2015) https://inforrm.org/2015/05/14/data-privacy-and-intermediary-liability-striking-a-balance-between-privacy-reputation-innovation-and-freedom-of-expression-part-1-ashley-hurst/.

[96] Google Op. Par. 77; *but compare* Opinion of Advocate General Szpunar, *Case C-136/17, G.C., A.F., B.H., E.D. v. Commission nationale de l'informatique et des libertés (CNIL),* (January 10, 2019) Par. 103 (noting variation in EU Member State law regarding the "Right to Be Forgotten" and criminal offenses)

leaves individual Member States to "reconcile the right to the protection of personal data … with the right to freedom of expression and information."[97] Member State law varies widely in the balance it strikes between those rights.[98] Unless this divergence in Member State law changes, there will always be cases where one country would uphold a "Right to Be Forgotten" claim, while another would reject it. Enforcing Sweden's laws in Hungary, or vice versa, inevitably risks, in the AG's words, "tak[ing] into account only one side of the coin."[99]

That's an unavoidable dilemma, not just for Europe, but the world. Countries that respect international human rights law frequently balance or interpret rights in different but equally permissible, ways. To respect other countries' sovereignty and principles of comity, the AG suggests, national courts and platforms can rely on geoblocking to prevent people from seeing particular content in countries where it is unlawful—while leaving it online in countries where the same content is lawful expression. That's not a perfect solution, because determined users can circumvent the blocks using VPNs or other tools. But that imperfection is not, in itself, reason to insist that one country's preferred balance between competing fundamental rights prevails over another's.[100]

It's hard to get away from prioritizing one set of rights or another, though. The AG implicitly does so in discussing the burdens of litigation. "[S]hould a claimant," he asks, "be required, in spite of the practical difficulties," to prove that content is forbidden under "all the potentially applicable laws" in the world?[101] That's a sympathetic argument. But if plaintiffs are spared this burden, then it will fall on defendants. One could equally ask if defendants should be required to prove that expression is lawful in other countries. Major platforms

---

http://curia.europa.eu/juris/document/document.jsf?text=&docid=209686&pageIndex=0&doclang=en&mode=req&dir=&occ=first&part=1&cid=6147463.
[97] GDPR Art. 85; *see also* Keller 2018, supra note 51 at 349.
[98] David Erdos, *Fundamentally Off Balance: European Union Data Protection Law and Media Expression* (2015); University of Cambridge Faculty of Law Research Paper No. 42/2014. https://ssrn.com/abstract=2471531.
[99] Google Op. Par. 36.
[100] Par. 100-101; Google Op Par. 75-76.
[101] Par. 97.

like Facebook might have resources to litigate that kind of thing—though there is little reason to think they *will*, outside of rare test cases. But smaller defendants won't. And as discussed in Section II of this White Paper, publishers and other proponents of expression or information rights may already face major disadvantages by not being participants in Intermediary Liability disputes in the first place.

Ultimately, in both the Google and Facebook cases, the AG lays out competing considerations, counsels against readily issuing global removal orders, but accepts that they may be appropriate in some cases.[102] What cases might those be? The AG understandably doesn't say, but a passage in his Google Opinion suggests an answer, or part of one. Under ECtHR case law, he notes, orders with extraterritorial effects are permissible when courts are protecting "human rights which form the basis of any State governed by the rule of law and from which no derogation is possible."[103] That's similar to the position urged by human rights organizations when Canada's Supreme Court considered similar issues.[104] Essentially, if the international human rights standard favors one party, and no human rights-compliant state could hold otherwise, then global takedown orders are appropriate.

The Court should be cognizant of the limitations created by comity and international human rights law, and of the important divergences between EU Member States' laws on both defamation and data protection, in resolving the jurisdiction issues presented in these cases.

---

[102] Par. 100; Google Op. Par. 62.
[103] Google Op. Par. 56.
[104] Factum of the Interveners, Human Rights Watch, Article 19, Open Net (Korea), Software Freedom Law Centre and Center for Technology and Society, *Google v. Equustek*, Supreme Court of Canada, (October 4, 2016) https://cis-static.law.stanford.edu/cis/downloads/HRW%20Equustek.pdf.

V. Conclusion

This is a complex case. Both the underlying factual questions about filters and online expression, and the core legal questions about the eCommerce Directive, fundamental rights, and jurisdiction, require careful exploration. Unfortunately, because of the case's speedy progression and the limited expert intervention, the Court and AG have received only relatively cursory briefing.

That background would counsel caution in any case, and particularly in a case so intertwined with current EU political debates and lawmaking. The Court should not approve the hastily-issued Austrian filtering injunction, and it should not provide sweeping support for global content takedown orders. Its analysis should emphasize the need for strong factual showings about filters' function, and the importance of avoiding undue burdens on privacy, expression, and other fundamental rights of third-party Internet users. Thoughtful and precise reasoning in this case will provide guidance to both the Austrian courts and the EU's political branches in key public decisions over the coming years.

About the Author

Daphne Keller is the Director of Intermediary Liability at Stanford's Center for Internet and Society. Her work focuses on platform regulation and Internet users' rights. She has published both academically and in popular press; testified and participated in legislative processes; and taught and lectured extensively. Her recent work focuses on legal protections for users' free expression rights when state and private power intersect, particularly through platforms' enforcement of Terms of Service or use of algorithmic ranking and recommendations. Until 2015 Daphne was Associate General Counsel for Google, where she had primary responsibility for the company's search products. She worked on groundbreaking Intermediary Liability litigation and legislation around the world and counseled both overall product development and individual content takedown decisions.

About the Center for Internet and Society

The Center for Internet and Society (CIS) is a public interest technology law and policy program at Stanford Law School and a part of Law, Science and Technology Program at Stanford Law School. CIS brings together scholars, academics, legislators, students, programmers, security researchers, and scientists to study the interaction of new technologies and the law and to examine how the synergy between the two can either promote or harm public goods like free speech, innovation, privacy, public commons, diversity, and scientific inquiry. CIS strives to improve both technology and law, encouraging decision makers to design both as a means to further democratic values. CIS provides law students and the general public with educational resources and analyses of policy issues arising at the intersection of law, technology and the public interest. CIS also sponsors a range of public events including a speakers series, conferences and workshops. CIS was founded by Lawrence Lessig in 2000.