Moderation Made in Europe

A Look into the Future of Social Media Content Moderation Litigation

Lorenzo Gradoni & Pietro Ortolani

02.04.2024 doi: 10.59704/e2e54bd5b465b8cf

The EU's Digital Services Act (DSA) has been fully applicable for a little more than a month now. The conditions are thus in place for the emergence of the out-of-court dispute settlement (ODS) ecosystem envisaged in Article 21 DSA, arguably the DSA's most original contribution to securing digital platform users' rights. In this post, we try to envision the shape such an ecosystem might take over the next few years in the key area of social media content moderation (SMCM).

We argue that the peculiarities of SMCM litigation, the way the DSA frames it, the interests of social media platforms and their considerable power – which includes the right *not* to enforce the rulings made by the ODS bodies established in accordance with the DSA – could lead to the emergence of an adjudication system dominated by a few ODS providers backed by public-private partnerships and ready to work in concert with the complaint-handling mechanisms set up by the platforms themselves. Our analysis singles out the most sophisticated of such mechanisms, that of Meta, the largest social media provider in Europe.

Filling the remedial gap

What triggers SMCM litigation is, typically, a platform's decision to take down or leave up a post or comment. The algorithms that run the platforms make millions of such decisions every day, many of which turn out to be controversial. In response, social media companies have set up complaint-handling mechanism dispensing justice in the form of quick rulings, machine-made or issued by overworked human moderators.

While SMCM litigation's sheer volume and generally low stakes makes it scarcely amenable to court rituals, it nonetheless calls for remedies suited to the sensitivity of its subject matter, often involving the alleged breach of human and constitutional rights. Out of this realisation came Mr Zuckerberg's decision to enable the creation of the Oversight Board, an entity with last-instance jurisdiction over cases introduced by users of Facebook, Instagram and Threads, and whose rulings bind Meta. The Board checks Meta's conduct against internationally recognised human rights and adjudicates in the style of a human rights court. However, it can only afford to deal with a handful of cases it deems critical: in the second semester of 2023, of the 158,786 complaints it received, the Board selected 15, which makes less than one in 10,000. The remedies offered by the platforms are therefore either perfunctory or inaccessible to most users. It is to close this remedial gap that the EU lawmakers sought to foster the emergence of an ODS ecosystem independent of platforms and standing as a cost-effective alternative to ordinary judicial remedies.

Under the DSA, all social media users in the EU have the right to challenge a platform's decision through ODS procedures working within comfortable timeframes and yielding reasoned rulings "free of charge or at a nominal fee" (all quotations in this post are from Article 21 DSA). How does the DSA ensure that such a service is provided virtually for free?

Platforms pay but may decline to comply

Member States may establish and operate ODS bodies or support their activities, but they have no obligation to do so. The DSA just requires them to designate a Digital Services Coordinator (DSC) responsible for EU-wide accreditation of ODS bodies based on common criteria, including the possession of adequate legal and technical expertise and guarantees of independence and impartiality. Against this background, the DSA

makes (virtually) free access to ODS possible by shifting procedural costs to the platforms, regardless of the outcome of the case. If the platform wins, it still pays, unless the ODS body finds that that the plaintiff "manifestly acted in bad faith" – an almost prohibitive threshold. If the plaintiff wins, the platform in addition will have to reimburse the nominal fee charged by the ODS body, if any, and "any other reasonable expenses" the plaintiff incurred "in relation to the dispute settlement" (presumably lawyer's fees commensurate with the complexity of the case and not exceeding local standard tariffs). To ensure that ODS bodies do not abuse the system – possibly in cahoots with bad-faith claimants – the DSA caps the fees charged to the platform at "costs incurred" and requires disclosure thereof prior to the start of the proceedings.

Non-compliance is the only weapon platforms have against an ODS body that they find too expensive, or that adjudicates disputes in a manner they dislike. This option is, to a certain extent, within a platform's rights: as the DSA emphatically states, "[t]he certified out-of-court dispute settlement body shall not have the power to impose a binding settlement of the dispute on the parties". However, platforms are not at liberty to systematically ignore ODS procedures and outcomes: the DSA requires them to engage in good faith with whatever accredited ODS body a user picks from an online register maintained by the European Commission.

It can be expected that platforms will use their right of non-performance strategically. Repeated non-compliance with rulings made by a specific ODS body may put the latter in serious trouble, as users will not go to an operator whose decisions a platform tends to disregard. However, generalised non-compliance would be at variance with the obligation of good-faith engagement spelt out in the DSA, potentially triggering the platform's liability under Article 54 DSA. All in all, it should be in the platforms' best interest to contribute to the smooth and cost-effective functioning of an ODS system whose operating costs they have to pay anyway. In this respect, the wisest course of action for platforms may turn out to be automatic compliance with rulings made by certain ODS bodies, identified by the platforms themselves as trustworthy.

Dual accreditation and conditional automatic compliance

Platforms pour a huge amount of money into in-house content moderation. Incidentally, the cost of an external watchdog like the Oversight Board – entirely borne by Meta – is comparable to that of the International Court of Justice. The ODS system envisaged by the DSA adds to such costs those of screening ODS bodies' decisions to sort out those to be complied with from those to be turned down. This process will be all the more cumbersome and expensive the less standardised the ODS bodies' output will be in terms of which law(s) they apply, their style of interpretation and modes of reasoning, and the format of their rulings. Interestingly, the former director of the Oversight Board administration expressed a similar concern.

To drastically lower such costs, platforms may establish their own accreditation schemes, parallel to the one governed by the DSA. The lynchpin of such systems would be automatic enforcement of rulings made by ODS bodies holding dual accreditation, i.e. being both certified under Article 21 DSA, and regarded as *a priori* trustworthy by the relevant platform. The remaining rulings would still have to go through a screening process, as their automatic rejection would be inconsistent with Article 21's good-faith engagement rule. Platforms, in other words, would grant accreditation to ODS bodies that commit to high standards of decision-making quality and cost-efficiency. For instance, for disputes concerning Facebook or Instagram Community Standards, Meta could require the embedding of the Oversight Board's case law – which in turn incorporates internationally-recognised human rights – into the decision-making standards of ODS bodies seeking accreditation. It could even entrust the Board with the task of screening accreditation applications. Meta would then contractually agree to automatically implement the rulings of the accredited ODS bodies. In exceptionally serious cases, Meta, or the Board itself by way of self-referral, could suspend enforcement and refer the case to the

Board, whose ruling would prevail over those rendered by the ODS body.

The upshot of such arrangements would be the formation of an integrated adjudication system based on the DSA and a network of international contracts connecting platforms and EU-based ODS bodies. This conglomeration of public and private law would achieve something – the mandatory nature of ODS rulings – which eluded the EU lawmakers. Such a system would likely become more complex over time. Initially, the size of the caseload that accredited ODS bodies will likely have to handle would make it unwise to enable them to request preliminary rulings from the Oversight Board (along the lines of the CJEU's preliminary reference procedure). However, conditions conducive to such developments may arise if the ODS ecosystem achieves a high degree of concentration and organisation.

Moderation made in Europe

The SMCM litigation market that the DSA conjures up will likely attract, at first, a host of firms with online dispute resolution expertise. This fragmentation would not facilitate cooperation with platforms, on which, however, the system's effectiveness ultimately depends. The recourse by platforms to accreditation systems of their own, coupled with conditional automatic compliance, would contribute to streamline the market, as users will gravitate towards ODS bodies that can promise that their rulings will be complied with, sidelining the rest.

Since platforms would likely adopt a wait-and-see strategy as they watch an EU-wide ODS system take shape, it will be up to European civil society to make the first move. An interesting opening would consist in the foundation of a European Centre for the Settlement of Disputes on Social Media Content Moderation. One could also think of a plurality of Centres whose respective jurisdictions would mirror intra-EU linguistic faultlines. We call "the Centre" an imagined institutional complex that could be either plural, unified or networked. For instance, the Centre could take the legal form of a non-profit spinoff of a publicly funded university or research centre, i.e. a setting bringing together various kinds of expertise – legal, sociological, linguistic, IT and most crucially AI – and cultivating an interdisciplinary culture open to experimentation.

To keep up with the platforms' ultrafast decision-making processes, the Centre would need to be able to manage a very large number of cases with the help of specially developed AI-powered language models capable of co-producing, in dialogue with human operators, rulings comparable in quality to a human rights court's – and this, well inside the timeframes set by the DSA. This would also make the Centre a major testing ground for human-machine interaction in the field of quasi-judicial decision-making, its justification and communication. The substantial public investment required to set up the Centre would be compensated by a steady flow of payments from platforms as they begin to be summoned before it, and by the economic and political benefits that the presence of such a decision-making institution in Europe would bring.

Ideally, the Centre would do on a large scale something that the Oversight Board can only perform in a tiny number of cases, i.e. handling SMCM litigation with the care it deserves, given the weighty questions of principle it often raises. The Centre would therefore make an ideal candidate for accreditation by the platforms. Such accreditation, one might think, would amount to a kind of vassalage bond, especially if – a suggested above – it should entail that primacy is attached to the rulings of a body such as the Oversight Board. We think otherwise.

To begin with, it would be a mistake to assume that the Board operates under Meta's thumb. Guarantees of its independence are robust: Meta's substantial funding is irrevocable and currently covers more than a decade of operation. The Board's members are selected (and will start self-selecting) from among people who are unlikely to turn into corporate puppets. Practice confirms that. In the second half of 2023, the Board overturned 82% of Meta's decisions it reviewed. Over the same period, Meta admitted error in 53 of the 75 cases that the Board shortlisted and notified to it. A similar pattern of prompt recognition of claimants' rights may arise as soon as Meta begins

to receive notices of appeal from the Centre. It can also be expected that the Centre's large case-law will impact the Board's at least as much as the Board's will contribute to shape the Centre's. The shared language of human rights will enable a two-way avenue of communication and mutual influence. In this way, Europe could speak with a still more authoritative voice in the global conversation on standards applicable to social media content moderation.

ABOUT THE AUTHORS

Lorenzo Gradoni
Lorenzo Gradoni holds a PhD in European Union Law (Bologna). He is currently a member of the Luxembourg Centre for European Law (LCEL).Lorenzo Gradoni holds a PhD in European Union Law (Bologna). Formerly Associate Professor of International Law at the University of Bologna and Senior Research Fellow at the Max Planck Institute Luxembourg for Procedural Law, where he also served as the Chief Scientific Advisor to the Max Planck Encyclopedia of International Procedural Law. He currently works at the University of Luxembourg.
Pietro Ortolani
Pietro Ortolani is a full professor of digital conflict resolution at Radboud University in Nijmegen, the Netherlands.

SUGGESTED CITATION

Gradoni, Lorenzo; Ortolani, Pietro: Moderation Made in Europe: A Look into the Future of Social Media Content Moderation Litigation, Verfassungsblog, 02.04.2024, doi: 10.59704/e2e54bd5b465b8cf.

Licensed under CC BY-SA 4.0.

Langzeitarchiviert bei IntRechtDok unter: doi.org/10.17176/20240402-230440-0.