Expression as the Reorganization of a Shared Condition: Freedom, Restriction, and the Participation Criterion

Pablo Dietrich
Master's student in Law, Estácio de Sá University (UNESA), Brazil. Attorney (OAB/SC 41,212). ORCID: https://orcid.org/0009-0003-5447-0550. E-mail: pablodietrichh@hotmail.com

Abstract: Debates on the limits of free speech tend to follow one of two shortcuts: from the finding that speech has altered something to the authorization to contain it, or from the protection of expression to the conclusion that nothing is owed to those it reaches. Both are refused here on the basis of a single criterion, that of participation: expression that reorganizes conditions others live from calls for justification before them; the intervention that restricts such expression also reorganizes conditions of participation and must meet its own requirements of justification, competence, and control. The criterion is unfolded into three questions — of diagnosis, of justification, and of response and competence — which separate the duty to answer from the authority to restrict and subject the restrictor to the same examination as the speaker. Drawing on three earlier studies on reception, political myth, and the authorization of action, the argument shows what reorganization alone does not authorize, how the available responses compare, and what each intervention must satisfy. It then applies the criterion to Brazilian law, recording what that law already secures and what remains open, especially in the regime of Article 19 of the Marco Civil da Internet after the Supreme Federal Court's final ruling in Themes 987 and 533. A comparison with the European Union's Digital Services Act and the United Kingdom's Online Safety Act shows that the gaps in the Brazilian regime — reasons for reductions in reach and for moderation under terms of service, and a review body independent of the provider — have normative counterparts elsewhere, which widen the institutional alternatives under discussion.

Keywords: freedom of expression; participation; justification; content moderation; comparative law; Brazilian Internet Civil Framework.

Introduction

Two moves recur in the debate on the limits of freedom of expression. One points to a harm and concludes in favour of containment; the other points to constitutional protection and concludes that nothing is owed to those whom speech reaches. Neither belongs to a single political camp, and each appears, depending on the occasion, in both. The two moves are shortcuts, and they are symmetrical. The first leaps from the finding that a piece of speech has altered something to the authorization to contain it, as if identifying a harm were the same as showing who may respond to it and how. The second leaps from the protection of speech to the dispensation from any justification, as if protecting an act were the same as declaring it to have no addressees. Both shortcuts run through the regimes that, in the last decade, have redesigned the liability of platforms — the European Union's Digital Services Act, the United Kingdom's Online Safety Act, the ruling of Brazil's Supreme Federal Court and the decree that implements it — and in each of them they have practitioners who read the other side as a political adversary. What follows does not choose a side in that dispute; it puts the same

questions to both positions, and the symmetry between them must appear in the application, not in the declaration.

The instrument is a criterion formulated elsewhere and already applied to action in general: that of participation. An action that reorganizes conditions others live from has those persons within it by virtue of the position they occupy, and it therefore owes them an answer (Dietrich, 2026a). Applied to speech, the criterion yields a result that neither shortcut accepts: expression that reorganizes conditions of participation calls for justification before those who live from them, but that requirement does not by itself become an authorization to restrict; and the intervention that restricts expression is itself a reorganization of conditions of participation, subject to the same questions it puts to the speaker.

Three earlier studies, on reception in online networks, on political myth, and on the authorization of action, prepared the ground without crossing it (Dietrich, 2026b, 2026c, 2026d); none derived from harm an authorization to restrict, and none will be derived from them here. What is added is the distinction the three did not draw: between the answer a piece of speech owes and the measure someone may impose on it there are three steps — diagnosis, justification, and the choice of a response by someone competent to adopt it — and each requires its own argument. The method is conceptual reconstruction submitted to test: the criterion is formulated, confronted with the strongest objections, and applied to two documented cases, to a hypothetical case with variations, to current Brazilian law, read in the text of the statutes and of the judicial ruling, and, by comparison, to the European and British regimes. Brazil is the main case because there the redesign was the most recent and the most abrupt, carried out by a court and completed by decree.

The ground to be covered thus runs from what is already established to what is not: from the three earlier results to the criterion and its questions; from what reorganization alone does not authorize to the comparison between responses and the requirements of each intervention; from intervention as reorganization to current Brazilian law — the Constitution, the criminal offences of defamation and incitement, Law 14,197/2021, and the final ruling of the Supreme Federal Court in Themes 987 and 533 — recording what that law already secures and what remains open; and, finally, to the comparison with the European and British regimes.

1. Three earlier results

The three studies that precede this one examine dimensions of a single phenomenon — how speech is received, what circulates, what is authorized — and each grounds a different piece of the argument. Taking them as a block would erase the difference and lend the whole a force that none of them has on its own.

The first described reception. When someone's success online is taken as fitness for office, it is not only the candidate who displays competence in a new way; it is the voter who comes to believe he possesses the competence to judge him. Effective communication is taken as enlightenment when it confirms expectations, offers an immediately reproducible answer, and produces a sense of mastery, without sufficient examination (Dietrich, 2026b). That result is hermeneutic, not injurious. No one is treated as a non-participant for receiving a message in this way, and no position is obstructed. What the study grounds is that speech can form in its recipient a judgment without examination; what it does not ground is any response beyond the one addressed to examination itself. Its hypotheses, moreover, describe a mechanism and do not count as empirical evidence that the mechanism operates in a given case.

The second dealt with content. Political myth, in the sense that Barthes, Edelman, Flood, and Bottici allow us to reconstruct, integrates different events into a narrative of shared political meaning, and the label derived from it replaces examination with belonging (Dietrich, 2026c). Yet the conclusion of that study was deliberately opposed to restriction: the most offensive interpretation of an ambiguous phrase is not, by itself, enough to close the square. The German Federal Constitutional Court, ruling on the phrase "soldiers are murderers", did not certify that the phrase was true or that it was fair; it held that the square in which it could be examined had to remain open (BVerfGE 93, 266). Legal protection of speech does not certify its judgments, and the degradation of debate by myth does not strip anyone of the standing of participant. What the second study grounds is a negative result: there is a class of speech that degrades the square without closing it, and the answer owed to it is examination.

The third dealt with authorization. An understanding formed under influence can turn a response the agent held to be forbidden into one he comes to hold as required, and the agent can do wrong without recognizing it as such. The cases of Southport, the Capitol, and Brasília showed the cycle — influence, understanding, action — and showed where it can be interrupted with the least harm: in timely public correction by whoever holds the information, in responsibility concentrated on the few who form the understanding of the many, and in a

narrowly bounded measure of reach reduction for content already verified as false during a declared crisis (Dietrich, 2026d). That study acknowledged the objection that the evidence that disinformation turns citizens into rioters is weak, and that a power to demote content in a crisis will be used outside it (Reilly, 2024), and it confined its proposal to limits the objection did not reach. What it grounds is that speech can reorganize the public assurance that a group will not be attacked for what it is (Waldron, 2012), and that there are responses to this that do not enter the cycle. What it does not ground is that the limits of that intervention are conditions of every legitimate response to expression: incitement can use true information, a threat can occur outside a crisis, and requiring a declared crisis hands to the authority that declares it part of the power to define when protection diminishes.

The order in which the three studies were written suggests a sequence: reception without examination would make the label effective, and the label would make authorization possible. The sequence is useful for organizing the analysis and misleading if read as a causal chain. None of the three showed that the first moment produces the second or that the second produces the third; each described a mechanism and gathered evidence of circulation — views, repetitions, calls to action — which is not evidence of effect. Four hundred and twenty thousand views of a false name and two hotels attacked in the following days are facts of the same episode, not a demonstrated causal relation between a post and an attack. What follows therefore keeps three things apart: the interpretation of the mechanism, the evidence of circulation, and the demonstration of effects. Only the last would authorize causal claims about a specific post, and it is rarely available.

2. The criterion: participation, justification, and competence

A person receives a message saying that her polling station has moved. The information is false and was spread to send her to the wrong address. No one declared that she had lost the right to vote. Her name remains on the register, her standing as a voter is still recognized, and the rules that secure her participation have not changed. Even so, the message can prevent her from exercising that position. Recognition of a right and the conditions for exercising it are different things.

Consider now an allegation of irregularity in the vote count. Its circulation may reduce confidence in the procedure and provoke challenges. That is not enough to place it alongside the previous message. The allegation may reveal a real defect, call for an investigation, or present a suspicion that is not confirmed. The reliability of a procedure includes the

possibility of examining its errors. Preserving it does not mean keeping intact the confidence people place in it. There is, however, another kind of allegation: one that remains unchanged after the count, the forensic examination, and the appeal, and that serves to mobilize rather than to examine. That allegation closes the very possibility it invokes. The difference between the two lies not in their content, which may be identical, but in the relation each bears to the examination it calls for: one opens it, the other replaces it.

The two examples show the difficulty of taking the alteration of a shared condition as an immediate ground for restriction. Expression modifies understandings, expectations, and relations. Some of these changes make it possible to correct injustices; others hinder the participation of those who already held a position. Identifying a transformation is where assessment begins.

A shared condition, in the domain considered here, is one on which several people depend for the exercise of their participation: the possibility of accessing operational information about an election, of contesting a decision, or of moving about without being subjected to targeted threats. Its existence does not require consensus on what is true or just. People who disagree profoundly may depend on the same conditions to sustain their disagreement. Nor is every shared condition presumed legitimate. An exclusionary practice may be stable, old, and widely accepted; those features do not justify preserving it. What distinguishes it from a legitimate condition is precisely what the criterion examines: the exclusionary practice treats some of those who depend on it as if they did not participate in it, and for that reason the second question, below, reaches it just as it reaches speech.

Participation designates a determinate relation with the condition under examination. The voter participates in the electoral procedure; the resident of a place depends on the conditions that make it habitable; whoever lodges a challenge participates in the practice in which reasons can be offered and answered. Identifying these relations requires description and argument. It is not enough to call someone a participant, nor to say that everyone reached by a consequence occupies the same position. The starting point is that of the book from which the criterion derives: no one begins by producing what he lives from, and receiving a condition is a mode of participating in it, not a lesser degree of participation (Dietrich, 2026a, chs. 1, 4, 8); and the field in which positions are distributed precedes the point of view of whoever acts upon it (Dietrich, 2026a, ch. 15).

The first question is diagnostic: what did the speech do — or give sufficient reason to fear it would do — to a condition of participation? The answer must identify the act, the condition, and the persons whose position is at stake. Those persons are within the act by

virtue of the position they occupy, not by virtue of the speaker's intention: the direction of an action does not arise from the agent's intention but from the place on which the action falls (Dietrich, 2026a, ch. 11). The answer must also distinguish a transformation that has already occurred from the risk of one. A message that never reaches its addressees, a threat that immediately alters their conduct, and a call to action that meets a mobilized audience present different problems. The abstract possibility of producing consequences does not allow them to be treated as equivalent.

That description does not yet contain the judgment of injustice. An allegation may make the continuation of a practice impossible because it revealed its abuses. True information may frustrate expectations whose maintenance was not owed. Criticism may shake the authority of someone who depended on the absence of challenge. The transformation of prior relations calls for examination; it does not establish a general priority for preserving them.

The second question is normative: why is the identified transformation justifiable or unjustifiable before those whose participation it reaches? The assumed ground is that whoever intervenes in the conditions of others' participation cannot treat that intervention as exclusively his own affair (Dietrich, 2026a, ch. 13). The premise is assumed, and its defence can be given in a few lines. Whoever acts upon a shared condition lives from it as much as those whose position he alters; treating the alteration as his affair alone would be to use what belongs to everyone as if it were his own, claiming for himself a use he denies to others. The premise refuses that asymmetry, and whoever rejects it must explain why a condition everyone lives from may be reorganized by one person without reasons addressed to the others. It requires reasons addressed to the positions affected. It does not require that every utterance receive prior authorization, nor that each affected person hold a veto over its circulation. The formulation comes close to Habermas's discourse principle, according to which those norms of action are valid to which all possibly affected persons could agree as participants in rational discourses, and it departs from it on two points. Habermas defines the affected party by the interests touched by the consequences of a regulation; here, the addressee is defined by the position occupied in the altered condition. The difference is not merely verbal: the advertiser whose revenue falls when a piece of content is removed has an interest affected and no position in the condition of participation at stake; the voter who never received the false message has no individual interest affected and nonetheless occupies a position in the procedure whose reliability the message affected. And the discourse principle submits norms to hypothetical assent, whereas the premise assumed here requires reasons

addressed to the affected positions without making their assent a condition of action. The proximity has a consequence that matters below: for Habermas, the discourse that validates norms presupposes the free flow of topics, information, and arguments in a public space (Habermas, 1997, vol. 1, p. 142), and the intervention that restricts expression bears on those very conditions.

The proximity is greater with Forst, for whom there should be no political or social relations of rule that cannot be adequately justified to those affected by them, by reasons meeting the criteria of reciprocity and generality, and for whom the core of injustice is arbitrariness, rule without legitimate grounds (Forst, 2012, pp. 1–2, 20–21). One feature separates him from the criterion adopted here, and it is that feature the argument must defend: Forst grounds the right to justification in the dignity of a being who can ask for and give reasons; the premise assumed here grounds it in the position occupied in a shared condition. His notion of an order of justification — any context in which norms and institutions govern a common life — nevertheless has a consequence that matters to the argument: the platform that sets and applies rules of circulation for millions of people governs, and whoever governs owes reasons to the governed.

The distinction between denying a claim and denying someone's standing as a participant helps in this examination. Contesting a proposal is not the same as asserting that its author has ceased to be someone to whom reasons are owed. Yet the first operation may also be unjustified: a claim may be rejected for false, irrelevant, or discriminatory reasons. And exclusion may occur without any express declaration of inferiority. The message about the polling address obstructs an exercise while formally preserving the recognition of the person who should exercise it.

For that reason, the denial of standing and the obstruction of exercise are indicators of problems that need to be examined. Neither, on its own, determines the appropriate measure. The assessment depends on what was denied, on the mode of the intervention, on the reasons given, and on the evidence available. A collective reason may reach each participant, but invoking a general benefit does not dispense with explaining the position concretely imposed on each of them.

The third question concerns response and competence. Recognizing a piece of speech as unjustified does not confer on just anyone the power to suppress it. Contesting, correcting, repairing, reducing reach, removing content, and punishing are different responses. Each intervenes in existing relations in its own way and requires a corresponding justification. The choice must consider the measure's suitability for addressing the problem, the available

alternatives, its extent and duration, and the conditions for challenge and review. A lawyer will recognize in these requirements the proportionality test — suitability, necessity, and balancing in the strict sense (Alexy, 2008) — and the resemblance is real. The participation criterion does not propose a different test, and the review of restrictions on rights has long been a subject of constitutional theory (Alexy, 2008; Silva, 2002). What it offers is a common foundation for examining both the speech and the intervention: it identifies the addressees of both — the persons whose positions each alters — and requires, for both, the possibility of effective challenge in good time. Whether this reading makes visible something that an application confined to the content of speech would fail to examine has to be shown in a case. There is also a limit the book formulates that applies to the response more than to the speech: the response cannot destroy the very thing with respect to which it is owed (Dietrich, 2026a, ch. 18). A response that eliminates the practice in which reasons can be offered and answered destroys the condition it invoked to justify itself.

Whoever restricts expression also reorganizes a condition of participation. Blocking a channel may interrupt a threat and prevent other expression; removal may protect someone and eliminate material needed to criticize the decision itself. These effects do not make every intervention inadmissible. They make it insufficient to justify an intervention merely by identifying a defect in the prior speech.

The requirement of justification thus reaches both the speaker and the restrictor, without making their positions identical. An authority that can impose decisions has competences and means that an isolated participant lacks. A platform that controls access and circulation can create practical dependencies even when its rules announce safeguards. What matters is whether challenge works in good time and whether the decision-maker remains subject to effective control.

The questions apply at two levels, which should not be confused. At the level of the regime — the statutory, judicial, or contractual rules that organize expression and its moderation — one asks how those rules distribute the possibilities of speaking, receiving, challenging, and obtaining review. At the level of the particular intervention, one asks whether the application of those rules to the case is justified, competent, and open to challenge in good time. The distinction tracks the one Rawls drew between justifying a practice and justifying an action falling under it (Rawls, 1955, pp. 3–7), without reducing one level to the other: a rule may be unjust and correctly applied, and a justifiable rule may be wrongly applied. In moderation at scale, examining isolated decisions is not enough, because policies, automated systems, and the distribution of errors also reorganize participation

(Douek, 2022); nor can it be dispensed with, because it is in the particular decision that the error reaches someone. The same holds for the speaker: the legal permissibility of a piece of speech does not by itself make it justified before those it reaches, and the duty of justification does not turn into an obligation to account for every sentence before it circulates. It becomes exigible when speech reaches someone and alters, or gives sufficient reason to fear it will alter, a condition of participation.

The participation criterion organizes these requirements without replacing the arguments each step needs. Describing a transformation does not prove its injustice; injustice does not automatically determine a sanction; the legitimate aim of a response does not demonstrate the competence of whoever imposes it. Between speech and its restriction there are persons whose positions must remain identifiable, even when it is concluded that a claim of theirs must be rejected.

3. What reorganization does not authorize

Establishing that a piece of speech has altered a shared condition, and therefore owes an answer to those who live from it, is not enough to authorize a coercive intervention. The claim seems modest and is not: most restrictive arguments consist precisely in omitting it. It is worth going through, then, what falls short of the first two questions and what, even after crossing them, still indicates no measure at all.

Falling short, first, is opinion that asserts without imputing a crime, without calling to action, and without obstructing, even when it is false and even when it is hateful. A person who maintains that a certain policy is unjust, that a certain group is wrong, or that a certain institution does not deserve trust denies claims; by that act, he denies no one the standing of participant. The answer owed to such speech is contestation, and whoever contests it acknowledges, by the very act, that its author remains someone to whom reasons are owed. The falsity and offensiveness of a statement do not show, on their own, that it has obstructed anyone’s participation; they show that there is something to contest. At this point the criterion parts ways with Waldron, who argues for restricting published and lasting group defamation precisely because, through its permanence, it erodes the public assurance that serves here to describe the harm (Waldron, 2012). The criterion accepts the description and refuses the inference: the erosion of assurance is an alteration to be diagnosed, and only a showing that it has occurred, in a determinate form and to a determinate extent, authorizes moving on to the question of the measure. Kaufmann states the same point as a limit of tolerance: tolerance is

not unlimited, but it does not follow that intolerance must always be met with intolerance, and when intolerant speech does not endanger guaranteed liberty, there is no ground for denying it liberty (Kaufmann, 2004, pp. 494–495). If the offence is repeated, directed at a specific person, and organized so as to drive that person out of a space, what is at stake is no longer offence but obstruction, and the diagnosis of the previous section applies.

Falling short, second, is myth and the political label as a label. Calling an opponent a mass murderer, a thief, or a communist replaces examination with belonging, and the square in which examination could take place grows poorer. What the second study established holds here: the degradation of debate by myth calls for examination, and examination is the response that does not reproduce the myth with its sign reversed (Dietrich, 2026c). But the use of a label is not, by itself, a diagnosis. The same word may function as political criticism, as a determinate imputation of a crime, as an instrument of repeated harassment, or as part of a campaign of exclusion, and what it did to someone's participation depends on its use, its addressees, and its context. Being formally able to reply does not show that the exercise of participation remained unobstructed — the voter in the first example could also reply, and did not for that reason reach the right polling station. What can be asserted is narrower and safer: falsity, offensiveness, or the use of a label do not, on their own, show obstruction of participation or authorize restriction; what they show or fail to show is a matter for diagnosis, case by case.

Falling short, third, is ambiguity. A sentence that admits an offensive and a critical reading does not, merely by admitting the first, reorganize any condition; the first question requires that the act have done something, or given sufficient reason to fear it would, and the possibility of being read in a certain way is neither. The German Federal Constitutional Court did not decide that soldiers are not murderers; it decided that the most offensive interpretation of an ambiguous sentence does not authorize the state to choose it against the speaker (BVerfGE 93, 266). The rule has a corollary that the criterion accepts: where diagnosis cannot distinguish degradation from destruction, the response owed is the least intrusive one, because doubt about what the speech did cannot be resolved against the speaker. Hegel already saw the difficulty from both sides. He refused to define freedom of the press as the freedom to say and write whatever one pleases, a definition he considered as superficial as that of freedom as doing whatever one pleases; and he acknowledged that the indeterminacy of expression prevents laws from attaining the precision required of law, so that judgment about it tends to appear as a subjective decision (Hegel, 1997, § 319, pp. 290–291). The first side refuses immunity; the second has been, ever since, the problem of whoever restricts.

Falling short, finally, is what the first study described: the mechanism by which voters judge without examination and believe they possess the competence to do so does not show, on its own, that anyone has injured them. The mechanism describes a mode of reception, not a position reorganized against them, and the answer to a mode of reception is the offer of examination, not the containment of whoever prompted it.

Then there is what crosses the first two questions and still does not indicate the measure. Speech that verbalizes the exclusion of a group — that says, in so many words, that its members are owed nothing — demonstrates unjustifiability. It does not show whether the answer owed is public contestation, correction, reparation, or containment. That depends on what the third question examines: the context in which the speech was uttered, the reach it had, the risk it created, and the alternatives to the measure proposed. Mill had already observed it: the opinion that corn-dealers starve the poor may circulate freely through the press and justly incur punishment when delivered to an excited mob assembled before a corn-dealer's house (Mill, 2006, ch. III). The content is the same; the act is not.

Nor do the limits that the third study set for the intervention it proposed indicate the measure on their own. Verified falsity, a declared crisis, and the set of factors characterizing dangerous speech — speaker, audience, message, context, and medium (Benesch, 2013) — delimited a specific measure, the temporary reduction of the reach of content already debunked during the days of Southport. As conditions of that measure, they are reasonable. As requirements of every legitimate response to expression, they do not hold. Connolly's incitement used true information about the existence of hotels housing asylum seekers (Connolly v R, 2025); a threat directed at a specific person needs no crisis to alter immediately the conduct of whoever receives it; and requiring a declared crisis hands to whoever declares it the power to define when the protection of speech diminishes — a power that the criterion, applied to the restrictor, cannot grant without examination. The three limits remain a case, not a rule.

Nor, lastly, does the description that makes a measure seem light indicate the measure. Saying that reducing the reach of a post is not censorship, because it does not remove the post and merely declines to reward it, describes the mechanism and does not justify it. Reach reduction is an intervention in the condition of participation of whoever posted and of whoever would have read; its legitimacy depends on who decides it, on the effects it produces, and on the possibilities of challenging it, and those possibilities are in practice smaller than the ones that accompany removal, because reduction is less visible. An invisible intervention is not a lesser intervention; it is an intervention with less control.

What follows from diagnosis and justification, then, is that an answer is owed. It does not follow which answer, nor by whom.

4. Response, competence, and control

The responses to speech that has unjustifiably altered a shared condition are not equivalent to one another, and the difference between them is not merely one of intensity. Contesting publicly, correcting, repairing, reducing reach, removing, and punishing intervene in existing relations in different ways, require different titles of whoever adopts them, and admit different controls. What can be said of all of them is that each must be compared concretely with the alternatives — by its extent, its duration, its reversibility, the urgency that justifies it, its effects on third parties, and the possibility of reviewing it — and that whoever adopts one must explain why an alternative less burdensome for the speaker's participation, and sufficiently effective, would not have sufficed. It cannot be said that they always occupy the same order. Substantial damages may be more burdensome than the targeted removal of a threat; adding context to a post does not weigh the same as reducing its reach; an urgent threat may require removal without first passing through public correction. Preference for the least burdensome intervention is a requirement; a fixed sequence of steps is not.

It is still worth naming the responses, because each has a nature of its own. Public contestation is the response any participant can give and requires no title at all; a right of reply, when institutionally required, already imposes an obligation on a third party — the outlet, the platform — and is therefore not the same as contesting freely. Timely public correction by whoever holds the information is the response only a determinate authority can give: when a false attribution circulates and whoever could refute it stays silent, silence is not neutrality. Shklar called passive injustice the failure of those who could and should prevent a wrong to do so, and recalled that the public authority that does nothing is among the gravest cases of that failure (Shklar, 1990, pp. 40–46); the British inspectorate's review of Southport identified the information vacuum as a failure, which does not dispense with distinguishing what the law allowed to be said (HMICFRS, 2025). Civil reparation answers a past act without containing a future one. Reach reduction, labelling, and added context keep the speech available and alter its circulation, each in its own way. Removal withdraws the post; criminal sanction reaches the speaker. These differences in nature are what the concrete comparison must take into account, and they are also what makes each response a distinct

reorganization of the conditions of whoever speaks, whoever reads, and whoever will need to examine the decision.

The requirements that accompany any of these responses can be stated as six. The first is a demonstrated diagnosis: the expressive conduct identified, the condition identified, and sufficient evidence of the harm or risk the measure is meant to address, according to the nature of the intervention. Three things must be distinguished here, because they are often confused: proof that the speech occurred and of what it said; proof of the harm done or of the relevant risk it created; and the ground and requirements of the chosen measure. A threat may require interruption before the announced aggression, and requiring the final harm to have occurred as a condition of every measure would amount to forbidding prevention; what is required is that the risk be demonstrable, and demonstrated, to the extent of the intervention. Circulation, on its own, demonstrates none of these elements. When the requirement fails, the intervention responds to a class of content, not to an act — and responding to classes is the surest way of reaching those who did nothing. The second is a stated justification: the denial of standing or obstruction of exercise named, and the reason for the measure told to whoever bears it. A measure whose reason cannot be told to the person affected is a preference with force.

The third requirement is competence. Whoever decides must have a title to reorganize the speaker’s condition of participation, and the question is not whether there is a title but which one, within what limits, and under what controls. A court has it by mandate; an administrative authority, by the norm that confers it; a platform, by a more varied set of sources — the law that imposes duties on it, but also its contractual terms of service, the organization of a thematic forum, the fight against what degrades the service — and the relevant question is what powers those rules legitimate and to what scrutiny they themselves are subject. The European Digital Services Act allows decisions grounded both on the illegality of content and on its incompatibility with the terms of service, and subjects both to duties to state reasons (European Union, 2022, art. 17). That does not settle the normative question; it shows that excluding every platform policy as a ground would be a strong commitment, and that the problem lies in reasons and control, not in the origin of the rule. The requirement seems obvious and is the one most often omitted, because the gravity of a harm tends to be taken as a sufficient title for anyone to respond to it. It is not. The study on Coase showed, in another domain, that an administrator can have formal duties and retain unilateral practical control over the reorganization it administers (Dietrich, 2026e); the observation holds for whoever removes, blocks, or reduces reach. The fourth requirement is

that of safeguards and control: an adversarial hearing, appeal, time limits, publicity of the decision and of the standard it applied, and the possibility of challenging reach reduction as much as removal. Solum argued, with respect to adjudication, that the legitimacy of a binding decision depends on those affected being able to participate at a stage when participation still matters, and that this value is not reducible to the effect of participation on the outcome (Solum, 2004, p. 275). Solum restricts the thesis to final and binding decisions and expressly excludes temporary ones. Extending it to moderation is therefore an argument of the author's own, not Solum's: definitive removal comes close to the case he examines, and even a temporary measure can have a definitive effect when it strikes participation in a time-bound debate — the opportunity lost on the eve of an election is not recovered after it. Without control, the response becomes an exception, and the exception is the sign that the response has entered the cycle it set out to interrupt.

One consequence of this openness needs to be stated. A forum devoted to literature may remove electoral propaganda that is in no way unjust, simply because it is foreign to the subject that organizes that space. The removal is justified not by any injustice in the speech, of which there was none, but by the declared purpose of the space, which participants knew when they joined and which they also live from. The case separates two questions the argument has so far tended to bring together: why a piece of speech is unjustified, and why its exclusion from a given space is justified. The second can be answered without the first when the rule of the space is public, prior to the case, and uniformly applied, and when the speech retains other places in which to circulate. The answer changes as the space ceases to be one among many and becomes the one through which most circulation passes: the thematic purpose of a forum does not carry over to a platform that organizes a country's public debate, and the more participation depends on a space, the more the rule that organizes it comes to resemble a rule of the regime itself and the more justification it requires. The distinction tracks Post's between public discourse, whose protection serves to let people experience governing themselves, and expression circulating outside it, which is regulated for other purposes (Post, 2011, pp. 483–486). Post formulates the distinction for constitutional protection against the state; extending it to the duties of platforms is an argument of the author's own, resting on the criterion rather than on Post's classification: the thematic forum regulates expression according to its purpose, and the platform that has become the site of public debate, because it reorganizes the condition on which that debate depends, can no longer do so without the reasons required of whoever regulates it.

The fifth requirement is that of conditions for revision. A measure that leaves the speaker no possibility of revising what he said without confessing total guilt closes a path the response had reasons to keep open; the psychology of dissonance suggests that a penalty that closes it may reinforce the self-image it meant to correct (Tavris & Aronson, 2007), but the reaction to a measure may reinforce convictions, produce revision, or merely interrupt the conduct, and the legitimacy of a response does not depend on its succeeding in changing the agent's understanding. What is required is more modest: that the measure preserve conditions for revision, not that it guarantee them. The sixth is the distribution of responsibility. Reach and position are relevant factors — whoever forms the understanding of many answers for something that whoever merely repeats did not do, although whoever repeats, to the extent that he repeats, cannot call himself a stranger to what he repeats (Dietrich, 2026a, ch. 22) — but they do not replace knowledge, intention, actual contribution, and the possibilities of acting otherwise, which the third study's criterion of responsibility already brought together: effective access to the information that would refute the speech, capacity to examine it, reasons to doubt, and opportunities for examination (Dietrich, 2026d). A response that strikes the many and spares the originator has inverted the distribution; one that strikes the originator without examining what he knew and could do has inverted it in another way.

Two documented cases show the requirements at work, with uneven and partly inconclusive results; a third is only delimited. In Connolly's case, the condition was the public assurance of safety of the residents of hotels housing asylum seekers; the conduct, an unequivocal incitement to set them on fire, posted on the evening of 29 July 2024, hours after the attack and before the disorder that would begin the following day; the risk has to be assessed by the circumstances of that moment — the shock at the attack and the circulation of falsehoods about its perpetrator — without transferring later violence onto the audience; the measure, criminal sanction. The competence was that of an ordinary court, and control existed: the sentencing judge treated the timing of the post, on an evening of particular sensitivity, as an aggravating factor, and the Court of Appeal examined the sentence of thirty-one months, gave reasons for its adequacy, and refused leave to appeal against it, so that the sentence remained unchanged (Connolly v R, 2025). The case thus illustrates judicial reasons and appellate control — the third and fourth requirements. It does not, on its own, illustrate the others: the judgment does not compare imprisonment with less burdensome alternatives in the sense required here, and says nothing about the conditions for revision left to the convicted woman. Examined requirement by requirement, the case is conclusive as to two and inconclusive as to the rest, and that is how a case can test a criterion. The case of the false

name attributed to the Southport attacker is more complex, and for that reason more useful. The condition was the same; the conduct, a false attribution of a crime to a category of people; the false name was viewed some four hundred and twenty thousand times, a figure that measures potential reach, not persons convinced (House of Commons, 2025a, para. 8). The response of those who held the information was limited in different ways. The name of the accused, who was seventeen, was protected by legal restrictions — the rules on contempt of court and those that protect the identity of minors accused of crimes; other corrective information could be released and was, such as his place of birth and the denial of the false name; on still other information, such as his religion, there was divergent advice from different levels of the Crown Prosecution Service; and the Chief Constable of Merseyside could not tell the parliamentary committee whether releasing his religion would have prevented further disorder. The committee recorded that these restrictions created a vacuum in which disinformation grew (House of Commons, 2025b). The case is therefore neither one of simple silence nor one of imposed silence. It is a conflict between the public assurance of the threatened community and the procedural protection of the accused, both of them participants in the condition at stake, and the criterion requires each category of information to be examined before both. The passive injustice Shklar speaks of reaches only the part of the omission not covered by a competing duty; Shklar herself measures the responsibility of those with competing duties in the light of those duties (Shklar, 1990, pp. 44–45). What the case allows us to say is narrow. Where correction was legally possible and came late, there was an answer owed that did not arrive in time. Where the law imposed restraint, the question shifts to the regime — the rules on anonymity of minors — and to how to preserve them without leaving the vacuum to be filled by falsehood. On neither level does the case show that the post, or the lack of correction, caused the attacks. In the case of the calls to action that preceded 8 January 2023 in Brasília, the condition was institutional; the conduct, calls to prevent the functioning of the institutions; the response, holding those who issued the calls responsible, in proceedings before the Supreme Federal Court. The case, however, needs a precision that the formula conceals. By January 2023 the electoral procedure was complete; the condition at stake was respect for the proclaimed result and the functioning of the institutions that follow from it, and the calls that preceded the storming of the seats of the three branches of government were directed against that functioning. Assessing the response by the six requirements would call for the examination of identified decisions, one by one, and the existence of proceedings and of published decisions does not show, by itself, that the

requirements were met. The case is recorded as the one in which the criterion would have to be applied with the greatest caution, because in it symmetry is hardest to sustain.

The six requirements can be put as five questions to any intervention. What condition was altered, by what conduct, with what evidence of harm or risk? Does the alteration deny standing or obstruct exercise, and why? What response, and why would a less burdensome and sufficiently effective alternative not have sufficed? Does whoever adopts the measure have a title to reorganize the speaker's condition, and which one? What safeguards and control accompany the measure, including when it is a reduction of reach? Put to the regime, the same questions fall on the rules: what conditions they allow to be altered, what reasons support them, what measures they authorize, to whom they give title, and what review they guarantee — including over the distribution of errors that no rule applied at scale can avoid. The first four questions were present, in scattered form, in the three earlier studies. The fifth was in none of them.

5. Intervention as reorganization

When a profile is blocked, a post removed, or its reach reduced, something has been reorganized. Not only the speaker's situation: also that of those who read him, of those who would have answered him, of those who would need the removed material to assess whether the removal was fair. Restricting expression is an action upon a shared condition — the square in which reasons can be offered and answered — and, like every action of that kind, it has within it, by virtue of the position they occupy, the persons who live from that condition. The criterion that was put to the speaker is therefore put to whoever restricts him. Not because the positions are identical, but because the field is the same.

The difference between the positions is real and must be stated. An authority that can impose decisions has competences and means that an isolated participant lacks; a platform that controls access and circulation can create practical dependencies that no speaker creates. That difference does not lessen the requirement of justification; it increases it. Whoever can do more reorganizes more, and the more one reorganizes, the more one is asked. The book from which the criterion derives distinguishes three ways of being within a single reorganization — that of whoever acts, that of whoever holds the capacity to reorganize and sustains it, and that of whoever repeats it — and observes that the answer of whoever holds that capacity cannot consist in good intentions (Dietrich, 2026a, ch. 22). The restrictor occupies the second. Applied to the restrictor, the criterion's three questions take the

following form. What shared condition does the intervention reorganize, and with what evidence that the reorganization was necessary? Before whom must the intervention be justified — and the answer includes the speaker, who does not cease to be a participant for having spoken badly? Who controls the decision of whoever intervened, within what time, and with what publicity?

The last question is decisive, and it is the one the study on Coase prepared. There it was shown that the existence of formal duties — to inform, to consult, to give account — does not prevent whoever administers a reorganization from retaining practical and unilateral control over it, if those duties do not give whoever invokes them a means of altering the decision in good time (Dietrich, 2026e). The same holds here. A judicial ruling may impose on platforms duties of transparency, service channels, and periodic reports; the criterion's question is not whether those duties exist, but whether they give whoever suffers a removal or a reach reduction a means of challenging it that works before the harm to his participation is complete, or whether they merely document, afterwards, the control of whoever adopted it. An annual report on removals is not a remedy; it is a statistic.

The distinction between justifying a practice and justifying a particular action falling under it helps to sharpen the point. Rawls formulated it with respect to punishment: to those who objected that a certain doctrine would authorize convicting an innocent person to deter others, he replied that one had to say who the "we" is that would decide to do so, with what authority and under what rules, because that authority would have to be written into the institution; and an institution that granted its officials the power to decide, case by case and according to their judgment of the common good, who would be convicted could hardly be justified (Rawls, 1955, pp. 10–13). The question holds for the restriction of expression. It is not enough to show that, in this case, it would be good for someone to remove this speech. One must describe the practice that authorizes someone to remove it — who, under what rule, with what control — and ask whether that practice is justified before those who participate in it, including the speakers whose speech it will reach by mistake.

Reach reduction is the most acute case. It is not invisible by nature: it can be communicated to whoever suffers it, accompanied by reasons, and subjected to challenge, and the European Digital Services Act requires a statement of reasons for certain restrictions of visibility, demotion included (European Union, 2022, art. 17). The problem lies in opaque demotion — demotion that is not notified, does not disclose the standard applied, and offers no means of challenge. In that form, the measure that seems least is the one that admits least control. Whoever describes it as a mere refusal to reward describes its mechanics; the

question remains who decides what to reward, before whom the decision is justified, and who reviews it.

The symmetry test to which the criterion submits has, then, two axes. On the first, it must produce the same result when the speaker belongs to the political camp with which whoever applies the criterion identifies. Two speech acts with the same structure — a false attribution of a crime to a category of people, a call to prevent a procedure — receive the same five questions, and the answers do not change with the colour of the speaker. On the second axis, which the three earlier studies did not travel, the criterion must reach whoever restricts: the state that blocks a profile, the platform that reduces reach, the court that imposes a measure. If the questions held for the call to action and not for the exception that answers it, the criterion would cease to examine the common field and would come to serve one of the sides.

There is an objection to this point that deserves to be met on its own terms. It says that submitting the restrictor to the same examination as the speaker equates the arsonist and the firefighter, and that symmetry is a form of paralysis. The reply is that the symmetry lies not in the measures but in the questions. The arsonist and the firefighter do not receive the same answer; they receive the same requirement to justify what they did to the house, and the firefighter, who entered with a title the arsonist lacked, has more to say, not less. A restrictor who can answer the five questions is not paralysed by them; he is legitimated. What symmetry paralyses is the restrictor who cannot answer them, and that is exactly the one who should be paralysed.

A hypothetical case shows what this reading adds. On the eve of an election, a message circulates in a city’s groups announcing that certain polling stations have been moved to a distant address. An analysis confined to content asks whether the message is false and, if so, whether it may be removed. The participation reading asks other questions, and in the order in which they matter. Who is at stake: the voters of those stations, whose standing as voters remains intact and whose exercise may be obstructed; but also the voters of any station that has in fact been moved, if there is one, for whom a true notice of similar content is indispensable. Which response meets the risk within the time in which it exists: correction by the Electoral Justice — the branch of the Brazilian judiciary that runs elections and holds the information on polling places — published through its own channels before the polls open, reaches the addressees without depending on removal. If removal is adopted, who decides it and with what control: a platform that removes en masse, by textual similarity, messages about changes of polling station may also remove the true notice and obstruct, by excess, the

very exercise it meant to protect; and a decision that can only be challenged after the election cannot be challenged in good time. What removal must preserve: the evidence of the conduct, for the eventual liability of whoever hindered the exercise of the vote (Electoral Code, art. 297). None of these questions is foreign to proportionality review. What the participation reading makes visible is that the addressee of the speech and the addressee of the measure must be identified before the measure is chosen, and that the same error — obstructing an exercise — can be committed by the speaker and by whoever restricts him.

The case changes when the conditions change, and it is in the variations that the criterion does its work. If official correction does not reach the addressees — because the message circulates in closed groups, because time is short, because the electoral authority does not have their trust — the least burdensome alternative ceases to be sufficient, and removal or limits on forwarding may become the answer owed, without the questions of title and review disappearing. If communication is open, correction alongside the message may suffice; if it is encrypted, not even the platform sees the content — and the Brazilian regime excludes interpersonal communications on messaging services from the duties of care and notice (Decree 8,771/2016, art. 16-O) — and the possible response shifts to whoever can speak to the addressees. If the risk is remote — the message circulates weeks earlier, when there is time to refute it — the urgency that would justify removal is lacking. And if the notice was true, removal is itself the obstruction, and the criterion turns entirely against whoever restricted. In none of the variations does the criterion decide on its own; in all of them, it indicates who must be taken into account and what must be shown before the measure.

There follows, finally, a limit that applies to the restrictor with greater force than to the speaker: the response cannot destroy the very thing with respect to which it is owed (Dietrich, 2026a, ch. 18). The limit needs to be read carefully, because the image of closing the square conflates three things of different scope. Removing a post is not excluding a person from a space, and neither is eliminating the possibility of challenging the intervention itself. A post can be removed with the evidence preserved, a reasoned decision, and an effective appeal, and none of that destroys the square. Excluding a person alters his participation more extensively and durably, and requires more. The intervention that suppresses the means of challenging it — because it is not notified, because it eliminates the material on which criticism would depend, because the appeal arrives after the harm to participation is complete — is the one that strikes the condition it invoked to justify itself. It is in this third sense that the response closes the square. That is why the fifth question — who controls, within what time, with what

publicity — is not a procedural add-on; it is the form the criterion takes when it turns to whoever applies it.

6. Brazilian law under the criterion's questions

Brazilian law distinguishes, without naming it, what the criterion distinguishes. It protects the expression of thought and of ideas, free from censorship or licence (Federal Constitution, art. 5, IV and IX; art. 220); it provides for reply and reparation before any containment (art. 5, V and X); and it responds to acts — incitement, discrimination, the violent abolition of the democratic state — not to opinions. The examination that follows puts the criterion's questions to each source, records what the normative text already answers, and delimits what remains open.

The first source is the Constitution itself, and the first question is whether it already contains the criterion. The prohibition of anonymity (art. 5, IV) is usually read as a condition of subsequent liability (art. 5, V and X); it can also be read — and this is the reading proposed here — as a recognition that expression has addressees before whom its author answers. The right of reply (art. 5, V) and the inviolability of honour (art. 5, X) fix the first-degree response — speech against speech — before any other. And the decision of the Supreme Federal Court (STF) that set aside the Press Law established the rule that freedom of expression takes precedence, with subsequent liability rather than prior restraint (STF, ADPF 130, 2009). The question to put to constitutional doctrine is whether this combination — addressees before whom one answers, reply before containment, liability afterwards and not before — amounts to the comparison between responses required above, or whether the precedence of freedom is read as a dispensation from justification. The answer determines whether the criterion adds something to the Constitution or merely reads it.

The second source is the criminal offences, and here the question is twofold. Crimes against honour (Criminal Code, arts. 138 to 140) respond, as a rule, to speech that strikes at someone's claim to the regard of others, and prosecution depends, in most cases, on the initiative of the person offended. The classification, however, is not uniform: aggravated insult under art. 140, § 3, of the Criminal Code and racial insult, now an offence under Law 7,716/1989 (art. 2-A, added by Law 14,532/2023) and prosecuted publicly, strike precisely at standing. The Supreme Federal Court maintained, for crimes against honour, the requirement of a court order as a condition of platform liability, and one may ask whether that regime also reaches the offences that deny standing. Public incitement to crime (Criminal Code, art. 286)

and incitement to discrimination (Law 7,716/1989, art. 20) respond to acts of mobilization and of denial of standing, and the decision that upheld the conviction of a publisher of antisemitic works held that freedom of expression does not extend to incitement to racism, in terms that come close to the public assurance that a group will not be attacked for what it is (STF, HC 82,424, 2003; Waldron, 2012). The boundary between advocating the change of a norm and inciting crime was drawn by the Court in the "cannabis march" case (STF, ADPF 187, 2011), and it is there, more than in abstract formulations, that the distinction between denying a claim and mobilizing others is tested in Brazilian law. The criterion's question to these offences is whether their application online requires sufficient evidence of the conduct and of the harm or risk the offence presupposes, and not merely of circulation, or whether the gravity of the offence has been taken as a dispensation from diagnosis. The question does not assume that the harmful outcome must occur: incitement offences exist precisely to reach risk, and what the criterion asks is that the risk be demonstrated to the extent of the response.

The third source is the protection of the electoral procedure and of the democratic order. Law 14,197/2021 created the offences of violent abolition of the democratic rule of law and of coup d'état, both requiring violence or serious threat, and that of interrupting the electoral process by unduly breaching the security mechanisms of the electronic voting system (Criminal Code, arts. 359-L to 359-N). None of these offences describes the call to action as such. The same law, however, added to art. 286 of the Criminal Code a sole paragraph punishing public incitement of animosity between the Armed Forces, or of the Armed Forces against the constitutional powers, civil institutions, or society; by that route, by incitement to crime in general (head of the article), or by participation in the acts, speech that mobilizes enters the criminal domain, and the distinction matters, because the response to a call depends on what it called for and how. The same law contained an offence aimed at whoever promoted or financed a campaign to spread facts known to be untrue and capable of compromising the integrity of the electoral process (art. 359-O); the provision was vetoed. The absence of that offence does not leave the voter in the first example unprotected: the Electoral Code punishes whoever prevents or hinders the exercise of the vote (art. 297), and also punishes the dissemination, in campaign propaganda or during the campaign, of facts known to be untrue about parties or candidates (art. 323), an offence which, given its object, probably does not reach the message about the polling place. The Supreme Federal Court has, moreover, examined a power of rapid removal of disinformation about the electoral process conferred on the Electoral Justice (Superior Electoral Court Resolution 23,714/2022) and upheld it, first by endorsing the refusal of interim relief on 26 October 2022, and later by

dismissing the action on the merits (STF, ADI 7261); it is in that judgment that the questions of title, time, and control were actually confronted in Brazilian law, and the criterion must be measured against it. The criterion's question thus becomes narrower: whether hindrance by deception — the false message about the polling place — falls within the reach of that offence and of the instruments of the Electoral Justice, and with what requirements of diagnosis, or whether mass obstruction of exercise was left without a proportionate response after the veto.

The fourth source is the platform liability regime, and it is the one that most requires reading. Article 19 of the Marco Civil da Internet — Brazil's 2014 statute on internet rights and duties — made the provider's civil liability conditional on non-compliance with a specific court order. In June 2025, the Supreme Federal Court declared the provision partially and progressively unconstitutional and established a tiered regime (STF, RE 1,037,396 and RE 1,057,258, Themes 987 and 533); on 17 June 2026, ruling on motions for clarification, it consolidated the final thesis, with immediate finality, effects running from 5 August 2025, and a sixty-day period for structural obligations (STF, 2026). The final thesis distinguishes regimes: for crimes and unlawful acts in general, extrajudicial notice suffices, and the liability of a provider that does not remove is joint and several, unless it shows reasonable doubt as to unlawfulness after diligent and qualified analysis; for crimes against honour, a court order remains the condition of liability, with the express proviso that removal upon extrajudicial notice remains possible; for a closed list of serious offences, a duty of care with active monitoring is imposed, whose systemic failure gives rise to liability; for advertisements, paid boosting, and artificial distribution networks, fault is presumed, rebuttably, by diligence within a reasonable time. Alongside the thesis, two decrees of 20 May 2026, in force since 20 July, make up the normative framework, with distinct objects: Decree 12,975 amends Decree 8,771/2016, which regulates the Marco Civil, and Decree 12,976 deals with the protection of women in the digital environment. The first bears directly on the questions about the restrictor: it entrusts the National Data Protection Agency with supervising providers' duties (Decree 8,771/2016, art. 19-A, as amended by Decree 12,975/2026) and reserves to the Attorney General's Office of the Union a narrow power of notice, limited to misleading, abusive, or fraudulent advertising related to public policies (art. 16-N, § 1). Its validity is being challenged before the Supreme Federal Court (ADI 7981).

Read in the text of the decree that implements it, this regime gives partial answers to the criterion's questions, and it is worth recording them before pointing out what is missing. As to diagnosis, the duty of care does not respond to isolated posts: the provider is liable for systemic failure to take down a list of criminal conducts — terrorism, inducement to suicide,

incitement to discrimination, crimes against women, sexual crimes against vulnerable persons, human trafficking, and the offences of incitement and of attack on the democratic state (Criminal Code, art. 286, sole paragraph, and arts. 359-L, 359-M, 359-N, 359-P, and 359-R); the isolated presence of unlawful content does not constitute the failure; and the authority may neither hold the provider liable solely for keeping or removing an isolated piece of content nor notify it to moderate isolated content (Decree 8,771/2016, arts. 16-B, § 3, and 16-I, as amended by Decree 12,975/2026). The regime thus separates the two levels distinguished above: the system is examined by the authority; the particular decision, by the provider, upon notice. As to justification, removal upon notice must be communicated to the user with the specific ground and the means of challenging it, a decision to keep the content must be communicated to the notifier on the same terms, and reconsideration after challenge must be reasoned to both parties (art. 16-E); terms of use must provide for due process and annual transparency reports (art. 20-A). On this point the regime meets the requirement of reasons addressed to those whom the measure reaches. As to competence, the provider may keep the content when, after diligent and reasoned analysis, it concludes that there is reasonable doubt as to its criminal character, weighing the doubt against the gravity of the crime, and it must take into account the context and the informative, critical, satirical, or parodic purpose of the post (art. 16-G, §§ 1 and 2). The degree of scrutiny grows with what is at stake, as in the epistemic responsibility described by Amaya (2011, pp. 316–318); and, as in her account, the diligence shown in the process serves to assess the decision-maker — the provider — while the correctness of the decision to keep or remove remains a distinct question (Amaya, 2011, p. 318, n. 11). The provider must also curb abusive use of notices against freedom of expression (art. 16-F), a duty addressed to the restrictor itself.

What remains open is narrower than the initial question. The decree does not set time limits for challenges; it authorizes the competent authority to regulate them (art. 16-D, sole paragraph), and challenge in good time depends, among other things, on such limits existing and being short. The safeguards of art. 16-E are triggered by notice; the decree does not say whether they reach the moderation the provider exercises under its terms of use, nor does it deal with reach reduction, which remains the least visible and least controlled measure. Administrative control of diligence is systemic, exercised by the data protection authority (art. 19-A). Control of the particular decision is judicial and is provided for in the thesis itself: whoever has had content removed may seek its restoration by showing the absence of unlawfulness, without the provider being ordered to pay damages, and the provider or the author of the post may seek interim relief to prevent removal (STF, 2026, items 5.5 and 5.6).

The question ceases to be whether individual control exists and becomes whether it is sufficient, accessible, and timely. Restoration without damages also raises a problem that bears directly on the criterion: restoring the post does not restore the opportunity for participation lost while it was down, and a message returned after the debate for which it was intended returns to a field that has already reorganized without it. Finally, advertisements, paid boosting, and artificial networks are subject to a presumption of liability rebuttable by diligent action (art. 16-L), which concentrates responsibility on whoever amplifies; but the joint and several liability established by the thesis, which Justice Mendonça's dissent identified as a source of chilling effect on users' expression (Migalhas, 2026), reorganizes the condition of every speaker. Whether the safeguards of items 5.5 and 5.6 suffice to justify that reorganization before speakers, or whether the absence of reparation for lost participation leaves part of it unjustified, is the question the thesis leaves to the criterion; answering it would require reconstructing the grounds of the individual opinions, which is not undertaken here.

7. Three regimes under the same criterion

The Brazilian regime is not the only one to have redesigned platform liability in the last decade. The comparison with two others — the European Union's Digital Services Act and the United Kingdom's Online Safety Act — serves a determinate purpose: to check whether the criterion's questions distinguish between regimes, and whether the gaps found in the Brazilian one are unavoidable or merely unresolved.

On diagnosis, the three converge in separating the system from the particular decision. The European regulation combines a notice-and-action mechanism, applicable to each item of content (art. 16), with the duty of very large online platforms to assess and mitigate systemic risks, including negative effects on freedom of expression, civic discourse, and electoral processes (arts. 34 and 35) (European Union, 2022). The Online Safety Act imposes risk-assessment duties and duties of care concerning illegal content (ss. 9 and 10), supervised by Ofcom, the communications regulator, through codes of practice (United Kingdom, 2023). The Brazilian decree holds the provider liable for systemic failure and forbids penalties for isolated decisions. In all three, the regime is examined as a regime, and the particular decision as an application.

On justification addressed to the person affected, the difference is clear. The European regulation requires a clear and specific statement of reasons to every recipient affected by a

restriction — removal, disabling of access, demotion and other restrictions of visibility, suspension of payments, of the service, or of the account — both when the decision rests on the illegality of the content and when it rests on its incompatibility with the terms and conditions (art. 17). It further requires platforms to apply their terms and conditions in a diligent, objective, and proportionate manner, with due regard to fundamental rights, freedom of expression among them (art. 14(4)). The Brazilian regime requires reasons for removal upon notice and is silent on reach reduction and on moderation under terms of use. What appears in Brazil as an open point has, in the European Union, an express rule.

On control, the difference lies in the interval between the provider and the judge. The European regulation imposes an internal complaint-handling system, free of charge and available for at least six months after the decision (art. 20), and opens the route of certified out-of-court dispute settlement bodies, without prejudice to recourse to the courts (art. 21); both obligations do not apply, subject to the provisos laid down there, to micro and small enterprises (art. 19). The time limit of art. 21 shows, however, that the out-of-court route does not by itself resolve the loss of participation in a time-bound debate: a decision may take up to ninety days, extendable by a further ninety in highly complex disputes. The criterion thus reaches the European model as well. The Online Safety Act requires complaints procedures that cover, among others, users whose content has been taken down or restricted (s. 21). The Brazilian regime offers challenge before the provider itself and the judicial remedies of items 5.5 and 5.6 of the thesis, but no intermediate body. It is in that interval that challenge in good time is decided: the internal route is decided by whoever restricted, and the delay of any external route, judicial or out-of-court, counts against the participation it was meant to protect. Whether an intermediate body would be faster in Brazil is an empirical question that the normative comparison does not resolve.

The Online Safety Act adds a feature that the other two do not have in the same form. Services classified in Category 1, a legal category defined by the regime itself, must operate systems that take into account, in moderation decisions, the importance of the free expression of content of democratic importance and of journalistic content (ss. 17 and 19), and all providers must have particular regard to freedom of expression when fulfilling their duties (s. 22). Read by approximation — an interpretive one, not a demonstrated influence — the provision treats the rule that organizes public discourse as more demanding than the one that organizes a thematic space, which comes close to Post’s distinction. The Brazilian decree approaches this when it requires consideration of the informative, critical, satirical, or parodic

purpose of a post (art. 16-G, § 2), but it does so as a criterion for examining content, not as a duty to protect a category.

The comparison does not show that one regime is just and the others are not; the criterion does not grade whole regimes with a single mark. It shows three more precise things. The criterion's questions distinguish the regimes: the three converge on diagnosis and diverge on justification and control. The open points of the Brazilian regime — reasons for reach reduction and for moderation under terms of use, and a review body independent of the provider — have normative counterparts in other legal orders; the mechanisms identified widen the institutional alternatives available for discussion, without showing that their absence in Brazil was a deliberate choice or that they would suffice there. And filling a gap in the text does not guarantee that it works: whether statements of reasons are read, whether internal complaints reverse decisions, whether out-of-court bodies decide in time are empirical questions that the normative comparison does not answer and that the criterion merely formulates.

Conclusion

The square stays open, and what keeps it open is not the absence of answers. Speech that reorganizes conditions others live from — the assurance that a group will not be attacked for what it is, the reliability of a procedure, access to the exercise of a position — owes an answer to those persons, by virtue of the position they occupy and not of the speaker's intention. But the answer owed does not, by itself, become an authorization to restrict. Between the shortcut that runs from harm to containment and the one that runs from protection to the dispensation from justification there is a longer path, which both avoid.

The path has three steps, and each calls for its own argument. Diagnosis identifies the conduct, the condition, and the persons, and distinguishes proof of what was said from proof of harm or risk. Justification says why the alteration was unjustified, by denial of standing or obstruction of exercise, without the use of a label, falsity, or offence settling the question on their own. The response is chosen by concrete comparison with the alternatives — extent, duration, reversibility, urgency, effects on third parties — and requires of whoever adopts it a title and a control. The three earlier studies supplied elements for the first two steps; the third called for the construction attempted here.

The intervention that restricts expression reorganizes the square as much as speech does — more, when whoever intervenes can impose decisions — and answers to the same

questions: what condition it altered, before whom it is justified, who controls it. The symmetry lies not in the measures, which are unequal because the positions are, but in the questions. Removing a post with the evidence preserved, the reasons stated, and an appeal available in good time does not close the square; suppressing the means of challenging the suppression itself does. Read through these questions, Brazilian law appears closer to the criterion than the debate around it suggests — it responds to acts, provides for reply before containment, requires speakers to be identifiable, punishes whoever hinders the vote — and, in the regulation of the Marco Civil, already requires reasons addressed to the affected user, means of challenge, and scrutiny proportionate to the gravity of the case. What remains open is more precise: time limits for challenges, which the decree does not set; the sufficiency of judicial remedies in the face of lost participation; safeguards for moderation under terms of use; control of reach reduction; and the reach of electoral protection against mass deception. The comparison with the European and British regimes shows that some of these points have already received a normative answer elsewhere, which widens the institutional alternatives under discussion, without showing that they would suffice in Brazil.

The result is argumentative and has three parts. The protection of expression does not dispense whoever intervenes, by speaking, in the conditions of others' participation from justification: in that sense, freedom of expression is not unrestricted, because it is not an immunity from the reasons owed to those whom speech reaches. That duty of justification does not, by itself, authorize any restriction: owing reasons to someone and being authorized to restrict someone are different relations. And restriction, when it comes, requires a sufficient material ground — either protection against unjustified harm or risk, or the justifiable organization of a space with a determinate purpose — comparison with the less burdensome alternatives, competence on the part of whoever imposes it, and effective control in good time; none of these conditions suffices without the others. Left aside, because they require another method and other evidence, are the demonstration that a given piece of speech caused a given harm and the decision-by-decision assessment of the institutional response to 8 January.

References

Alexy, R. (2008). *Teoria dos direitos fundamentais* [A theory of constitutional rights] (V. A. da Silva, Trans.). Malheiros. (Original work published 1985)

Amaya, A. (2011). Legal justification by optimal coherence. *Ratio Juris, 24*(3), 304–329.

Benesch, S. (2013). *Dangerous speech: A proposal to prevent group violence*. Dangerous Speech Project.

Brazil. (1940). Decree-Law No. 2,848 of 7 December 1940 (Criminal Code), arts. 138–140, 286, 359-L to 359-N, 359-P, and 359-R.

Brazil. (1965). Law No. 4,737 of 15 July 1965 (Electoral Code), arts. 297 and 323.

Brazil. (1988). Constitution of the Federative Republic of Brazil.

Brazil. (1989). Law No. 7,716 of 5 January 1989, arts. 2-A and 20.

Brazil. (2014). Law No. 12,965 of 23 April 2014 (Marco Civil da Internet), arts. 19 and 21.

Brazil. (2016). Decree No. 8,771 of 11 May 2016, as amended by Decree No. 12,975 of 20 May 2026.

Brazil. (2021). Law No. 14,197 of 1 September 2021.

Brazil. (2023). Law No. 14,532 of 11 January 2023.

Brazil. (2026a). Decree No. 12,975 of 20 May 2026. Amends Decree No. 8,771 of 11 May 2016, which regulates Law No. 12,965 of 23 April 2014.

Brazil. (2026b). Decree No. 12,976 of 20 May 2026. On the protection of women in the digital environment.

Brazil, Superior Electoral Court. (2022). Resolution No. 23,714 of 20 October 2022.

Connolly v R [2025] EWCA Crim 657 (Court of Appeal, Criminal Division, England and Wales).

Dietrich, P. (2026a). *O amanhã é possível?* [Is tomorrow possible?]. Autografia.

Dietrich, P. (2026b). Da visibilidade à competência presumida: desempenho digital e atribuição de aptidão política [From visibility to presumed competence: Digital performance and the attribution of political fitness]. *Revista Tópicos*. https://doi.org/10.70773/revistatopicos/792268195

Dietrich, P. (2026c). *Quando a praça vira trincheira: mitos políticos e o enfraquecimento do debate democrático* [When the square becomes a trench: Political myths and the weakening of democratic debate] [Unpublished manuscript].

Dietrich, P. (2026d). *Doing wrong in the name of good* [Preprint]. Zenodo. https://doi.org/10.5281/zenodo.23744884

Dietrich, P. (2026e). *O dano recíproco e a condição comum: Coase diante da pergunta sobre quem reorganiza* [Reciprocal harm and the common condition: Coase and the question of who reorganizes] [Unpublished manuscript].

Douek, E. (2022). Content moderation as systems thinking. *Harvard Law Review*, *136*(2), 526–607.

European Union. (2022). Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services (Digital Services Act), arts. 14, 16, 17, 19, 20, 21, 34, and 35.

Federal Constitutional Court of Germany. (1995). BVerfGE 93, 266 ("Soldaten sind Mörder"), decision of 10 October 1995.

Forst, R. (2012). *The right to justification: Elements of a constructivist theory of justice* (J. Flynn, Trans.). Columbia University Press.

Habermas, J. (1997). *Direito e democracia: entre facticidade e validade* [Between facts and norms] (F. B. Siebeneichler, Trans.; Vol. 1). Tempo Brasileiro. (Original work published 1992)

Hegel, G. W. F. (1997). *Princípios da filosofia do direito* [Elements of the philosophy of right] (O. Vitorino, Trans.). Martins Fontes. (Original work published 1820)

HMICFRS. (2025, May 7). *An inspection of the police response to the public disorder in July and August 2024: Tranche 2*. His Majesty's Inspectorate of Constabulary and Fire & Rescue Services.

House of Commons, Science, Innovation and Technology Committee. (2025a). *Social media, misinformation and harmful algorithms* (Second Report of Session 2024–25, HC 441).

House of Commons, Home Affairs Committee. (2025b). *Police response to the 2024 summer disorder* (Second Report of Session 2024–25, HC 381).

Kaufmann, A. (2004). *Filosofia do direito* [Philosophy of law] (A. U. Cortês, Trans.). Fundação Calouste Gulbenkian. (Original work published 1997)

Migalhas. (2026, June 11). Responsabilidade de redes: veja como votaram ministros do STF em embargos [Platform liability: How the STF justices voted on the motions for clarification]. *Migalhas*.

Mill, J. S. (2006). *Sobre a liberdade* [On liberty] (P. Madeira, Trans.). Edições 70. (Original work published 1859)

Post, R. (2011). Participatory democracy and free speech. *Virginia Law Review*, *97*(3), 477–489.

Rawls, J. (1955). Two concepts of rules. *The Philosophical Review*, *64*(1), 3–32.

Reilly, P. (2024). *Southport riots: Why social media's role in unrest is overblown*. The Conversation; University of Glasgow.

Shklar, J. N. (1990). *The faces of injustice*. Yale University Press.

Silva, V. A. da. (2002). O proporcional e o razoável [The proportional and the reasonable]. *Revista dos Tribunais, 798*, 23–50.

Solum, L. B. (2004). Procedural justice. *Southern California Law Review, 78*, 181–321.

STF (Supreme Federal Court of Brazil). (2003). HC 82,424 (Rapporteur for the judgment: Justice Maurício Corrêa). Judgment of 17 September 2003.

STF (Supreme Federal Court of Brazil). (2009). ADPF 130 (Rapporteur: Justice Ayres Britto). Judgment of 30 April 2009.

STF (Supreme Federal Court of Brazil). (2011). ADPF 187 (Rapporteur: Justice Celso de Mello). Judgment of 15 June 2011.

STF (Supreme Federal Court of Brazil). (2022). ADI 7261 (Rapporteur: Justice Edson Fachin). Endorsement of the interim decision, 26 October 2022.

STF (Supreme Federal Court of Brazil). (2026). RE 1,037,396 (Theme 987) and RE 1,057,258 (Theme 533) (Rapporteurs: Justices Dias Toffoli and Luiz Fux). Judgment of 26 June 2025; motions for clarification decided 17 June 2026. *Informativo STF*, 1222, 30 June 2026.

STF (Supreme Federal Court of Brazil). ADI 7981. Pending.

Tavris, C., & Aronson, E. (2007). *Mistakes were made (but not by me): Why we justify foolish beliefs, bad decisions, and hurtful acts*. Harcourt.

United Kingdom. (2023). Online Safety Act 2023, c. 50, ss. 9, 10, 17, 19, 21, and 22.

Waldron, J. (2012). *The harm in hate speech*. Harvard University Press.