

Moderation and complaints procedure

In short

- Anyone can report content. With an account via the menu next to the content, without an account via the [public reporting point](#).
- **Software runs alongside** that flags possible breaches and, in a small number of very clear-cut cases, temporarily hides a post. What that software does and does not do is set out below.
- **Removing content, issuing a warning or restricting an account is always done by a person.** Never a computer on its own.
- For every decision about your content or account you will be told what was decided, which rule it concerned, what we saw, and whether software was involved.
- **[Lodging a complaint]/(bezwaar) is possible for six months and costs nothing.** Someone other than the person who made the decision then looks at it again.
- If you cannot resolve it with us, there are three routes outside us: an independent dispute settlement body, the supervisory authority, and the courts.

Who moderates

In brief: a small team of people. Debaty is small, and that has consequences you should know about here.

Debaty is moderated by people who work at Debaty. There is no external moderation company and there are no volunteers with moderation rights.

Debaty is a small service, and that means two things. The good: every notice reaches someone who knows the product, and there is no call centre in between. The less good: there are few of us, so a notice may sit for a few days if there is nothing urgent about it. Notices relating to safety or the law take priority.

Reporting

In brief: with or without an account, confidential in both cases. We ask for the link and for an explanation of why it is not acceptable.

If you come across content that breaches the [guidelines](#) or the law, report it.

- **With an account:** via the menu next to the post, the comment or the profile, or via the [reporting page](#).
- **Without an account:** via the [public reporting point](#). You provide the link to the content, explain why it crosses the line, and leave your name and e-mail address so that we can let you know the outcome. In doing so you confirm that you are making the notice in good faith.

One exception to that e-mail address: a notice about child sexual abuse material may be anonymous. No one should feel a barrier for that reason.

A notice is confidential. The person it concerns is not told who reported them — not in the decision, not in the statement of reasons, and not in the public log.

What you add helps. "This is not right" is hard to assess; "this is my address, I did not ask for it to be here" is immediately clear.

Notices about the same post become one case

In brief: if ten people report the same post, that is one case with ten notices, and not the same work ten times over.

As soon as a notice comes in about a piece of content, a **case** is opened for it. Every subsequent notice about that same post is added to that case. The moderator sees one row, with the counter alongside it and with all the stated reasons side by side.

This is not administrative tidiness but the reason the process is sound. One decision about one piece of content means that everyone who reported it gets the same answer, and that it does not matter whether you were the first or the tenth. If your notice came in after a decision had already been made, it opens a new case — because something new is then going on.

What the software does, and what it does not do

In brief: it flags and queues. In a small number of very clear-cut cases it temporarily hides a post. It never removes anything, it never restricts an account, and it never decides anything about private messages.

This chapter is more detailed than the rest, because it is the most uncomfortable part and you are entitled to it.

Two kinds of automated tools are running.

- **Rules and behavioural signals.** Word lists, and patterns such as the same post appearing in many places within a short time.
- **Language models that give a text a score** for how hurtful or threatening it is. For Dutch we use RobBERT, for other languages a multilingual model. Both run **on our own servers within the EU**; your text is not sent to another company for this and is not used to train those models further.

What happens with that score. There are two thresholds, and the whole difference lies between the two:

Score	What happens
Below the low threshold	Nothing.
Above the low threshold	The content stays visible and enters the queue, so that a moderator looks at it.
Above the high threshold	The content is temporarily hidden until a moderator has looked at it.

That temporary hiding is a **provisional measure, not a decision**. If the moderator looks at it and the notice does not hold up, the content automatically returns, and the same signal cannot hide it again afterwards.

Four limits the software does not cross:

- **Only the Dutch model may temporarily hide.** The multilingual model advises the moderator and never hides anything on its own, because it is not equally reliable in every language.
- **Private messages are never hidden automatically.** Only the human route applies there, and only after a notice.
- **Removing, warning, restricting, suspending and terminating is always done by a person.** There is no path by which software can do that.
- **Every flag and every temporary hiding is recorded**, together with the score and the threshold used, so that afterwards it can be traced why something happened.

Why there is software at all, then. Because otherwise only what someone reports would be moderated, and the worst things are often not reported by the people they affect. That is the trade-off; we consider it defensible, and that is why it is here rather than in a footnote.

What it is not. This is not decision-making by a computer within the meaning of the privacy law (AVG Article 22 — the provision that protects you against a computer deciding about you on its own). There is always a person in between. If there is software in the run-up to a decision about you, that is stated in the statement of reasons you receive. That is what the Digital Services Act, the European regulation on online services, requires on this point (Article 14 and Article 17).

How a moderator decides

In brief: they look at the content and at the context, and choose the lightest measure that solves the problem.

The moderator sees the reported post, the comments around it, what the reporters themselves wrote, and whether anything has been decided about this account before. After that there is one of five outcomes:

Outcome	What it means
Do nothing	The notice does not hold up. The content stays.

Warning	You are told what went wrong; the content stays.
Remove content	That one post is taken down. Your account stays.
Restriction or suspension	No posting or commenting for a time, or the account temporarily silenced.
Termination	Only in the most serious case.

We choose the **lightest measure that works**, and the heaviest only for something serious or for repetition. The same ladder is set out in the [terms of use](#) and in the [guidelines](#).

A moderator who starts on a case makes that visible to the others, so that two people do not assess the same thing independently of each other.

For a notice about a private message the moderator gets the reported message with a handful of messages around it as context — not the whole conversation and not the rest of your messages. This is recorded in our internal log: who looked, when, and in which case.

Thirty seconds in which no one hears anything

In brief: after a decision there is half a minute in which the moderator can undo it. During that time nothing has yet been sent and nothing has yet been irreversibly discarded.

A moderator can make a mistake, and the most painful moment is the moment just after the click. That is why not everything happens at once.

- **Immediately:** the decision itself. The content is gone or the account is silenced, and the case is closed. Waiting would mean harmful content stays up for another half minute, and that is the wrong side to err on.
- **Within thirty seconds:** the moderator can undo it. Then the world returns to its old state and nothing goes out.
- **After that:** everything a person gets to hear — the message to you, the message to whoever reported — and the permanent clearing-out of files.

This means that an undone mistake does not reach you, and that is the intention. It also means that a decision that does reach you was already made thirty seconds earlier.

The decision and the explanation you receive

In brief: you receive a notice in the app with what was decided, which rule it concerned, what we saw, whether software was involved, and how to lodge a complaint.

If we make a decision about your content or your account, you receive an explanation with four parts (Digital Services Act Article 17):

- **the basis** — which guideline or which legal rule is concerned;

- **the facts** — what we specifically saw;
- **whether software was involved in the flagging**;
- **that and how you can lodge a complaint.**

All your decisions are together on the [complaints page](#), including the older ones.

What a reporter is told

In brief: that your notice has been assessed. Not what happened to the other person's account.

If you reported something, you will hear from us that your notice has been dealt with — even if the outcome was "do nothing", and even if you were one of ten reporters.

What you do **not** hear is which measure someone else received. That is that person's case. You are entitled to an answer, not to another member's file.

Complaint

In brief: six months' time, free of charge, and someone other than the person who decided looks again. If it is upheld, we reverse the decision.

If you disagree with a decision about your content or your account, go to the [complaints page](#) and tell us why. You have **six months** for this from the decision — the period that the Digital Services Act prescribes. Lodging a complaint costs nothing, and you do not need to bring in a lawyer.

Who assesses it. Your complaint goes to a moderator who was not involved in the original decision. That is not just an agreement: our own systems refuse to accept the complaint from the person who made the decision. The only exception is the situation in which Debaty has only one moderator at that time — then the rule would mean that your complaint cannot be dealt with at all, and that is worse for you than a reassessment by the same person.

What can come out of it.

- **Upheld:** the decision is reversed. Removed content returns, a suspension is lifted, a warning lapses.
- **Rejected:** the decision stands, and you are told why.

In both cases you receive notice in the app.

How long it takes. We aim to assess a notice within seven days and a complaint within fourteen days, counted from the moment it comes in. Anything relating to safety or the law takes priority. These are our own target periods and not legal ones: there is no automatic clock on it, and at busy times it may take longer. If it takes unreasonably long, let us know via [contact](#).

If you cannot resolve it with us

In brief: three routes outside us, and all three are open.

- **An independent dispute settlement body.** You can submit your case to a dispute settlement body that has been certified for this by a supervisory authority (Digital Services Act Article 21). We cooperate with this and are bound by the outcome as the law provides. Which bodies these are is kept by the European Commission; we may not designate any ourselves.
- **The supervisory authority.** If you consider that we are not complying with the rules, you can report that to the **Autoriteit Consument & Markt** in the Netherlands, or to the supervisory authority for digital services in your own country.
- **The courts.** That route is always open, and the two above are not a precondition for it.

If your complaint is about your personal data and not about a moderation decision, it runs differently: see [your data rights](#) and the [privacy statement](#).

Serious and criminal matters

In brief: where there is a suspicion that someone's life or safety is at risk, we inform the police or another competent authority.

Where there is a suspicion of a criminal offence that threatens someone's life or safety, we inform the competent authorities (Digital Services Act Article 18). Child sexual abuse material always goes that route, and is removed immediately.

Authorities that wish to contact us have a [dedicated contact point](#). We do not hand over any data about members without a valid legal basis, and we record every request.

If someone turns out to be under 16

In brief: we first ask for confirmation, and only remove the account if it fails to come. Children are not simply removed from their account on a suspicion.

Debaty is intended for people aged 16 and over (AVG Article 8, as applied in the Dutch UAVG) and is not aimed at children. If you suspect that an account belongs to someone younger, report it with the reason "**Suspected minor**".

Such a notice takes a separate route, because it is not about what is written but about who posted it:

- **We ask for confirmation.** The member is asked, via a message or e-mail, to confirm within **14 days** that they are 16 or older.
- **If that confirmation fails to come, we suspend the account.**
- **After that we remove it**, with erasure of the personal data via the ordinary AVG deletion procedure. An administrator confirms the final removal.

Why not remove it immediately? Because a suspicion is not a finding, and an unjustified removal on the basis of an anonymous notice would be an attractive weapon. And why not check everyone in advance? Because that would mean we would have to see proof of identity from every single member. The full reasoning is set out in the [minors policy](#).

What we keep of a decision

In brief: moderation decisions are kept, even after an account is gone — otherwise no complaint is possible and no checking afterwards.

A notice, the decision on it and the explanation are kept. That is also necessary: without a file a complaint cannot be dealt with and it cannot be verified afterwards whether we complied with our own rules. If an account is removed, the decision continues to exist without the person behind it.

The precise retention periods per type of data are set out in the [privacy statement](#), in the chapter on retention.

Public log and figures

In brief: removals of public content appear in a public log, without names.

If a moderator takes down public content, that appears in a public log: the type of decision, the reason and the time. **Never** the name of whoever reported, and never the name of whoever posted it. This makes it possible to see that and how moderation takes place without it becoming a pillory.

Warnings, restrictions and suspensions are not in it — they are about a person and not about a piece of content.

On the [transparency page](#) we explain this and periodically publish the figures added together.

About this policy

In brief: it changes along with the service and with the rules. We announce important changes in advance.

We adjust this policy when the platform or the regulations call for it. We announce important changes in advance, and the date at the bottom shows when this page last changed.

See also the [guidelines](#) for what is and is not allowed, the [terms of use](#) for the arrangements around it, and the [minors policy](#).

Version and changes

In brief: this is version 1.0 of this document, adopted on 23 July 2026.

- **v1.0 — 23 July 2026.** First adopted, complete version of this document.

Where to take a complaint

In brief: come to us first. If that does not resolve it, the choice is yours: our Dutch supervisory authority, or the one in your own country.

We are a Dutch company. Our own supervisory authorities are therefore the **Autoriteit Persoonsgegevens** (the Dutch data protection authority) and the **Autoriteit Consument & Markt** (the Dutch authority for digital services). You can turn to either one, wherever you live.

If you live in the EU or the EEA, you do not have to come to the Netherlands for that. You can lodge your complaint with the supervisory authority of the country where you live or work, in your own language. That authority handles it or passes it on, and remains your point of contact. The data protection authorities are listed by the European Data Protection Board (edpb.europa.eu); the Digital Services Coordinators are listed by the European Commission.

If you live outside the EU and the EEA, you can still turn to our Dutch supervisory authorities. On top of that, the law of your own country may give you an authority of your own; a complaint that reaches us through such an authority is handled in exactly the same way.

Why no foreign authority is named here. The language you are reading this in does not say which country you live in. Which supervisory authority is yours depends on where you live, not on the language you read — and sending you to the wrong desk costs you the very time a complaint does not leave you.

And the courts are always open. If you live in the EU, that may be the court in your own country, where consumer law there allows you to do so.

Language and validity

In brief: the version you read is the one that applies to you — a translation included.

This document is available on Debaty in fourteen languages. **The version in your language applies to you, and it binds us too.** Where a translation differs from the Dutch text, the reading that is most favourable to you applies: we cannot hold our own translation error against you.

Dutch is the language in which we write changes and against which we check translations. That is a working arrangement inside our company, not a ranking between readers. If a translation lags behind the Dutch text, we say so at the top of the page, with both dates.

Wherever you live, the mandatory consumer law of your own country continues to apply, whatever else is written here.