Paddy Leerssen
Nieuwe Achtergracht 166
1018WV Amsterdam
The Netherlands

Parliament of the Republic of Austria
Dr. Karl Renner-Ring 3
1017 Vienna
Austria

Amsterdam, 15 October 2020

Re.: Consultation Response 49/ME

Dear sir/madam,

This document contains my submission for the Austrian Parliament's public consultation on the proposed KoPIG law (49/ME).

This submission is based on my expertise as a PhD candidate in information law at the University of Amsterdam, and as a non-residential fellow at Stanford University's Center for Internet and Society. I have also served as a researcher for the Transatlantic Working Group on Content Moderation Online and Freedom of Expression, where I co-authored reports on, *inter alia*, automation in online content moderation, and on Germany's *Netzwerkdurchsetzungsgesetz*.

Though this contribution is informed by professional experiences as a researcher, I submit it in a strictly personal capacity and solely on my own behalf.

My analysis is based on the English translation of the draft law, obtained from the website of the European Commission.[1]

I am grateful for the opportunity to respond to this consultation, and hope that my submission can be of some value in your deliberations,

Sincerely,

Paddy Leerssen

---

[1] https://ec.europa.eu/growth/tools-databases/tris/en/search/?trisaction=search.detail&year=2020&num=544

1

CONSULTATION RESPONSE

This submission consists of two parts. The first discusses the matter of transparency. The second discusses the design of reporting mechanisms.

Please note that this submission is not exhaustive of my concerns about the KopiG, and should not be taken as an endorsement of any other aspect of this instrument which I leave unaddressed (e.g. the design of the complaints office and the definition of regulated services, etc.).

PART I: TRANSPARENCY

Transparency is an essential element of legislation such as the KoPIG. Precisely because the KoPIG restricts (if not infringes) the fundamental right to freedom of expression and information, and since it relies heavily on powerful, foreign platform services to enforce its provisions, it is essential that the democratic public is able to assess how it is enforced in practice. Without adequate transparency safeguards, this is difficult or even impossible.

Consequently, the reporting obligation included under §4 of the draft law (*Berichtsplicht*, not to be confused in the English translation with the 'reporting' of content by users, or *Meldeverfahren,* under §3) is an essential aspect of the law and must be retained in the final version. In particular, the KoPIG offers an important improvement over its precursor, Germany's *Netzwerkdurchsetzungsgesetz* ('NetzDG'), by specifying unambiguously that reporting obligations also apply to content reported under the law which is removed on the basis of Community Guidelines and other Terms of Service restrictions, rather than on the basis of illegality under national law. However, there is also room for improvement.

The Reporting Obligation under §4 should address the use of automated systems

Modern content moderation is not performed exclusively by humans, but increasingly by complex automated systems – operating either autonomously, or interactively in conjunction with human moderators. Indeed, during the Covid-19 crisis, this trend has accelerated as platforms including YouTube and Facebook have been forced to rely even more extensively on autonomous systems as their human workforces are waylaid.[2] An extensive academic literature points to the potential for bias and inaccuracy in these automated moderation systems, creating new and heightened risks to fundamental rights.[3]

At present, §4(2)(6) only makes reference to 'technical equipment' as a topic of disclosure, but this insufficiently precise to guarantee that meaningful information is shared about these complex automated systems. A more robust provision would demand that platforms describe and explain, at a minimum, what types of automation they employ; how these systems are guided by and interact with human operators; whether they employ machine-learning, and, if so, what types of training data have been used to develop these systems; what the estimated false positive and false negative rates of these systems are; and what safeguarding measures have been deployed to correct for bias and inaccuracy in these systems.

---

[2] https://www.wired.co.uk/article/coronavirus-facts-moderators-facebook-youtube
[3] e.g. Robert Gorwa, Reuben Binns and Christian Katzenbach (2020). "Algorithmic content moderation: Technical and political challenges in the automation of platform governance". Big Data & Society 7:1. Hannah Bloch-Wehba (2020). "Automation in Moderation". *Cornell International Law Journal,* forthcoming. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3521619. Emma Llanso, Joris van Hoboken, Paddy Leerssen and Jaron Harambam (2019), "Artificial Intelligence, Content Moderation and Freedom of Expression". Transatlantic Working Group. Available at: https://www.ivir.nl/publicaties/download/AI-Llanso-Van-Hoboken-Feb-2020.pdf

The KoPIG must be complemented by an independent research access framework

Whilst the reporting obligations in KoPIG are a useful first step towards much-needed transparency, I must emphasize that they remain grossly insufficient by themselves. My concern here is that the KoPIG, much like the NetzDG, continues to focus its reporting obligations on *aggregate takedown statistics*, even though the value of this data for independent research and, ultimately, accountability in content moderation, is minimal.

Aggregate takedown statistics tell us something about the overall scale of the KoPIG's enforcement, but they tell us very little about the *quality* of this enforcement. They tell us how much content is removed, and on what basis, but they cannot explain whether these decisions were in fact correct – whether they accurately targeted illegal content, or instead censored constitutionally protected speech.[4] In other words, the most meaningful questions about the quality of content moderation, and, ultimately, the protection of free expression, remain unanswered.

To make such assessments about the quality of content moderation, researchers need insight into the actual content that is being removed (or not being removed), and neither the NetzDG nor KoPIG provide this insight. Since the content at issue often privacy-sensitive and/or unlawful, this information cannot be made publicly available. Instead, therefore, it is essential that laws on content moderation such as KoPI-G are complemented by efforts to create confidential data access frameworks for independent researchers.

This is not a novel idea in public policy but one that has already been implemented in other sectors such as public health.[5] Through a combination of technical protection measures (such as secure operating environments) and legal conditions (sanctions for abuse of data), trusted researchers can be offered access to sensitive data such as potentially illegal user-generated content whilst minimizing the risk of abuse. In collaboration with researchers from AlgorithmWatch and the University of Amsterdam, we have develop detailed best practices for governments wishing to develop such a framework.[6]

Creating confidential access regimes for sensitive data is a task of significant complexity, and I therefore recognize that it may be too late to incorporate this into the KoPIG legislation. Nonetheless, I take this consultation as an opportunity to emphasize that legislation such as KoPIG will remain deeply circumspect from a rule of law and fundamental rights perspective so long as adequate content-level transparency is impossible.

If the KoPIG is indeed passed, the Austrian government should therefore pursue as an immediate policy priority, both at national and EU level, the creation of confidential access frameworks for independent research into the content processed under this law.

PART II: DESIGN OF THE REPORTING PROCEDURE (§3(2))

Anyone who has studied the NetzDG in detail understands how important the design of reporting procedures is. As is by now well-known, Facebook received significantly fewer complaints under

---

[4] Daphne Keller & Paddy Leerssen (2020), "Facts and Where To Find Them: Empirical Research on Internet Platforms and Content Moderation". Joshua Tucker & Nathaniel Persily (eds.), *Social Media and Democracy: The State of the Field and Prospects for Reform*. Cambridge University Press. Available at: https://www.cambridge.org/core/books/social-media-and-democracy/E79E2BBF03C18C3A56A5CC393698F117

[5] Jef Ausloos, Paddy Leerssen & Pim ten Thije (2020). "Operationalizing Research Access in Platform Governance What to learn from other industries?". AlgorithmWatch. Available at: https://www.ivir.nl/publicaties/download/GoverningPlatforms_IViR_study_June2020-AlgorithmWatch-2020-06-24.pdf

[6] Id.

NetzDG than Twitter and Google due to the fact that their reporting mechanism was not as easily accessible to users.[7]

Section § 3(2) of the KoPIG seeks to regulate the design of reporting procedures by requiring, inter alia, that they are "easy to find, constantly available and easy to use". This is an important addition and ought to be maintained in the final text. However, these standards remain rather general and open to interpretation, and further detail could help to prevent foreseeable abuses.

The Austrian legislator might draw inspiration here from ongoing reforms in Germany, where the *Draft Law to Amend NetzDG* attempts to address the same problem.[8] This proposal contains a similar general standard about the mechanism being "easy to find, constantly available and easy to use", but it also contains two specifications which would be useful in the Austrian context.

Firstly, the *Draft Law to Amend NetzDG* specifies that the review procedure must be available "while viewing" the content in question. This proximity between content and reporting mechanisms is already standard practice for most self-regulatory flagging systems operated by major online platforms. Requiring the same of KoPIG's reporting mechanisms will help to ensure that it remains equally visible, and that users seeking content removal are made aware of this option as an alternative to self-regulatory remedies. If the KoPIG's reporting mechanism is only visible on a different page, or hidden at the bottom of the page far removed from the content at issue, there is a risk that users will rely solely on the platform's self-regulatory mechanisms.

Secondly, the *Draft Law to Amend NetzDG* specifies, in its *Gesetzesfolgen*, that the requirement of accessibility includes the principle that the reporting mechanism must not contain statements that are "unnecessarily deterring" (*unnötig abschreckend*), including legal warnings and disclaimers such as an instruction that he reporting user should specify a specific clause from national statutes. This is an important addition because Facebook has, in the context of NetzDG, included various disclaimers in their reporting mechanism which were likely to discourage users from submitting a complaint in this way.[9] Similarly, Facebook also warns users on that "allegations of illegality are a serious matter, and you may wish to seek guidance from an attorney or regulatory body before submitting a report."[10] The KoPIG should consequently include a clear prohibition against language that has the likely result of creating an unnecessary chilling effect on usage of the complaint mechanism.

---

[7] E.g. William Echikson & Olivia Knodt (2018). "Germany's NetzDG: A Key Test for Combatting Online Hate". CEPS Policy Insight. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3300636. Heidi Tworek & Paddy Leerssen (2019). "An Analysis of Germany's NetzDG Law". Transatlantic High Level Working Group. Available at: https://www.ivir.nl/publicaties/download/NetzDG_Tworek_Leerssen_April_2019.pdf. Amelie Pia Heldt (2019). "Reading between the lines and the numbers: an analysis of the first NetzDG reports". *Internet Policy Review* 8(2). Available at: https://policyreview.info/articles/analysis/reading-between-lines-and-numbers-analysis-first-netzdg-reports

[8] www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/RegE_Aenderung_NetzDG.pdf;jsessionid=D89 52A9C53BA2715E7C504D2C0BE7A44.2_cid297?__blob=publicationFile&v=2

[9] Amélie Pia Heldt (2019). "Reading between the lines and the numbers: an analysis of the first NetzDG reports". *Internet Policy Review* 8(2). Available at: https://policyreview.info/articles/analysis/reading-between-lines-and-numbers-analysis-first-netzdg-reports

[10] https://help.instagram.com/877138215797916?helpref=search&sr=13&query=How%20do%20I%20report%20a %20deceased%20person%27s%20account%20on%20Instagram%3F

CONCLUSION

The KopIG is an experiment in legislation with potentially profound impacts on fundamental rights. It is therefore crucial that the public has the right data to assess the outcomes of this experiment. The current proposal falls short in this regard.

- The list of reporting obligations in §4(2) KopIG should be expanded to require detailed information about the use of automated systems.
- The KoPIG should be complemented by a legally mandated data access framework enabling independent research into the content processed under this law.
- The requirements for the design of the reporting procedure § 3(2) should be further specified based on the example of the *Draft Law to Amend NetzDG.*