High-quality standards and good governance tools for compliance with the AI Act

eu-LISA Industry Roundtable
Budapest, 12-13 Nov 2024

Dr Jochen Friedrich
Technical Relations Executive

IBM

Overview

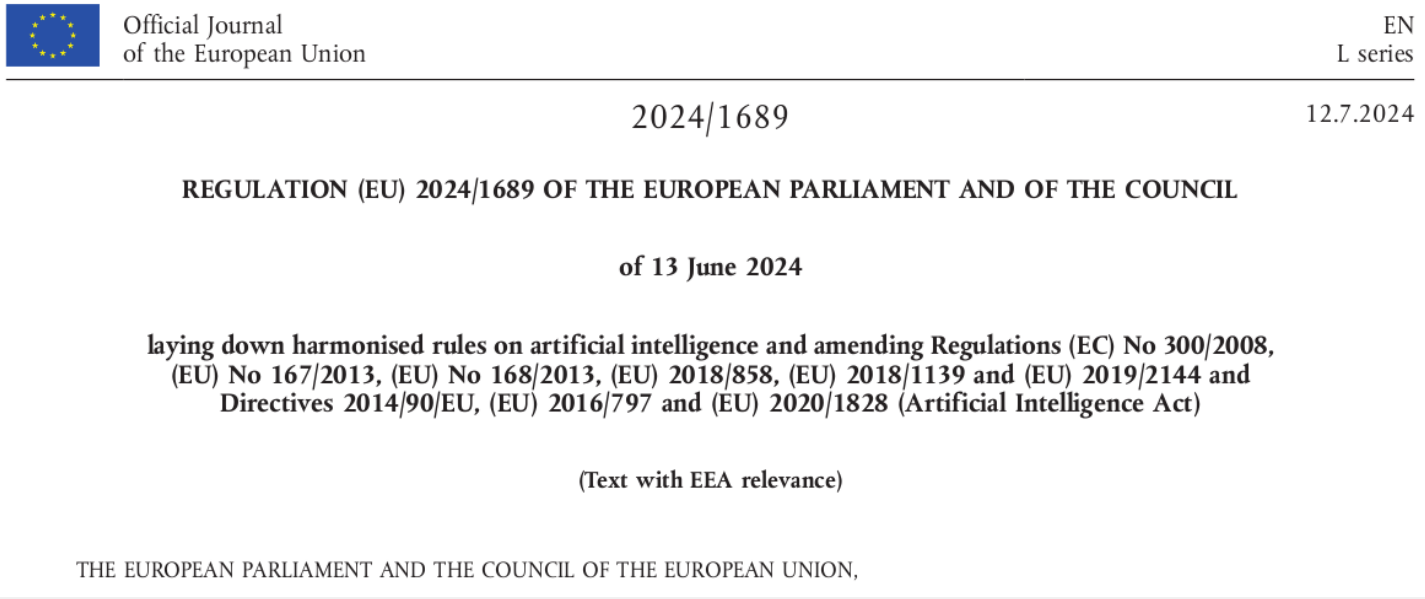
1 The AI Act – a typical technical regulation

2 Harmonised European standards for compliance with the AI Act

3 AI Governance Tools – watsonx.governance

4 Open Source AI – IBM Granite

The AI Act – Rules for trustworthy technology in operation in the EU



Official Journal of the European Union

EN
L series

2024/1689

12.7.2024

REGULATION (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

of 13 June 2024

laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)

(Text with EEA relevance)

THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION,

Scope

Promote the uptake of human centric and trustworthy artificial intelligence, while ensuring a high level of protection of health, safety, fundamental rights

Harmonised rules for the placing on the market, the putting into service and the use of artificial intelligence systems (‘AI systems’)

Prohibitions of certain artificial intelligence practices;

Specific requirements for high-risk AI systems and obligations for operators of such systems;

Harmonised transparency rules for certain AI systems;

Harmonised rules for the placing on the market of general-purpose AI models.

In force since 1 August 2024.

Transition time for compliance started.

AI Act – a typical EU technical regulation

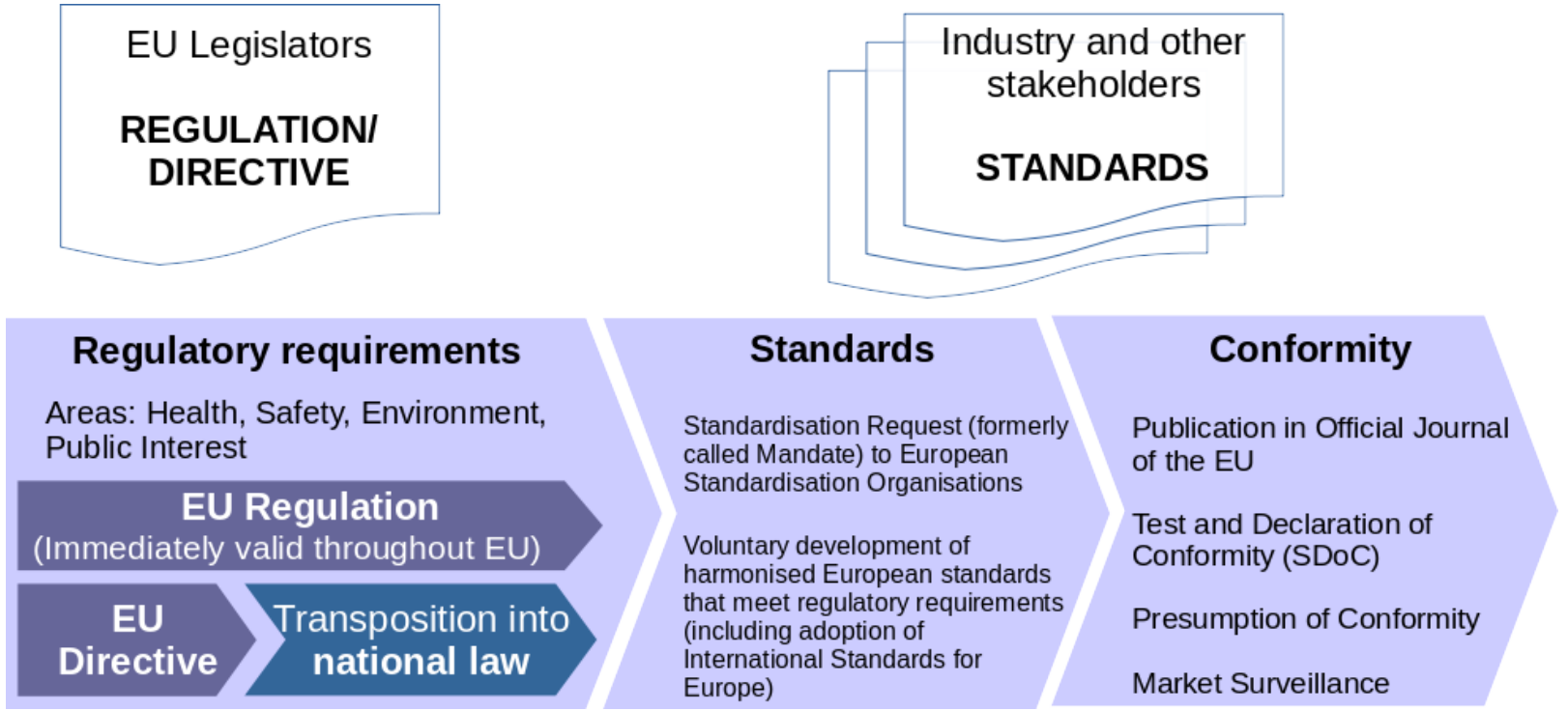
Basic Principle ("EU New Legislative Framework"):

Legal acts lay down the essential requirements and define safety objectives.

Harmonised European Standards define the technical way how to fulfil the legal requirements and be compliant with the safety objectives.

Compliance is mandatory for market access.

Harmonised standards are developed in one or more of the European standardisation organisations and are based on formal EU Standardisation Requests.



- Key element of the EU Single Market
- Innovation-friendly
- Laid down in Articles 40, 53 and 55 of the AI Act
- IBM has decades-long experience of working in this regulatory environment and its processes

In the hands of the private sector

European Commission: Right of initiative for legal acts

European Council and European Parliament: Two chambers negotiating and agreeing on final legal act

LEGISLATIVE PROCESS COMPLETED WITH LISTING IN OFFICIAL JOURNAL OF EU

With the listing of the AI Act in the Official Journal of the EU – the legal process is done.

Development of the harmonised European Standards required for compliance is in the hands of the private sector.

In the hands of the private sector

With the listing of the AI Act in the Official Journal of the EU – the legal process is done.

Development of the harmonised European Standards required for compliance is in the hands of the private sector.

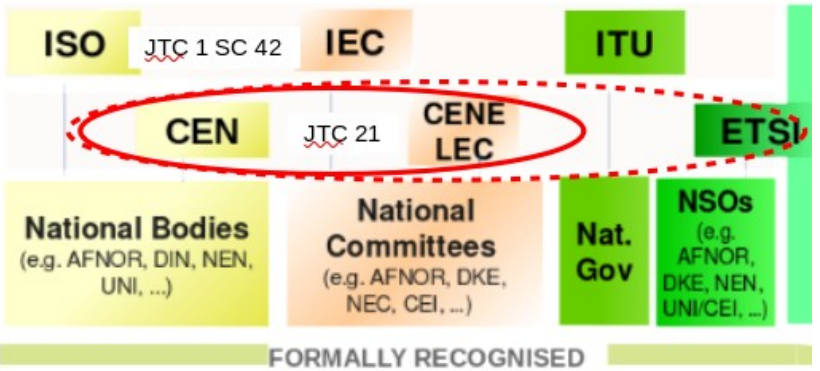
European Commission: Right of initiative for legal acts

European Council and European Parliament: Two chambers negotiating and agreeing on final legal act

LEGISLATIVE PROCESS COMPLETED WITH LISTING IN OFFICIAL JOURNAL OF EU

PRIVATE SECTOR ENTITLED TO DEVELOP HARMONISED EUROPEAN STANDARDS

Industry
Civil Society
Research
Academia
Administrations
...



CONFORMITY ASSESSMENT

Option A: Build up test lab for self-assessment internally

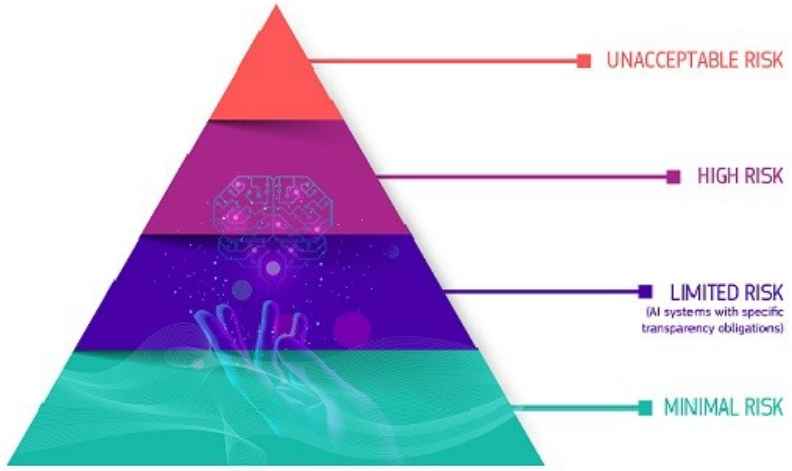
Option B: Work with notified bodies

EU MEMBER STATES

Accreditation of notified bodies

Building up of market surveillance

Essential Requirements as laid down in the legal act and requested standards



Graphic taken from the
website on the AI Act of the European Commission

Essential Requirement for High risk AI Systems (Chapter III, Section 2, Articles 8ff.):

Risk management system

Data and data governance

Technical documentation

Record-keeping

Transparency and provision of information to users

Human oversight
Accuracy, robustness and cybersecurity

EU Standardisation Request

Risk management system for AI systems

Governance and quality of datasets used to build AI systems

Record keeping through built-in logging capabilities in AI systems

Transparency and information to the users of AI systems

Human oversight of AI systems

Accuracy specifications for AI systems

Robustness specifications for AI systems

Cybersecurity specifications for AI systems

Quality management system

Conformity assessment for AI systems

European standards – Example: Cybersecurity for AI Systems

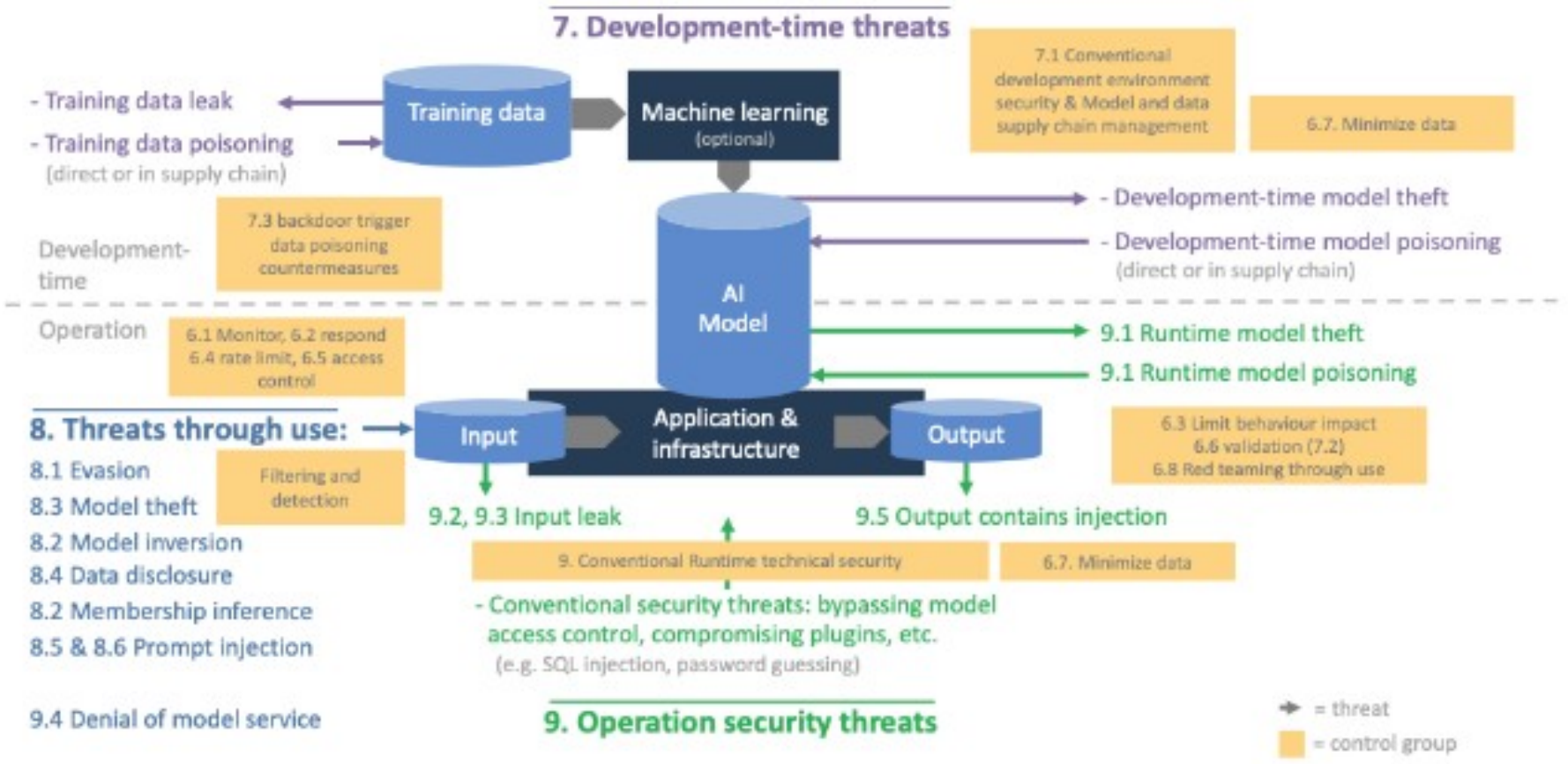
From the Standardisation Request:

… resilience against attempts to alter use, behaviour, or performance or to compromise security properties

… prevent and control cyberattacks

trying to manipulate AI specific assets, such as training data sets (e.g. data poisoning) or trained models (e.g. adversarial examples),

or trying to exploit vulnerabilities in an AI system'sdigital assets or the underlying ICT infrastructure.



European standards – Example: Cybersecurity for AI Systems

From the Standardisation Request:

… resilience against attempts to alter use, behaviour, or performance or to compromise security properties

… prevent and control cyberattacks

trying to manipulate AI specific assets, such as training data sets (e.g. data poisoning) or trained models (e.g. adversarial examples),

or trying to exploit vulnerabilities in an AI system'sdigital assets or the underlying ICT infrastructure.

Currently in progress: Definition of respective controls:

Resilience controls
- Rate limiting
- Model access control
- Input validation
- Data poison detect and respond (embedded in 7.3)

Other preventative controls
- Hide confidence in output (embedded in 8.3)
- Prompt input segregation (embedded in 8.7)

Supervision controls
- Logging
- Monitoring (OR: input detect and respond)
- Supply chain management (embedded in 7.1)
- AI asset management
- Change management (unsure) (new)

Quality assurance controls
- Model performance validation
- Security testing through input

Impact limitation controls
- Data minimization, obfuscation and exposure minimization (Former 6.7,)
- Impact limitation of unwanted model behavior

Driving Trustworthy and Ethical Use of AI

IBM Policy Lab: Precision Regulation for Artificial Intelligence

Designate a lead AI ethics official

Different rules for different risks

Don’t hide your AI

Explain your AI

Test your AI for bias

https://www.ibm.com/blogs/policy/ai-precision-regulation/

EU High Level Expert Group on AI: Ethics Guidelines on AI

Human agency and oversight

Technical robustness and safety

Privacy and data governance

Transparency

Diversity, non-discrimination and fairness

Societal and environmental well being

Accountability

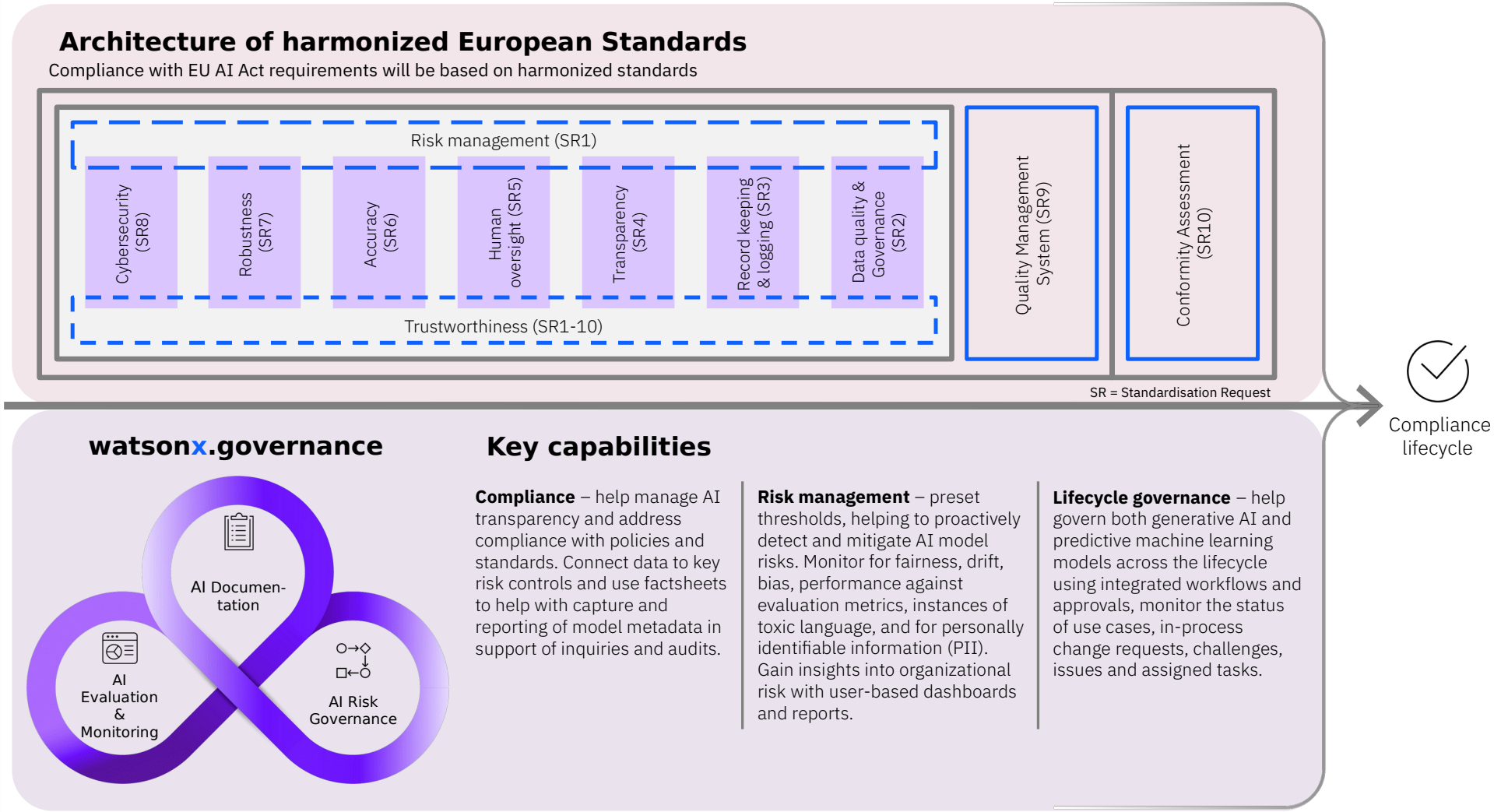
IBM has been a leader on trustworthy AI and the ethical use of AI for many years

IBM actively engages in European standardisation

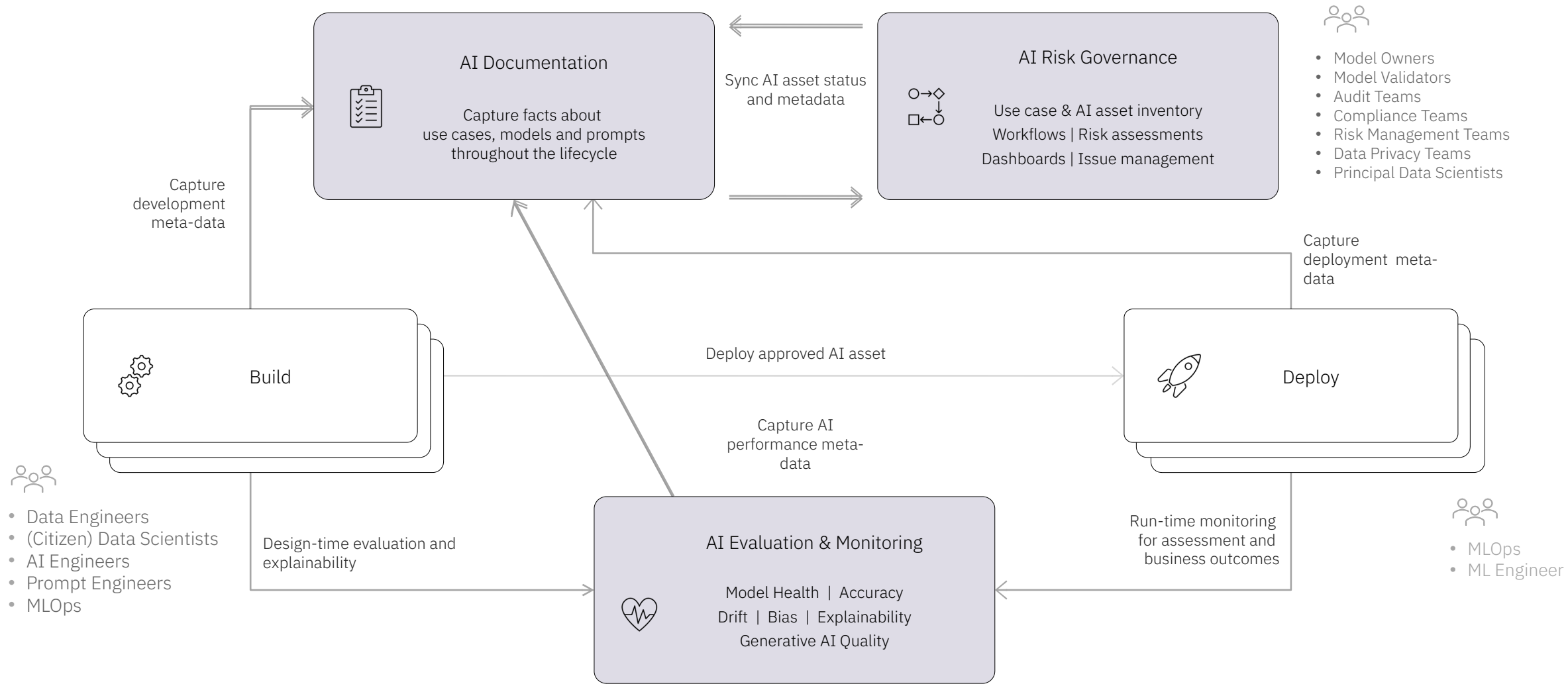
IBM governance tools aligned with European standards

IBM’s watsonx.governance toolset is informed by the technical requirements laid down in the harmonised European standards under development.

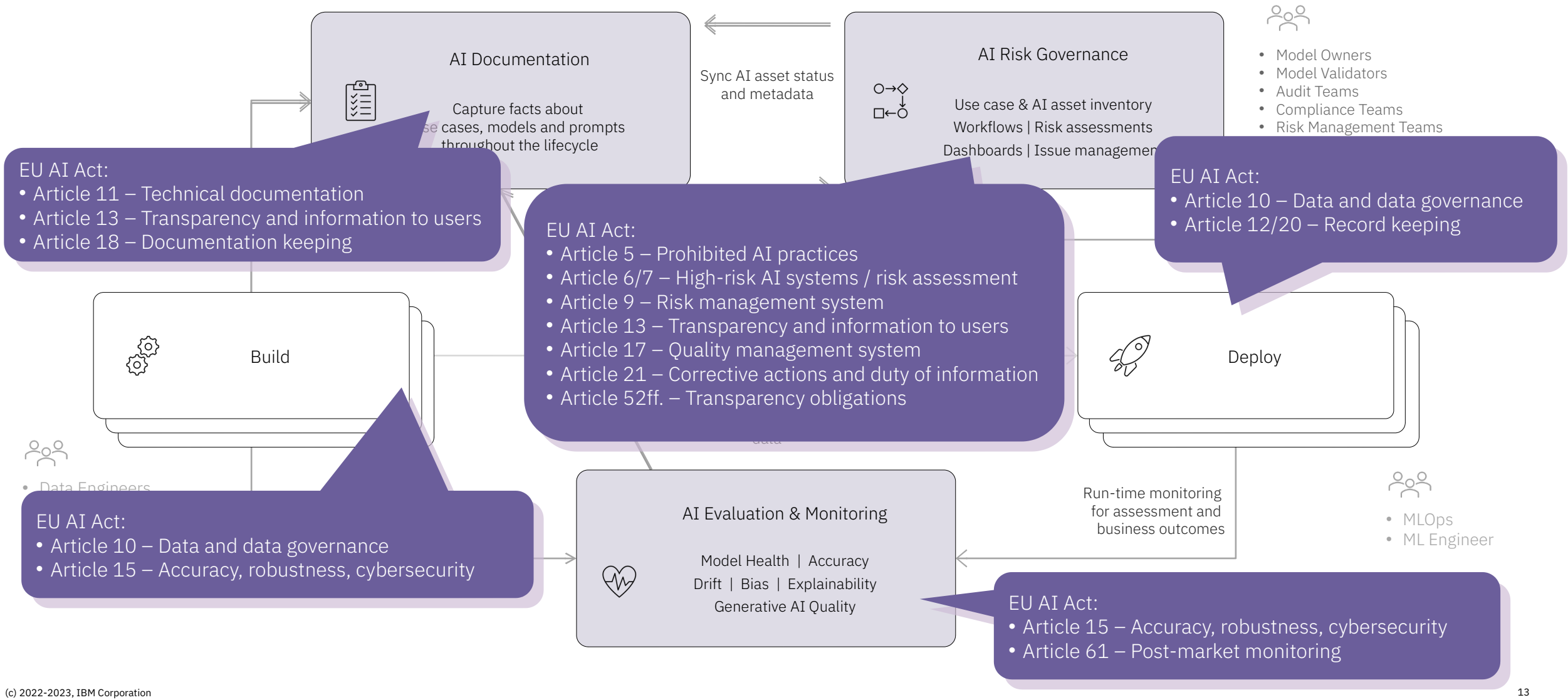
watsonx.governance enables performing the respective system management tasks, tests and lifecycle monitoring, inter alia by providing comprehensive dashboards.



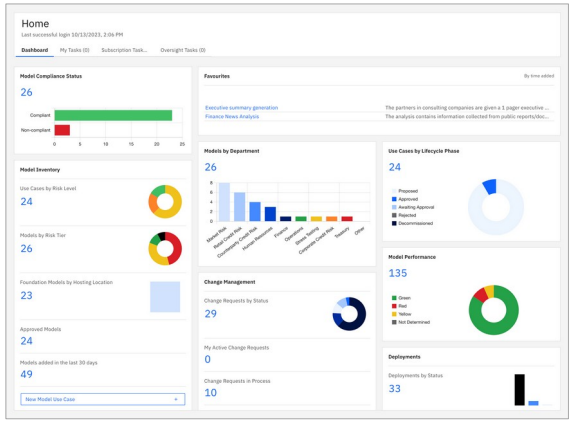
watsonx.governance - conceptual mapping of potentially relevant provisions of the EU AI Act

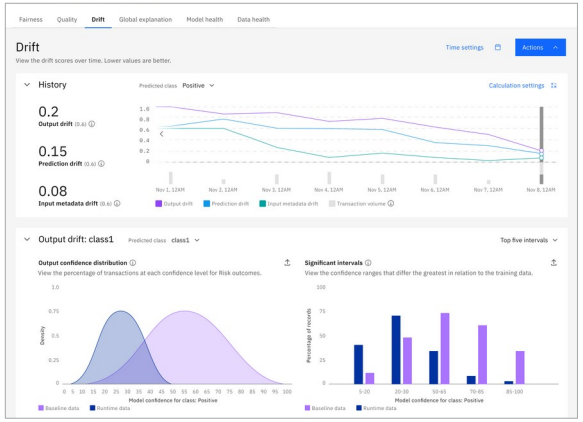


watsonx.governance - conceptual mapping of potentially relevant provisions of the EU AI Act



watsonx.governance support for compliance with the EU AI Act





Applicability and Risk Categorisations
[Articles 5,6,7]

Watsonx.governance provides EU AI Applicability and Risk Categorisation Assessment Questionaire. Updates on the act will be considered via regularly provided updates on these questionaires.

Governance Console

Compliance Requirements for High-Risk AI Systems
[Article 8]

watsonx.governance is a compliance tracking system that monitors and ensures adherence to the EU AI Act and other relevant regulations including a unified documentation and reporting system

Governance Console, AI Documentation

Risk management system for high-risk AI systems
[Article 9]

watsonx.governance is a comprehensive risk management system for AI systems, identifying, analyzing, estimating, and evaluating risks, as well as adopting appropriate risk management measures.

Governance Console

Need for high-quality data sets and robust data governance practices
[Article 10]

Part of the development process identify bias in training data etc.

Evaluation & Monitoring

Technical documentation
[Article 11]

Workflows in the Governance console ensure the documentation is in place before a model is put into production.
All metadata of a model and its development and monitoring activities are automatically captured in the AI factsheet.

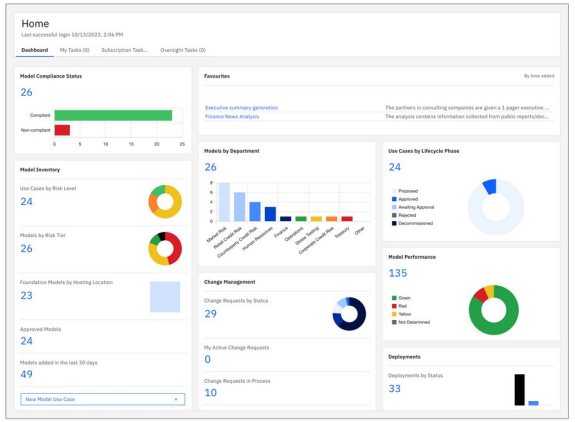
Governance Console, AI Documentation

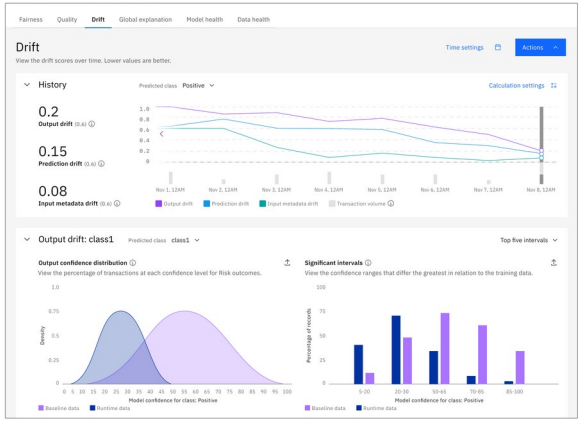
Record-keeping
[Article 12, 20]

Model inferences are logged through the Monitoring service, regardless of the deployment environment, in order to record risk-related events and increase transparency

Evaluation & Monitoring

watsonx.governance support for compliance with the EU AI Act





Transparency and Provision of Information to Deployers of High-Risk AI Systems
[Article 13]

Model cards (watsonx.ai)

Human Oversight of High-Risk AI Systems
[Article 14]

watsonx.governance allows effective oversight by humans during their use of high-risk systems

 Governance Console

Accuracy, robustness and cybersecurity
[Article 15]

Monitor for various accuracy metrics through the model lifecycle, preventing drift and bias

 Evaluation & Monitoring

Documentation keeping
[Article 18]

AI factsheets remain in the system through the lifecycle of a use case and can be exported as PDF documents to be stored in dedicated folders / archives

 AI Documentation

IBM Granite – performant and trusted AI model. Open Source

Open

Choose the right model, from sub-billion to 34B parameters, open-sourced under Apache 2.0.

Performant

Don’t sacrifice performance for cost. Granite performance is proven across a variety of enterprise tasks.

Trusted

Build responsible AI with a comprehensive set of risk and harm detection capabilities, transparency, and IP protection.

The future of AI is Open.

Concluding Remarks

The EU AI Act is a typical EU safety regulation. Harmonised Europeans standards are key for compliance with the legal requirements.

IBM has decade long experience in operating under the EU New Legislative Framework and providing standards to meet EU safety and security objectives.

IBM actively contributes to standardisation work at international and European level bringing in its expertise and supporting the implementation of the AI Act.

IBM has long advocated for trustworthy AI and offers governance tools (watsonx.governance) for managing trustworthy AI in line with the requirements of the European standards for the AI Act.

IBM drives open source in the area of AI – inter alia with the the IBM Granite models, open, performant and trusted.

Many thanks for your attention...

... see you at our demo both

Dr. Jochen Friedrich
IBM Technical Relations Executive