Digital Policy Alert

How Governments Address Discrimination and Bias in AI

Tommaso Giardini and Nils Deeg

November 2026

An initiative of the
St.Gallen Endowment
for Prosperity through Trade

How Governments Address Discrimination and Bias in AI

The risk that AI systems reproduce or amplify existing bias and lead to discrimination is particularly acute where AI informs high-stakes decisions, including areas such as housing, credit, hiring, and healthcare. Examples of discrimination cut across these areas, as evidenced by cases in the US: They include a settlement between the US Justice Department and Meta over claims that its housing advertising algorithm discriminated by race, sex and national origin, a settlement between the Massachusetts Attorney General and a student lender concerning allegations that its AI underwriting disadvantaged Black and Hispanic applicants, a US federal court allowing race, age and disability claims against Workday's AI hiring tools to proceed, and a pharmacy chain agreeing to a five-year ban on using facial recognition after the Federal Trade Commission suggested that it produced false matches more often in stores in Black and Asian neighbourhoods. Governments worldwide are aiming to prevent such cases and ensure redress when they occur.

Domestic Regulation

Discrimination and bias is an emerging concern in AI regulation: The DPA has documented 115 domestic policy and enforcement developments on discrimination and bias in AI since January 2020. The DPA's systematic taxonomy reveals the following trends in AI discrimination and bias regulation:

- The US is the most active jurisdiction, with 32 state-level and 13 federal developments, followed by Argentina (9), Saudi Arabia (7) and the EU (6).
- The current status of developments is spread: 44 developments are in force or adopted, 59 are currently under deliberation and 12 are concluded, rejected or revoked.
- Most discrimination and bias rules are found in binding laws (60) or binding orders (15), compared to 32 non-binding guidelines. Only 8 enforcement cases and inquiries focus on AI discrimination and bias.

Recently, California's Governor signed a bill on AI and health care services requiring both developers and deployers to make reasonable efforts to identify systems at risk of biased impacts when used in healthcare, while Colorado's Governor signed a Bill on automated decision-making in consequential areas that establishes liabilities under anti-discrimination laws for automated decisions. Further, the legislature of the Republic of Korea introduced a bill adding anti-discrimination standards to the AI Basic Act and Thailand's Electronic Transactions Development Agency closed a consultation on a proposed AI Act that would prohibit AI systems whose use results in broad discrimination.

Our text-level analysis of 84 rulebooks across 19 jurisdictions reveals how governments address discrimination and bias in AI in detail. We find that governments use three different regulatory requirements, appearing 50 times, to pursue non-discrimination and address bias in domestic rules, namely non-discrimination obligations requiring AI to be fair and banning discriminatory systems (24 provisions), mitigating bias (13), and detecting bias in datasets and outputs (13).

International Coordination

Our text-level analysis of nine international AI frameworks shows that all cover discrimination and bias across 22 provisions.

International frameworks address all three discrimination and bias requirements. Bias mitigation is the most commonly mentioned requirement (7), followed by non-discrimination (6) and bias detection (6). Five frameworks address all three requirements, while the OECD AI Principles address none.

Framework	Bias mitigation	Non-discrimination	Bias detection
ASEAN Expanded Guide (Generative AI)	■	■	■
ASEAN Guide	■	■	■
Council of Europe Framework Convention	–	■	–
G7 Hiroshima Code of Conduct	■	–	■
G7 Hiroshima Guiding Principles	■	–	–
Joint Declaration	■	■	■
OECD AI Principles	–	–	–
UNESCO Recommendation	■	■	■
UNGA Resolution	■	■	■

International AI frameworks address discrimination and bias with various degrees of focus and actionability. Of the 22 provisions, nine focus on discrimination and bias as the core concern, while 13 only touch upon it tangentially. Also, 16 provisions across seven frameworks provide actionable language, and six address discrimination and bias in non-actionable language (in the Joint Declaration, the OECD AI Principles, the Council of Europe Framework Convention and the G7 Code of Conduct).

Alignment Between Domestic and International Action

Domestic regulations and international frameworks thus weigh discrimination and bias differently. Domestically, non-discrimination dominates, with nearly twice as many provisions as either bias mitigation or bias detection. Internationally, seven frameworks include bias mitigation, and six each cover non-discrimination and bias detection. Finally, the international language behind this overlap is mostly concrete but often indirect. Most provisions are actionable, but fewer than half focus on discrimination and bias.

To understand how governments tackle discrimination and bias in AI in depth, readers can explore our AI Governance Hub's discrimination and bias page. It shows the relevant DPA entries and the full-text provisions of domestic AI regulations and international AI frameworks, structured using our taxonomy. The page draws data from the DPA's Activity Tracker, which provides daily updates on digital policy and provides tailored notifications, as well as our full-text analysis tool Clairk.

Digital Policy Alert

Contact
Written by Tommaso Giardini and Nils Deeg
Please send questions and suggestions to nils.deeg@sgept.org

The independent Digital Policy Alert is a pillar of the Swiss-based St. Gallen Endowment for Prosperity Through Trade

St.Gallen Endowment
for Prosperity through Trade