stonetusker

Stonetusker Insights · Technical note on artificial intelligence, machine learning and data platforms

Can You Name the Data Behind Each Production Model? The European Union Dates Are Now Fixed

Regulation (EU) 2026/1744 fixed the dates for the high-risk rules of the European Union Artificial Intelligence Act. Much of what those rules ask for is model lineage. This note shows how to check it for one model in 30 minutes.

For: Teams that build or ship artificial intelligence features to European Union customers.

Skip if: Your product contains no artificial intelligence features.

By Subeesh Sivanandan, Founder, Stonetusker Systems · Updated 29 October 2026 · About 3 minutes to read · PDF version

Technical and engineering guidance. Please confirm legal obligations and the classification of your systems with your own counsel.

The short version

- Regulation (EU) 2026/1744 entered into force on 27 July 2026 and fixed the dates for high-risk systems: 2 December 2027 for Annex III uses and 2 August 2028 for regulated products under Annex I.[1,2,3]
- The rules ask for governed training data, up-to-date technical documentation and automatic logs kept for at least six months.[4]
- Most of this is model lineage: the record of which data, code and evaluation produced each model in production.
- Lineage costs little to record at training time and a great deal to reconstruct later. One model and five questions show where you stand.

2 December 2027
High-risk obligations apply to Annex III systems[2]

2 August 2028
High-risk obligations apply to Annex I systems[2]

6 months
Minimum retention for automatically generated logs[4]

A note to forward to your team

Model lineage check

The European Union has fixed the dates for high-risk artificial intelligence systems: 2 December 2027 for Annex III uses and 2 August 2028 for Annex I products. The rules ask for governed training data, technical documentation and logs kept for at least six months.

Please run the check in this note by [date] and reply with three numbers:
1. Models in production
2. Models we can trace to a data set version from stored records
3. Models whose inference logs record the model version

Classification reviewed with counsel: [yes or no]
Owner: [name]
Note: https://stonetusker.com/insights/eu-ai-act-model-lineage.html

Why it matters now

Regulation (EU) 2026/1744 was published in the Official Journal on 24 July 2026 and entered into force on 27 July 2026. It moved the high-risk dates and made them firm.[1,3] Annex III covers uses such as recruitment, credit scoring and life and health insurance pricing. Annex I covers regulated products such as medical devices.

Three groups of requirements fall on the engineering organisation.[4] Article 10 asks for training, validation and testing data under governance practices that cover data origin, labelling, cleaning and examination for possible biases. Article 11 and Annex IV ask for technical documentation that describes the training data, how it was obtained and selected, the testing metrics and the software versions. Articles 12 and 19 ask for automatic logs kept for at least six months.

Fines for most provider obligations reach 15 million euros or 3% of worldwide annual turnover, whichever is higher. For small and medium-sized enterprises, start-ups and small

mid-cap enterprises, the lower amount applies.[4] A nearer date also applies: generative systems placed on the market before 2 August 2026 must mark their output in a machine-readable format by 2 December 2026.[3,4]

Check one model in 30 minutes

Step 1. Choose the model and find what is deployed

Pick the model that serves the most traffic. If it runs on Kubernetes, list what is deployed:

```
kubectl get deployments --all-namespaces \
  -o 'custom-
columns=NAMESPACE:.metadata.namespace,NAME:.metadata.name,IMAGE:.spe
c.template.spec.containers[*].image'
```

Step 2. Answer five questions from stored records only

- Which model version serves production, and where is that recorded?
- Which training data set version produced it, and can you retrieve that exact snapshot today?
- Which code commit and configuration trained it?
- Which evaluation results justified its promotion, and who approved it?
- Do inference logs record the model version, and for how long are they kept?

Step 3. Count the answers

Write down how many of the five you answered from a registry, a log or a versioned store, without asking anyone to remember.

What the result means

What you find	What it means	Typical effort
Five answers from stored records	Lineage exists. The documentation can become an export from the pipeline	Days, to repeat the check on the other models
Four answers	The release pipeline records code and model, and the data side needs work	A few weeks of platform work
Three or fewer	The documentation would start as a reconstruction project	A pipeline change; about eight weeks for a first set of models

A model release moves three things: the code, the data and the trained weights, plus the evaluation behind the promotion. When only the code is under version control, every question from a customer or an auditor becomes a search through old notebooks and storage paths.

Recording lineage at training time, step by step

1. **List production models.** Record each model in production, the product it serves and its intended use, and agree the risk classification of each use with counsel.
2. **Version the data.** Store training, validation and testing data as immutable snapshots, with origin, labelling and cleaning steps recorded for each version.
3. **Register every model.** Use a model registry that links each model version to its training run, data set version and code commit.
4. **Gate promotion on evaluation.** Evaluate candidates on held-out data in the pipeline, store the results with the model version and block promotion below the agreed threshold.
5. **Log inference with the model version.** Add the model version to every prediction log and keep logs for at least six months for high-risk uses.

6. **Generate documentation from records.** Produce the technical documentation from the registry and data set metadata, so each new model version updates it.

The same records help outside the regulation: reproducing a regression, rolling back a model with confidence and explaining a drift incident to a customer.

What is still open

- The European Commission published draft guidelines on the classification of high-risk systems on 19 May 2026. Please base the classification of a specific system on the final guidelines and on advice from counsel.[5]
- Many systems are outside the high-risk categories. The transparency rules in Article 50 may still apply to generative and interactive systems.

Questions

When do the high-risk obligations apply?

From 2 December 2027 for systems listed in Annex III and from 2 August 2028 for artificial intelligence in regulated products under Annex I. Regulation (EU) 2026/1744, in force since 27 July 2026, set these dates.

Is our system high-risk?

It depends on its use. Annex III lists areas such as recruitment, credit scoring and life and health insurance pricing, and Annex I covers products such as medical devices. The Commission published draft classification guidelines in May 2026. Please confirm with counsel.

How long must logs be kept?

For a period appropriate to the intended purpose of the system, and at least six months.

Do smaller companies receive any relief?

Small and medium-sized enterprises, start-ups and small mid-cap enterprises may provide the technical documentation in a simplified form, and the lower of the two fine amounts applies to them.

Want a second opinion?

Run the check yourself first. If you would like a second opinion, reply with two counts only: the models you run in production and how many of them you can trace to a data set version. No model names, data, configuration or command output are needed. Within five working days you receive a short written review with the likely gaps and an order of work for closing them. It is free and confidential, and a call is optional.

Ask for a second opinion

The button opens an email to subeesh@stonetusker.com with the subject and the numbers to fill in. You can also use the contact page, or reply to the email or LinkedIn message that brought you here.

Next page for your situation

- Coding assistants and your delivery baseline: measure what coding assistants changed in your delivery.
- Checklist for artificial intelligence, machine learning and data teams: eight plain questions and the dates that matter.

Sources

1. Official Journal of the European Union: Regulation (EU) 2026/1744 (Digital Omnibus on artificial intelligence)
2. European Commission AI Act Service Desk: Article 113, entry into force and application
3. White and Case: EU AI Omnibus enters into force (4 August 2026)
4. European Commission AI Act Service Desk: AI Act Explorer, consolidated text as at 27 July 2026

5. European Commission: Draft guidelines on the classification of high-risk AI systems (19 May 2026)

About the author

Subeesh Sivanandan is the Founder and CEO of Stonetusker Systems. He has worked for more than 27 years in release engineering, DevOps, platform engineering & embedded Linux. His career includes serving as the Principal Engineer & Manager of Release Engineering at Stryker, alongside earlier roles at Nokia, VeriSign, and IP Infusion. LinkedIn profile.

Revision history (recent 3 updates)

Version 2.1, 29 October 2026. Every fact on this page is checked against a minimum of two sources each month.

Version	Date	Change
1.1	16 October 2026	Rewritten as an article for quick readers: introduction, summary of all major changes, what it means and first steps; one figure; commands folded for platform leads; print option and direct email offer
2.0	25 October 2026	Simpler layout and shorter text; the review offer now asks for a few counts only
2.1	29 October 2026	Who-this-is-for box, note to forward moved up, email offer with the numbers listed, and next pages