Proposed Frontier Model Transparency Framework

I. **Scope**
   A. Applies to AI companies that are developing frontier models, defined as a combination of thresholds for computing power, computing cost, evaluation performance, annual revenue and R&D.
   B. Should not apply to start-ups and very small companies, based on either an annual revenue or aggregate R&D expenditure threshold.

II. **Pre-Deployment Requirements**
   A. **Develop and Implement Secure Development Framework**. Require covered AI companies to develop and follow Secure Development Frameworks (SDFs) prior to deployment that set out how they assess and mitigate unreasonable Catastrophic Risks from covered models.
   B. **Catastrophic Risks Defined**. Catastrophic Risks refers to specific threats from Chemical, Biological, Radiological, and Nuclear (CBRN) and models that cause significant harm by acting autonomously in ways contrary to the intent of their developers and users.
   C. **Secure Development Framework Minimum Standards.**
      1. Identify the model(s) to which it applies
      2. Describe how the covered AI company plans to assess and mitigate Catastrophic Risks, including standards, capability evaluations, and mitigations
      3. Address process for modifying the SDF modification
      4. Identify a primarily responsible corporate officer for SDF compliance and implementation
      5. Describe whistleblower processes in place for employees to raise concerns about SDF content and implementation, and protections from retaliation
      6. Require the covered AI company to confirm separately that it has implemented its SDF and relevant policies and procedures prior to frontier model deployment
      7. Retain copies of SDFs and updates for at least five years

III. **Minimum Transparency Requirements**
   A. Covered companies must:
      1. Disclose the SDF under which they are currently developing and evaluating models in a readily accessible format on a public-facing website registered to and maintained by the AI company
      2. Publish a system card or similar documentation at the time of deployment of a new model or the addition of a substantial and new capability to an existing model. This documentation should summarize model testing and evaluation procedures, results, and required mitigations under the SDF

3. Certify compliance with SDF requirements and describe any mitigations implemented as required by the SDF and disclose compliance on a publicly accessible website

B. Covered companies may redact information required for the SDF and/or system card-type documentation that constitutes a trade secret, confidential business information, or information that would materially compromise public safety or the security of the model. Any redactions or omissions shall be briefly identified and justified in the public version.

IV. Enforcement

A. Prohibit intentionally false or materially misleading statements related to SDF compliance
B. Authorize the attorney general to seek civil penalties for material violations
C. 30-day right to cure