GETTING READY FOR TRUSTWORTHY AI REGULATION

HOW TO ANTICIPATE THE EU REGULATORY FRAMEWORK FOR AI
As the market is experiencing exponential digital intelligence advancements with AI systems becoming more complex, the European Union seeks to promote an unprecedented legislative framework that will provide a solid legal basis and a coordinated European approach for designing and developing trustworthy human-centered AI systems.

On the 21st of April 2021, the European Commission published the long-awaited regulation proposal, known as the AI Act, governing the use of AI systems regarding their level of risk (prohibited uses, high-risk, limited-risk, and low risk). The proposal establishes horizontal proportional requirements that will ensure the proper development of trustworthy AI systems.

This regulation has the ambition to harmonize the existing EU laws to facilitate investment and innovation in AI while protecting the individuals’ fundamental rights and principles on which the EU is founded when developing AI systems.

TO WHOM IS THIS REGULATION APPLICABLE?

The AI Act applies to everyone who wants to develop, distribute and import AI systems in the EU market or to systems that affect users located in the EU territory.
Today’s Challenges

The AI Act identifies eight different mandatory requirements to design and develop trustworthy High-Risk AI systems:

- Risk Management (Art. 9)
- Data and Data Governance (Art. 10)
- Record-keeping (Art. 12)
- Transparency and Provision of Information to Users (Art. 13)
- Human Oversight (Art. 14)
- Accuracy, Robustness and Cybersecurity (Art. 15)
- Technical Documentation (Art. 11)
- Conformity Assessment (Art. 19)

Approaching all the AI Act’s requirements can be overwhelming, as each one of them presents different complexities and require unique approaches to comply with the regulation successfully.

The Regulation also encourages providers of non-high-risk AI systems to implement a code of conduct to voluntarily apply the mandatory requirements for high-risk AI systems.

AI REGULATION ALIGNMENT, TO BE A LEADER OR A FOLLOWER?

At everis, we are aware of the dilemma an organization faces in overcoming these challenges, especially when adapting the AI systems’ design and development processes to the new regulatory provisions.

For that reason, we look forward to supporting organizations to start ahead at defining the Trustworthy AI journey and transform the challenges into a new competitive advantage by harnessing our Responsible AI-driven solutions and methodologies with an end-to-end approach aligned with the provisions of the new AI regulation.

Moreover, organizations will benefit from boosting their ongoing AI performance, maximizing and scaling the value of their AI systems, spreading the best practices on AI across the company, and positioning as a market leader on state-of-the-art technologies.
In order to develop a trustworthy AI system free of errors and perils, the regulation pays special attention to the proper management of risks and biases throughout the entire AI system’s lifecycle, from the birth of the AI-driven initiative until the monitoring phase. Thus, according to this provision, organizations and providers of high-risk AI systems that look forward to remaining efficient and risk-free must leverage a Risk Management System, which must include the following:

**01 Identification and analysis** of the known and foreseeable risks associated with each high-risk AI system

**02 Evaluation of other possibly arising risks** based on the analysis of data gathered from the post-market monitoring system

**03 Estimation and evaluation** of the risks when the high-risk AI system is used in accordance with its intended purpose

**04 Adoption of suitable risk management measures**, such as:
- Elimination or reduction of risks through design and development
- Implementation of adequate mitigation and control measures
- Implementation of transparency measures

Successfully deploying AI compliance initiatives requires AI stakeholders’ strategic alignment to develop an end-to-end AI risk management system. As a starting point, designing an AI risk classification should serve as the backbone for the compliance strategy, defining the mechanisms to be implemented across the AI lifecycle.

To support our clients, we’ve structured a Risk Management Methodology embedded in our AI Governance Practice to identify, evaluate, mitigate existing and future-prone risks and biases.

Operating a risk management system on an ongoing basis not only requires deploying specific control mechanisms and documentation but also should serve as a tool that is further iterated throughout the design, development, and monitoring of AI systems.
Data and Data Governance

REQUIREMENT BREAKDOWN

The new AI regulation emphasizes a relevant aspect for building trustworthy AI models with reliable outcomes: Data and Data Governance.

This provision defines the elements and characteristics to be considered for achieving high-quality data when creating your training and testing sets.

Additionally, it demands organizations to deploy a responsible data governance that oversees the end-to-end data lifecycle to ensure risk-free and trustworthy data sets to build resilient AI models and with reliable solutions.

THE FOUR KEY CHARACTERISTICS OF DATA & DATA GOVERNANCE:

- **RELEVANT**
  Relevant, representative, free of errors and complete data.

- **SUITABLE**
  Integrate appropriate statistical properties regarding the individuals on which the HRAIS* is intended to be used.

- **INCLUSIVE**
  Include the specific geographical, behavioral or functional characteristics or elements within which the HRAIS is intended to be used.

- **COMPLIANT**
  Comply with the rest of the EU laws (such as GDPR) to appropriate safeguard the individuals’ fundamental rights and freedoms.

*HRAIS: High Risk AI System

DATA TRANSFORMATION FRAMEWORK FOR AI

Define the standards, techniques and documentation needs to be incorporated during AI development.

The Feature store serves as a central repository for Data Scientist to: discover, define, store, share and reuse features.

**FEATURE STORE**

**FEATURE TRANSFORMATION**

**DATA PREPROCESSING**

FEATURE STORE

Prevents from:
- Inability to reproduce results
- Redundant features from multiple AI teams
- Inadequate features transformation
- Lower AI teams productivity related to high data transformation workload

Fosters:
- Centralized access to data for AI applications
- Stable Data Pipeline for feature serving
- Single source of truth for features definition
- Data lineage preservation

This framework should also be complemented with Fairness & XAI methodologies, to be incorporated during the development process favoring data representativeness as well as supporting model’s interpretability.
As the AI regulation states, it is paramount that in the design and development stages of High-risk AI systems, organizations include capabilities for enabling the automatic recording of events and activities (‘logs’) while the AI system is operating.

Facilitating audit trails for AI systems requires dedicated toolchain and processes to be integrated within the AI architecture platform, enabling the recording and monitoring of all the events generated by the model throughout its lifecycle.

In order to deploy efficient log management, KPIs need to be defined at an early stage to create actionable notifications and alerts based on these metrics.

Within our AI architecture practice, we recommend implementing the centralization of logs and its management within the Audits & Alert module of the MLOps Architecture, facilitating its automated end-to-end tracking.
Transparency and Provisions of Information to Users

REQUIREMENT BREAKDOWN

The Regulation looks forward to guaranteeing that the AI system’s operations are sufficiently transparent and explainable for both developers and users to interpret the system’s output and use it appropriately. Furthermore, to attain higher transparency levels, high-risk AI systems must attach complete instructions or manuals of use with all the relevant information to ensure the appropriate usage of the AI system.

AI SYSTEMS INSTRUCTIONS FOR USE

WHAT TO INCLUDE?

- Identity and the contact details of the provider
- Characteristics, capabilities and limitations of performance
- Description of the changes & performance
- Description of the human oversight measures
- Expected lifetime of the system
- Description of maintenance measures to ensure the proper functioning

KEY STEPS TOWARDS COMPLIANCE

Documentation about AI systems should not only target AI teams but also address the specific needs related to the interpretability by AI systems users.

This approach to transparency encompasses both explainability techniques (global & local) as well as a methodology to deliver the AI systems “instructions”. These should be systematic and structured on a document to be applied to all AI systems developments.

As an everis AI development best practice, we’ve have defined a “model development manual”, an internal documentation standard to be applied across all AI Lab experiments, boosting efficiencies and promoting reproducibilities of AI initiatives.
Human Oversight

REQUIREMENT BREAKDOWN
With the advances in technology, AI systems’ decision-making processes are becoming increasingly autonomous, leading to little to no human intervention required along its operating lifetime.

However, the regulation urges to keep humans in the loop, not as an alternative, but as a necessity to oversee the system’s operations, verify its outcomes, and intervene when necessary.

As a result, AI systems’ design, development and operations must avoid causing harm to the health and safety of people, minimizing the potential risks throughout its entire lifecycle.

The Human Oversight approach to regulation:
According to the AI Regulation, the proper implementation of Human Oversight measures should lead to:
- Fully understand the capacities, limitations, and risks of the AI system
- Correctly interpret the high-risk AI system’s output
- Comprehend whether to use the AI system
- Wisely interrupt the system through a “stop” button

KEY STEPS TOWARDS COMPLIANCE
Defining an AI Accountability Framework is recommended for organizations to clearly identify stakeholders’ roles and responsibilities and assign adequate human-machine interface tools since the birth of the initiative.

Such structure should be scalable by nature and designed to keep in step with gains in maturity in the overall AI Teams structure.

AI ACCOUNTABILITY FRAMEWORK
The new AI regulation looks forward to building resilient AI models that are protected against risks, inconsistencies and external vulnerabilities when running operations.

For that reason, it demands organizations and providers of high-risk AI systems to establish a homogeneous threshold by defining a set of measures and KPIs regarding accuracy, robustness and cybersecurity:

**Accuracy, Robustness and Cybersecurity**

**REQUIREMENT BREAKDOWN**

The new AI regulation looks forward to building resilient AI models that are protected against risks, inconsistencies and external vulnerabilities when running operations.

For that reason, it demands organizations and providers of high-risk AI systems to establish a homogeneous threshold by defining a set of measures and KPIs regarding accuracy, robustness and cybersecurity:

**TRUSTWORTHY AI MODELS**

**ACCURACY**
Specify the relevant accuracy metrics used to overcome errors, faults, or inconsistencies that may occur across the AI Model.

**ROBUSTNESS**
Place technical redundancy solutions:
- Backup plans
- Fail-safe plans
Place appropriate mitigation measures to avoid biased outputs due to feedback loops, that is, inaccurate outputs used as an input for future operations.

**CYBERSECURITY**
Place technical solutions to address vulnerabilities:
Measures to prevent and control for attacks trying to manipulate the system output (‘data poisoning’, adversarial examples).

**AI-TO-PRODUCTION VALIDATION TOOLS & METHODOLOGIES**

**ACCURACY**
Ensuring performance of the models deployed to prevent models staleness and biased outputs.

**ROBUSTNESS**
Monitoring health of the systems where the models are deployed prepared with technical redundancy mechanisms.

**CYBERSECURITY**
Guaranteeing security of the AI systems addressing models vulnerabilities in regards to potential manipulation or attacks.

To reinforce quality and security standards for AI systems, organizations seeking to scale AI across production environments need to align guarantee AI lifecycle quality with **AI-to-Production Validation Tools and Methodologies** that ensure a safe and robust AI lifecycle management.

An AI-to-Production standard integrates customized metrics and processes to define a homogeneous compliance threshold. This standard also involves solutions to prevent external threats and tools for enabling reproducibility.
To demonstrate high-risk AI system compliance with the mandatory requirements, the Regulation proposes to document the technical processes involved throughout the end-to-end AI systems lifecycle.

For that reason, organizations must place the technical documentation as a compass when designing and developing High-risk AI systems, supporting them to:
- Remain compliant with the mandatory requirements
- Ensure a homogeneous methodology for the proper documentation of all AI systems across the organization
- Facilitate traceability, monitoring & auditability

A detailed description of the HRAIS
A detailed description of the elements of the AI system and of the process for its development
Detailed information about the monitoring, functioning and control of the AI system
A detailed description of the risk management system

A general description of the system through its lifecycle
A list of the harmonized standards applied
A copy of the EU declaration of conformity
A description of the system performance in the post-market phase, including the post-market monitoring plan

A description of any change made to the system

A centralised model registry pursues three main objectives at an organizational level:
- AI Democratization: Model registries facilitate sharing of expertise and knowledge across teams by making AI models more discoverable.
- Agile Innovation: Generates efficiency through models reusability.
- Enhance Traceability: Provides full visibility and enables governance by keeping track of each model’s activity centralizing insights about “how, when, and why”.

**HRAIS: High Risk AI System**
Organizations and providers of high-risk AI systems must verify that their AI systems successfully meet all provisions and standards, not only prior to putting it into service, but also after being distributed in the marketplace.

Thus, the Conformity Assessment is a written verification process of compliance that agrees and ensures with all the following:

**KEY STEPS TOWARDS COMPLIANCE**

As established within the AI regulation proposal, aligning compliance requirements with the defined risk management system should leverage internal control mechanisms to ensure ongoing compliance.

Also, developing an internal audit procedure translates higher transparency and consistency to the internal verification methodology.
A Guide to start
DEFINING THE TRUSTWORTHY AI JOURNEY

FIRST STEPS TOWARDS A TRUSTWORTHY AI JOURNEY

HIGH-RISK AI SYSTEMS
The methodology seeks progressive compliance with upcoming regulation.

NON-HIGH-RISK AI SYSTEMS
The methodology seeks a voluntary alignment through a code of conduct and self-regulation.

AI Compliance Assessment
In depth-evaluation of compliance requirements with high-risk AI system standards.

AI Compliance Action Plan
Define the compliance roadmap and action plan of initiatives to align technical needs and capabilities.

Compliant Organization & Culture
Identify and align compliance responsibilities within the organization, fostering internal culture and change management to promote adoption of the Trustworthy AI vision.

REGULATORY REQUIREMENTS WRAP UP

Risk Management
Risk Classification
Accountability Framework
Human Oversight
Data and Data Governance
Model Registry
Technical Documentation
Record-keeping
Audit Procedure
Verification of Conformity
Data Transformation Framework for AI
AI-to-Production Validation
Accuracy, Robustness and Cybersecurity
Model Development Manual
Transparency and Provision of Information to Users
GETTING READY FOR TRUSTWORTHY AI REGULATION

About everis NTT Data

everis is part of NTT Data, named a Challenger by Gartner in its 2020 Magic Quadrant for Data and Analytics Service Providers Worldwide.

The company shares the Innovation DNA as part of NTT Group, accelerates open ecosystems and contributes to fostering Responsible AI across its operations.

As a trusted global innovator, our values come from “consistent belief” to shape the future society with clients and “courage to change” the world with innovative digital technologies.

For more information

To find out more about how everis NTT Data can help your business: https://www.everis.com/global/en/trustworthy-ai-regulation