Eduardo Alencar

POLICY BRIEF

# Deepfakes: rules advanced, verification has not

What ANPD's Technology Radar no. 6 shows about Brazil, what the design is missing, and how to implement it

**39%**

OF DEEPFAKE CASES DETECTED IN LATIN AMERICA IN 2025 WERE IN BRAZIL (SUMSUB, AS CITED BY ANPD)

**61%**

OF AI-CREATED POLITICAL PROFILES DO NOT SAY THEY ARE AI (IA NAS ELEIÇÕES OBSERVATORY, AS CITED BY ANPD)

**72 H**

BEFORE AND 24 H AFTER THE VOTE: WINDOW BANNING NEW SYNTHETIC CONTENT (TSE, AS CITED BY ANPD)

NOVEMBER 2026

Eduardo Alencar

## Executive Summary

Brazil's National Data Protection Authority (ANPD) has published *Technology Radar no. 6* on deepfakes, a detailed technical and legal diagnosis. This brief starts from that document to answer an implementation question: what does a public manager or regulator need in place for Brazil's 2026 rules to work in practice?

### 01 Brazil moved responsibility from the author to the provider

Two decrees of 20 May 2026 bar providers from generating third parties' intimate content with artificial intelligence (AI) and presume provider liability for illicit content in ads, paid boosting and artificial content-distribution networks. The Digital Statute of Children and Adolescents (ECA Digital) already imposed a removal duty on suppliers. The change is one of design: the target is no longer only whoever produces the deepfake.

### 02 In elections the rule is strict, but transparency fails

According to ANPD, Brazil's Superior Electoral Court (TSE) bans, in 2026, the publication and paid boosting of new synthetic content using the image or voice of candidates or public figures from 72 hours before to 24 hours after the vote, even if labelled. Yet a survey cited in the Radar found that 61% of AI-created political profiles did not disclose that condition.

### 03 Detection cannot yet support enforcement on its own

The Radar reports that three detection tools compared in a 2025 study did not reliably identify the cases, and concludes that detectors gain legal validity only as support for human experts. A rule without reliable verification becomes a rule without evidence.

### 04 The gaps are in implementation, not in intent

Missing are: a regulation saying who the competent authority is and how providers' duties scale; a national standard for digital evidence; a specific criminal offence (none exists today, according to ANPD); and public measurement of effectiveness.

**DESTAQUE** The decisive question is not whether Brazil has a rule on deepfakes. It is who verifies authenticity, by what method, and with what right to contest.

## Contents

**01** The public problem

**02** Method and limitations

**03** What the rules do today

**04** Case study: synthetic characters that look like citizens

**05** Critical success factors for implementation

**06** Gaps and proposed contributions to ANPD and other actors

**07** Prioritised recommendations

**08** Conclusion

**09** Notes, sources and credits

## In numbers

| Indicator | Result | Source and reading |
|---|---|---|
| Brazil's share of deepfake cases detected in Latin America (2025) | 39% | Sumsub, as cited by ANPD (p. 114). Data from an identity-verification firm, not re-verified here |
| Growth in attacks using deepfakes and synthetic identities, 2024 to 2025 | 126% | Sumsub, as cited by ANPD (p. 114) |
| AI-created political profiles that do not disclose being AI | 61% | IA nas Eleições Observatory, as cited by ANPD (p. 109) |
| Electoral ban window on new synthetic content | 72 h before, 24 h after | TSE Resolution 23,755/2026, according to ANPD (p. 127); TSE text to be confirmed |
| US states with synthetic-media laws for campaigns (2019 to 2025) | 28 of 50 | Fukuha, Fürst and Archegas (2026), as cited by ANPD (p. 127) |

## 1. The public problem

Deepfakes are synthetic image, voice or video content generated or manipulated by AI with a high degree of realism. ANPD's Radar shows that they depend on data that identify people, such as facial images and voice, which is why they matter for data protection (p. 145). The harms it cites are unevenly distributed: they fall above all on women and girls (non-consensual synthetic pornography), on electoral debate, on consumers (financial scams) and on racialised groups (p. 145-146).

**Evidence.** According to ANPD, the deepfake "has ceased to be a futuristic concern" and has become "a current problem of economic and social impact" in Brazil (p. 97; translation by the author).

**Inference.** When visual and audio evidence no longer proves authenticity, the cost of verification falls on victims, platforms and the courts. Public policy has to decide who bears that cost.

## 2. Method and limitations

This brief analyses ANPD's Technology Radar no. 6 (2026, Portuguese edition) and cross-checks the Brazilian norms it cites against the official texts published by the Presidency's legal portal and the Chamber of Deputies' open-data API.

- The sections read were those on identification (3.3), the Brazilian context (5.3 to 5.5), the regulatory outlook and mitigation (6.1 to 6.3) and the concluding remarks. The remaining sections were consulted only through the table of contents.
- Statistics from companies and observatories are cited **as reported by ANPD**; they were not re-verified at their original sources.
- TSE resolutions are cited **according to ANPD**: the TSE website could not be consulted, and the official text must be checked before any legal use.
- Reading standard: **evidence** (a fact supported by a source), **inference** (the author's reading) and **recommendation** (a proposed normative choice).
- Page numbers refer to the Portuguese edition of the Radar; an English translation (160 pages) also exists.

**DEFINIÇÃO** Deepfake: synthetic content that simulates the image or voice of a real person with realism. Cheapfakes and shallowfakes, simpler alterations, appear in the Radar as a separate category (section 2.2.3).

## 3. What the rules do today

### Three international patterns and Brazil's position

The Radar identifies three regulatory patterns centred on risk rather than technology: transparency (labelling), bans in electoral contexts, and criminalisation of sexual abuse (p. 124).

| Pattern | Example cited by ANPD | Brazil |
|---|---|---|
| Transparency | EU AI Act, Art. 50: synthetic content marked and detectable | TSE requires an explicit notice in propaganda using synthetic content (p. 126) |
| Electoral ban | South Korea bans campaign deepfakes from 90 days before to election day (p. 125) | TSE: bans since 2024 and a 72 h before to 24 h after window in 2026 (p. 125-127) |
| Criminalisation | UK, France, Australia and the US (*Take It Down Act*, 2025) criminalise sexual deepfakes (p. 128-131) | No specific offence; fitted to Art. 218-C and a penalty increase under Art. 147-B of the Penal Code (p. 131-132) |

### Brazilian norms, checked against the official text

| Norm | What it establishes | Verification |
|---|---|---|
| Decree 12,975/2026 (20 May 2026) | Inserts into Decree 8,771/2016 a presumption of provider liability when illicit content circulates in ads, paid boosting or artificial distribution networks, regardless of notice (Art. 16-L); the provider is exempt if it proves it acted diligently and within a reasonable time | Official text read on 21 Nov 2026 |
| Decree 12,976/2026 (20 May 2026) | Bars providers from generating and modifying third parties' intimate content with AI (Art. 9) and requires technical safeguards to identify and block such requests, "as established in regulation by the competent authority" (Art. 10), scaled by traffic volume and risk level | Official text read on 21 Nov 2026 |

| Norm | What it establishes | Verification |
|---|---|---|
| Law 15,211/2025 (ECA Digital) | Duty to remove and report content of apparent sexual exploitation and abuse of children and adolescents (Art. 27); bans monetisation and boosting of eroticised content depicting children (Art. 23) | Official text read on 21 Nov 2026 |
| Law 15,123/2025 | Increases by half the penalty for psychological violence against women (Art. 147-B of the Penal Code) when committed with AI or technology that alters image or sound | Official text read on 21 Nov 2026 |
| TSE Resolutions 23,732/2024 and 23,755/2026 | Ban on deepfakes in propaganda; ban on new synthetic content using the image or voice of candidates and public figures in the window from 72 h before to 24 h after the vote | Cited by ANPD (p. 125-127); TSE text to be confirmed |
| Bill (PL) 2,338/2023 (AI framework) | The Deepfakes Working Group of the Secretariat for Digital Policies proposed adding a concept of synthetic content, transparency and traceability, and a specific risk assessment; a non-binding recommendation (ANPD, p. 135-136) | Chamber API on 21 Nov 2026: under consideration, status "Awaiting Opinion", latest procedural step on 13 Nov 2026 |

**Inference.** The Brazilian design combines three instruments (bans, provider duties and targeted punishment), in line with the Radar's recommendation of multi-faceted models (p. 133).

## 4. Case study: synthetic characters that look like citizens

The Radar describes profiles such as "Dona Maria" and "Sô Zé da Feira", characters entirely generated by AI that simulate ordinary people and comment on politics (p. 107-109). Instead of faking a politician, they build false spokespeople for society.

| Signal | What the Radar reports | Why it matters for implementation |
|---|---|---|
| Transparency | 61% of AI-created political profiles do not disclose that condition, according to the IA nas Eleições Observatory (p. 109) | The labelling rule exists; compliance cannot be observed without monitoring |
| Content | More than half of the posts analysed contained some form of disinformation (p. 109) | Labelling is not enough if the content deceives |
| Reach | Manipulated content can reach millions before it is identified, corrected or removed (p. 108) | Response time matters more than later sanctions |

**ALERTA** The case identifies a pattern of labelling non-compliance in one observatory's sample. It does not allow attributing authorship or measuring any effect on the vote.

## 5. Critical success factors for implementation

1. **Regulate by risk and harm, not by technology.** The Radar shows that technology-based norms become obsolete and are circumvented by model opacity (p. 123-124).
2. **A provider duty with measurable scaling.** Decree 12,976 provides for implementation scaled by volume and risk; without a public metric there is no way to know whether the scaling was met.
3. **Verification with human expertise and chain of custody.** Detectors are black boxes, lose accuracy outside the lab and have legal validity only as support for experts (p. 69-70, 142).
4. **Provenance at the source.** Cryptographic signatures, watermarks and standards such as C2PA document the content's production chain (p. 141), but still require broad testing.
5. **Coordination among institutions.** The Radar points to the link between data protection, electoral regulation and accountability as increasingly necessary (p. 112).
6. **Literacy as a complement.** The Radar itself states that its effect is limited given growing realism and the power asymmetry (p. 140).

## 6. Gaps and proposed contributions to ANPD and other actors

This block is a collaborative contribution: it points to what the diagnosis could incorporate to support implementation.

| Gap | Proposed contribution |
|---|---|
| Who the "competent authority" in Art. 10 of Decree 12,976 is, and when the regulation will issue | Map the open point and propose scaling criteria (volume, risk, type of service) |
| Effectiveness of the 2026 electoral rules | Propose an evaluation with 2026 election data (labelling, removals, response time), with the electoral authorities and platforms |
| Digital evidence standard | Propose a national protocol for expert examination and chain of custody for deepfakes, consistent with using detectors only as support |
| A specific criminal offence | Compare the Radar's dilemma (over-criminalisation versus impunity, p. 128) with a narrow offence for extortion and abuse of intimate images |
| Cost and capacity | Estimate the cost of verification for municipalities, small agencies and small firms, currently outside the diagnosis |
| Victims' perspective and the Global South | Include reporting channels, time to redress and evidence from middle-income countries |

## 7. Prioritised recommendations

### 01 Regulate Art. 10 of Decree 12,976

The competent authority should publish scaling criteria and public metrics on blocked requests. Risk of inaction: a duty without a parameter, hard to enforce.

### 02 Establish a national digital-evidence standard for deepfakes

Courts, forensic services and police should adopt a common protocol for verification and chain of custody. Risk of inaction: divergent rulings and easy challenges to the evidence.

### 03 Publish a labelling and removal dashboard

Electoral authorities and platforms should disclose labelling met, removals and response time for each election. Risk of inaction: no learning from 2026 to 2028.

### 04 Sign official communication with provenance

Public bodies should adopt provenance signatures on institutional video and audio, so that true content is verifiable. This is the author's proposal, to be tested in a pilot.

### 05 Decide the criminal offence with proportionality

If Congress opts for a specific offence, it should be limited to extortion and abuse of intimate images, as the Radar suggests (p. 133), and tied to Bill 2,338/2023.

## Conclusion

In 2026 Brazil built a framework that combines electoral bans, provider duties and targeted punishment. What does not yet exist is the verification infrastructure that gives those rules meaning. If image and voice no longer prove authenticity, who becomes the arbiter of evidence, and with what guarantee of the right to reply?

## Notes, sources and credits

**Main source.** ANPD. *Radar Tecnológico nº 6: deepfakes* (Technology Radar no. 6). Brasília, 2026 (authorship: Gabriela Campos Alkmin, Ingrid David Alves de Carvalho, Roseane Salvio, Roberta Fernandes Mendiondo Nunes, Mariana Luisa da Costa Lage and Ticianne de Gois Ribeiro Darin). Pages follow the printed numbering of the Portuguese edition. Quotations from the Radar are translated by the author.

**Official texts consulted on 21 Nov 2026.**

- Decree no. 12,975 of 20 May 2026: <https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2026/decreto/D12975.htm>
- Decree no. 12,976 of 20 May 2026: <https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2026/decreto/D12976.htm>
- Law no. 15,211 of 17 September 2025 (ECA Digital): <https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2025/lei/L15211.htm>
- Law no. 15,123 of 24 April 2025: <https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2025/lei/L15123.htm>
- Chamber of Deputies, open-data API, Bill 2,338/2023: <https://dadosabertos.camara.leg.br/api/v2/proposicoes/2487262>

**Third-party data cited by ANPD, not re-verified:** Sumsub (*Identity Fraud Report 2025-2026*), IA nas Eleições Observatory, Fukuha, Fürst and Archegas (2026), Moraes and Batista (2025). TSE Resolutions 23,732/2024 and 23,755/2026: check the official TSE text.

**Authorship.** Eduardo Alencar. Independent analysis, with no affiliation to ANPD. The gaps identified are contributions to the debate.