ALGORITHM WATCH

/ Submission by AlgorithmWatch

On the European Commission's "Digital Services Act" (DSA)

September 2020

As an organization committed to upholding democratic values and fundamental rights, we see an urgent need to commit internet platforms to a higher level of transparency. We welcome the European Commission's proposed Digital Services Act (DSA) as an opportunity to hold powerful platforms to account; and we strongly urge EU policymakers to put effective transparency at the heart of this "Magna Carta for Digital Services" by (1) maintaining the limited liability regime outlined in the e-Commerce Directive; (2) developing approaches to effectively audit algorithms; and (3) introducing binding data access frameworks for public interest research, including (4) comprehensive advertisement libraries.

I. Context

Internet platforms play a central and ever-expanding role in modern society

In the spring of 2020, AlgorithmWatch found clear indications that Instagram's newsfeed algorithm was nudging female content creators to reveal more skin, by making semi-nude photos more visible in their followers' feeds[1]. When we confronted Facebook, Instagram's parent company, with our findings, they declined to answer our questions, and accused us of "misunderstanding how Instagram works." They also rejected any requests to back up this statement with their own data, making it impossible to validate their claims.

Instagram's response to our investigation is emblematic of a much deeper problem: algorithmically-driven internet platforms play a central and ever-expanding role in modern society. They are inextricably linked to how we coordinate remotely at school or work, how we find and consume information or products, and how we organize our social movements or exercise key democratic rights. From vacation home rental services to social networking sites, large parts of our commerce and communications infrastructure are governed by opaque and proprietary algorithmic curation, optimization, and recommendations systems.

[1] Judith Duportail et al (2020): Undress or fail: Instagram's algorithm strong-arms users into showing skin.

Platforms rely on extremely intrusive data collection practices to power these systems, but when independent journalists, activists or academics try to understand the effects of these practices, data for public interest scrutiny is a scarce resource[2].

In recent years, platforms have further restricted access to their public Application Programming Interfaces (APIs), making it nearly impossible to hold companies accountable for illegal or unethical behavior[3].

The trend is especially worrisome with regard to online platforms that serve as content hosting providers. Today, these intermediaries play a key role in democratic deliberation. Algorithmically-driven content curation, on social media platforms in particular, can introduce a host of risks that affect the functionality of communication processes necessary for democracy[4], ranging from hate speech, to extremist content to algorithmic curation that discriminates against certain groups/parts of society.

Towards accountability in the digital public sphere

The e-Commerce Directive from 2000 cannot address the power that online platforms exercise in today's digital society, let alone make the digital ecosystem fit for the future.

When updating the e-Commerce directive, we need to move beyond approaches that rely on online platforms' self-regulation. Self-regulatory transparency frameworks have been found "incomplete, ineffective, unmethodical, and unreliable"[5]. And although improved user-facing transparency can offer much-needed insight into the personalized results presented to individual users, it cannot provide insight into the collective influence of platforms. To assess and monitor how platforms apply their community standards, or address collective societal risks like disinformation, polarization, and bias, we must rely on evidence from independent third parties/audits.

From Germany's Interstate Media Treaty to France's Avia Law to the EU's most recent proposal on preventing the dissemination of terrorist content online, previous approaches to tackling platform power have been overlapping, incoherent, or otherwise fragmented[6]. This member state-driven patchwork approach has created insecurity for businesses and weakened civil

[2] Madeline Brady (2020): Lessons Learned: Social Media Monitoring during Elections: Case Studies from five EU Elections 2019-2020 (Democracy reporting International.

[3] Researchers depend on private data sharing partnerships and privileged access to platform data which has the effect of further entrenching platform power, leading to chilling effects amongst researchers, who are afraid to lose access to platform data. Researchers who depend on what little data is available complain about its poor quality. Frequently, data is not available in machine-readable format or it is clearly inaccurate. The consequences are twofold. The impact of platforms on society remains severely understudied at a systemic level, and the research that exists skews heavily towards the most transparent platforms, causing substantial distortions. For further details see Nikolas Kayser-Bril (2020) For researchers, accessing data is one thing. Assessing its quality another; and Under the Twitter streetlight: How data scarcity distorts research.

[4] Birgit Stark et al (2020): Are Algorithms a Threat to Democracy? The Rise of Intermediaries: A Challenge for Public Discourse.

[5] For further elaboration on the recommendations see Jef Ausloos et al (2020): Operationalizing Research Access in Platform Governance: What to Learn from Other Industries?.

[6] Matthias Cornils et al (2020): Designing Platform Governance: A Normative Perspective on Regulatory Needs, Strategies, and Tools to Enhance the Information Function of Intermediaries.

society's collective power. Accountability can only be achieved through a consolidated European approach.

The Digital Services Act (DSA) is a great opportunity to fix these regulatory shortcomings and pave the way for a more accountable digital public sphere.

II. Policy Recommendations

Below, we outline our key demands for the DSA. The obligations and enforcement powers of the proposed institution should differentiate between major players and smaller intermediaries. We propose that the scope of the recommendations be limited to dominant platforms[7].

1. Maintain the limited liability regime

Public authorities bound by fundamental rights cannot ban "low-quality" content or demand its suppression as long as it is legal[8]. This is a fundamental basis of freedom of expression and must not be undermined. The limited liability framework outlined in the E-Commerce directive is, thus, the right approach to dealing with illegal user-generated content. However, this framework must be enhanced and refined. Instead of introducing measures that oblige or encourage platforms to proactively monitor speech, the **existing limited liability regime should be enhanced through more rigorous, and clear procedural standards for notice and action**. Such standards should include complaint management and redress mechanisms, including put-back obligations. When embedded in a co-regulatory approach, independent dispute settlement bodies, such as those proposed by European Digital Rights (EDRi)[9], can play an important complementary role in ensuring that users' fundamental rights are upheld. Moreover, an **updated limited liability regime needs to be complemented by transparency frameworks and systems to effectively audit algorithmic systems** that empower independent third parties to hold platforms to account and thereby help to, at least partly, relieve the individual user from the burden of proof.

2. Develop approaches to effectively audit algorithmic systems

AlgorithmWatch **calls on the EU Commission to initiate a stakeholder engagement process aimed at developing effective measures for the auditing** of **algorithmic systems**. We understand auditing in this context in accordance with ISO's definition as a "systematic,

[7] For nuanced criteria that characterize dominant platforms/intermediaries, see EDRi (2020): Platform Regulation Done Right. EDRi Position Paper on the EU Digital Services Act, p.16.

[8] Matthias Cornils et al (2020): Designing Platform Governance: A Normative Perspective on Regulatory Needs, Strategies, and Tools to Enhance the Information Function of Intermediaries; at the same time international human rights law (e.g. Art. 15 ECHR) puts very strict requirements for the conditions under which states can restrict freedom of expression and information, notably the principles of legality, necessity and proportionality and legitimacy.

[9] EDRi (2020): Platform Regulation Done Right. EDRi Position Paper on the EU Digital Services Act.

independent and documented process for obtaining objective evidence and evaluating it objectively to determine the extent to which the audit criteria are fulfilled."[10]

Specifically, the **Commission should leverage stakeholder input to determine**:

(a) **audit criteria** against which to assess systems. They can be laws or ordinances, the online services' own terms of service or content moderation guidelines[11], technical standards or even broader societal norms

(b) **appropriate auditing procedures** of examination and assessment against these criteria.

Depending on the purpose of the audit and the system audited, different approaches can be appropriate, including code reviews, conformity assessment, examination of training data, "black box testing" / audit studies or other means. Auditing can also be done in the form of risk / impact assessments, especially in cases where algorithmic systems may produce negative externalities (e.g. a vacation rental online marketplace driving gentrification of communities, information intermediaries' automated content moderation systems putting certain communities at risk).

Both criteria and procedures should be further developed following a multi-stakeholder approach that actively takes into consideration the disproportionate affect ADM systems have on vulnerable groups and solicits their participation. We therefore ask the Commission and Member States to make available sources of funding aimed at enabling participation by stakeholders who have so far been inadequately represented.

The **Commission should clarify**:

- Who / what (services / platforms / products) should be audited? How to customize the auditing systems to type of platform / type of service?
- When should an audit be undertaken by a public institution (at EU Level, national level, local level), when can it be done by private entities/experts (business, civil society, researchers)?
- How to clarify the distinction between assessing impact ex ante (i.e. in the design phase) and ex post (i.e. in operation) and the respective challenges
- How to assess trade-offs in the different virtues and vices of auditability: e.g. simplicity, generality, applicability, precision, flexibility, interpretability, privacy, efficacy of an auditing procedure may be in tension
- Which information needs to be available for an audit to be effective and reliable (source code, training data, documentation)? Do auditors need to have physical access to systems during operation in order to audit effectively?
- What obligation to produce proof is necessary and proportionate for vendors / service providers?
- How can we ensure the auditing is possible? Do auditing requirements need to be considered in the design of algorithmic systems ("auditable by construction"?)

[10] ISO (2018): Guidelines for auditing management systems (199011).

[11] In general, online services and platforms should be free to choose the terms by which they offer their services and the guidelines they use to moderate content. But they e.g. have an obligation to treat all users equally, so that this treatment needs to be accounted for.

- Rules for publicity: When an audit is negative, and the problems are not solved, what should be the behavior of the auditor, in what way can it be made public that a failure occurred?
- Who audits the auditors? How to make sure the auditors are held accountable?

3. Introduce comprehensive data access frameworks for public interest research

The Commission must urgently address the auditing questions outlined above. In the meantime, it is crystal clear that any successful accountability framework will require improved transparency and access to platform data. That's why **AlgorithmWatch proposes that the DSA introduce mandatory data access regimes for public-interest research that empower civil society and pave the way for true accountability.**

Learning from best practices in privacy-respecting data-sharing governance models, the DSA should enshrine:

(a) **Binding rules outlining *who* can directly access data or can apply for access, what specific data can be accessed**[12] **and *how and by whom* that data is to be gathered and checked before disclosure.**

- Disclosure obligations should be based on the technical functionalities of the platform service, rather than more ambiguous and politically-charged conceptions of harm such as 'disinformation', 'political advertising', and 'hate speech'.
- Technical features might include: high-level aggregate audience metrics; advertising and micro-targeting; search features; feeds, ranking and recommendation; and content moderation (including removal but also other measures such as demonetization or fact-checking).

(b) **An EU institution with a clear legal mandate to enable access to data and to enforce transparency obligations in case of non-compliance across the EU27.**

- The institution should **act as an arbiter in deciding on requests for confidentiality from the disclosing party** (based on e.g. intellectual property or data protection law). Barriers to gaining access to pre-defined data should be minimized. The institution should maintain relevant access infrastructures such as virtual secure operating environments, public databases, websites and forums. It should also be tasked with pre-processing and periodically auditing disclosing parties to verify the accuracy of disclosures.
- Furthermore, the mandate shall comprise **collaboration with multiple EU and national-level competent authorities such as data protection authorities, cyber-**

[12] It is essential that disclosure rules remain flexible and subject to updates and revisions by the proposed independent institution.

security agencies and media regulators to minimize the risk of capture or negligence. The legal framework should explicitly outline different levels of oversight and how they interact. Because trust in government bodies differs widely across Member States, installing tiered safeguards and guarantees for independence is critical. To prevent competence issues and minimize the politicization of the framework, it is advisable that the role of such an institution be limited to the role of a 'transparency facilitator.'

- The institution shall **proactively support relevant stakeholders.** The freedom of scientific research must be explicitly enshrined. In this spirit, the proposed institution must proactively **facilitate uptake, tools and know-how among stakeholders including journalists, regulators, academics, and civil society**. The institution might also explore the possibility of **engaging the broader European public in the development of research** agendas (see e.g. lessons from the Dutch National Research Agenda[13]) or by incubating pilot projects that explore the possibility of connecting users and researchers through fiduciary models. **Independent centers of expertise on AI/ADM at national level**, as proposed by AlgorithmWatch and Access Now[14], could play a key role in this regard and support building the capacity of existing regulators, government and industry bodies.

(c) Provisions that ensure data collection is privacy-respecting and GDPR compliant

- Because of the sensitive nature of certain types of data, there are legitimate concerns to be raised regarding threats to user privacy. The Cambridge Analytica scandal should serve as a cautionary tale, and any misuse of data by researchers would severely undermine the integrity of any transparency framework.
- It is **imperative that the institution upholds the GDPR's data protection principles** including (a) lawfulness, fairness and transparency; (b) purpose limitation; (c) data minimization; (d) accuracy; (e) storage limitation and (f) integrity and confidentiality.
- The proposed data access entity should take inspiration from existing institutions like the Finnish health data framework Findata[15] which **integrates necessary safeguards (both technical and procedural) for data subjects**, including online rights management systems that allow citizens to exercise their data subject rights in an easy manner.
- **Granular data access should only be enabled within a closed virtual environment**, controlled by the independent body. As was the case with the Findata framework, it is advisable for the Commission to consider testing key components of the framework in pilot phases.

[13] Beatrice de Graaf et al (2017): The Dutch National Research Agenda in Perspective: A Reflection on Research and Science Policy in Practice.

[14] AlgorithmWatch (2020): Our response to the European Commission's consultation on AI.

[15] See Findata case study in Jef Ausloos et al (2020): Operationalizing Research Access in Platform Governance: What to Learn from Other Industries?.

4. Establish mandatory and comprehensive universal advertisement libraries

What kind of data access could such a framework help facilitate? A promising approach in the area of online advertisement are advertisement libraries ("ad-libraries") for online platforms. These ad-libraries disclose relevant information/data about advertisement on a particular platform. At the moment, such libraries are provided by select platforms on a voluntary basis, but as many studies on the implementation of the EU Code of Practice against Disinformation have shown, false negatives and false positives were rife in the political ad libraries of the signatories of the code: non-political advertisements were erroneously included in the libraries, while many political ads were excluded. The lack of a comprehensive repository of all ads made it impossible to verify whether all political ads were included in the libraries, and the political ad libraries and labelling missed a lot of sponsored content. In a situation where it is difficult to police the labelling of political ads, it is ultimately necessary to ensure the transparency of all ads. This is why we join the European Partnership for Democracy (EPD) and many other partners[16] **in calling for the introduction of comprehensive advertisement libraries**. Together we are **asking that the European Commission develop and legally enact a set of minimum technical standards and protocols** with which any ad-library has to comply. This ensures accessibility and compatibility across ad-libraries. As a guidance we **suggest to consider the criteria outlined in our joint statement.**

[16] See our joint statement here; in a similar vein Mozilla has called for ad archives.