To: **The Department of Justice and Constitutional Development**
Adv M Khosa and Adv LT Sebelemetja
Email: MaKhosa@justice.gov.za/ LSebelemetja@justice.gov.za

28 January 2026

Campaign for Free Expression (CFE) Comments to the Draft Regulations under the Prevention and Combating of Hate Crimes and Hate Speech Act, 2023

1. Introduction and framing

The Campaign for Free Expression (CFE) supports effective and constitutionally compliant measures to prevent and combat hate crimes and hate speech. We recognise the importance of reliable data in understanding patterns of harm and informing appropriate policy and institutional interventions. Accordingly, we welcome the opportunity to make comment to the draft regulations under the Prevention and Combating of Hate Crimes and Hate Speech Act, 2023 (the Act).

Section 8 of the Act authorises the prescription of information for the limited and specific purpose of enabling "effective monitoring, analysis of trends and interventions". Regulations made under this provision must therefore be narrowly tailored, analytically sound, and designed to generate data that is both reliable and meaningful.

CFE's concerns with the draft Regulations do not relate to the legitimacy of monitoring hate crimes or hate speech as such, rather they relate to whether the regulatory scheme, as currently designed, is fit for purpose, constitutionally proportionate, and capable of supporting coherent analysis and appropriate intervention.

2. Scope of information collected: grounds versus identifiability

CFE accepts that, in order to fulfil the purpose of section 8, it is both legitimate and necessary for the prescribed forms to record the prohibited ground or grounds in respect of which a hate crime or hate speech offence is alleged to have been committed. Monitoring which grounds most frequently motivate such offences ,or are referenced in the perpetration of such offences, and how those patterns fluctuate over time, is significant in understanding trends and designing interventions.

CFE does not object to the fact that a reported offence may be recorded as having been committed in respect of grounds such as race, religion, HIV status, or political affiliation. Properly understood, this does not require confirmation that a complainant or victim of such an offence in fact belongs to that category, and does not in itself entail intrusive personal-data processing.

The difficulty arises from the extent of personally identifiable information that the Regulations require to be collected, collated, and transmitted to the Director-General alongside this grounds-based information.

For the purposes contemplated in section 8, specifically, monitoring, trend analysis and intervention, it is unnecessary for the Director-General to receive or retain full identifying particulars of complainants or victims, accused persons or convicted persons – such as their names, identity or passport numbers, residential addresses, or fingerprint numbers. Those identifiers serve operational criminal-justice functions governed elsewhere in law, but are not required to achieve the objectives of section 8.

By mandating the routine central collation of case-level identifying information for section 8 reporting purposes, the Regulations go beyond what is necessary to give effect to the Act.

CFE accordingly submits that:

> information transmitted to the Director-General for section 8 purposes should, as a default, be aggregated or de-identified; and
>
> a clear separation should be maintained between operational justice records and the monitoring and reporting framework contemplated by section 8.

3. Doctrinal uncertainty and the risk of analytically incoherent data

CFE is further concerned that the Regulations proceed on an implicit assumption that “hate speech” and “hate crimes” constitute settled, uniform legal categories, capable of consistent identification and recording across institutions. That assumption does not reflect the current state of South African jurisprudence.

Hate-speech doctrine in particular remains confused and contradictory, with divergent approaches in our courts, and ongoing uncertainty regarding the boundary between:

constitutionally protected but offensive expression;

expression that may attract civil liability under equality legislation; and

expression meeting the criminal threshold under the Act.

In this context, there is a real risk that the database envisaged by the Regulations will aggregate materially unlike cases — grouping together different forms of expression, harm, and legal characterisation under a single label of "hate speech". Where different police stations, prosecutors, and courts apply different understandings of the offence, trend data may reflect inconsistency in classification rather than genuine changes in prevalence or severity.

This is not a merely technical concern. Analytically incoherent data risks:

misleading Parliament and Chapter 9 institutions;

distorting public understanding; and

informing interventions that are poorly calibrated to the actual nature of the problem.

CFE submits that, until greater clarity is achieved through judicial interpretation of the Act, the section 8 monitoring framework should be designed with doctrinal uncertainty in mind, and should:

prioritise aggregated data over case-level records; and

incorporate outcome markers (such as withdrawal, acquittal, or reversal on appeal) so that trend analysis does not conflate allegations with legally sustained findings.

4. Static form design and evolving modes of harm, particularly online

While the prescribed forms are clearly designed to assist in identifying which prohibited grounds attract more or less hate speech or hate crime, they are largely structured as a check-box exercise and are not designed to reflect how such conduct is committed.

This limitation is particularly acute in respect of digital hate speech.

Form 3 appropriately recognises the importance of electronic communications by creating a distinct category for offences under section 4(1)(b) of the Act. However, the form does not allow for the recording of information that would reflect:

platform-specific dynamics;

algorithmic amplification;

coordinated or networked dissemination; or

the use of emerging technologies, including generative artificial-intelligence tools embedded within digital platforms.

As a result, the monitoring framework may record that hate speech occurred online, while failing to capture how it was generated, amplified, or incentivised. This is a material shortcoming, given the growing recognition, both locally and internationally, that certain platform architectures and technologies actively shape the scale and intensity of online harm.

Without the ability to capture evolving modes of perpetration, the data collected under section 8 will be of limited value in informing meaningful or differentiated interventions. Interventions appropriate to one-to-one communication, for example, differ fundamentally from those required to address algorithmically amplified or AI-assisted abuse.

CFE does not suggest that the Regulations must exhaustively anticipate all future technological developments. However, they should be designed to be adaptable, either by:

permitting periodic revision or supplementation of prescribed forms following consultation; or

allowing for categorical or descriptive fields capable of capturing emerging patterns of harm.

5. Conclusion

CFE submits that, in their current form, the draft Regulations risk producing data that is:

unnecessarily identifiable;

analytically incoherent in light of unsettled jurisprudence; and

insufficiently responsive to evolving forms of harm, particularly online.

These weaknesses undermine the stated purpose of section 8 of the Act. CFE accordingly urges the Department to revise the Regulations so that they are narrowly tailored, analytically robust, and genuinely capable of supporting effective monitoring, trend analysis, and intervention.

Nicole Fritz
Executive Director
Campaign for Free Expression

nicolelfritz@outlook.com
0824523909