Building Global Trust and Security: A Framework for AI-Driven Criminal Scoring in Immigration Systems

A.I. Safety Initiative
AI Governance Policy Hackathon
Spring 2025

Group Members:

Buğrahan Namdar

James Hamraie

Michael Bolleddu

Table of Contents

I. Executive Summary

The accelerating use of artificial intelligence (AI) in immigration and visa systems, especially for criminal history scoring, poses a critical global governance challenge. Without a multinational, privacy-preserving, and interoperable framework, AI-driven criminal scoring risks violating human rights, eroding international trust, and creating unequal, opaque immigration outcomes.

While banning such systems outright may hinder national security interests and technological progress, the absence of harmonized legal standards, privacy protocols, and oversight mechanisms could result in fragmented, unfair, and potentially discriminatory practices across countries.

This policy brief recommends the creation of a legally binding **multilateral treaty** that establishes:

1. **An International Oversight Framework:** Including a Legal Design Commission, AI Engineers Working Group, and Legal Oversight Committee with dispute resolution powers modeled after the WTO.

2. **A Three-Tiered Criminal Scoring System:** Combining Domestic, International, and Comparative Crime Scores to ensure legal contextualization, fairness, and transparency in cross-border visa decisions.

3. **Interoperable Data Standards and Privacy Protections:** Using pseudonymization, encryption, access controls, and centralized auditing to safeguard sensitive information.

4. **Training, Transparency, and Appeals Mechanisms:** Mandating explainable AI, independent audits, and applicant rights to contest or appeal scores.

5. **Strong Human Rights Commitments:** Preventing the misuse of scores for surveillance or discrimination, while ensuring due process and anti-bias protections.

6. **Integration with Existing Governance Models:** Aligning with GDPR, the EU AI Act, OECD AI Principles, and INTERPOL protocols for regulatory coherence and legitimacy.

An implementation plan includes treaty drafting, early state adoption, and phased rollout of legal and technical structures within 12 months. By proactively establishing ethical and interoperable AI systems, the international community can protect human mobility rights while maintaining national and global security.

Without robust policy frameworks and international cooperation, such tools risk amplifying discrimination, violating privacy rights, and generating opaque, unaccountable decisions.

This policy brief proposes an international treaty-based or cooperative framework to govern the development, deployment, and oversight of these AI criminal scoring systems. The brief outlines technical safeguards, human rights protections, and mechanisms for cross-border data sharing, transparency, and appeal. We advocate for an adaptive, treaty-backed governance framework with stakeholder input from national governments, legal experts, technologists, and civil society. The aim is to balance security and mobility interests while preventing misuse of algorithmic tools.

II. Background and Problem Statement

The lack of a multinational, privacy-preserving, and interoperable framework for AI-driven criminal scoring risks undermining the fairness and integrity of global migration systems. Without strong oversight, safeguards, and public accountability, such technologies could erode human rights, international trust, and ethical AI governance.

However, merely banning these systems outright may be counterproductive, posing concerns for national security and technological innovation. Instead, establishing an international standard is crucial to prevent conflicts arising from incompatible data formats, privacy rules, or legal standards. Without a shared framework, governments may adopt conflicting systems or misuse criminal scores, harming international cooperation, data protection, and human mobility rights.

A. Background Information

Governments worldwide are increasingly incorporating AI into immigration, border control, and national security systems, offering efficiencies and predictive capabilities. However, these benefits come with significant risks, especially when AI is deployed without safeguards, explainability, or alignment with democratic values. AI-driven criminal scoring systems, which automate the process of assessing criminal histories for immigration decisions, raise challenges due to differing legal definitions, sentencing policies, and expungement rules across countries. Criminal data has long been a policy input in immigration decisions, but AI's potential to standardize this data across jurisdictions introduces concerns about fairness and legal comparability.

Currently, there is no global legal framework for sharing, interpreting, or standardizing criminal data for AI use, creating governance gaps. Existing agreements like INTERPOL's I-24/7 and the Budapest Convention do not address AI scoring or data governance issues, leaving room for ad hoc systems without oversight. This lack of interoperable standards could lead to geopolitical imbalances or misuse.

AI's use in visa processing also raises concerns about explainability and bias. Denied applicants may not understand the reasoning behind decisions, especially when based on foreign or unverified data. Additionally, biases in AI models can perpetuate local enforcement disparities. Growing interest in AI for migration management highlights the urgent need for international cooperation to establish ethical and legal safeguards and avoid a fragmented, inequitable landscape.

B. Problem Statement: Status Quo Challenges for Visa Processing and Potential of Implementing AI Criminal Scoring Systems

Governments around the world are increasingly turning to artificial intelligence (AI) to streamline and enhance immigration decision-making processes. The increasing mobility of individuals with varying criminal backgrounds, and the use of AI in migration-related contexts, raises the question of how to responsibly integrate criminal history information into visa decisions. The growing use of AI in migration decisions, particularly for generating criminal risk scores promises consistency and efficiency, but pose significant legal, ethical, and technical challenges

1. Cross-Border Governance, Privacy, and Comparative Law

AI-generated criminal scores may reflect inconsistent legal definitions across countries, creating fairness issues. Varying crime classifications and privacy laws make score comparisons potentially misleading or unjust without harmonized standards. Lack of transparency and safeguards could also lead to wrongful visa denial.

2. Geopolitical Challenges

Differing privacy laws, trust deficits, and reluctance to share sensitive data complicate cross-border cooperation. Sovereignty concerns and resistance from powerful states further hinder efforts to align the international AI system.

3. Fairness and Transparency

AI's "black box" nature risks systemic discrimination, particularly when algorithms are trained on biased or uneven data. Lack of explainability and procedural fairness could deny individuals the right to understand or contest critical decisions.

4. Abuse Safeguards

Without robust safeguards, governments could misuse criminal scoring systems for exclusionary or surveillance purposes, including targeting political dissidents or marginalized groups.

5. Risk of Discrimination or Bias

While AI promises efficient risk management, its use in immigration decisions could exacerbate inequalities, particularly when trained on biased or incomplete data. This could lead to unjust outcomes, such as family separation or persecution risks.

III. Recommendations & Implementation Plan

This section contains a set of policy recommendations for rules and legal frameworks governing AI criminal score systems followed by a detailed description of an implementation plan for national governments, legislatures with treaty ratification authority, and members of transnational organizations, such as the United Nations.

A. Policy Recommendations

1. Establish a Multilateral Treaty and Cooperation Body

In order to form the international ai based criminal scoring system we need to have an international body that has the following sub-groups. All signatory states will have members in each body.

a. **Legal Design Commission(LDC):** Legal commission will be tasked to identify possible legal jargon definitions, compose main crime categories, and discuss which national criminal code articles fall under proposed categories. This body shall include 2 legal representatives from each signatory state.
b. **AI Engineers Working Group(AEWG):** This group will be working to design, assess and improve the AI. Shall include 1 government assigned expert from each country.

c. **Legal Oversight Committee(LOC):** This commission shall be responsible to oversee national uses of the proposed AI to avoid abuses and follow compliance to privacy laws. Members of the oversight committee shall be composed of 1 national constitutional judge, 1 national public prosecutor, 1 foreign constitutional judge for each member country.

I. Under this committee we will have a **dispute settlement body** that uses the same structure as the WTO dispute settlement body. Individuals can apply to their constitutional courts or prosecutors to bring individual cases to this body, which will be evaluated in bulk rather than singularly.

II. This committee will host **bi-annual meetings** to improve the legal system of the project, **privacy laws** and **interoperability framework**, **regulate and propose resolutions** to improve the design. Resolutions adopted by the Legal Oversight Committee shall apply to operations of both Legal Design commissions and AI Engineers working groups.

III. Legal Design commissions and AI Engineers working group will have **2 representatives each** in the Legal Oversight Committee; will have the right **to propose resolutions and take part in discussions, and vote**.

2. Develop a Three-Tiered Criminal Scoring System

Primarily being the task of the training process of Criminal Score AI, this task will be checked and re-evaluated by the Legal Design Commission and national legal entities. It is the cornerstone of the proposed framework is a harmonized, three-tiered scoring mechanism that evaluates criminal histories across jurisdictions. This model aims to balance domestic sovereignty, international objectivity, and legal contextualization.

The **Domestic Criminal Score (DCS)** is generated based on the individuals' criminal background evaluated under the national laws and sentencing guidelines of their country of origin. This allows for national legal variance while maintaining a foundational baseline. The weights assigned to domestic score shall primarily be through auto weighting by the AI, and corrected by national legal entities responsible for criminal law. The national entity's formation is up to signatory nations' decision makers; it is recommended to be a commission of judges and lawyers from that country. The base model is as:

$$DCS = \sum_{i=1}^{n} q_i * w_i * f(t_i) + \delta \cdot S$$

Where

$$S = \sum_{i=1}^{n} ([r_i > 1] * w_i * (r_i - 1)) + \sum_{i \epsilon D} (w_i * [D > 1]$$ And $$f(t_i) = e^{-\alpha t_i}$$

- qi = commitment quantity of crimes of type
- wi= national weight for crime type
- f(ti) = decay function since last crime commitment
- e= 2.72 (Euler's number for natural decay)
- α = national decay rate

- ti= time passed since the crime occurred
- n = total number of crime types
- δ=1 if the person is a serial offender, if not 0 (so S doesn't get triggered)
- S = serial crime penalty
- ri = number of times crime type iii is repeated
- D = set of distinct crime types committed
- [ri>1] = Iverson bracket: equals 1 if condition is true, 0 otherwise (so repeated crime condition doesn't get triggered)
- D = number of distinct crime types

The **International Criminal Score (ICS)** is calculated using harmonized criteria agreed upon by consortium members, such as crime type, crime severity and time elapsed since sentencing. It serves as a neutral benchmark for comparison across legal systems and cultures. International Criminal Score will be average Weights assigned by all of the signatory countries' domestic score systems. Will define an average of signatories.

$$ICS = \frac{1}{m}\sum_{j=1}^{m} DCSj$$

- m: Number of signatory countries
- DCSj: The Domestic Crime Score computed by country j based on harmonized criteria (e.g., crime type, severity, time elapsed)

The **Comparative Crime Score(CCS)** provides a contextual analysis of how the crime is viewed and penalized in both the origin and destination countries. This score is shown as domestic scores of both countries, their average and the weighted average based on international score model.

$$CCS = \frac{1}{2}(DCSorigin + DCSdestionation) + \frac{1}{2}ICS$$

Combines both domestic average and international criminal score

The International Score is shown to all signatories. National scores are only available to citizens' own countries' legal bodies. Comparative score is not visible to bureaucrats but only available to system in which bureaucrats submit documents to, and dispute settlement body if demanded.

By layering these scores, the framework avoids one-size-fits-all assumptions and provides a nuanced tool for immigration and visa assessments.

3. Create Interoperable Data Standards and Legal Safeguards

For the system to function effectively across borders, technical interoperability and legal safeguards must be foundational. Shared data fields and formats should be developed to enable the secure exchange of anonymized criminal records. Data should be **pseudonymized upon transfer** and re-identified **only by authorized visa adjudicator system,** bureaucrats in consulates shall not be able to view this information directly. **Cryptographic protocols** must be used to ensure **confidentiality and integrity** during transmission.

Legal safeguards should include clear limitations on data use. Only **consular and visa authorities' computerized systems** should have access to criminal scores, and applicants

must be granted the **right to appeal their scores**. Appeals should include access to a clear explanation of the scoring inputs and outputs. A **centralized logging system** should record every score request, score generation, and access event to support **transparency and auditability.**

4. Require Training and Transparency for AI Systems

To ensure **trustworthiness and accountability**, AI models used to generate criminal scores must meet rigorous training and explainability standards. These systems should be trained on **both domestic legal texts** (criminal codes, sentencing guidelines, and case outcomes) and an **international corpus** that enables comparative evaluation.

Furthermore, the system should support regular **audits by independent oversight bodies**. Training materials should also include guidance for edge cases and emphasize the importance of **human review**.

5. Enshrine Human Rights and Privacy Protections

Any international framework must be grounded in core **human rights n**orms. The criminal scoring system should be used **strictly for visa and immigration-related purposes**. Participating states must formally commit not to repurpose the scores for domestic surveillance, employment screening, rights violations or policing. This limitation would be enforced through treaty language or a binding consortium charter.

The framework must also include **anti-discrimination and due process** provisions. These include the **right to contest** adverse scoring outcomes, **appeal** denials, and **correct inaccuracies**. Special safeguards should be built in to prevent disproportionate harm to **marginalized groups, refugees. Transparency** mechanisms**,** such as public reporting on use cases and aggregate scoring trends**,** should further reinforce **democratic accountability.**

6. Incorporate Existing Governance Frameworks

The proposed treaty bodies should **leverage existing AI and data governance frameworks**. The EU's General Data Protection Regulation (GDPR) provides a robust model for data minimization, purpose limitation, and user rights. The EU AI Act offers valuable rules for high-risk AI systems, particularly in sensitive areas like public decision-making. The OECD AI Principles provide high-level guidance on transparency, fairness, and accountability. Similarly, INTERPOL and the Council of Europe offer existing protocols on biometric and criminal data exchange that can be used to structure secure data sharing processes. Aligning the proposed system with these frameworks ensures interoperability, reduces regulatory friction, and facilitates broader adoption.

B. Implementation Timeline

Objective: Build political, legal, and institutional foundation

Phase	Time Frame	Key Activities
Phase 1	Months 1–3	Treaty negotiation and establishment of international body with sub-groups
Phase 2	Months 4–6	Recruitment of members and formation of Legal Design Commission, AI Working Group, and Oversight Committee
Phase 3	Months 7–12	Define legal taxonomy, develop crime categories, national crime code mappings; start AI prototype
Phase 4	Year 2	Test AI scoring model domestically, design DCS model; begin legal and privacy framework discussions
Phase 5	Year 3	Harmonize ICS model; launch pilot ICS scoring between few nations; initiate Comparative Score testing
Phase 6	Year 4	Expand ICS usage to all signatories; resolve pilot disputes via WTO-modeled body; begin full AI audit trails
Phase 7	Year 5	Finalize interoperability protocols, start bi-annual legal review cycles, resolve edge case applications

Phase 1: Foundational Framework (Months 0–12)

- **Month 0–3**
  - Draft initial **Multilateral Treaty** proposal
  - Identify and invite potential **signatory states**
  - Establish **provisional secretariat** to manage coordination

- **Month 4–6**
  - Ratify treaty among early adopters (at least 5–10 states)
  - Form **Interim Committees**: Legal Design Commission (LDC), AI Engineers Working Group (AEWG), Legal Oversight Committee (LOC)

- **Month 7–12**
  - Draft standardized **crime typology** and category framework
  - Begin **legal mapping** across jurisdictions
  - Define initial **data interoperability schema** and privacy principles
  - Identify open-source AI frameworks for prototyping

Phase 2: System Development & AI Prototyping

Objective: Build technical infrastructure and develop scoring models

- **Month 13–18**
  - Develop **DCS and ICS models** (based on existing laws and weights)
  - Prototype **AI scoring engine** trained on anonymized case data
  - Implement **pseudonymization** and **encryption protocols**
  - Define **decay functions**, serial crime penalties, dispute flagging logic
- **Month 19–24**
  - Run **sandbox trials** with simulated data across 3–5 legal systems
  - Finalize the **Comparative Score model**
  - Design interfaces for **automated visa systems**
  - Complete first round of **model audits and fairness evaluations**

Phase 3: Legal Integration & Pilots

Objective: Begin phased adoption within legal and consular systems

- **Month 25–30**
  - National legal entities validate and calibrate **DCS weights**
  - Finalize **interoperability API** for government use
  - Form **Dispute Settlement Body** and prepare intake protocols
- **Month 31–36**
  - Launch **pilot integration** in 2–3 willing countries (for visa processing only)
  - Begin **training programs** for visa officials and legal monitors
  - Publish **first transparency report** (aggregate scores, appeals, etc.)

Phase 4: Full Legal & Consular Rollout

Objective: Deploy full system in signatory countries

- **Month 37–42**
  - Adopt and integrate **Legal Oversight Committee resolutions**

- ➢ Expand deployment to remaining signatory states
- ➢ Establish national **score auditing commissions**

❖ **Month 43–48**

- ➢ Conduct **cross-border interoperability testing**
- ➢ Launch **public-facing rights portal** (appeals, transparency, score correction)
- ➢ Apply GDPR-style privacy review mechanisms
- ➢ Host **bi-annual summit** to adjust weightings and models

Phase 5: Consolidation & Expansion

Objective: Strengthen compliance, review efficacy, and expand adoption

❖ **Month 49–54**

- ➢ Publish independent **impact assessment**
- ➢ Refine AI using **real-world data + dispute outcomes**
- ➢ Begin negotiation with **non-signatory states** for observer status or entry

❖ **Month 55–60**

- ➢ Expand training programs to judges, prosecutors, and oversight personnel
- ➢ Codify model into broader frameworks (OECD AI, UN digital governance)
- ➢ Release **global implementation report** and propose Phase 2 treaty upgrades

IV. Impact Assessment

This section evaluates the benefits of a standardized international criminal scoring framework for migration and visa systems, followed by an overview of key risks and mitigation strategies.

A. Benefits and Opportunities

A harmonized criminal scoring system could enhance global governance by improving visa fairness, public safety, and cross-border legal coordination. It also promotes responsible AI use in sensitive decision-making domains.

1. Legal Alignment and Data Sharing.

The framework encourages consistent criminal assessment across borders by mapping national laws to a shared scoring rubric. While preserving national sovereignty, this standardization would enable more equitable visa determinations and facilitate broader cooperation in extradition, human rights, and law enforcement.

2. Responsible AI Innovation.

Embedding safeguards such as explainability, data minimization, and auditability enables innovation while ensuring AI systems remain fair and accountable. Requiring diverse training datasets and comparative scoring models supports legally and culturally grounded decision-making.

3. Geopolitical Cooperation.

By offering pseudonymized data-sharing formats and restricted access protocols, the framework reduces diplomatic friction. An opt-in consortium model with flexible participation could help overcome national reluctance and promote incremental alignment.

4. Rehabilitation and Reintegration.

Shifting from rigid criminal labels to nuanced scoring based on proportionality and recidivism affirms principles of second chances, dignity, and non-discrimination. This could temper punitive migration practices and improve societal reintegration of reformed individuals.

B. Risks and Mitigation

Deploying a global criminal scoring framework in immigration decisions introduces substantial risks that must be proactively managed. Without clear safeguards, such systems could lead to rights violations, discrimination, diplomatic friction, and loss of public trust. Key risks and mitigation strategies are discussed below.

1. Overreliance on Automated Scores and Lack of Transparency

Visa officers may defer too heavily to AI-generated scores, treating them as objective rather than interpretive tools, and without transparency, applicants may not understand or challenge their scores. To mitigate this, the framework should require human-in-the-loop review for all decisions, empower officers to override scores, and mandate intelligible explanations and meaningful appeal rights, supported by regular audits and independent oversight bodies to ensure appropriate use and fairness.

2. Algorithmic Bias and Structural Discrimination

AI systems trained on biased or incomplete datasets may disproportionately flag certain groups, reinforcing systemic inequalities and global patterns of exclusion. This risk can be mitigated by using representative, internationally vetted datasets; conducting fairness audits across demographics; excluding non-criminal factors such as political views or immigration status; and enabling independent challenge mechanisms to correct systemic disparities.

3. Data Privacy and Function Creep

Collected data may be repurposed for surveillance or domestic enforcement, particularly in weak-rule-of-law environments, undermining trust and violating individual rights. To address this, the framework should impose strict purpose limitations, require encrypted and pseudonymized data exchanges, limit access via audit logs, and ensure alignment with international privacy standards such as the OECD Guidelines or Convention 108+.

4. Governmental Resistance and Uneven Adoption

States may resist participation due to concerns about sovereignty, political sensitivities, or reputational risk, resulting in fragmented or inconsistent implementation. To mitigate this, the framework should be modular and opt-in, allowing phased adoption with technical support, and should be co-developed with input from both Global North and Global South stakeholders to foster legitimacy, flexibility, and mutual benefit.

5. Erosion of Procedural Justice and Human Rights Norms

Routine use of algorithmic tools in immigration contexts may erode due process and human rights protections, particularly for vulnerable populations lacking legal recourse. Mitigation requires embedding robust procedural safeguards—such as notice, explanation, and appeal rights—aligned with international human rights standards (e.g., ICCPR), and establishing

independent review bodies along with regular human rights impact assessments to ensure fairness and accountability over time.

References

- Cihon, P. (2019). *Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development*. Future of Humanity Institute, University of Oxford.
- Council of Europe (2020). *Consultative Committee Convention 108 – Guidelines on Artificial Intelligence and Data Protection*.
- European Commission. (2021). *Proposal for a Regulation laying down harmonised rules on artificial intelligence (EU AI Act)*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206
- European Commission. *Handbook on European Data Protection Law*. (2022). https://fra.europa.eu/en/publication/2022/handbook-european-data-protection-law
- INTERPOL. (2021). *INTERPOL's Framework for Responsible Use of Biometrics and AI*.
- OECD (2021). *Framework for the Classification of AI Systems*. https://www.oecd.org/going-digital/ai/classification/
- OECD (2019). *Recommendation of the Council on Artificial Intelligence*. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449
- UNESCO (2021). *Recommendation on the Ethics of Artificial Intelligence*. https://unesdoc.unesco.org/ark:/48223/pf0000381137
- United Nations. (n.d.). *UN Structure*. United Nations. https://www.un.org/en/model-united-nations/un-structure
- WTO Dispute Settlement Understanding (DSU) – WTO Legal Texts. https://www.wto.org/english/docs_e/legal_e/ursum_e.htm