Artificial Intelligence and IT Audit

Frank van Praat
NOREA Algorithm Assurance working group
KPMG Responsible AI
16 November 2023

What we will discuss today

Digital Disruption, Data Science and Artificial Intelligence

Audit *of* Data Analytics

D&A as audit subject - assurance on data & data driven technologies

Audit *with* Data Analytics

D&A as audit tool - using data & data driven technologies to give assurance

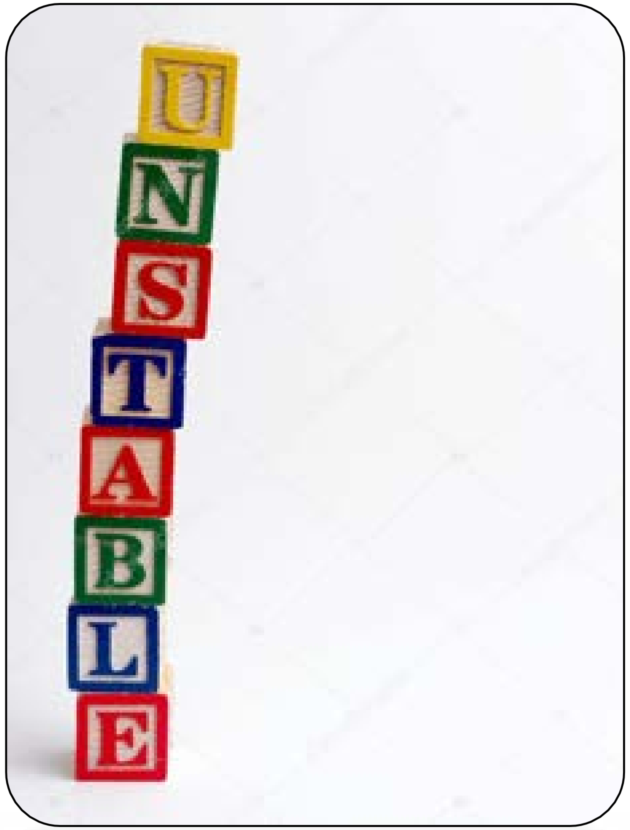
(IT) Auditing Theory and Professional Practice

Challenges and risks of Artificial Intelligence


© Warneken & Tomasello

Challenges 1 and 2: intuitive psychology and intuitive physics

Intuitive understanding of the reward function motivating someone's behaviour.



Broader understanding of the Physical environment

Challenge 3: Induction and inductive bias

Machine learning is based on assumptions about the training and testing data. These assumptions typically do not hold! And it is not the data that is at fault.

Assumption: Training, testing, and production data are independent and identically distributed samples taken from the exact same causal mechanism.

Unless told otherwise the algorithm is free to guess what the causal mechanism is.

If you ignore differences between samples you incorporate inductive biases in your model. These will manifest as systematic error no matter which algorithm and sampling approach you use.

This is often misunderstood (in the context of fairness) as meaning that trained models often only reflect a history of already existing bias.

Solution is to understand the causal mechanisms that create your data very well.

Challenge 4: Underspecification

Not properly thinking through your problem conceptualization can manifest as underspecification problems.

Ill-defined problem: Some real world problem, finding the solution has business value

Problem specification: Well-defined problem

- A mathematical abstraction of the problem reduced to finding a causal mechanism in data and a standard for rationally exploiting that mechanism
- We know the space of hypotheses we must test to find the solution
- We can find a good solution with a good loss function and good test criteria

Manifestation of underspecification: depending on algorithm and sampling methods used we find many different solutions (causal mechanisms), each leading to different types of systematic errors.

Solution is to specify better what causal mechanisms/systematic errors we don’t want to see (fault models), monitor for these faults, and to devise targeted stress tests for them.

To some extent underspecification is unescapable. If we exactly understand our problem we don’t have a case for AI technology, but we should avoid uncritically believing in AI snake oil as well.

Auditing applications of AI

Four audit approaches to assess algorithms and AI

Object of investigation

Algorithm overall control environment

Algorithm design and maintenance

Algorithm output

Indirect audit approaches. Evaluate the algorithm and its control environment.

Direct audit approaches. Test the output of the algorithm and/or controls regarding algorithm output

Audit approach

Evaluate algorithm entity level controls

— Evaluate if entity level controls are in place to ensure algorithms are built in a controlled environment.

Test the model

— Perform an in-depth assessment to determine if the algorithm performs in line with relevant criteria (including GITC when testing ToE).

Test monitoring controls

— Test if internal controls are in place to monitor the transactions performed by the algorithm and mitigate the risks of algorithm failure.

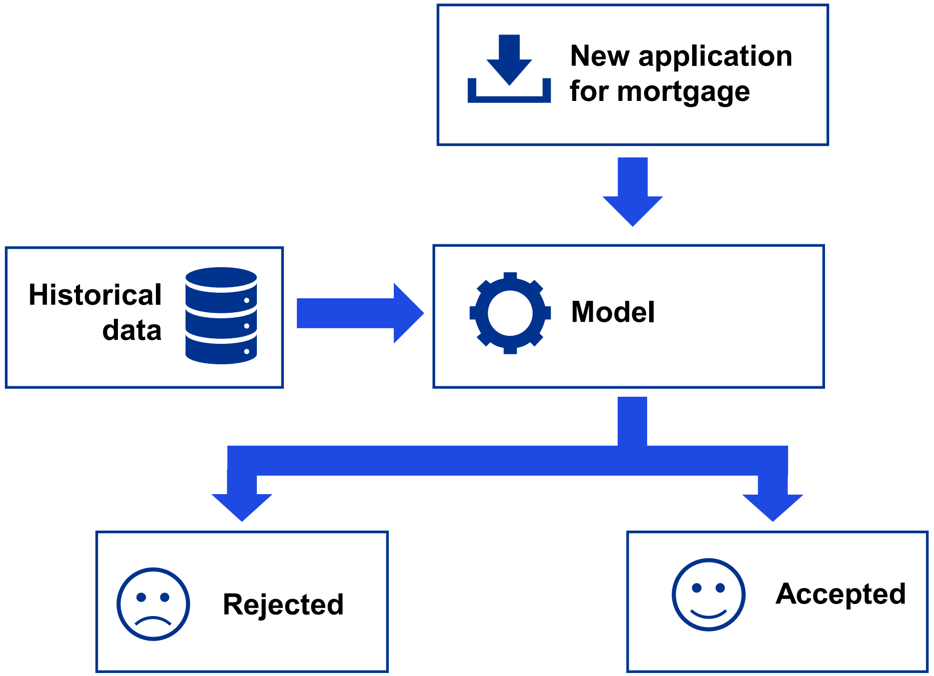
Substantive procedures

— Test if (a sample of) the transactions were processed by the algorithm in line with relevant criteria.

…differing in feasibility and level of assurance provided

The problem of testing AI/algorithm output

Example: how to assess if your model has accurately predicted loan default probability?



Would the applicant really have defaulted?

Wait 30 years to verify model prediction?

Why testing output is not straightforward

— Is the outcome already available (e.g. 30 year mortgage loan)?

— Has the algorithm prediction resulted in a decision affecting the outcome (e.g. rejecting a job or loan application)?

— Is reperformance by a human possible (e.g. search engine)?

— Is reperformance by a human feasible (e.g. fraud detection)?

The problem of a ‘test of design’ for AI

Traditional analytics and queries

Explicit business rules (requirements)

- e.g. exclude al items larger than EUR 100

Are the business rules properly translated into the design / code?

Design documentation / code

- e.g. SELECT items WHERE value<=100

Machine learning (/AI) models

Training and test data

- e.g. database of real estate sales including various property characteristics

No direct link between model (trained parameters) and input data. How to test???

ML model

- e.g. property_price = 782,45 x lot_size + 1115,47 x floor_space

… an AI audit must include the design process

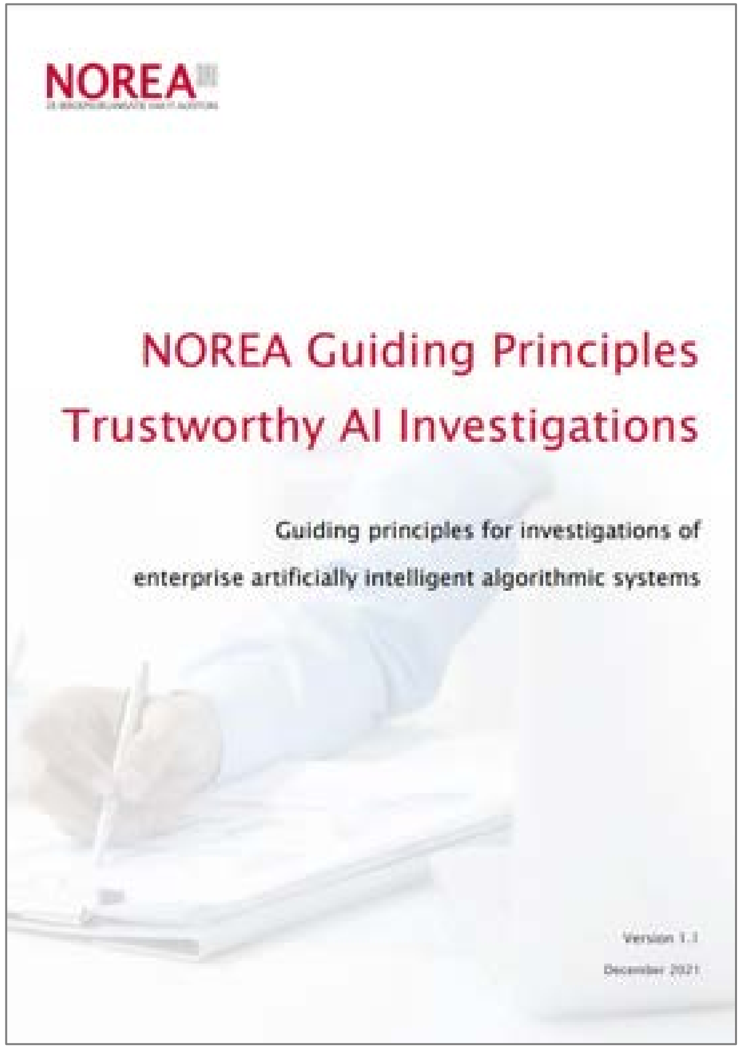
Control-based testing

1. Numerical summary of Guiding Principles

The table below summarizes the Guiding Principles. The Principles contain 119 key considerations for Trustworthy AI investigations, categorized into 5 risk categories and 6 CRISP-DM phases + an added Governance phase. Table 1 presents a descriptive numerical summary of the Guiding Principles, followed by the detailed framework.

CRISP-DM Phase	Risk categories	# Key considerations
1. Business Understanding	Governance	3
	Ethics	16
	Privacy	4
	Performance	4
	Security	1
2. Data Understanding	Ethics	5
	Privacy	4
	Performance	6
	Security	2
3. Data Preparation	Ethics	1
	Privacy	4
	Performance	1
	Security	5
4. Modeling	Ethics	4
	Privacy	2
	Performance	8
	Security	6
5. Evaluation	Ethics	2
	Performance	22
	Security	1
6. Deployment	Ethics	1
	Performance	8
	Security	2
Added Phase		
Governance	Roles & responsibilities	2
	Ethics	1
	Privacy	2
	Performance	2
Total		**119**

Table 1: numerical summary of NOREA Guiding Principles



The process that we typically follow when auditing AI

	Design & Implementation			Operating effectiveness
	Data	**Algorithm design**	**Algorithm implementation**	**Operational controls**
Control objectives	— The source data used to develop the AI-model is of sufficient quality and quantity. — Data quality is ensured throughout the data preparation phase.	— The real-world decision problem has been properly translated into an AI-problem with clear success criteria (‘definition of success’)	— The process of building, testing and optimizing the AI-model was methodologically sound.	— Controls are in place to ensure continuous operation of the algorithm in line with the ‘definition of success’. — Relevant when operating effectiveness must be established.
How to test	— ‘Traditional’ testing of data integrity controls (e.g. in data conversions) — Additional AI-specific data controls. — Inspection or reperformance of the exploratory data analyses (EDA).	— ‘Peer review’ by KPMG data science experts on key assumptions and design decisions.	— Testing procedures with different levels of depth. Ranging from inspection of the build and test process to complete replication of the model.	— Testing of the ‘traditional’ GITC regarding access management and change management. — Additional AI-specific operational controls for monitoring performance and retraining.

Steps are fairly similar to an application control audit

One level deeper on the AI development lifecycle

1

BUSINESS UNDERSTANDING
Select one algorithm and follow the process
- SIA

Talk to the main people involved:
- External audit department, Developers, Data privacy, Ethics, GDPR, CIO CTO, Information manager and Security officers.

2

DATA UNDERSTANDING
Follow the flow of the data:
- Where is the data coming from?
- How does the data flow through the model?
- To what standards does the data needs to comply and how is this ensured?

Interviews with developer but also with the other key people with a relation to data usage.

3

DATA PREPERATION
Understand the data preparation and to what extend it effects the model:
- What kind of trainings data is being used?
- How is the data quality monitored and to what extend is it cleaned?
- What effect does the data preparation has on the outcome of the model?

Interview with the developer.

4

MODELING
Gain an understanding of the model:
- How is the development of the model initiated and proceeded?
- How does the organization ensures that the model is correct and fulfills its purpose?
- Are there specific tests and checks performed to continue development?

Interviews with developer but also with other key people with a relation to project management

5

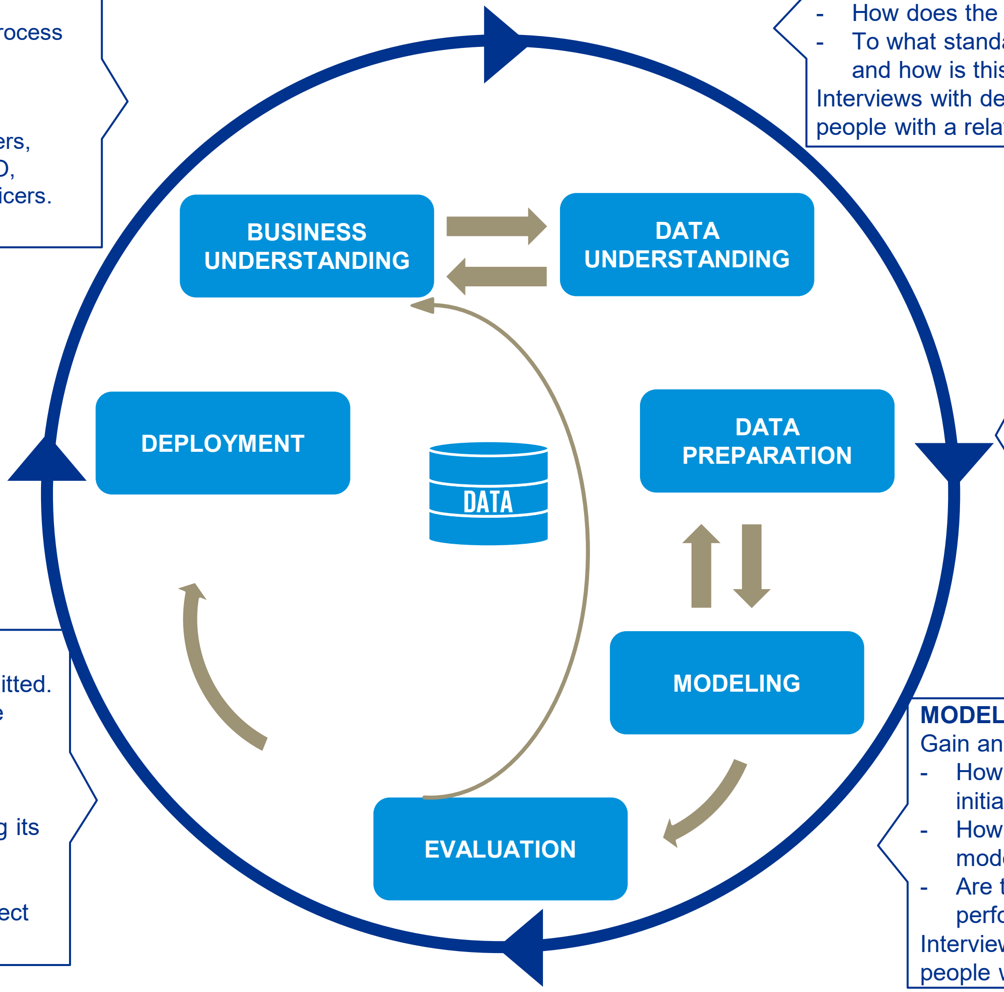
EVALUATION
To what evaluations is the solution submitted.
- How does the organization knows the described quality standards are met?
- How does the organization know the solution is secure?
- To what extend is the solution fulfilling its purpose?

Interviews with developer but also with other key people with a relation to project management

6

DEPLOYMENT
Deployed, what next?
- Standard GITC controls (change management, access management, business continuity etc.)
- AI specific components, for example monitoring

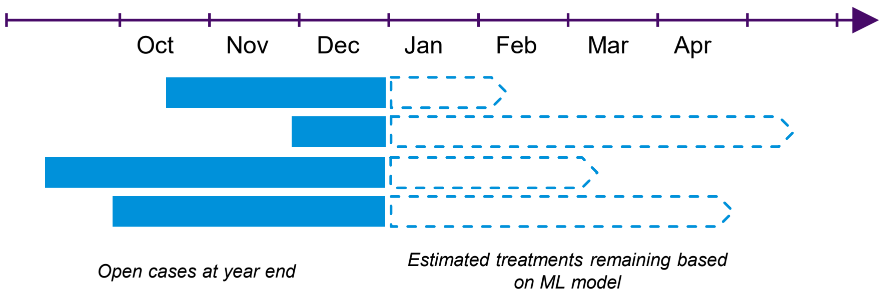
Interview with the developer.



Example case of an audit of an AI solution

Provision calculation for a health care institute

- Every year the organization receives a budget from healthcare insurers to cover all cases starting in that year
- Of course, at the end of the year not all cases are closed yet
- For the financial statement a provision must be calculated which estimates the costs of all remaining treatments required to close the open cases
- The organization developed an ML model to calculate the estimate, based on historical data in its databases using random forest regression and classification
- Output of the model is assessed by the Finance department. The Finance department makes the final decision regarding the size of the provision, based on model output and expert knowledge.



Risk profile from financial statement perspective

- High impact
- Low autonomy
- Medium complexity

Note: this is an 'easy' case. For a financial statement audit we do not need to worry about aspects such as fairness and explainability. Focus is on reliability only.

...How to approach such an audit?

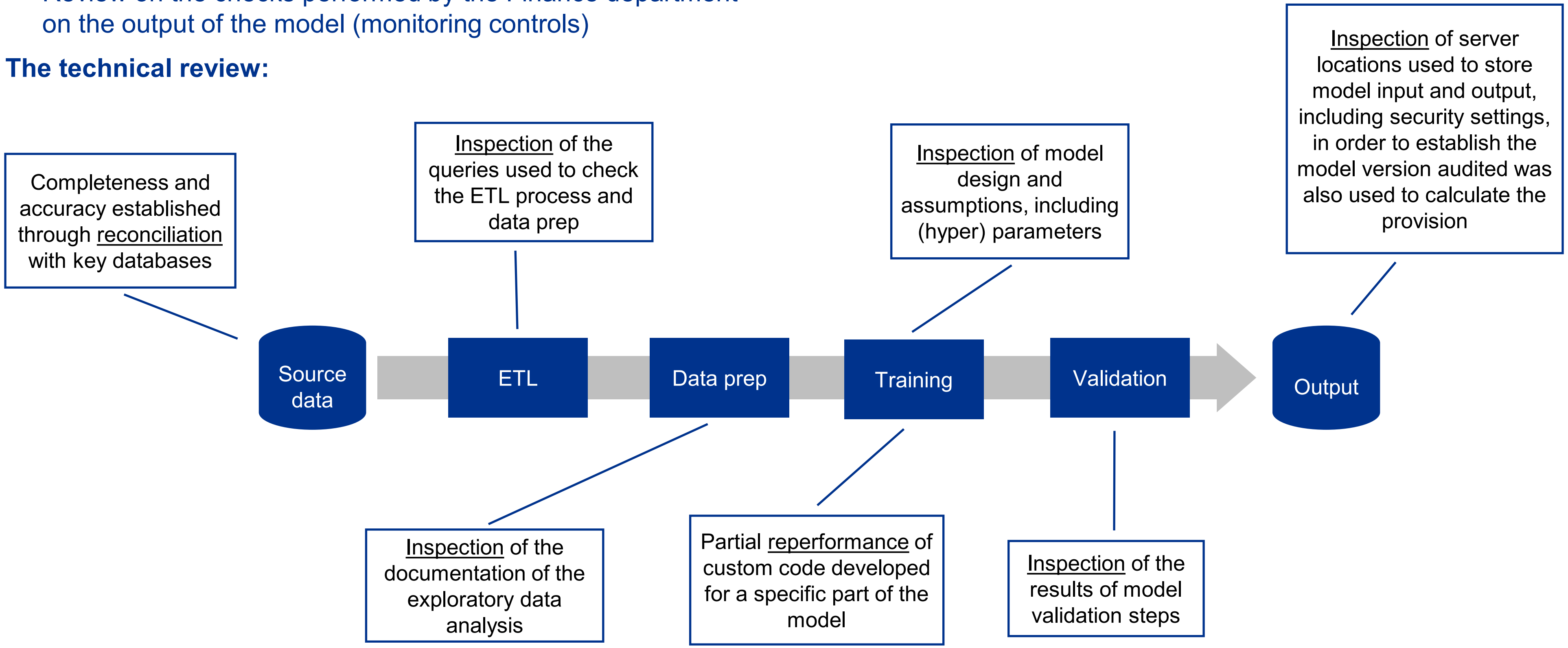
The approach on the case

Combined approach:

— Technical review on the design of the algorithm (testing the model)

— Review on the checks performed by the Finance department on the output of the model (monitoring controls)

The technical review:

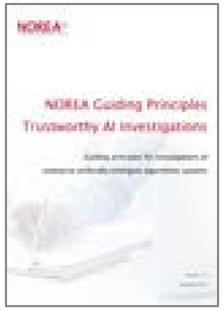


What is needed by (internal) Audit – and want to learn more

Want to learn more?

- Boer, A., de Beer, L., van Praat, F. (2023). Algorithm Assurance: Auditing Applications of Artificial Intelligence. In: Berghout, E., Fijneman, R., Hendriks, L., de Boer, M., Butijn, BJ. (eds) Advanced Digital Auditing. Progress in IS. Springer, Cham. https://doi.org/10.1007/978-3-031-11089-4_7

- NOREA Guiding Principles Trustworthy AI Investigations, v1.1 December 2021 https://www.norea.nl/nieuws/publicatie-norea-guiding-principles-trustworthy-ai-investigations-update


NOREA
NOREA Guiding Principles Trustworthy AI Investigations

Some or all of the services described herein may not be permissible for KPMG audit clients and their affiliates or related entities.

kpmg.com/socialmedia