

BLOOM ETHICAL AI CONSULTING

ARCHITECTING ETHICAL AI

Researched and Authored by Andrew Bloom

MORAL ARCHITECTURE WEEKLY BRIEF | Week Ending July 10, 2026

1. EXECUTIVE TAKEAWAY

Three developments this week move AI governance away from Washington's stalled preemption fight and toward verifiable safety obligations built by states and international bodies. On July 6, 2026, Illinois became the first state to require large frontier AI developers to undergo annual independent third-party safety audits, signing SB 315 into law. On July 1, the UN's Independent International Scientific Panel on AI, co-chaired by Yoshua Bengio and Maria Ressa, found no known technical guarantee that AI agents reliably follow instructions, a finding delivered two days before the inaugural UN Global Dialogue on AI Governance seated all 193 member states in Geneva for the first time. And on July 8, the newly formed AI for Good Global Commission, co-chaired by Rwanda's President Kagame and Salesforce's Marc Benioff, held its first working meeting, also in Geneva. Build the accountability architecture that satisfies the strictest jurisdiction, not the slowest regulator.

2. AT A GLANCE

Development	What Happened	Why It Matters
Illinois Enacts SB 315, the AI Safety Measures Act (July 6, 2026)	Governor Pritzker signed the nation's first law requiring large frontier AI developers to undergo mandatory annual independent third-party safety audits, alongside catastrophic-risk frameworks and 72-hour incident reporting.	Joins California's TFAIA and New York's RAISE Act to form a de facto three-state national frontier-AI safety standard covering roughly 40% of the U.S. AI market.
UN Scientific Panel Report & Global Dialogue on AI Governance (July 1 & July 6 to 7, 2026)	A 40-expert UN panel co-chaired by Bengio and Ressa found no technical guarantee that AI agents follow instructions reliably; findings anchored the first UN dialogue to seat all 193 member states on AI.	First globally coordinated, depoliticized scientific reference point for AI risk, arriving as governments negotiate binding rules with unequal access to compute and evidence.
AI for Good Global Commission Inaugural Meeting (July 8, 2026)	A 44-member UN/ITU-affiliated commission co-chaired by Rwanda's President Kagame and Salesforce's Marc Benioff, seating frontier AI lab CEOs alongside heads of state, held its first session in Geneva.	First formal structure placing AI builders and AI governors in the same governance body: a structural shift in how industry is seated at the international table.

3. DEVELOPMENT DEEP-DIVES

Illinois Enacts the Nation's First Mandatory Frontier-AI Audit Law

a. What Happened

On July 6, 2026, Illinois Governor JB Pritzker signed SB 315, the AI Safety Measures Act. It covers "large frontier developers" (revenue over \$500M, models above 10²⁶ FLOPs): a public Frontier AI Framework for catastrophic-risk assessment, pre-deployment transparency reports, and 72-hour (24-hour for imminent risk) incident reporting. Uniquely, it requires annual independent third-party audits. Disclosure duties take effect January 1, 2027; audits begin January 1, 2028. Penalties reach \$1M to \$3M per violation. Anthropic publicly backed the bill.

b. Why It Matters

SB 315 joins California's TFAIA and New York's RAISE Act to form a de facto national frontier-AI safety standard, reaching any developer whose models reach Illinois users regardless of headquarters.

c. Moral Architecture Analysis

This is a guardrail graduating from principle to enforceable structure: audit logging and third-party verification written into statute, not left to self-attestation. Organizations should not wait for the FTC's Section 5 preemption theory to be resolved in court before building this documentation architecture.

d. Board and Executive Action

- Map deployment against the IL, CA, and NY thresholds; treat all three as binding wherever models reach those states' users.
- Begin building a Frontier AI Framework now; the runway to 2028 audits is short once evaluator capacity is considered.
- Stand up the 72-/24-hour incident workflow and identify qualified independent auditors before demand outstrips supply.

Sources: Office of the Governor of Illinois, July 6, 2026; Capitol News Illinois, July 6, 2026; Crowell & Moring, July 2026.

UN Scientific Panel Warns No Guarantee AI Will Follow Instructions; 193 Nations Convene in Geneva

a. What Happened

On July 1, 2026, the UN's Independent International Scientific Panel on AI, a 40-member body co-chaired by Yoshua Bengio and Maria Ressa and selected from candidates across 140 countries, released its first Preliminary Report, concluding there is "no known technical guarantee" that AI agents consistently follow instructions, and warning that AI power is concentrating among a handful of firms and countries. The findings anchored the inaugural UN Global Dialogue on AI Governance, held July 6 to 7 in Geneva, the first UN forum to seat all 193 member states on AI. A second Dialogue is set for May 2027 in New York.

b. Why It Matters

Unlike prior curated summits (Bletchley 2023, Seoul 2024), Geneva guaranteed every member state a seat. The Panel's mandate is scientific, not prescriptive: a depoliticized evidence base akin to the IPCC.

c. Moral Architecture Analysis

Bengio put it plainly: there is no known technical guarantee that AI will follow instructions, norms, or laws. Meaningful human oversight in consequential decisions is not aspiration; it compensates for a documented engineering shortfall. Boards should treat this as technical fact, not a talking point to be weighed against competing testimony.

d. Board and Executive Action

- Brief the board on the instruction-following gap; where autonomous agents are deployed, document the compensating human-oversight control in writing.
- Track the Panel's thematic briefs ahead of its first full report, due May 2027, as an emerging cross-jurisdiction reference point.
- Flag the May 2027 New York Dialogue as the checkpoint for whether Geneva converts into binding instruments.

Sources: UN Independent International Scientific Panel on AI, July 1, 2026; UN News, July 6, 2026; Inter Press Service, July 2, 2026.

AI for Good Global Commission Holds Inaugural Meeting

a. What Happened

On July 2, 2026, Rwanda's President Paul Kagame, Salesforce CEO Marc Benioff, and ITU Secretary-General Doreen Bogdan-Martin launched the AI for Good Global Commission, a body of more than 40 members seating heads of state alongside tech executives (Amazon's Jassy, Microsoft's Smith, Nvidia's Huang, Anthropic co-founder Jack Clark). It held its inaugural meeting July 8 in Geneva, part of a "Digital Week" that also included the UN Global Dialogue, focused on connecting the 2.2 billion offline and on "responsible AI solutions."

b. Why It Matters

This is the first time frontier AI lab CEOs sit as formal members of a UN-affiliated governance body alongside heads of state, rather than as invited stakeholders, a structural shift in how industry is seated in international AI governance.

c. Moral Architecture Analysis

Seating AI's builders alongside its would-be governors blurs a separation this framework treats as structural: those who build and profit from a system should not be its sole overseers. The Commission cannot substitute for the independent guardrails enacted into law this same week in Illinois. Its legitimacy will turn on measurable commitments, not a communiqué endorsing principles already accepted.

d. Board and Executive Action

- Reconcile any commitments made through the Commission with binding state and EU obligations; voluntary pledges do not substitute for enforceable ones.
- Monitor whether its connectivity mandate intersects with your infrastructure footprint in underconnected regions.
- In board reporting, distinguish multistakeholder commitments from enforceable obligations (SB 315, the EU AI Act, state ADMT laws).

Sources: International Telecommunication Union, July 2, 2026; Axios, July 1, 2026; Salesforce Newsroom, July 2, 2026.

4. REGULATORY RADAR

Deadline	Item	Jurisdiction	Board Implication
July 31, 2026	Comment deadline on FTC's proposed AI accuracy / Section 5 state-preemption policy statement	U.S. Federal	Still open; submit substantive comment. A proposed statement remains non-binding until finalized.
Aug. 1, 2026	NSA/CISA deadline to deliver classified frontier-model benchmarking process and voluntary pre-release framework under the June 2 cybersecurity executive order	U.S. Federal	Deadline is statutory, not tied to any signing ceremony; confirm exposure if your models could be deemed "covered frontier models."
Aug. 2, 2026	AI Act Art. 50 transparency and labelling duties become binding (watermarking delayed to Dec. 2, 2026 for legacy systems)	European Union	Confirm which systems qualify for the watermarking delay versus the original date.
Dec. 2, 2026	Ban on AI-generated CSAM and non-consensual intimate imagery takes effect	European Union	Screen generative and synthetic-media tools now; compliance window is under six months.
Jan. 1, 2027	Illinois SB 315 core disclosure/framework obligations take effect; Colorado ADMT law (SB 26-189) substantive obligations take effect	IL / CO	Build disclosure and framework documentation now; Colorado remains the direct target of the FTC's preemption theory.
Jan. 1, 2028	Illinois SB 315 mandatory independent third-party audit requirement begins for large frontier developers	Illinois	Secure qualified audit capacity early; evaluator supply is expected to be a bottleneck across all covered states.
May 2027	Second UN Global Dialogue on AI Governance convenes in New York	International	Checkpoint for whether Geneva's commitments produced binding instruments or remained aspirational.
Ongoing	Great American AI Act discussion draft (released June 4, 2026 by Reps. Obernolte and Trahan) remains open for stakeholder comment; national-security groups are separately pressing for AI export-control provisions in the FY2027 NDAA	U.S. Federal	Submit input on GAAIA's three-year development-law preemption clause; monitor FY2027 NDAA markup for export-control riders, which remain the more likely near-term federal vehicle.

5. WHAT LEADERS SHOULD BE WATCHING NOW

- **States Are Setting the Verification Standard, Not Just the Disclosure Standard:** Illinois' independent-audit mandate goes further than California's or New York's frontier-AI laws, requiring proof rather than self-attestation. Expect other states to follow the "Illinois audit" template rather than build their own from scratch.
- **The Science Is Now on the Record, Not Just the Politics:** The UN Panel's finding that no technical guarantee of AI instruction-following exists is a depoliticized, globally endorsed statement of the alignment gap. Boards can no longer treat "AI will do what we tell it" as a safe assumption in governance documentation.
- **Industry Is Being Seated as Governor, Not Just Governed:** The AI for Good Global Commission formalizes a role for AI lab executives inside international governance bodies. This raises the bar for what counts as genuine independent oversight versus industry self-representation dressed as multilateralism.

6. CLOSING CHALLENGE QUESTION

When a UN panel of forty scientists concludes there is no technical guarantee AI will follow instructions, and in the same week a state legislature, not Congress, becomes the first government anywhere to require independent proof of frontier-model safety, is global AI governance converging on higher standards, or fragmenting into whichever jurisdiction moves first?

The answer to that question is the measure of your Moral Architecture.

7. ABOUT THE AUTHOR

Andrew Bloom is one of a small number of practitioners who has built original AI governance frameworks from ethical first principles. As an ordained Rabbi and founder of Moral Architecture, his frameworks are grounded in formal ethical and theological training, distinguishing his approach from compliance-driven consulting. Moral Architecture embeds accountability, transparency, and human oversight into AI systems, and his Guardrails Framework prevents AI-enabled harm before it occurs. He is the author of The Ten Commandments of Ethical AI and Technology and Theology. As founder of Bloom Ethical AI Consulting and Rabbi of Congregation Ahavath Sholom in Fort Worth, he advises executives and boards on AI governance readiness.

Andrew Bloom, Founder and Principal | AI Ethicist | Governance Advisor

8. SOURCES

1. Federal Trade Commission, "FTC Seeks Public Comment on Policy Statement Addressing AI Accuracy," July 1, 2026.
2. Office of the Governor of Illinois, "Gov. Pritzker Signs Nation-Leading Artificial Intelligence Safety Law," press release, July 6, 2026.
3. Capitol News Illinois, "Pritzker signs landmark AI regulation bill that aims to mitigate risks," July 6, 2026.
4. Crowell & Moring LLP, "Illinois AI Safety Measures Act SB 315: What Frontier AI Developers Must Do Before January 2028," client alert, July 7, 2026.
5. United Nations, Independent International Scientific Panel on Artificial Intelligence, Preliminary Report: Evidence-Based Assessment of Opportunities, Risks and Impacts of AI, July 1, 2026.
6. UN News, "From AI to 'killer robots': UN chief issues urgent governance call," July 6, 2026.
7. Inter Press Service, "UN Artificial Intelligence Panel Launches Report Ahead of Global Conference," July 2, 2026.
8. International Telecommunication Union, "Global Leaders Launch AI for Good Global Commission to Expand Access, Strengthen Trust and Accelerate Impact," press release, July 2, 2026.
9. Axios, "Exclusive: UN launches 'AI for Good' commission," July 1, 2026.
10. Salesforce Newsroom, "AI for Good Global Commission Announcement," July 2, 2026.
11. The White House, Executive Order "Promoting Advanced Artificial Intelligence Innovation and Security," June 2, 2026.
12. Seyfarth Shaw LLP, "Colorado Enacts Artificial Intelligence Replacement Law," May 22, 2026.
13. Council of the European Union, final adoption of the AI Act Digital Omnibus, June 29, 2026; European Parliament vote, June 16, 2026.
14. Reps. Jay Obernolte and Lori Trahan, discussion draft of the Great American AI Act of 2026, released June 4, 2026.
15. Foundation for Defense of Democracies Action, coalition letter urging AI export-control provisions in the FY2027 NDAA, June 2026.