

**Data and Artificial Intelligence
for Internal Audit**

-

Fahd KAATICH

-

10th Audit Masters Conference 2025

21/05/2025

Introduction and main objectives

A few words about me:



Fahd KAATICH

Head of Audit

Assignment @ CNP

Assurances

- Engineer and Actuary by training, Internal auditor by vocation;
- ex-Modeler in the banking industry : Counterparty, credit, and balance sheet risks;
- Head of Audit assignment in different corporations covering quantitative risks in the finance and risk management functions;
- As such contributed to multiple AI tools internal to the audit function.

And about the topic at hand:

The use of artificial intelligence, enabled by technical advances in data collection and treatment has proven to be a disruptor in many industries:

- Internal audit can benefit from this new set of tools to improve effectiveness, efficiency and efficacy.
- Internal audit must harness these tools to ensure that they remain on par with other lines of defense.

➔ Tools, use cases and methodologies have to be developed to promote Internal Audit's capabilities in terms of making use of, and auditing AI audit objects.

Outline

- 1.**
Introduction 03
- 2.**
A brief overview of processes 04
- 3.**
A walk through various methods and use cases 06
- 4.**
Closing remarks 16

1 – Introduction: regulatory context

Non-sector specific in the EU:

- AI Act
- GDPR
- Local legislations

Sector-specific:

Banking oriented regulations:

- Risk related : BCBS 239 for risk data quality, Model Risk Management, risk specific regulations, indicators, etc.
- Compliance function : KYC data, and KYC data quality.

Audit function-specific:

- 2024 Global Internal Audit Standards of the IIA

Key takeaways:

A complex environment with a wide range of regulatory texts that:

- set boundaries on methods allowed for different use cases;
 - defines which data can be used for which use case;
 - standards of documentation, transparency, and outcomes;
- ➔ All of which have to be taken into account as audit points but also when developing own tools.

1 – Introduction: definitions

The EU Artificial Intelligence Act offers a useful definition:

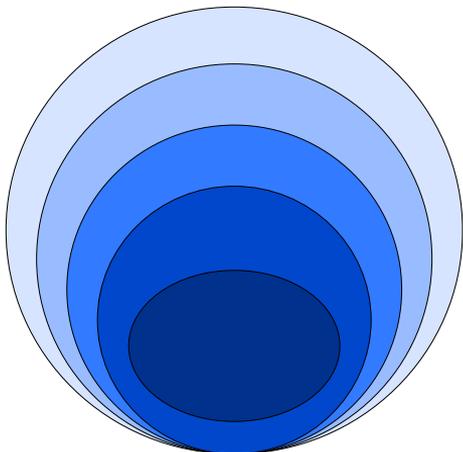
- **Artificial Intelligence system :**

For the purposes of this Regulation, the following definitions apply:

- (1) ‘AI system’ means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.

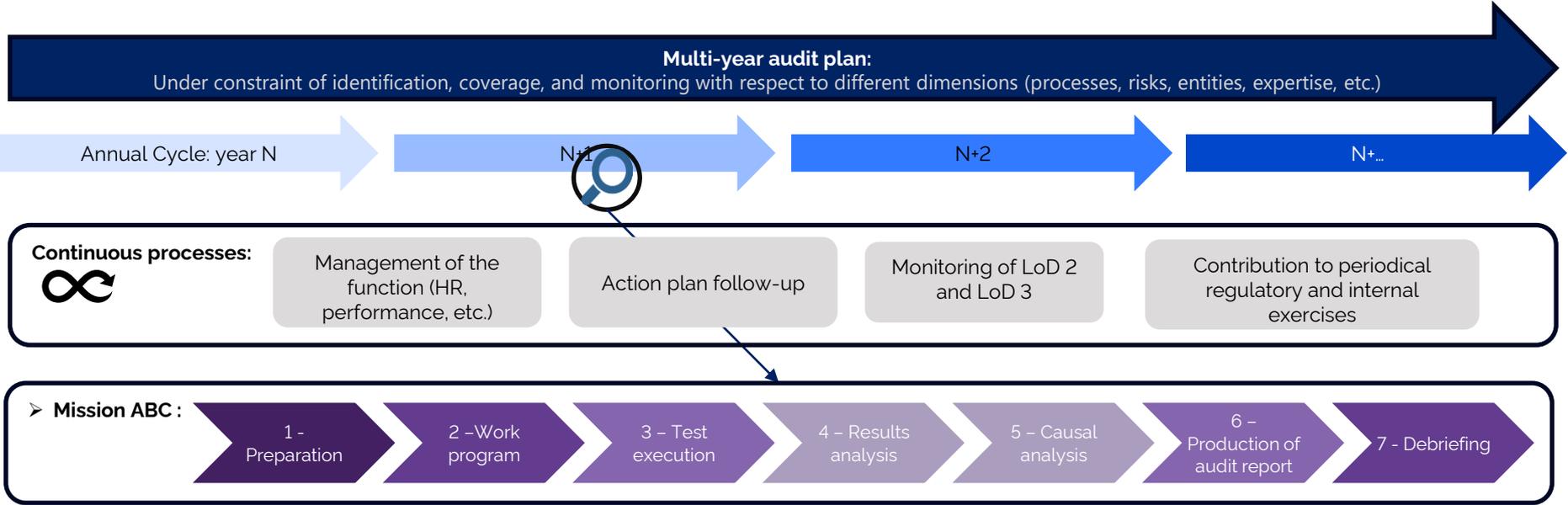
EU Artificial Intelligence Act Chapter 1, Article 113 (a)

A generic term, referring to a spectrum of techniques:



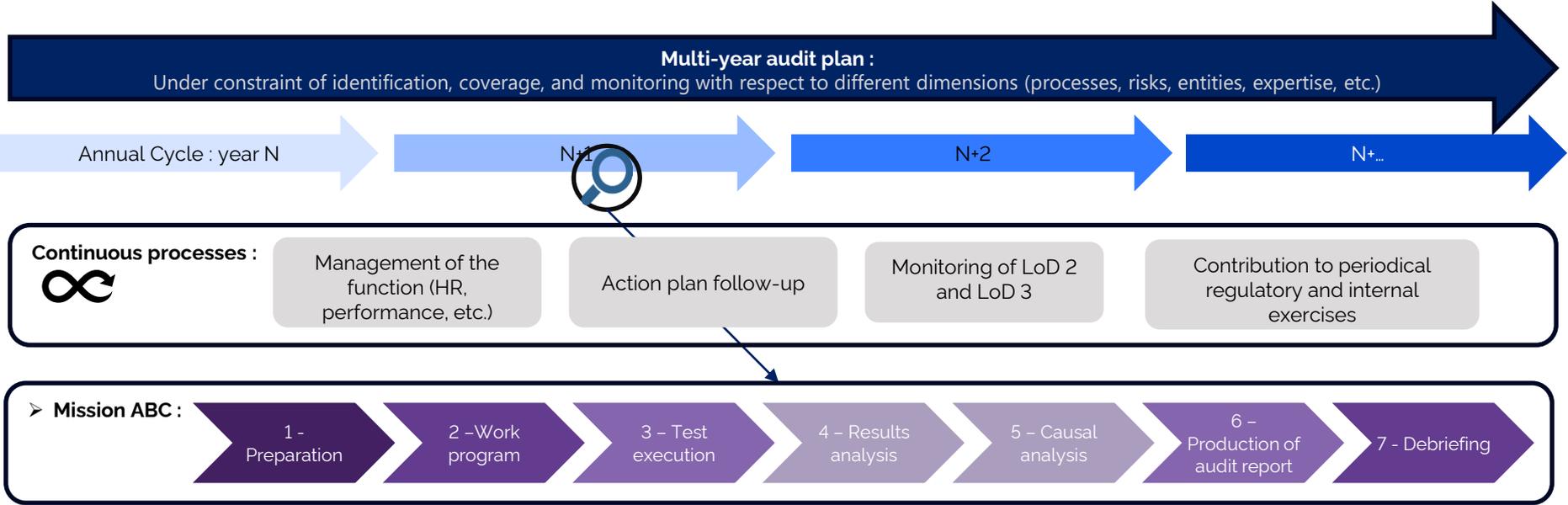
- **Artificial Intelligence:** Broad term referring to techniques that aim to replicate human actions, e.g.: expert systems, min-max, search methods.
- **Machine Learning:** set of programs that aim to execute a task without being explicitly programmed to for said tasks, e.g.: logistic or linear regression, SVM.
- **Deep Learning:** set of machine learning methods that focuses specifically on utilizing neural networks with multiple layers, e.g.: Deep neural nets, convolutional networks.
- **Generative AI:** models that can generate content in the form of text, images, sound, videos, etc.
- **Large Language models:** text generators trained on large corpus of documents (Chat GPT, Claude, Grok...)

2 – Relevance for internal audit processes



For which one of these processes can AI prove helpful?

2 – Relevance for internal audit processes



For which one of these processes is AI helpful?



3 – Use cases (1) – Decision support for audit plan construction

Scenario:

Corporations or audit departments specifically, usually have a taxonomy of the audit universe across different dimensions : entities x risks, or processes x risks or entities x processes x risks x information systems, etc. How do we make sure we prioritize the right audit objects within the audit plan?

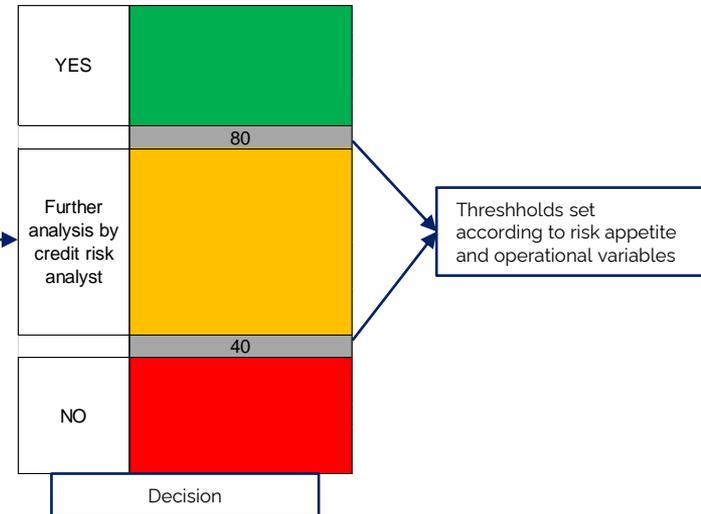
Usually → Judgement selection by supervisors challenged and validated by the Audit Director.

Can this decision-making process be facilitated? → Yes

The idea: Statistical credit scorecards

Characteristic	Bin	Score
Age	0 - 18	10
	18 - 32	20
	32 - 40	45
	40 - 55	50
	55 - 67	40
	67 +	30
Professional activity	Unemployed or in training	0
	Self-employed	30
	Public sector	45
	Private sector	35
Primary residence	Rented	5
	owned - house	25
	owned - apartment	20

Client 100001 :
- Age 42 → Score = 110
- Private Sector
- owned house



Retained variables and scorecard is product of a thorough modeling process

3 – Use cases (1) – Decision support for audit plan construction

Let's transpose this methodology to an audit space. For the sake of this example, let's consider IT systems.

Base data:

Any data expressing base and residual risk exposition

Useful data
Various criticality levels (DORA, prudential, risk based...)
Incident related data
IT asset management data
lod 2 related data
lod 3 related data
rating of provider
cloud based vs on premise
availability statistics

Modeling process (iterative):
Variable selection, Binning,
model selection, training, etc.

Target variable should be something indicative of residual risk. e.g. : was an action plan regarding the IT system issued in the previous audit cycle ?

Outcome:

With the scorecard, one can order IT systems by score, which describes a certain measure of residual risk to be covered by Internal Audit. The top n systems, with n chosen depending on available resources, can be prioritized for the next audit plan.

Even if the model is not statistically representative, the method is still useful to help prioritize by risk.

3 – Use cases (2) - Test points on complete data to cover high impact low probability risks

Scenario:

A process is being audited with a risk with low probability of occurrence but very high impact. For example, identifying failures in KYC processes for sanctions and embargoes.

Classical audit methods → A large enough sample to gather statistically meaningful information.

What level of assurance at what price?

Population size: 10^4
Intuitive proportion: 50%

Sample size:

Marge d'erreur	Niveau de confiance										
	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,1	
0,99	6239	2932	1557	940	623	441	328	253	201	164	
0,98	5751	2528	1307	780	514	363	269	208	165	134	
0,97	5406	2274	1157	686	450	317	235	181	144	117	
0,96	5133	2087	1049	619	405	285	211	163	129	105	
0,95	4899	1937	965	567	370	260	193	148	118	96	
0,94	4694	1811	895	524	342	240	178	137	108	88	
0,93	4508	1703	836	488	318	223	165	127	101	80	
0,92	4338	1608	785	454	298	209	154	119	94	77	
0,91	4182	1523	740	430	280	196	145	112	88	72	
0,9	4035	1447	700	408	264	185	137	105	83	68	
0,89	3898	1377	663	384	250	175	129	99	79	64	
0,88	3767	1313	630	364	237	166	122	94	75	61	
0,87	3644	1254	599	346	225	157	116	89	71	57	
0,86	3526	1199	571	330	214	149	110	85	67	55	
0,85	3413	1147	545	314	204	142	105	81	64	52	
0,84	3305	1099	520	300	194	136	100	77	61	50	
0,83	3201	1053	498	286	185	130	96	74	58	47	
0,82	3101	1011	476	274	177	124	91	70	56	45	
0,81	3005	970	456	262	169	118	87	67	53	43	
0,8	2911	931	437	251	162	113	84	64	51	41	
0,79	2821	895	419	240	155	108	80	62	49	40	
0,78	2734	860	402	230	149	104	77	59	47	38	
0,77	2649	827	385	221	143	100	73	56	45	36	
0,76	2566	795	370	212	137	95	70	54	43	35	
0,75	2486	764	355	203	131	91	68	52	41	33	

Population size: 10^6
Intuitive proportion: 20%

Sample size:

Marge d'erreur	Niveau de confiance										
	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,1	
0,99	10905	2647	1179	664	425	295	217	166	132	107	
0,98	8845	2161	963	541	347	243	177	136	107	87	
0,97	7479	1881	837	471	302	210	154	118	94	76	
0,96	6704	1685	750	422	270	188	138	106	84	68	
0,95	6109	1535	683	384	246	171	126	97	76	62	
0,94	5628	1413	629	354	227	158	116	89	70	57	
0,93	5226	1312	584	329	211	146	108	83	65	53	
0,92	4880	1225	545	307	197	137	101	77	61	50	
0,91	4578	1149	511	288	184	128	94	72	57	46	
0,9	4311	1082	481	271	174	121	89	68	54	44	
0,89	4071	1021	454	256	164	114	84	64	51	41	
0,88	3853	966	430	242	155	108	79	61	48	38	
0,87	3655	917	408	230	147	102	75	58	46	37	
0,86	3473	871	388	218	140	97	72	55	44	35	
0,85	3305	829	369	208	133	93	68	52	41	34	
0,84	3149	790	351	198	127	88	65	50	39	33	
0,83	3004	753	335	189	121	84	62	48	38	31	
0,82	2868	719	320	180	116	80	59	45	36	29	
0,81	2741	687	306	172	110	77	57	43	34	28	
0,8	2621	657	292	165	106	73	54	42	33	27	
0,79	2508	628	280	158	101	70	52	40	32	26	
0,78	2402	602	268	151	97	67	50	38	30	25	
0,77	2301	577	257	145	93	65	48	37	29	24	
0,76	2205	552	246	139	89	62	46	35	28	23	
0,75	2113	530	236	133	85	59	44	34	27	22	

If any failure of the process is critical, sampling is not sufficient.

3 – Use cases (2) - Test points on complete data to cover high impact low probability risks

Solution:

Test 1: Is there one document of each required type attached to the file of each client (Identity, domicile, occupation, etc.)? → Simple procedure

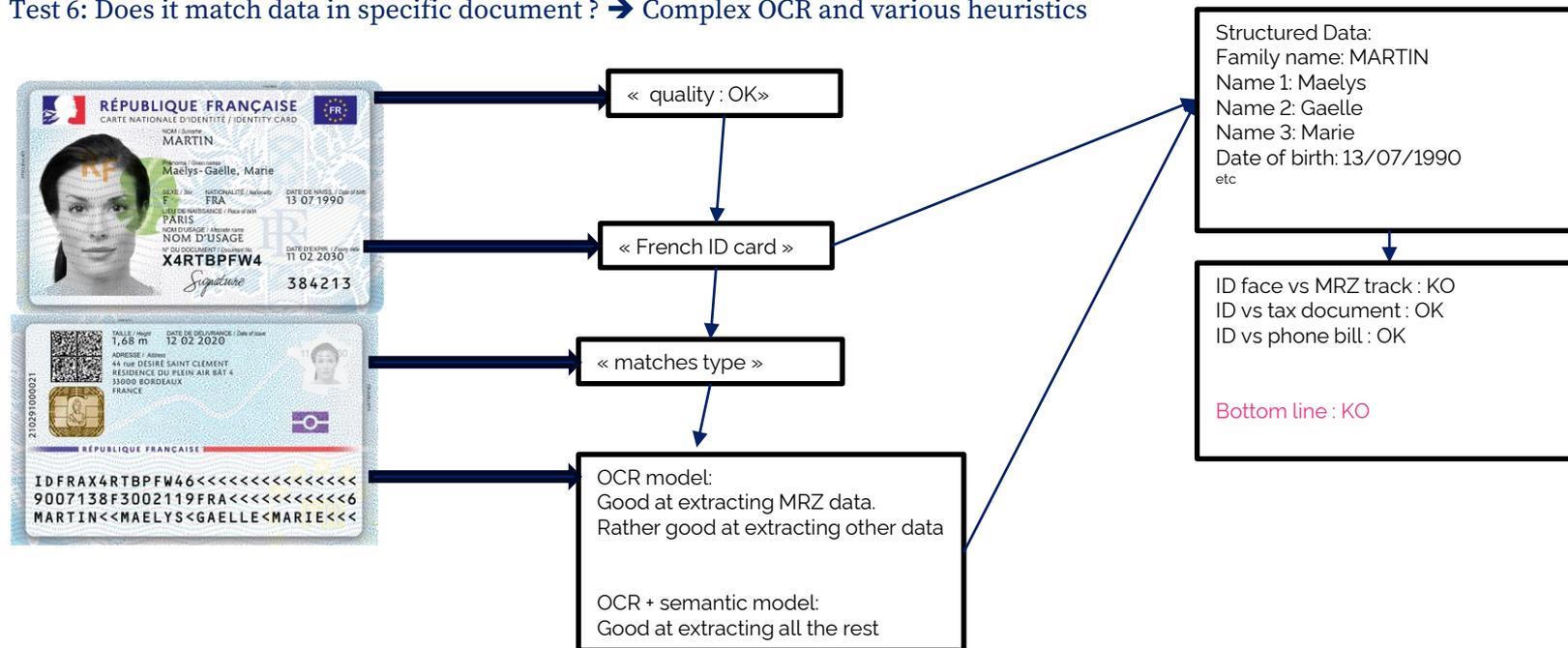
Test 2: Is it readable? → Classification algorithm

Test 3: Does the type of each document in the metadata match the type of the scanned document? → Classification algorithm

Test 4: Does it match KYC Data? → Complex OCR and various heuristics

Test 5: Does the content of the scanned document match from document to document ? → Complex OCR and various heuristics

Test 6: Does it match data in specific document ? → Complex OCR and various heuristics



3 – Use case (3) – Modeling operational risk

Scenario:

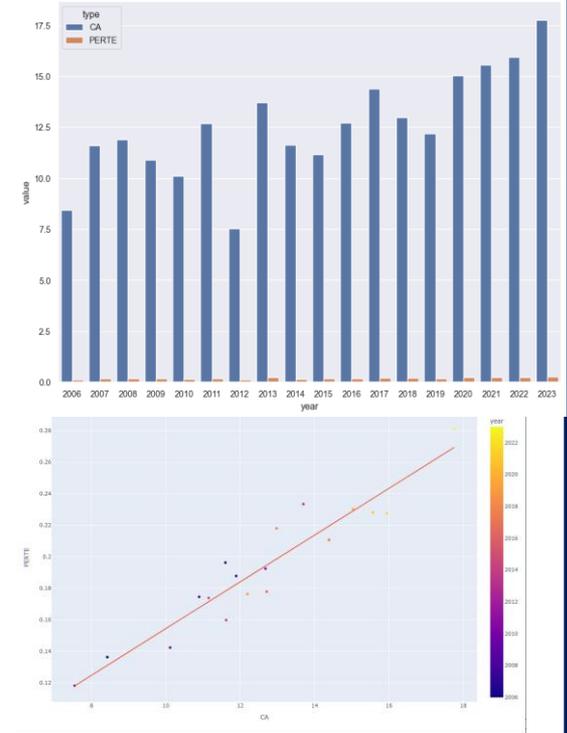
A manufacturing company uses an internal incident management process.

A specific production line faces incidents which need to be minimized in an efficacious fashion..

- Access to the database of incidents allows us to observe the following datapoints :
- By modeling the link between production volumes and losses we find the following equation:
$$\text{Losses} = 0,0148 \times \text{production} + 0,0061 \text{ (in M€)}$$
- Root cause analysis allows us to identify one of the causes. It is linked to a specific operation in the manufacturing process. The action plan calls for a risk management measure to be deployed. In the following year, we observe:

	Year 2026	
	Production	Operational loss
Without the measure (projected)		0,28
With the measure (observed)	18,6	0,15

- The linear parameter has therefore gone from 1,48% to 0,77% thereby lowering the probability of occurrence of the risk by 50%.



3 – Use case (4) – Challenging an indicator and quantifying action plan efficacy

Scenario:

Let us consider that a trading desk manages the risk linked to a simple portfolio composed of 1 ticker. The risk is quantified via an indicator and a limit.

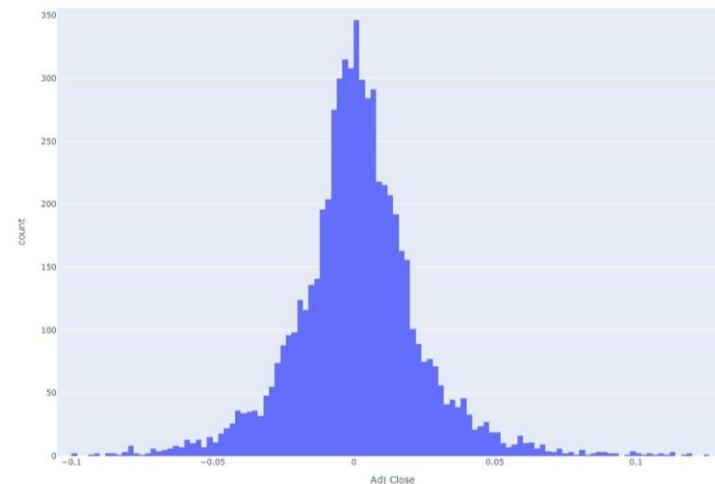
- The audit mission identifies that the indicator chosen is not compliant with regulatory requirements (documentation-based audit). The desk uses daily VaR (95%).
- But only a data-based analysis can allow us to quantify how inadequate the indicator is.
- Let us consider a more complete indicator. Expected shortfall with the same level of appetite (95%).

Calculations based on historical AAPL prices yield:

VaR 95%: 19500€

ES 95%: 29300€

- The indicator underestimates as high as 10^4 € worth of risk, which represents 1% of the value of the portfolio.



Distribution of daily returns of AAPL between "2000-01-01" and "2024-11-01"; Credits : Fahd Kaatich

3 – Use case (5) – Bibliography and introduction to technical topics

Scenario:

Auditor needs an introduction to a model that the audited department happens to be using.

Example:

give me a quick presentation and a bibliography on the SABR pricing model



Quick Presentation on the SABR Pricing Model

Introduction:

The **SABR (Stochastic Alpha, Beta, Rho)** model is a stochastic volatility model used primarily in the derivatives markets to capture the volatility smile. Developed by Patrick S. Hagan, Deep Kumar, Andrew Lesniewski, and Diana Woodward in 2002 ^①, it has become an industry standard for pricing options, especially in interest rate derivatives.

Key Components:

1. **Alpha (α):** Represents the volatility of volatility.
2. **Beta (β):** Controls the elasticity of the volatility with respect to the underlying asset price.
3. **Rho (ρ):** Correlation between the underlying asset price and its volatility.

3 – Use case (6) – Everything regarding text treatment

Scenario:

Auditor needs to compare the content of 2 distinct texts (e.g.: an international accounting standard and an internal accounting norms or procedure) and identify differences.

Example:

Compare Marlowe's Faust and Goethe's Faust

 Copilot

Certainly! Both Christopher Marlowe's "Doctor Faustus" and Johann Wolfgang von Goethe's "Faust" are seminal works that explore the legend of Faust, a scholar who makes a pact with the devil. However, they differ significantly in themes, characterization, and philosophical outlook.

Themes and Ideas:

Ambition and Knowledge:

- Marlowe's "Doctor Faustus": Focuses on the Renaissance theme of the quest for knowledge and power. Faustus's ambition leads him to make a pact with Mephistopheles, seeking unlimited knowledge and worldly pleasures. His journey is marked by hubris and the tragic consequences of overreaching ambition.
- Goethe's "Faust": Explores a broader range of themes, including the quest for meaning, the nature of human striving, and redemption. Goethe's Faust is driven by a deep dissatisfaction with the limits of human knowledge and experience, seeking a more profound understanding of life.

translate dante's inferno canto 26 into english, but use italian american slang

 Copilot

Alright, let's give this a fun twist! Here's a translation of a part of Dante's Inferno, Canto 26 using some Italian-American slang:

Yo, so Dante and Virgil, they're cruising through the eighth circle, right? They hit this spot called the Evil Counselors.

Dante's like, "Hey, Virg, what's with all these flames?"

Virgil goes, "Fuhgeddaboutit, each flame's got a sinner inside, like a wiseguy who gave bad advice."

summarize julius caesar's commentaries on the gallic war

 Copilot

Certainly! Julius Caesar's "Commentarii de Bello Gallico" (Commentaries on the Gallic War) is a firsthand account of his campaigns in Gaul from 58 BCE to 50 BCE. Here's a brief summary of the key points:

Structure:

The work is divided into **eight books**, each detailing a year of the campaigns.

Key Points:

Book 1 (58 BCE):

- **Helvetii Campaign:** Caesar describes his defeat of the Helvetii tribe, who were attempting to migrate through Roman territory.
- **Ariovistus:** Caesar confronts and defeats the Germanic leader Ariovistus, securing Roman control over Gaul.

3 – Use case (7) – Generating code that implements a control point

Scenario:

We need to run a test to check if internal procedures have been respected. This section of the work program was assigned to an auditor not acquainted with the programming language used in the transaction IT system (or with prog lang in general). **Do we need to resort to a sample test?**

→ **No. Thanks to Generative AI, the test can be conducted on the full dataset by anyone:**

Solution: A prompt describing the test to conduct allows us to implement simple testing procedures.

Input

Generate python code that goes through a table called "card_transactions" and checks if the amount of the transaction, represented in the column "amount" is higher than 10000€, if it is, check if there was a manual AML check on the transaction. If there was return OK if not return the line as an anomaly.

However, outputs have to be verified through test cases!!!

Output

```
# Function to run checks
def run_checks():
    with engine.connect() as connection:
        # Fetch all transactions
        transactions = connection.execute(select([card_transactions])).fetchall()

        anomalies = []

        for transaction in transactions:
            amount = transaction['amount']
            aml_check = transaction['aml_check'] # Assuming 'aml_check' column exists
            and indicates manual AML check

            if amount > 10000:
                if aml_check:
                    print(f"Transaction ID {transaction['id']}: OK")
                else:
                    anomalies.append(transaction)

        return anomalies
```

3 – Use case (7) – Generating code that implements a control point

Particularly useful for programming languages that have almost disappeared from the labor market:

Input

translate the code to COBOL

Output

```
CHECK-TRANSACTION.  
  IF AMOUNT > 10000  
    IF AML-CHECK = 'Y'  
      DISPLAY "Transaction ID " TRANSACTION-ID " : OK"  
    ELSE  
      ADD 1 TO WS-ANOMALY-COUNT  
      DISPLAY "Anomaly found in transaction ID " TRANSACTION-ID  
    END-IF  
  END-IF.
```

4 – Closing remarks

Many other applications:

- The good: translation, audit program ideas, AI agent does the counter-modeling.
- The bad: Automatic root cause analysis, asking AI agents to think critically or to make risk-based decisions, minutes of meetings or anything that requires understanding or flair.
- The ugly: Automatic note taking during meetings, report drafting, etc.

Access to tools:

- Can be a pain for ready use tools like Chat-GPT and depends on IT departments and overall, AI strategy of the company you operate in.
- However, good old fashioned (python or substitutes) + local models goes a long way and is usually available for all.

The right tool for the right auditor:

- Teams should be layered structure: experts, occasional contributors, users.
- Generalizable use cases have to be generalized.
- Make reusable tools as simple to use as possible.

AI capabilities unlock new audit objects:

- We are making use of a new technology to make our job easier, but we also must act as LOD 3 regarding this new technology.
- Mature enough use cases, as long as they contribute to managing some risk and are used periodically, should be shipped to LOD 2 or even LOD 1.

A few words to keep in mind:

- Always keep a critical distance with generative AI outputs. Never take anything at face value.
- The best questions and use cases are the ones you can verify immediately (i.e.: AI writes some code you cannot write yourself, or could but want to be quick about it, code must be verified).
- Some information can be wrong (hallucination) or outdated (datasets are not updated as frequently as you would think). This is especially relevant when you check for regulatory texts.

The golden rule:

- Always tinker, implement and test things out.

For further questions :
kaatich.fahd@protonmail.ch