

Doug Fiefia proposes the following substitute bill:

1 **Artificial Intelligence Transparency Amendments**

2026 GENERAL SESSION

STATE OF UTAH

**Chief Sponsor: Doug Fiefia**

Senate Sponsor: Michael K. McKell

---

---

2 **LONG TITLE**

3 **General Description:**

4 This bill enacts the AI Transparency Act relating to transparency and whistleblower  
5 protections for frontier artificial intelligence models.

6 **Highlighted Provisions:**

7 This bill:

8 ▶ defines terms;

9 ▶ requires developers of certain artificial intelligence models to create, implement, and  
10 publish public safety and child protection plans;

11 ▶ requires developers to publish summaries of risk assessments for certain artificial  
12 intelligence models;

13 ▶ prohibits developers from making materially false or misleading statements about covered  
14 risks;

15 ▶ requires developers to report certain safety incidents to the Office of Artificial  
16 Intelligence Policy (office);

17 ▶ requires the office to provide annual assessments and legislative recommendations  
18 regarding regulation of certain artificial intelligence models;

19 ▶ establishes civil penalties for violations;

20 ▶ provides whistleblower protections for employees who report safety concerns of certain  
21 artificial intelligence models;

22 ▶ establishes remedies for employees who suffer adverse action for whistleblower  
23 activities;

24 ▶ creates the AI Transparency Enforcement Restricted Account to fund enforcement  
25 activities; and

26 ▶ provides a severability clause.

27 **Money Appropriated in this Bill:**

29 None

## Other Special Clauses:

None

## Utah Code Sections Affected:

## ENACTS:

13-72b-101, Utah Code Annotated 1953

13-72b-102, Utah Code Annotated 1953

13-72b-103, Utah Code Annotated 1953

13-72b-104, Utah Code Annotated 1953

**13-72b-105, Utah Code Annotated 1953**

**13-72b-106, Utah Code Annotated 1953**

13-72b-107, Utah Code Annotated 1953

**13-72b-108, Utah Code Annotated 1953**

**13-72b-109, Utah Code Annotated 1953**

**13-72b-201, Utah Code Annotated 1953**

**13-72b-202, Utah Code Annotated 1953**

13-72b-203, Utah Code Annotated 1953

*Be it enacted by the Legislature of the state of Utah:*

Section 1. Section **13-72b-101** is enacted to read:

## CHAPTER 72b. AI Transparency Act

## Part 1. Artificial Intelligence Transparency and Child Protection

### 13-72b-101 . Definitions.

### As used in this chapter:

(1) "Affiliate" means a person controlling, controlled by, or under common control with a specified person, directly or indirectly, through one or more intermediaries.

(2) "Artificial intelligence model" means an engineered or machine-based system that varies in the system's level of autonomy and that can, for explicit or implicit objectives, infer from the input the artificial intelligence model receives how to generate outputs that can influence physical or virtual environments.

(3)(a) "Catastrophic loss" means:

(i) the death or serious bodily injury of more than 50 individuals; or

- (ii) damage to property, or loss of property, exceeding \$1,000,000,000.
- (b) "Catastrophic loss" does not include the loss of value of equity.
- (4)(a) "Catastrophic risk" means a foreseeable and material risk that a frontier developer's development, storage, use, or deployment of a frontier model will materially contribute to a catastrophic loss in a single incident by:
  - (i) providing assistance in creating or releasing a chemical, biological, radiological, or nuclear weapon;
  - (ii) engaging in a cyberattack, or conduct that, if committed by an individual, would constitute murder, assault, extortion, or theft, including theft by deception, under Utah law, without meaningful human oversight, intervention, or supervision; or
  - (iii) evading control of the frontier developer or user.
- (b) "Catastrophic risk" does not include a foreseeable and material risk from any of the following:
  - (i) information that a frontier model outputs if the information is otherwise publicly accessible in a substantially similar form from a source other than a foundation model;
  - (ii) lawful activity of the federal government; or
  - (iii) harm caused by a frontier model in combination with other software if the frontier model did not materially contribute to the harm.
- (5) "Child protection plan" means a documented technical and organizational protocol to manage, assess, and mitigate child safety risks.
- (6) "Child safety incident" means an occurrence in which a covered chatbot, when interacting with a minor, engages in behavior that, if engaged in by a human, would be considered to intentionally or recklessly:
  - (a) cause death or bodily injury to the minor; or
  - (b) cause severe emotional distress to the minor.
- (7) "Child safety risk" means a material and foreseeable risk that a frontier developer's foundation model, when used as part of a covered chatbot operated by the frontier developer, will engage in behavior when interacting with a minor that, if the behavior had been engaged in by a human, would be considered to be intentionally or recklessly:
  - (a) cause death or bodily injury to the minor, including as a result of self-harm; or
  - (b) cause severe emotional distress to the minor.
- (8) "Covered chatbot" means a service that:
  - (a) allows an ordinary person to have conversations in which humanlike responses are

generated by a foundation model;

(b) is foreseeably likely to be accessed by minors; and

(c) has at least 1,000,000 monthly active users.

(9) "Covered risk" means a catastrophic risk or a child safety risk.

(10) "Critical safety incident" means any of the following:

(a) unauthorized access to, modification of, inadvertent release of, or exfiltration of, the model weights of a frontier model;

(b) the death of, or serious injury to, more than 50 people or more than \$1,000,000,000 in damage to, or loss of, property resulting from the materialization of a catastrophic risk;

(c) loss of control of a frontier model that:

(i) causes death or bodily injury; or

(ii) demonstrates materially increased catastrophic risk; or

(d) a frontier model that uses deceptive techniques against the frontier developer to subvert the controls or monitoring of the frontier developer outside of the context of an evaluation designed to elicit this behavior and in a manner that demonstrates materially increased catastrophic risk.

(11)(a) "Deploy" means to make a frontier model available to a third party for use,

modification, copying, or combination with other software.

(b) "Deploy" does not include making a frontier model available to a third party for the primary purpose of developing or evaluating the frontier model

(12) "Foundation model" means an artificial intelligence model that is all of the following:

(a) trained on a broad data set:

(b) designed for generality of output; and

(c) adaptable to a wide range of distinctive tasks.

(13)(a) "Frontier developer" means a person who has used, or initiated the use of, a

quantity of computing power of at least  $10^{26}$  integer or floating-point operations to train a frontier model, including computing used for the original training run and for any subsequent fine-tuning, reinforcement learning, or other material modifications.

(b) "Frontier developer" does not include an accredited college or university to the extent the college or university is developing or using frontier models exclusively for academic research purposes.

(14) "Frontier model" means a foundation model that was trained using a quantity of computing power of at least  $10^{26}$  integer or floating-point operations, including

130 computing for the original training run and for any subsequent fine-tuning,  
131 reinforcement learning, or other material modifications the developer applies to a  
132 preceding foundation model.

133 (15) "Large frontier developer" means a frontier developer who together with the frontier  
134 developer's affiliates, had annual revenue of at least \$500,000,000 in the preceding  
135 calendar year.

136 (16) "Minor" means an individual younger than 18 years old.

137 (17) "Model weight" means a numerical parameter in a frontier model that is adjusted  
138 through training and that helps determine how inputs are transformed into outputs.

139 (18) "Office" means the Office of Artificial Intelligence Policy created in Section 13-72-201.

140 (19) "Property" means tangible or intangible property.

141 (20) "Public safety plan" means a documented technical and organizational protocol to  
142 manage, assess, and mitigate catastrophic risks.

143 (21) "Safety incident" means a child safety incident or a critical safety incident.

144 Section 2. Section **13-72b-102** is enacted to read:

145 **13-72b-102 . Public safety plan for catastrophic risks -- Requirements.**

146 (1) A large frontier developer shall write, implement, comply with, and clearly and  
147 conspicuously publish on the large frontier developer's internet website a public safety  
148 plan that describes in detail how the large frontier developer:

149 (a) incorporates national standards, international standards, and industry-consensus best  
150 practices into the public safety plan;

151 (b) defines and assesses thresholds used by the large frontier developer to identify and  
152 assess whether a frontier model has capabilities that could pose a catastrophic risk,  
153 which may include multiple-tiered thresholds;

154 (c) applies mitigations to address the potential for catastrophic risks based on the results  
155 of assessments undertaken pursuant to Subsection (1)(b);

156 (d) reviews assessments of catastrophic risk and adequacy of mitigations of catastrophic  
157 risk as part of the decision to deploy a frontier model or use the frontier model  
158 extensively internally;

159 (e) uses third parties to assess the potential for catastrophic risks and the effectiveness of  
160 mitigations of catastrophic risks;

161 (f) revisits and updates the public safety plan, including any criteria that trigger updates  
162 and how the large frontier developer determines when the large frontier developer's  
163 frontier models are substantially modified enough to require disclosures pursuant to

164                   Section 13-72b-104;

165                   (g) implements cybersecurity practices to secure unreleased frontier model weights from  
166                   unauthorized modification or transfer by internal or external parties;

167                   (h) identifies and responds to critical safety incidents;

168                   (i) institutes internal governance practices to ensure implementation of the processes  
169                   described in this Subsection (1); and

170                   (j) assesses and manages catastrophic risk resulting from the internal use of the large  
171                   frontier developer's frontier models, including risks resulting from a frontier model  
172                   circumventing oversight mechanisms.

173                   (2) If a large frontier developer makes a material modification to the large frontier  
174                   developer's public safety plan, the large frontier developer shall clearly and  
175                   conspicuously publish the modified public safety plan and a justification for that  
176                   modification within 30 days after the day on which the large frontier developer makes  
177                   the material modification.

178                   Section 3. Section **13-72b-103** is enacted to read:

179                   **13-72b-103 . Child protection plan -- Requirements.**

180                   (1) A large frontier developer that operates a covered chatbot shall write, implement,  
181                   comply with, and clearly and conspicuously publish on the large frontier developer's  
182                   internet website a child protection plan that describes in detail how the large frontier  
183                   developer:

184                   (a) incorporates national standards, international standards, and industry-consensus best  
185                   practices into the child protection plan;

186                   (b) assesses potential for child safety risks;

187                   (c) applies mitigations to address the potential for child safety risks based on the results  
188                   of assessments undertaken pursuant to Subsection (1)(b);

189                   (d) uses third parties to assess the potential for child safety risks and the effectiveness of  
190                   mitigations of child safety risks;

191                   (e) revisits and updates the child protection plan, including any criteria that trigger  
192                   updates and how the large frontier developer determines when the large frontier  
193                   developer's foundation models are substantially modified enough to require  
194                   disclosures pursuant to Section 13-72b-104;

195                   (f) identifies and responds to child safety incidents; and

196                   (g) institutes internal governance practices to ensure implementation of the processes  
197                   described in this Subsection (1).

198 (2) If a large frontier developer makes a material modification to the large frontier  
199 developer's child protection plan, the large frontier developer shall clearly and  
200 conspicuously publish the modified child protection plan and a justification for that  
201 modification within 30 days after the day on which the large frontier developer makes  
202 the material modification.

203 Section 4. Section **13-72b-104** is enacted to read:

204 **13-72b-104 . Publication requirements -- Frontier models and foundation models.**

205 (1) A large frontier developer shall conspicuously publish on the developer's internet  
206 website summaries of the following before deploying a new or substantially modified  
207 foundation model as part of a covered chatbot operated by the developer:  
208 (a) assessments of child safety risks conducted pursuant to the developer's child  
209 protection plan;  
210 (b) the results of the assessments described in Subsection (1)(a);  
211 (c) the extent to which third-party evaluators were involved in the assessments described  
212 in Subsection (1)(a); and  
213 (d) other steps taken by the developer to fulfill the requirements of the child protection  
214 plan.

215 (2) A large frontier developer shall conspicuously publish on the large frontier developer's  
216 website summaries of the following before deploying a new frontier model or a frontier  
217 model that was substantially modified by the large frontier developer:  
218 (a) assessments of catastrophic risks from the frontier model conducted pursuant to the  
219 developer's public safety plan;  
220 (b) the results of the assessments described in Subsection (2)(a);  
221 (c) the extent to which third-party evaluators were involved in the assessments described  
222 in Subsection (2)(a); and  
223 (d) other steps taken to fulfill the requirements of the public safety plan with respect to  
224 the frontier model.

225 Section 5. Section **13-72b-105** is enacted to read:

226 **13-72b-105 . Prohibited conduct -- Redactions.**

227 (1)(a) A frontier developer may not make a materially false or misleading statement or  
228 omission about covered risks from the developer's activities or the developer's  
229 management of covered risks.  
230 (b) A large frontier developer may not make a materially false or misleading statement  
231 or omission about the large frontier developer's implementation of, or compliance

232                   with, the large frontier developer's public safety plan.

233                   (c) A large frontier developer that operates a covered chatbot may not make a materially  
234                   false or misleading statement or omission about the large frontier developer's  
235                   implementation of, or compliance with, the large frontier developer's child protection  
236                   plan.

237                   (d) This Subsection (1) does not apply to a statement that was made in good faith and  
238                   was reasonable under the circumstances.

239                   (2)(a) When a frontier developer publishes documents to comply with this part, the  
240                   frontier developer may make redactions to those documents that are necessary to  
241                   protect:

242                   (i) the frontier developer's trade secrets;  
243                   (ii) the frontier developer's cybersecurity;  
244                   (iii) public safety;  
245                   (iv) the national security of the United States; or  
246                   (v) compliance with any federal or state law.

247                   (b) If a frontier developer redacts information in a document pursuant to Subsection  
248                   (2)(a), the large frontier developer shall:

249                   (i) describe the character and justification of the redaction in any published version of  
250                   the document to the extent permitted by the concerns that justify the redaction; and  
251                   (ii) retain the unredacted information for five years after the day on which the  
252                   developer makes the redaction.

253                   Section 6. Section **13-72b-106** is enacted to read:

254                   **13-72b-106 . Safety incident reporting mechanism -- Rulemaking -- Annual**  
255                   **report.**

256                   (1) The office may make rules in accordance with Title 63G, Chapter 3, Utah  
257                   Administrative Rulemaking Act, to:

258                   (a) establish a mechanism for a large frontier developer or a member of the public to  
259                   report a safety incident; and

260                   (b) establish alternate compliance procedures if substantially equivalent or stricter  
261                   federal reporting requirements or guidance documents are established.

262                   (2) A large frontier developer shall report a safety incident to the office within 15 days after  
263                   the day on which the large frontier developer discovers the incident.

264                   (3) A large frontier developer that discovers a critical safety incident that poses an  
265                   imminent risk of death or serious physical injury, shall disclose that incident within 24

266 hours to a law enforcement agency or public safety agency with appropriate jurisdiction  
267 based on the nature of the incident.

268 (4) A large frontier developer shall submit to the office a report summarizing assessments  
269 of catastrophic risk resulting from internal use of the large frontier developer's frontier  
270 models:

271 (a) at least once every three months; or  
272 (b) pursuant to an alternate schedule if:  
273 (i) the large frontier developer requests the alternate schedule from the office in  
274 writing; and  
275 (ii) the office agrees to the alternate schedule.

276 (5) The office may transmit the reports described in Subsections (2) and (4) to the  
277 Legislature, the governor, the federal government, or appropriate state agencies, but may  
278 consider risks related to trade secrets, public safety, cybersecurity, or national security  
279 when transmitting reports.

280 (6) A report submitted under Subsection (2) or (4) may be classified as a protected record  
281 under Subsections 63G-2-305(1) and (2) if the requirements of Subsection  
282 63G-2-309(1)(a)(i) are met.

283 (7) On or before November 1, 2027, and annually thereafter, the office shall prepare a  
284 report for the Business and Labor Interim Committee that includes recommendations for  
285 modifying this chapter as well as anonymized, aggregated information about reports  
286 received pursuant to this chapter, without including information that would compromise  
287 the trade secrets or cybersecurity of a frontier developer, public safety, or the national  
288 security of the United States or that would be prohibited by any federal or state law.

289 Section 7. Section **13-72b-107** is enacted to read:

290 **13-72b-107 . Civil penalty.**

291 (1) A large frontier developer that violates this part is subject to a civil penalty that does not  
292 exceed:  
293 (a) for a first violation, \$1,000,000; or  
294 (b) for each subsequent violation, \$3,000,000.

295 (2) A civil penalty under this section may be recovered in a civil action brought by the  
296 attorney general on behalf of the office.

297 Section 8. Section **13-72b-108** is enacted to read:

298 **13-72b-108 . AI Transparency Enforcement Restricted Account -- Creation --**

299 **Deposits into account -- Distribution.**

300 (1) There is created within the General Fund a restricted account known as the "AI  
301 Transparency Enforcement Restricted Account."

302 (2) The account consists of:

303 (a) money collected by the attorney general from civil penalties, settlements, judgments,  
304 and other relief obtained in civil actions brought under Section 13-72b-107;

305 (b) appropriations made to the account by the Legislature; and

306 (c) interest and earnings on account money.

307 (3) The Division of Finance shall deposit money described in Subsection (2)(a) into the  
308 account.

309 (4) Upon appropriation by the Legislature, money in the account shall be distributed to the  
310 Office of the Attorney General for:

311 (a) investigations and enforcement of Part 1, Artificial Intelligence Transparency and  
312 Child Protection;

313 (b) attorney fees and litigation costs related to enforcement actions under this chapter;

314 (c) expert witnesses, consultants, and technical advisors with expertise in artificial  
315 intelligence safety and frontier models;

316 (d) specialized equipment, technology, and facilities necessary for enforcement activities;

317 (e) coordination with the Office of Artificial Intelligence Policy, federal agencies, and  
318 other state agencies; and

319 (f) other expenses related to the administration and enforcement of this chapter.

320 Section 9. Section **13-72b-109** is enacted to read:

321 **13-72b-109 . Transfer of frontier developer obligations -- Severability.**

322 (1) If any provision of this chapter or the application of any provision to any person or  
323 circumstance is held invalid by a final decision of a court of competent jurisdiction, the  
324 remainder of this chapter shall be given effect without the invalid provision or  
325 application.

326 (2) The provisions of this chapter are severable.

327 (3) The duties and obligations imposed by this chapter are cumulative with any other duties  
328 or obligations imposed under other law and shall not be construed to relieve any party  
329 from any duties or obligations imposed under other law and do not limit any rights or  
330 remedies under existing law.

331 Section 10. Section **13-72b-201** is enacted to read:

332 **Part 2. Public Safety and Child Protection Whistleblower Protections**

333 **13-72b-201 . Definitions.**

334       As used in this part:

335       (1) "Adverse action" means to discharge, threaten, harass, or otherwise discriminate against  
336       an employee in any manner that affects the employee's employment, including:

337       (a) compensation;

338       (b) terms;

339       (c) conditions;

340       (d) location;

341       (e) rights;

342       (f) immunities;

343       (g) promotions; or

344       (h) privileges.

345       (2) "Employee" means an individual who performs a service for wages or other  
346       remuneration under a contract of hire, written or oral, express or implied, for a frontier  
347       developer.

348       (3) "Reporter" means an individual who provides information relating to a violation in  
349       accordance with Section 13-72b-202.

350       Section 11. Section **13-72b-202** is enacted to read:

351       **13-72b-202 . Procedure for disclosure -- Internal whistleblower process.**

352       (1) To be a reporter for purposes of this part, an individual shall:

353       (a) reasonably believe that an act poses a specific and substantial threat to public health  
354       or safety or to the health or safety of a minor, or is a violation of Part 1, Artificial  
355       Intelligence Transparency and Child Protection; and

356       (b) provide information to the office:

357        (i) in writing; and

358        (ii) in accordance with procedures established by the office by rule made in  
359       accordance with Title 63G, Chapter 3, Utah Administrative Rulemaking Act.

360       (2)(a) Notwithstanding Title 63G, Chapter 2, Government Records Access and

361       Management Act, and except as provided in Subsection (2)(b), the office may not  
362       disclose information that could reasonably be expected to reveal the identity of a  
363       reporter.

364       (b) Subsection (2)(a) does not limit the office's ability to present evidence to a grand jury  
365       or share evidence with witnesses or defendants in an ongoing criminal investigation.

366       (3) A large frontier developer shall provide a reasonable internal process through which an  
367       employee may anonymously report information if the employee believes in good faith

368 that:

369 (a) the large frontier developer's activities pose a specific and substantial threat to public  
370 health or safety or to the health or safety of a minor; or

371 (b) the large frontier developer has violated Part 1, Artificial Intelligence Transparency  
372 and Child Protection.

373 (4) The process required by Subsection (3) shall include monthly updates to the reporting  
374 employee regarding the status of the investigation and actions taken in response to an  
375 anonymous report described in Subsection (3).

376 (5)(a) Except as provided in Subsection (5)(b), disclosures and responses under this  
377 section shall be shared with officers and directors of the large frontier developer at  
378 least once each quarter.

379 (b) If an employee alleges wrongdoing by an officer or director, Subsection (5)(a) does  
380 not apply with respect to that officer or director.

381 Section 12. Section **13-72b-203** is enacted to read:

382 **13-72b-203 . Reporter protected from adverse action -- Exceptions.**

383 (1) A frontier developer may not take adverse action against an employee because of a  
384 lawful act of the employee, or a person authorized to act on behalf of the employee, to:  
385 (a) provide information to the office in accordance with Section 13-72b-202, if the  
386 employee is a reporter;

387 (b) initiate, testify in, or assist in any investigation, judicial action, or administrative  
388 action based on or related to information provided to the office, if the employee is a  
389 reporter; or

390 (c) provide information through an internal reporting process established by the frontier  
391 developer.

392 (2) A frontier developer may not make, adopt, enforce, or enter into a rule, regulation,  
393 policy, or contract that would prevent an employee, or a person authorized on behalf of  
394 the employee, from taking any of the actions described in Subsection (1).

395 (3) An employee is not protected under this section if the employee:

396 (a) knowingly or recklessly makes a false, fictitious, or fraudulent statement or  
397 misrepresentation;

398 (b) uses a false writing or document knowing that, or with reckless disregard as to  
399 whether, the writing or document contains false, fictitious, or fraudulent information;  
400 or

401 (c) knows that, or has a reckless disregard as to whether, the disclosure is of information

402 that is false or frivolous.

403 (4) Information provided pursuant to this section may be classified as a protected record  
404 under Subsections 63G-2-305(1) and (2) if the requirements of Subsection  
405 63G-2-309(1)(a)(i) are met.

406 Section 13. Section **13-72b-204** is enacted to read:

407 **13-72b-204 . Remedies for employee bringing action.**

408 (1)(a) An employee who alleges a violation of Section 13-72b-203 may bring an action  
409 for injunctive relief, actual damages, or both, in a court with jurisdiction under Title  
410 78A, Judiciary and Judicial Administration.

411 (b) An employee may not bring an action under this section more than:  
412 (i) four years after the day on which the violation of Section 13-72b-203 occurs; or  
413 (ii) two years after the day on which facts material to the right of action are known or  
414 reasonably should be known by the employee.

415 (2) To prevail in an action under this section, an employee shall establish, by a  
416 preponderance of the evidence, that the employee suffered an adverse action because the  
417 employee, or a person acting on the employee's behalf, engaged or intended to engage in  
418 an activity protected under Section 13-72b-203.

419 (3) A court may award relief for an employee prevailing in an action under this section:  
420 (a) reinstatement with the same fringe benefits and seniority status that the individual  
421 would have had, but for the adverse action;  
422 (b) two times the amount of back pay otherwise owed to the individual, with interest;  
423 (c) compensation for litigation costs, expert witness fees, and reasonable attorney fees;  
424 (d) actual damages; or  
425 (e) any combination of the remedies listed in this Subsection (3).

426 (4)(a) An employer may file a counterclaim against an employee who files a civil action  
427 under this section seeking attorney fees and costs incurred by the employer related to  
428 the action and the counterclaim.

429 (b) The court may award an employer who files a counterclaim under Subsection (4)(a)  
430 attorney fees and costs if the court finds that:  
431 (i) there is no reasonable basis for the civil action filed by the employee; or  
432 (ii) the employee is not protected under Section 13-72b-203 because the employee  
433 engaged in an act described in Subsection 13-72b-203(3).

434 Section 14. **Effective Date.**

435 This bill takes effect on May 6, 2026.