2024

trail

AI Governance

An actionable framework

About This Framework

This document is based on different academic frameworks for AI governance and was created to derive a list of actionable steps to quickly get started with AI governance. References include the AIGA Framework, the EU AI Act, the NIST AI Risk Management Framework and IEEE standards.

How to best use this framework:

Additionally **request the accompanying Figma board** to gain an overview of all tasks and steps assigned to the responsible stakeholders and lifecycle phases. Tailor it to your organization and highlight which elements are already in place and which are still missing or require internal restructuring.

You can also find our checklist *"10 steps to get you started on AI governance"* in the appendix.

Table Of Contents

AI Governance

AI governance is a set of rules, processes, frameworks, and tools within an organization to ensure that the use or development of AI aligns with the organizational principles, legal requirements, as well as social and ethical standards. These governance measures aim to define **accountabilities** and seek to increase **quality** as well as **transparency**. Ultimately, this fosters trust in AI systems, enabling the large-scale adoption of AI.

The 5 Stakeholders in AI Governance:

AI Team

Can consist of Data Engineers, Data Scientists, AI Architects or ML Engineers, who execute AI projects

Dev Team

Software Engineering team, including team lead, who embed AI algorithms in applications

AI Lead

Owns and manages products or projects, communicates AI efforts ot other stakeholders

Compliance

Regulatory stakeholder, responsible to oversee regulatory compliance and documents*

Customer Success

Contact partner for customers, who interact with customers and receive claims or incidents

The AI Lifecycle:

AI governance applies to the whole AI lifecycle - from design to monitoring - and requires a joint effort of the stakeholders named above. In the accompanying Figma file, you can quickly gain an overview of all tasks per lifecycle stage as described in the following chapters.

#1 Planning and Design

#2 Data Preparation

#3 Model Development

#4 Model Testing

#5 Deployment

#6 Operation and Monitoring

**possibly also domain knowledge stakeholder (e.g. Marketing lead for marketing forecasting algorithm)*

AI Policy

Before we start:

Every organization needs to have an ***AI Policy*** in place when developing, deploying, or using AI systems or tools. The policy sets out rules, values and guidance for all employees of the organization engaging with AI.
The ***AI Policy*** should be co-created between the legal team and the AI team and signed off by C-level management.

1. Choose a framework

There is no one size fits all when it comes to AI Policy. Every organizations needs to tailor the policy to the individual values, compliance standards and risk appetite.

We recommend starting with a framework and working through the points. For example, the ***PwC Framework*** for Key Elements, or this template by ***Vischer.*** A more technical example can be found by ***Anthropic***.

2. Publish AI Policy

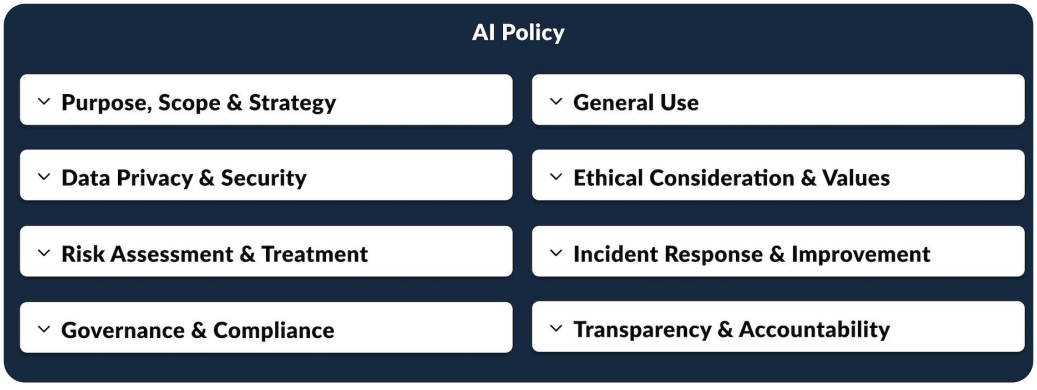
Make sure to publish the AI Policy in the organization and inform all employees with AI touchpoints. You could also think about releasing an externally-facing AI Policy to make your values and guidelines accessible to stakeholders outside the organization and to foster trust.

3. Train your employees

Make sure to train all employees that develop or use AI on your AI Policy. Keep track of all trained employees to ensure completeness. Keep the logs for potential audits. Make sure to re-train them in defined timeframes.

4. Keep up to date

Ensure you keep the policy up to date with relevant regulation and company standards. Monitor the AI use cases to make sure the policy is suitable to all new use cases.



How trail helps

trail's AI governance platform offers customizable ***AI Policy templates*** and according trainings for employees. Learn more ***here***.

AI Governance Of Procured Systems

If your organization is currently **deploying procured AI systems and tools only** instead of developing or finetuning own AI systems, you do not need to follow the extensive suggestions listed in this AI Governance Framework.

We aggregated all relevant measures for you on page 4-6 of this document.

Involved Stakeholders:

AI Team

Compliance

1. Set up an AI Policy and training

Design an organization-wide AI Policy and train your employees. Detailed information can be found on page 4 of this document.

2. Centralize AI administration

It is important to have one central responsible department in the organization (e. g. the IT department) to coordinate the release, use and control of AI systems and tools.

3. Implement an AI registry

To foster the centralization, aggregate all AI systems and tools in a registry (controlled by the responsible department from point 2) and make it accessible for all relevant stakeholders. Consult page 8 of this document for more information on the components of an AI registry.

4. Classify the risk category of the AI systems

Categorize your AI systems and use cases based on the risk categories in the EU AI Act (or other regulation) to identify your compliance requirements. This is also important if you are "just a deployer" and procured the AI system or tool.

5. Set up Usage Policies and training

Ensure you design according Usage Policies for each AI system or tool and foster awareness and compliance of these policies among the employees using the tool. Track if employees have understood these policies during their training prior first usage.

6. Monitor relevant regulation

7. Implement transparency measures

If you are deploying procured AI systems towards external end-users make sure to inform them they are interacting with an AI-based system. If the system is producing AI-generated content (e.g. a chatbot), make sure the content is labeled as such. This is, for instance, necessary under the EU AI Act.

AI Governance Of Procured Systems

In the EU AI Act:

Article 50 of the EU AI Act describes transparency obligations for providers and deployers of AI systems or GPAI models directly interacting with natural persons or generating content. Deployers shall ensure that users are informed about interacting with an AI system and (synthetic) outputs are marked as AI-generated in a machine-readable format.

trail helps you efficiently operationalize AI governance!

trail's AI governance platform offers a feature suite tailored to procured AI systems and tools. Centralize all AI systems in an AI registry, perform EU AI Act risk categorization, efficiently design your AI Policy and Usage Policies and assign roles and responsibilities within the organization. Orchestrate employee trainings and automate logging from one central hub.

Procure AI systems that supercharge your business and let trail take care of the governance!

Learn more *here* or contact *anna@trail-ml.com* directly.

Planning And Designing The AI System

Involved Stakeholders:

The first governance tasks begin with the kickoff of a new AI project. During this first planning phase, a ***Project Charter*** is created to centralize information about the AI project, goals and risks are aligned, and reporting responsibilities are assigned. Follow these initial steps:

1. Kickoff your project

Discuss the Project Charter to align goals and risks with all stakeholders. This also includes assigning reporting & documentation responsibilities.

2. Fill out Project Charter

3. Perform Risk Categorization

Categorize your AI use case based on the risk categories in the EU AI Act to identify compliance requirements.

4. Add system to AI registry

Each AI system or model needs to have assigned an unique ID, as well as a link to the source code and its data sources.

5. Set up development infrastructure

6. Check AI use-case against requirements

To uncover potential mismatches early on, immediately validate if organizational and legal requirements are (or can be) met.

💡 Project Charter

A document, also known as a “Model Card”, tailored to your organization that acts as the central source of information for your AI project. It includes various elements, such as project objective and scope but also lays down details on data, model or risks.

Planning And Designing The AI System

In the EU AI Act:

In this phase you will need to set-up a "Quality Management System" to ensure regulatory compliance next to keeping documentation. The documentation obligations in the AI Act lay down that system providers need to have a general as well as a detailed description of their AI system, consisting of e.g. intended use, system interactions or system design (see Article 11 and Annex IV in the AI Act). The Project Charter with its elements outlined below, can help you in providing this documentation. The documentation obligations vary based on the risk class of your AI system. Learn more about the EU AI Act in the **appendix**.*

The ***Project Charter*** shall include:

- Project background & vision;
- business value & impact;
- project objective & scope;
- ethical considerations;
- roles & responsibilities;
- system risks;
- transparency, explainability & contestability assessment;
- milestones;
- operating environment;
- information on data and model.

What is an AI Registry?

An overview of all AI systems developed and used in your organization. It references the repsonible stakeholder, business unit, system ID and risk class. It enables an holistic overview and tracking of all AI systems and their use cases in the organization.

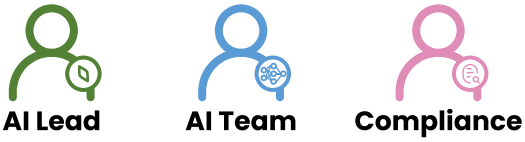
How trail helps

trail's AI governance platform offers customizable ***Project Charters,*** automated AI registry structures and risk classifications for each AI project or model. Learn more ***here***.

The requirements and obligations laid down by the EU AI Act and as mentioned in this framework have to be fulfilled mainly by organizations who provide high-risk AI systems or general purpose AI models. Please refer to the **latest EU AI Act version or the **appendix** for more details.*

Data Preparation

Involved Stakeholders:



A lot of work and time of AI projects is spent on data gathering, cleaning, labelling, and analysis. Data has the biggest impact on bias and therefore bias assessment efforts should be taken and documented among other data quality considerations.

1. Document data sources

Document them **for each dataset** (split by test, training, and validation dataset). Linking dataset iterations to model development allows you to trace back any changes in model performance due to changes in data. Use **dataset versioning tools** to facilitate the tracking, e.g. ***DVC***.

2. Document data rights

3. Define and document chosen data quality and bias metrics

Set up holistic catalogue with metrics you can choose from in each project, e.g. accuracy, completeness, consistency, timeliness, reliability, usability, etc.

4. Perform bias assessments

This helps you to understand if there are any unintended biases within the data distribution. Find a summary of relevant frameworks, tools and tests ***here***.

To interpret model predictions we can recommend these tools or frameworks: ***SHAP models***, ***Alibi***, ***LIME*** or ***TreeInterpreters***.

Or look into these bias and fairness tools: ***Ydata***, ***Aequitas***, ***Microsoft Fair Learn***.

5. Continuously assess and document selected metrics

Both during model development and before deployment. Make sure that you also abstract and explain the outcomes for less technical stakeholders on the project!

6. Check for data privacy concerns

7. Link results of all steps to the Project Charter to centralize information

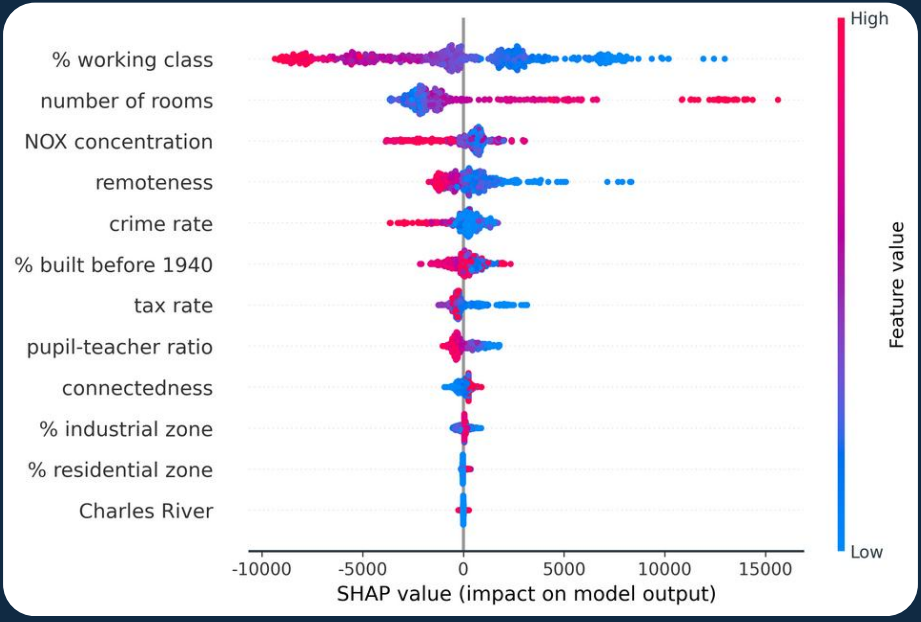
Data Preparation

In the EU AI Act:

Training, validation, and testing of data sets are subject to data governance and management practices under the AI Act. System providers will need to ensure that these practices are appropriate and relevant to the intended purpose, e.g. detecting biases which likely affect the safety and health of persons or forming assumptions about the data representation. All processes need to be documented (see Article 10 and Annex IV in the Act).

SHAP Values

SHAP or Shapley values can inform you about how important each input is to a machine learning model for its prediction. This method is commonly used to explain AI models. We can recommend reading this ***guide for non-technical stakeholders***.



Example for SHAP values

How trail helps

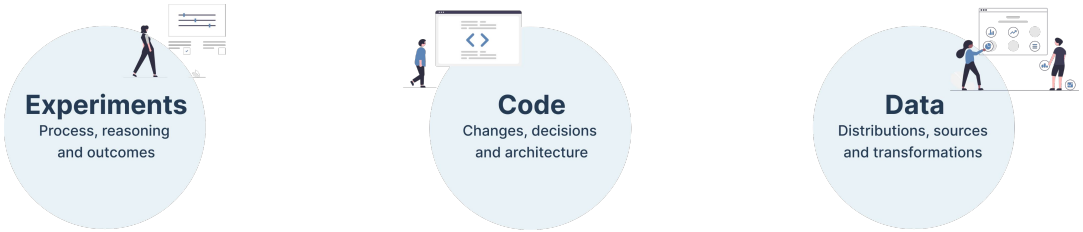
trail's AI governance platform links data sets to models and allows to trace back any changes made during experimentation (i.e. in data, model, or code). It also offers solutions for data analysis (i.e. continuous data statistics and bias assessment). Learn more ***here***.

Model Development

Involved Stakeholders:

AI Lead AI Team Compliance

During model development it is important to keep track of all experiments to ensure traceability. The main parts of technical documentation originate during this phase and close collaboration with compliance stakeholders is necessary to ensure conform model development.



Contents of technical documentation

1. Set thresholds for performance metrics and document them

2. Define & document bias metrics for model development

You can refer to the previously mentioned frameworks and tools under ***Data Preparation***.

3. Add each new model to AI registry under the applicable system

4. Document development

Use version control to keep track of code changes, e.g. ***Git***. Document technical decisions within code for a more comprehensive codebase, as well as architecture changes within the Project Charter.

5. Document ML experimentation

Use experiment tracking tools, such as ***MLflow***, to keep track of all iterations. Include any decisions you made during experimentation, as well as the rationale behind the runs.

6. Test against previously defined quality and bias measures and document them

7. Ensure the design choices comply with regulation or policies

Evaluate data and model choices regarding regulatory landscape and internal policies and document your considerations. **Tip**: Map your regulatory environment and any changes to it for all AI use cases in your organization and refer to this document to avoid compliance issues early on.

Model Development

In the EU AI Act:

As a high-risk AI system provider you will need to specify techniques, procedures and systematic actions to be used for the development, quality control and quality assurance within your quality management system (Article 17 in the AI Act). This also includes any standards that are applied.
Next to that, you are required to keep records of any events to ensure traceability, be transparent about the characteristics, capabilities and limitations of the performance of the AI system (e.g. defining metrics or known and foreseeable risks), use appropriate human-machine interface tools such that the AI system can be effectively overseen by a human and ensure consistent model performance using appropriate benchmarks (Article 12-15). This also means that you'll have a risk management system in place that runs continuously during the whole model lifecycle (Article 9).
Again, all of these processes need to be documented (see Article 11 and Annex IV in the Act).

How trail helps

trail's AI governance platform lets you set thresholds for chosen metrics and automatically tracks them for each experiment or data point. Save time on writing technical documentation with its automated documentation features deriving insights from code, experiments and data. Learn more ***here***.

Model Testing

Involved Stakeholders:

AI Lead

AI Team

In this phase you want to ensure your model is working as it is supposed to and meets your objectives. A simulation environment is set up, and the model is tested on representative datasets. Extreme tests are performed for subpopulations, and error analysis is conducted to improve the model.

1. Build simulation environment for testing

2. Conduct tests

Test on separate validation dataset, representative for target population (measure relevant metrics such as accuracy, precision, recall, F1-score, etc.). Minimize gaps in model predictions of training and testing. Further, conduct extreme tests for subpopulations and perform evaluations (e.g. cross-validation or bootstrapping). Document test outcomes, shortcomings and limitations.

3. Analyse errors

Examine edge cases (also look at cases that differ from human intuition). Identify patterns or common reasons for potential misclassifications and incorporate them for improvement.

In the EU AI Act:

Similar to the previous stage, you'll need to examine, test and validate procedures to be carried out after the development to ensure effective quality and risk management. Test your model for risks and as appropriate at any time during the development process, but latest prior deployment, under real world conditions. For that, define relevant metrics and probabilistic thresholds prior testing (Article 9 in the AI Act). Again, all of these processes need to be documented (see Article 11 and Annex IV in the Act) and test logs and reports need to be signed by the responsible person.

How trail helps

trail's AI governance platform integrates various tests required by AI certification bodies to facilitate audits. Our copilot suggests you the suitable state-of-the-art test for your use case. Learn more ***here***.

Deployment Decision

Involved Stakeholders:

This phase decides about the deployment of the AI model. The Project Charter and all documentation is revised, and the model's impact and bias assessments are documented. Regulatory compliance is verified, stakeholders are onboarded, and user transparency is implemented.

1. Revise and update Project Charter

Update Project Charter by ensuring that the latest changes to model and data version are included. Complete documentation on impact and bias assessments.

2. Revise model and data documentation

3. Set frequency and thresholds for health checks

Regularly reassess models in production for changes, drifts, updates, biases, etc. Make sure to assign the responsibilities to the relevant stakeholders.

4. Check and document regulatory compliance of final AI system

5. Prepare Usage Policies

Document Usage Policies for internal use cases and link them in the Project Charter and AI registry. If you deploy the AI system to external stakeholders make sure to provide transparent usage guidelines.

6. Onboard stakeholders

Plan an onboarding session with customer-facing and domain stakeholders to explain the new product or AI use-case. Prepare an easy-to-read, high-level overview for customers or non-technical stakeholders to ensure everyone understands the capabilities and limitations of the AI system.

7. Ensure transparency within user interface

Integrate the possibility for the user to easily report incidents within the application or interface. Establish an internal protocol for user interactions and incident reports. Provide insights into system explainability to users within that interface (i.e. highlight decisions made by an AI and explain why a certain outcome was derived).

Deployment Decision

In the EU AI Act:

Prior to deployment, make sure that you meet all applicable obligations under the EU AI Act, undergoing a conformity and a fundamental rights impact assessment (Article 16 and 27). For limited risk use-cases (i.e. AI-generated content or where a user interacts with the AI system) this also means to include the transparency information as outlined before or as seen in the example below to the user or to your employees (Article 50). Validate that you meet the requirements of the technical documentation and the quality management system. If necessary, i.e. if you have a high-risk AI system, register your use-case in the public database of the EU (Article 49).

An example how users can be informed about the AI's outcomes

How trail helps

trail's AI governance platform keeps any documentation automatically up to date, allowing for instant reports in different abstraction levels (tailored to technical or non-technical stakeholders and auditors). The integrated governance workflow ensures that all requirements are met before deployment. Learn more *here*.

Operation & Monitoring

Involved Stakeholders:

After the AI model is put in production it is crucial to monitor the systems to ensure governance. Regular health checks are additionally conducted to assess the model's performance and compliance. Documentation of requirements needs to be kept up to date and ready for audits by third party stakeholders.

1. Set responsibilities for monitoring the deployed AI system

2. Implement monitoring tool

You can look into ***Fiddler AI***, ***Mona*** or ***Evidently AI***, for instance.

3. Conduct regular health checks to identify potential issues early

Such health checks should observe the system for changes, drifts, biases, updates, etc.

4. Monitor regulatory landscape for changes

Keep track of the quickly evolving regulatory landscape and assess which obligations you may have to adhere to, as mentioned under ***Model Development***.

In the EU AI Act:

When making your AI system accessible to the EU market, you will need to ensure that your AI system behaves as it should during operation. If there are any significant changes to the system, incidents or if you believe your system is not compliant to the AI Act any more, you are required to take corrective actions, which need to be reported to both the public authorities as well as any other affected stakeholders, such as a deployer of your system (Article 20, 24, 82).

AI Governance Made Simple

AI governance sounds complex to you? Are you looking for more guidance?

We developed **trail** to make operationalizing AI governance easy and to enable trustworthy AI systems. Learn more *here*.

Our *AI Governance Copilot* supports you with AI Policy, AI registry, audit trails, automated documentation and complete AI governance workflows to increase transparency and quality, mitigate risks and ensure compliance with relevant regulation. All in one central place.

- **Easy and secure integration into your development environment**
- **Automated analysis of code and metadata to generate technical documentation for internal and external use**
- **Complete AI governance workflows tailored to your processes**
- **Saves time on compliance, documentation and collaboration**

Please **reach out** if you want **to discuss your individual governance needs** or if you have **feedback regarding this framework** to *anna@trail-ml.com*. We are looking forward to your messages!

Appendix

A 10-step checklist to get started, and an EU AI Act factsheet.

Checklist: 10 Steps To Get You Started On AI Governance

- [ ] Read through this document carefully and tailor the accompanying Figma process to your organization (request it *here*)
- [ ] Set up a central place to collect all important AI governance information, such as AI models deployed, documentation and supporting information (e.g. SharePoint folder, Notion, Confluence)
- [ ] Draft a general Project Charter for your organization
- [ ] Set up an AI registry and AI policy for your organization
- [ ] Assign a responsible person for each role
- [ ] Familiarize yourself with MLOps tools suggested in this framework and choose suitable ones (data versioning, experiment tracking, monitoring)
- [ ] Agree upon supporting tools used during the governance process (i.e. wiki/documentation, communication, project management)
- [ ] Set up an AI regulatory compass for your organizations' use cases to make sure you are aware of all relevant regulation
- [ ] Prepare a catalogue of data and model performance as well as bias metrics to choose from for future AI projects
- [ ] Decide on base quality requirements for documentation (see *trail* for automation)

Use trail for an all-in-one solution with automated AI governance workflows and reporting structures.

The EU AI Act

The **EU AI Act** is the world's first comprehensive AI law, regulating AI on a risk-based approach. Applicable to anyone providing internal or external AI systems in the EU.

Unacceptable Risk

Systems that are harmful, e.g. emotion recognition at the workplace or systems manipulating behavior and decision-making.

High Risk

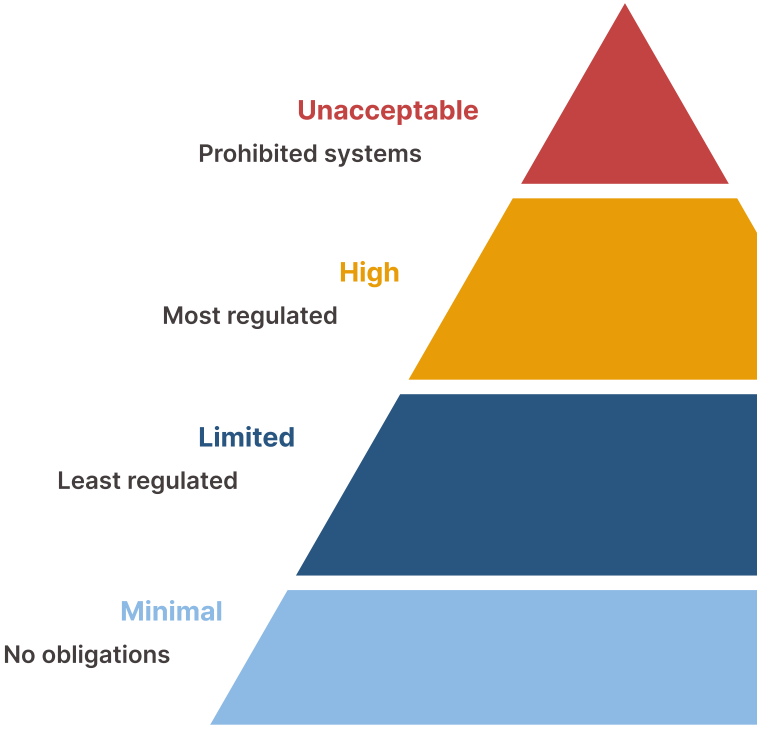
Systems that could be harmful if misused, perform profiling of persons, significantly influence decisions or that act as safety components, e.g. biometric identification, operation of critical infrastructure, recruiting tools, AI in medical devices or vehicles.

Limited Risk

Systems generating content or that interact with humans, e.g. customer service chatbot or an image creator.

Minimal Risk

All other systems, e.g. a spam filter.



What about GPAI?

General Purpose AI models will be classified as either non-systemic or systemic risk (i.e. those with high impact capabilities, i.e. computing power use for training of above 10^25 FLOPS or based on decision of European Commission using other factors, e.g. availability to 10.000 business users in the EU).

Obligations for high risk AI system providers include:

A quality management system (incl. transparent development procedures, risk management system, post-market monitoring system, accountability framework), extensive documentation of design, development and quality management, keeping auditable logs, fundamental rights impact assessment (How will the AI system be used? Who will be affected in what way?), conformity assessment and registration in public database.

Read about the EU AI Act and all obligations here.

trail

Contact us

hello@trail-ml.com

www.trail-ml.com