



Auditing AI Systems

A guide for public auditors



≡ Agenda

1. Introduction – A brief history of the whitepaper
2. Update
3. Key content of the guide
4. Summary



Introduction

A brief history of the whitepaper

≡ White Paper on Auditing Algorithms

→ Memorandum of Understanding (MoU) since 2017: international cooperation of SAIs on AI and data science

- Brazil,
- Finland,
- Netherlands,
- Norway,
- United Kingdom and
- Germany.



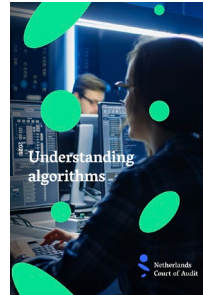
MoU 2025, image source: SAI Norway

→ In 2020: First version of white paper on auditing algorithms

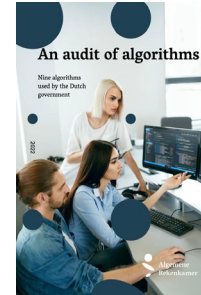
≡ Since then...

→ SAIs gain more experience with AI audits

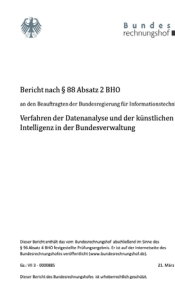
✓ Silent update 2023



2021



2022



2023



2024



2024

→ Changes of AI landscape:

- Rise of Large Language Models (LLMs) - ChatGPT launched 2022
- AI regulation
 - EU AI regulation proposal 2021,
 - AI Act published 2024
- ❖ Major update required



Update

≡ What we considered...

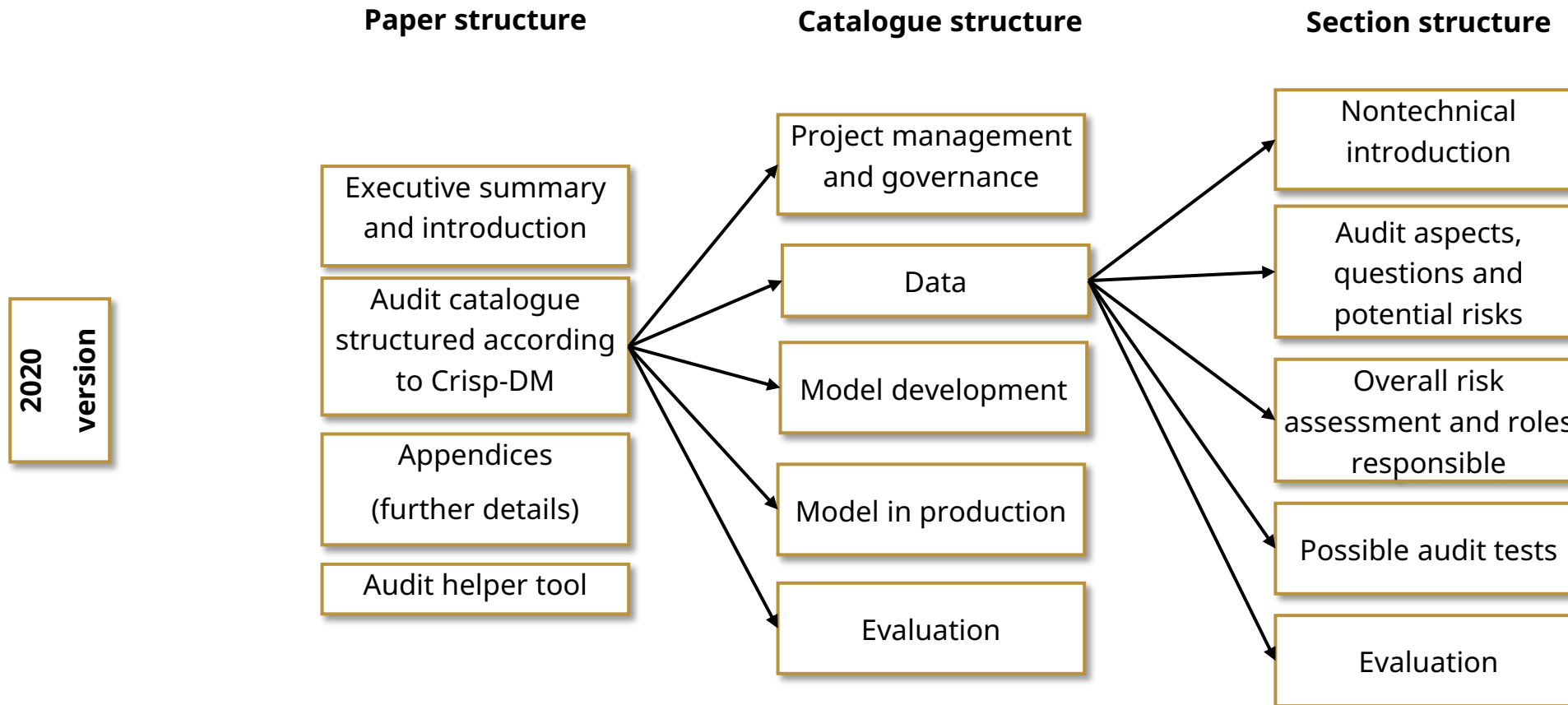
The 2020 paper in a nutshell:

- Paper sets out **key risks** related to the use of ML in public services
- **Audit catalogue** includes methodological approaches for auditing AI
- Suggestions are **based on MoU SAI's experience** from ML and software development audits
- SAIs should be able to **assess whether** the use of **ML contributes** to **efficient and effective public services**
- ML audits require **specific audit knowledge and skills**
- SAIs should **expand** their **capacities** to conduct more audits in the field of ML

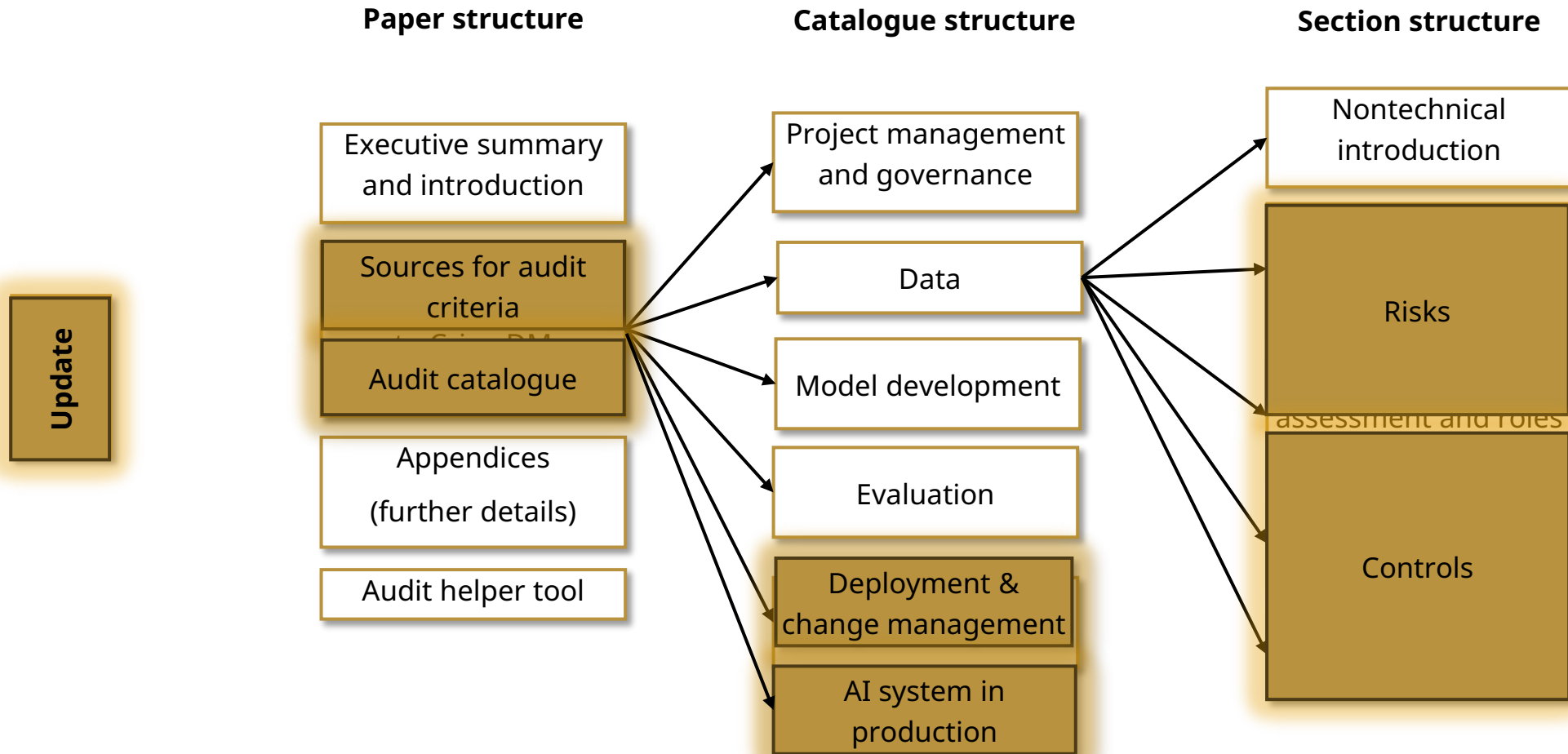
Main updates:

- **Generalisation** from ML models to AI systems
- Include **generative AI**
- Include **external development**: foundation models, purchased AI systems
- Include developments in **AI regulation**: EU AI Act

≡ What we changed ...



≡ What we changed ...





Key content of the guide

≡ General considerations I

→ Catalogue structured along development cycle

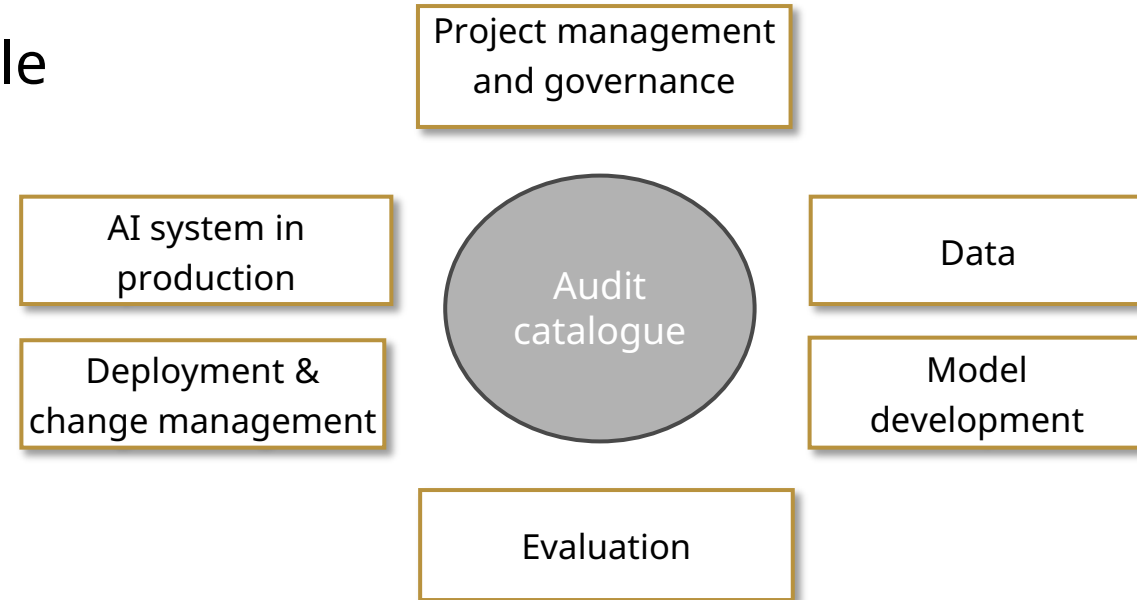
- ✓ Structured results
- ✓ Identification of root causes and dependencies

→ Key concepts throughout all stages:

- Fairness by design
- Transparency by design
- Privacy by design

→ Stages are slightly different for purchased systems, but still valid

- Example: In “model development” stage, model selection considerations are still relevant for purchased models.



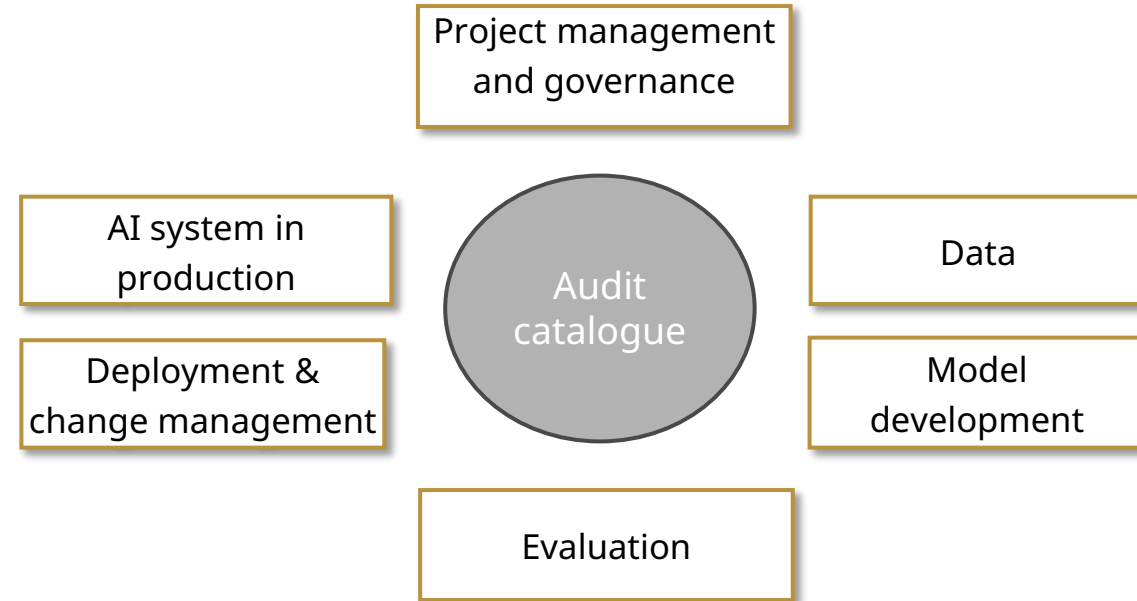
No linear sequence of single steps → Continuous loop of interacting stages

≡ General considerations II

→ Audit depths

- Baseline: document review and governance checks
- User access: Auditors run model and can check outputs for e.g. robustness and repeatability
- Full access:
 - Inspection of source code/weights/...
 - Reproduction of model training
 - Developing alternative approaches

→ Depends on auditor qualification and access granted by auditee



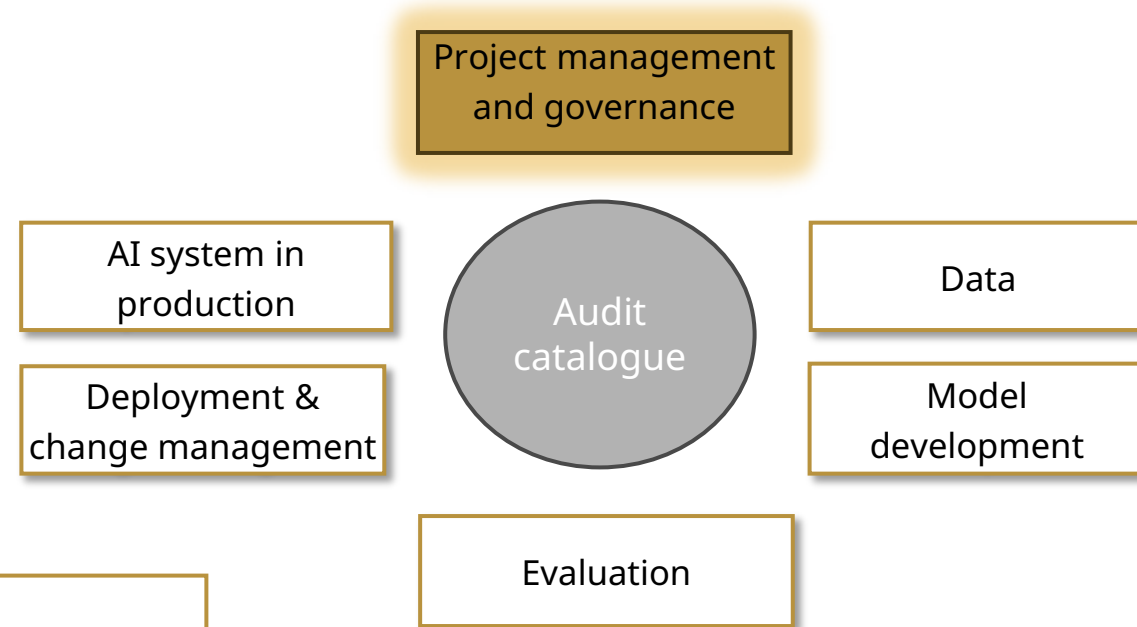
≡ Project management and governance

Key topics

- AI project proposal
- Jurisdiction-specific legal and ethical requirements
- AI lifecycle management
- AI risk assessment

Example audit questions

- In what business processes will the application be used?
- What evidence is there of a cost-benefit analysis?



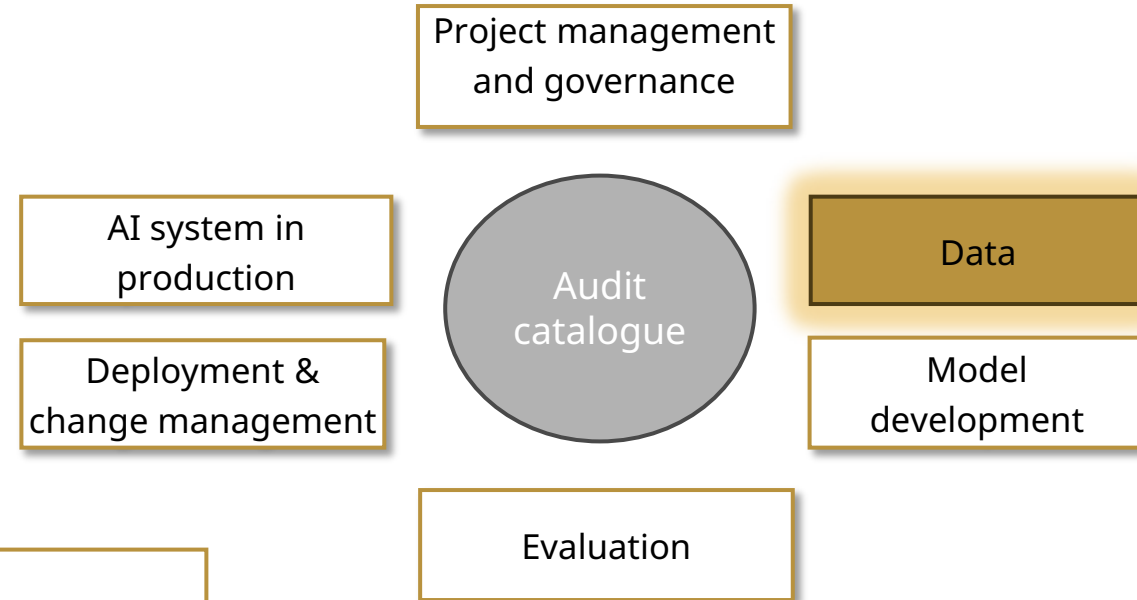
≡ Data (“Garbage in, Garbage out”)

Key topics

- Data quantity and quality
- Retrieval-augmented generation and large language models
- Privacy issues in the context of AI systems

Example audit questions

- What are the data sources?
- If applicable, how is the RAG knowledge base curated, validated, and maintained?



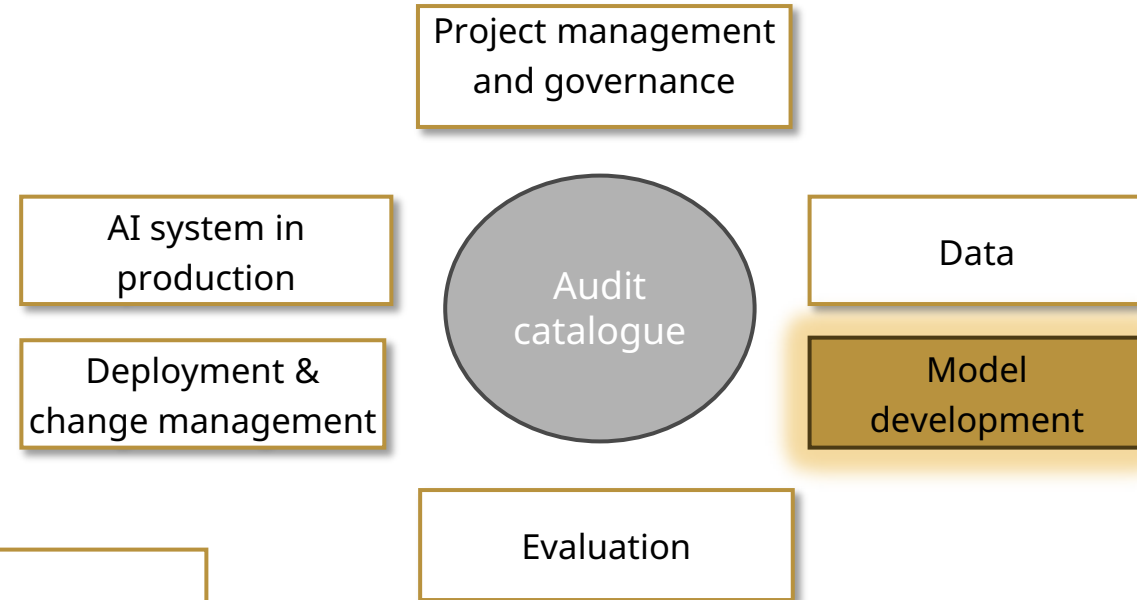
≡ Model development

Key topics

- Development process and performance characteristics
- Quality assurance, reliability and robustness in development
- Documentation and version control expectations

Example audit questions

- How was the model selected? Which alternative models were compared?
- What prompt engineering methodology and guardrails are implemented?



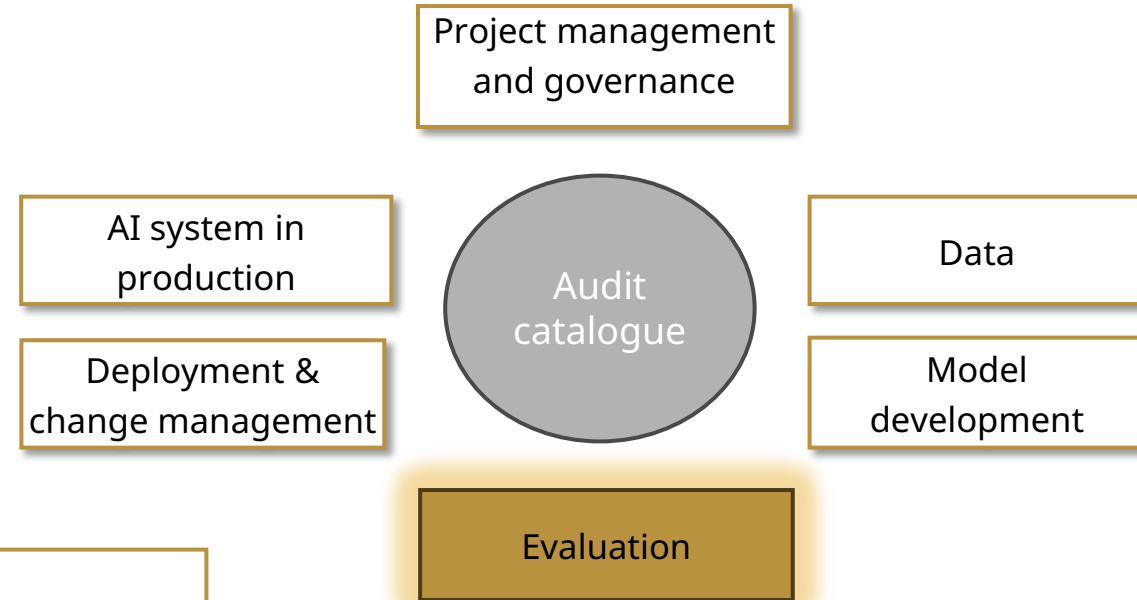
≡ Evaluation before deployment

Key topics

- Performance and acceptance testing
- Transparency and explainability
- Fairness and minimizing harm
- Security
- Environmental impact

Example audit questions

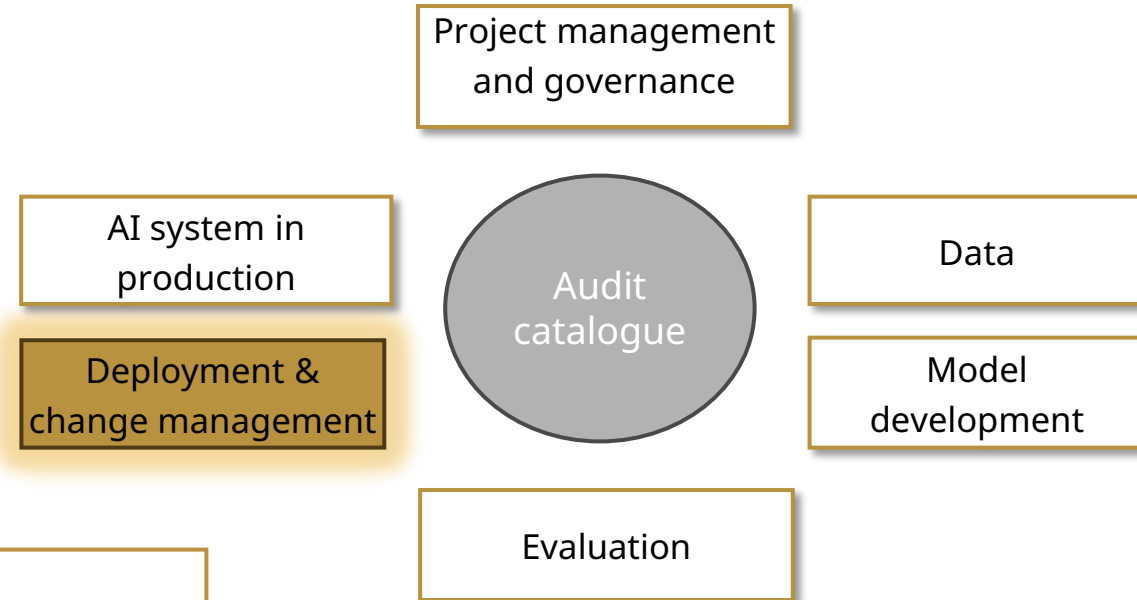
- How is model fairness guaranteed?
- How does the AI system respond to faulty or manipulated datasets?



Deployment & change management

Key topics

- Business readiness and change management
- Release criteria and gates



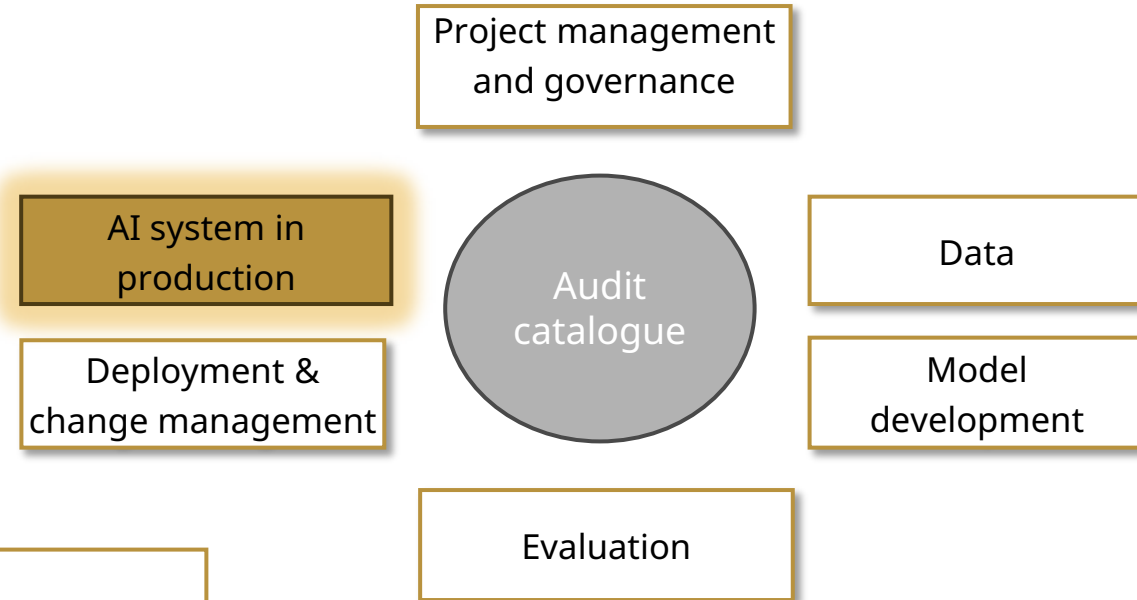
Example audit questions

- Does the system meet release criteria?
- Do users have the skills and understanding to use the system effectively?

≡ AI system in production

Key topics

- Monitoring and performance tracking
- Incident response and model updates
- Ongoing compliance and audit trails



Example audit questions

- How is AI system performance monitored?
- How are incidents related to the AI system detected and logged?

≡ Audit helper tool

→ List of audit questions grouped along development cycle

- Filter for questions
 - only relevant to “classical ML”, LLM/GenAI or both
 - relevant to externally developed models
- Additional information
 - Purpose of the question
 - Required evidence
 - Additional questions

No.	Relevance	Priority	Question	Purpose	Evidence Required	Model Development			
						Model Type	In-house	External	Related Questions
A6	Operation of the model and performance in production								
A6.001			What are the major features of human-machine interaction of the AI system?	Understand how the user may influence the AI system or rely on its results, how the user is informed about actions and results of the AI system and what autonomy the AI system may have.	- user interface specifications - user manual	All	WAHR	WAHR	A4.017, A5.007
A6.002			What qualifications do users of the AI system need?	- Understand what AI system-related knowledge users need to possess to appropriately assess the decisions made by the AI system. - Understand that users may know nothing at all about the AI system and its impact on the process.	- training documents (incl. training materials, training certificates, statistical data on training) - job description	All	WAHR	WAHR	A5.001, A6.003
A6.003			How can decisions or proposals made by the AI system be overruled by users? What level of autonomy is granted to the AI system, and how is its appropriateness evaluated?	Understand what autonomy the AI system has (BITKOM model on decision-making processes).	- procedural guidance	All	WAHR	WAHR	A4.016, A6.002
A6.004			What criteria govern decisions/proposals of the AI system that are submitted to the user?	Understand what decisions are submitted to the user and which are not.	- operational system - procedural guidance	All	WAHR	WAHR	A6.003
A6.005			How are the key performance indicators of the AI system provided to decision-makers?	Understand the extent to which decision-makers are informed about decision quality (or uncertainty) of the AI system.	- key performance indicators - logs - screenshots	All	WAHR	WAHR	A4.003, A6.007
A6.006			How is the data monitored and updated?	Confirm that the organisation has continuous monitoring and update processes for input data, feature pipelines, and reference datasets, with clear thresholds and governance.	Update procedures Data monitoring framework	All	WAHR	WAHR	A2.005, A6.007

Summary

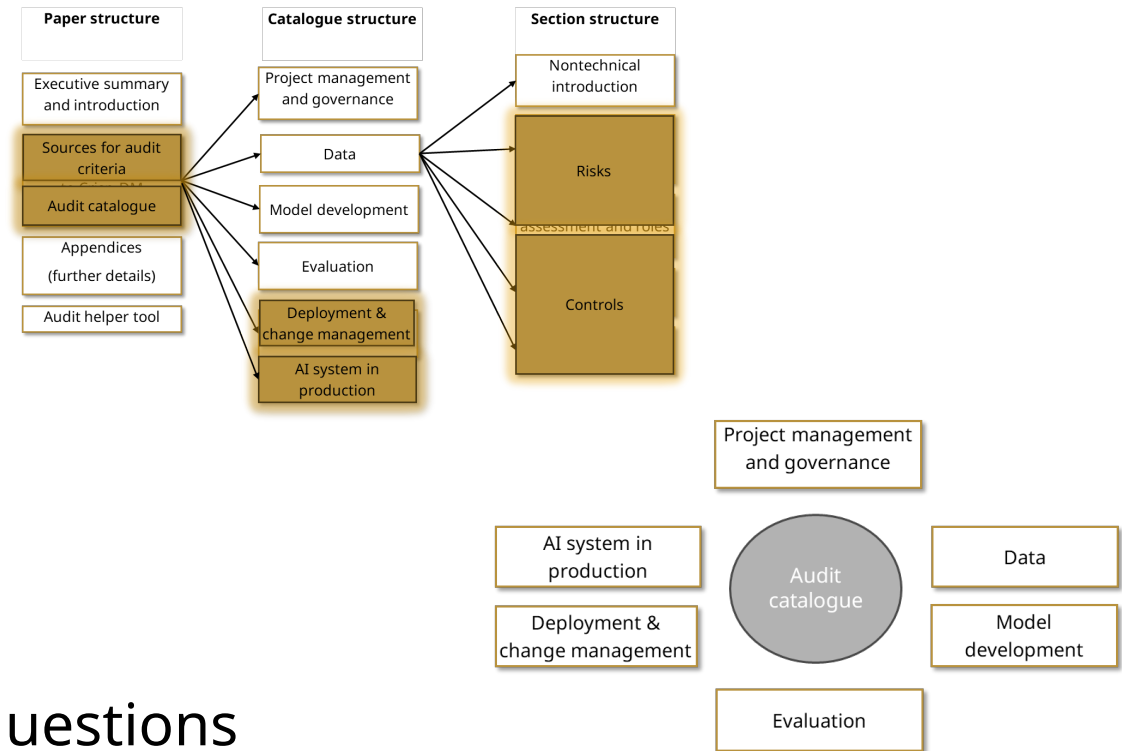
→ Guide updated to address

- new legislation,
- technical developments and
- experience from audits.

→ Focus on risks and controls

→ Audit helper tool provides audit questions

→ Audit depth depends on auditor qualification and access granted by auditee



≡ Thank you for your attention – Time for discussion

→ My questions:

- Who of you has read/seen the “old” white paper?
- Who of you has read/seen the updated guide?
- Do you have any feedback/comments?

→ Your questions?

→ Link to English version of the guide (German version coming soon)

