

HOUSE BILL 823

S1

5lr1136
CF 5lr1137

By: **Delegates Wu, Qi, Bagnall, Hill, Mireku–North, Phillips, Ruth, Schindler, and Ziegler**

Introduced and read first time: January 29, 2025

Assigned to: Economic Matters

A BILL ENTITLED

1 AN ACT concerning

2 **Generative Artificial Intelligence – Training Data Transparency**

3 FOR the purpose of requiring a developer of a generative artificial intelligence system, on
4 or before a certain date and before the developer releases or substantially modifies a
5 certain generative artificial intelligence system, to publish on the developer’s website
6 documentation detailing the data used to train the generative artificial intelligence
7 system; and generally relating to training data transparency for generative artificial
8 intelligence.

9 BY repealing and reenacting, with amendments,
10 Article – State Finance and Procurement
11 Section 3.5–801
12 Annotated Code of Maryland
13 (2021 Replacement Volume and 2024 Supplement)

14 BY adding to
15 Article – State Finance and Procurement
16 Section 3.5–807
17 Annotated Code of Maryland
18 (2021 Replacement Volume and 2024 Supplement)

19 SECTION 1. BE IT ENACTED BY THE GENERAL ASSEMBLY OF MARYLAND,
20 That the Laws of Maryland read as follows:

21 **Article – State Finance and Procurement**

22 3.5–801.

23 (a) In this subtitle the following words have the meanings indicated.

EXPLANATION: CAPITALS INDICATE MATTER ADDED TO EXISTING LAW.

[Brackets] indicate matter deleted from existing law.



(b) (1) “Algorithmic decision system” means a computational process that facilitates decision making.

(2) “Algorithmic decision system” includes decisions derived from machines, statistics, facial recognition, and decisions on paper.

(c) “Artificial intelligence” means a machine–based system that:

(1) can, for a given set of human–defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments;

(2) uses machine and human–based inputs to perceive real and virtual environments and abstracts those perceptions into models through analysis in an automated manner; and

(3) uses model inference to formulate options for information or action.

(d) “GENERATIVE ARTIFICIAL INTELLIGENCE” MEANS ARTIFICIAL INTELLIGENCE THAT CAN GENERATE DERIVED SYNTHETIC CONTENT, SUCH AS TEXT, IMAGES, VIDEO, AND AUDIO, THAT EMULATES THE STRUCTURE AND CHARACTERISTICS OF THE DATA USED TO TRAIN THE ARTIFICIAL INTELLIGENCE.

(E) (1) “High–risk artificial intelligence” means artificial intelligence that is a risk to individuals or communities, as defined under regulations adopted by the Department in consultation with the Governor’s Artificial Intelligence Subcabinet.

(2) “High–risk artificial intelligence” includes rights–impacting artificial intelligence and safety–impacting artificial intelligence.

[(e)] (F) “Impact assessment” means an assessment of artificial intelligence systems made under regulations adopted by the Department in consultation with the Governor’s Artificial Intelligence Subcabinet and required under § 3.5–803 of this subtitle.

[(f)] (G) “Public senior higher education institution” means:

(1) the constituent institutions of the University System of Maryland and the University of Maryland Center for Environmental Science;

(2) Morgan State University; or

(3) St. Mary’s College of Maryland.

[(g)] (H) “Rights–impacting artificial intelligence” means artificial intelligence whose output serves as a basis for decision or action that is significantly likely to affect civil rights, civil liberties, equal opportunities, access to critical resources, or privacy.

1 **[(h)] (I)** “Safety–impacting artificial intelligence” means artificial intelligence
2 that has the potential to significantly impact the safety of human life, well–being, or critical
3 infrastructure.

4 **3.5–807.**

5 **(A) (1)** **IN THIS SECTION THE FOLLOWING WORDS HAVE THE MEANINGS**
6 **INDICATED.**

7 **(2)** **“DEVELOPER” MEANS A PERSON OR A UNIT OF STATE OR LOCAL**
8 **GOVERNMENT THAT DESIGNS, CODES, PRODUCES, OR SUBSTANTIALLY MODIFIES A**
9 **GENERATIVE ARTIFICIAL INTELLIGENCE SYSTEM.**

10 **(3)** **“SUBSTANTIALLY MODIFIES” MEANS TO RELEASE A NEW VERSION**
11 **OF OR UPDATE TO A GENERATIVE ARTIFICIAL INTELLIGENCE SYSTEM THAT**
12 **MATERIALLY CHANGES THE FUNCTIONALITY, PERFORMANCE, OR TRAINING DATA**
13 **OF A GENERATIVE ARTIFICIAL INTELLIGENCE SYSTEM.**

14 **(B) (1)** **THIS SECTION APPLIES TO A GENERATIVE ARTIFICIAL**
15 **INTELLIGENCE SYSTEM THAT WAS RELEASED ON OR AFTER JANUARY 1, 2022, FOR**
16 **USE BY THE GENERAL PUBLIC IN THE STATE, REGARDLESS OF WHETHER THE TERMS**
17 **OF THAT USE REQUIRE COMPENSATION.**

18 **(2)** **THIS SECTION DOES NOT APPLY TO A GENERATIVE ARTIFICIAL**
19 **INTELLIGENCE SYSTEM THAT IS EXCLUSIVELY MADE FOR:**

20 **(I)** **USE BY HOSPITAL MEDICAL STAFF;**

21 **(II)** **USE BY AN AFFILIATE OF THE DEVELOPER OF THE**
22 **GENERATIVE ARTIFICIAL INTELLIGENCE SYSTEM;**

23 **(III)** **ENSURING PHYSICAL SAFETY OF AN INDIVIDUAL;**

24 **(IV)** **PROTECTING CONFIDENTIAL PERSONAL INFORMATION;**

25 **(V)** **DETECTING, RESISTING, OR PROSECUTING MALICIOUS,**
26 **DECEPTIVE, FRAUDULENT, OR ILLEGAL ACTIONS;**

27 **(VI)** **THE OPERATION OF AIRCRAFT IN THE NATIONAL AIRSPACE;**
28 **OR**

29 **(VII)** **USE BY A FEDERAL ENTITY TO SUPPORT NATIONAL**
30 **SECURITY, MILITARY, OR OTHER DEFENSE PURPOSES.**

(C) ON OR BEFORE JANUARY 1, 2026, AND BEFORE EACH TIME THEREAFTER A DEVELOPER RELEASES OR SUBSTANTIALLY MODIFIES A GENERATIVE ARTIFICIAL INTELLIGENCE SYSTEM, THE DEVELOPER OF THE GENERATIVE ARTIFICIAL INTELLIGENCE SYSTEM SHALL POST ON THE WEBSITE OF THE DEVELOPER DOCUMENTATION DETAILING THE DATA AND DATASETS USED TO TRAIN THE GENERATIVE ARTIFICIAL INTELLIGENCE SYSTEM, INCLUDING:

(1) THE SOURCES OR OWNERS OF THE DATA;

(2) A DESCRIPTION OF HOW THE DATA FURTHERS THE INTENDED PURPOSE OF THE GENERATIVE ARTIFICIAL INTELLIGENCE SYSTEM;

(3) IF A DATASET USED FOR TRAINING IS:

(I) STATIC, THE NUMBER OF DATA POINTS INCLUDED WITHIN THE DATASET; OR

(II) DYNAMIC, AN ESTIMATE OF THE NUMBER OF DATA POINTS INCLUDED WITHIN THE DATASET;

(4) IF A DATASET:

(I) INCLUDES LABELS, THE TYPES OF LABELS USED IN THE DATASET; OR

(II) DOES NOT INCLUDE LABELS, THE GENERAL CHARACTERISTICS OF THE DATAPPOINTS WITHIN THE DATASET;

(5) WHETHER ANY DATA ARE IN THE PUBLIC DOMAIN OR ARE PROTECTED BY COPYRIGHT, TRADEMARK, OR PATENT;

(6) WHETHER THE DEVELOPER PURCHASED OR LICENSED THE DATA;

(7) WHETHER THE DATA INCLUDE ANY PERSONAL INFORMATION;

(8) WHETHER THE DATA INCLUDE AGGREGATE CONSUMER INFORMATION;

(9) WHETHER THE DEVELOPER CLEANED, PROCESSED, OR MODIFIED THAT DATA AND WHY THE DEVELOPER CLEANED, PROCESSED, OR MODIFIED THE DATA;

1 **(10) (I) THE TIME PERIOD DURING WHICH THE DATA WERE**
2 **COLLECTED; AND**

3 **(II) WHETHER THE DATA COLLECTION IS STILL ONGOING;**

4 **(11) WHEN THE DATA WERE FIRST USED DURING THE DEVELOPMENT**
5 **OF THE GENERATIVE ARTIFICIAL INTELLIGENCE SYSTEM; AND**

6 **(12) WHETHER THE GENERATIVE ARTIFICIAL INTELLIGENCE SYSTEM**
7 **USED OR USES SYNTHETIC DATA GENERATION.**

8 SECTION 2. AND BE IT FURTHER ENACTED, That this Act shall take effect
9 October 1, 2025.