Eidgenössische Finanzmarktaufsicht FINMA
Autorité fédérale de surveillance des marchés financiers FINMA
Autorità federale di vigilanza sui mercati finanziari FINMA
Swiss Financial Market Supervisory Authority FINMA

FINMA Guidance 08/2024

Governance and risk management when using artificial intelligence

18 December 2024

Laupenstrasse 27, 3003 Bern, Tel. +41 (0)31 327 91 00, www.finma.ch

Contents

1 Introduction

The use of artificial intelligence (AI) in the financial market is increasing.[1] For supervised institutions, this is associated with both opportunities and risks. In this guidance, FINMA draws attention to the corresponding risks and the need to adequately identify, limit and control these risks.

To date, there is no AI-specific legislation in Switzerland. In financial market law, the technology-neutral, principle-based regulatory requirements for effective governance and risk management cover the risks arising from the use of AI. In line with international requirements, FINMA expects supervised institutions that use AI to actively consider the impact of this use on their risk profile and to align their governance, risk management and control systems accordingly. Besides the size, complexity, structure and risk profile of the supervised institutions, the materiality of the AI applications used and the probability that the risks resulting from the use of these applications will materialise must be taken into account. [2]

2 Findings from supervision

The risks from the use of AI are mainly in the area of operational risks,[3] in particular model risks (e.g. lack of robustness, correctness, bias and explainability) as well as IT and cyber risks. They also result from a growing dependence on third parties such as providers of hardware solutions, models or cloud services in an increasingly concentrated market.[4] Finally, there are legal and reputational risks as well as challenges in the allocation of responsibilities due to the autonomous and difficult-to-explain actions of these systems and scattered responsibilities for AI applications at supervised institutions.

The following are examples of measures to address specific risks resulting from AI applications that FINMA has observed as part of its ongoing supervision, namely during supervisory discussions and in initial specific on-

[1] On the adoption of AI in the financial market, see FSB, The Financial Stability Implications of Artificial Intelligence, 14 November 2024 (hereinafter: FSB), p. 3 ff.

[2] Possible (non-exhaustive) factors that influence the materiality of an application are: Significance for compliance with financial market legislation, financial impact on the company, legal and reputational risks, relevance of the product for the company, number of clients or investors affected, types of clients or investors (retail/institutional), importance of the product for clients or investors, consequences of errors or failure. Possible (non-exhaustive) factors that influence the probability of the events associated with the risks materialising are as follows: Complexity (e.g. explainability, predictability), type and amount of data used (e.g. unstructured data, integrity, appropriateness, personal data), unsuitable development or monitoring processes, degree of autonomy and process integration, dynamics (e.g. short calibration cycles), linkage of several models, potential for attacks or failures (e.g. increased due to outsourcing).

[3] See Art. 89 CAO: The term “operational risk” refers to the risk of loss resulting from the inappropriateness or failure of internal procedures, people or systems, or from external events.

[4] See also FSB, p. 16 ff.

site supervisory reviews. This is intended to support supervised institutions in identifying, assessing, managing and monitoring risks from internal and external AI applications.

2.1 Governance

FINMA observed that supervised institutions focus primarily on data protection risks, but less on model risks such as lack of robustness and correctness, bias, lack of stability and explainability. In addition, the development of AI applications is often decentralised, making it challenging to implement consistent standards, assign responsibilities clearly to employees with the appropriate skills and experience and address all relevant risks. In the case of externally purchased applications and services, the supervised institutions sometimes had difficulties determining whether AI is included, which data and methods are used and whether sufficient due diligence exists.

FINMA assessed whether supervised institutions with many or significant applications have AI governance in place, including a centrally managed inventory with a risk classification and resulting measures, the definition of responsibilities and accountabilities for the development, implementation, monitoring and use of AI, requirements for model testing and supporting system controls, documentation standards and broad training measures. In the case of outsourcing, it assessed whether the supervised institutions had implemented additional tests, controls and contractual clauses governing responsibilities and liability issues and ensured that the third parties entrusted with the outsourcing had the necessary skills and experience.

2.2 Inventory and risk classification

FINMA observed that some supervised institutions defined AI narrowly in order to focus on supposedly larger or new risks. For many supervisors, it was a challenge to ensure the completeness of inventories, as AI development and use is often widely spread across the organisation and, since the advent of generative AI, applications are accessible to everyone. Furthermore, not all supervised institutions had established consistent criteria for identifying applications that require special attention in risk management due to their materiality, specific risks and probability of these materialising.[5]

FINMA assessed whether the supervised institutions had a sufficiently broad definition of AI,[6] as traditional applications can also present similar risks and

[5] Risks tend to be higher when AI is used to comply with supervisory law or to perform critical functions, or when customers or employees are strongly affected by its results. The criteria for classification should be defined by the supervised institutions.

[6] See the OECD's definition approach: OECD, Explanatory Memorandum on the Updated OECD Definition of an AI System, OECD Artificial Intelligence Papers, March 2024 (No. 8).

the same risks must be addressed in the same way.[7] It then assessed the existence and completeness of AI inventories and the risk classification of AI applications.

2.3 Data quality

FINMA observed that some supervised institutions have not defined any requirements or controls to ensure data quality for AI applications.

AI applications often learn from data automatically and without human intervention. Data quality is therefore often more important than the selection of the specific model. At the same time, data can be incorrect, inconsistent, incomplete, unrepresentative or outdated and therefore of poor quality. Historical data may contain a bias that is carried forward into future forecasts, or it may no longer be representative of the forecast due to a change in the environment. In the case of purchased solutions, the supervised institutions often have no influence on or knowledge of the underlying data. This can lead to these not being suitable for the supervised institutions or the specific issue and the risk of the unconscious use of deliberately manipulated data increasing. Since the increased use of AI, more unstructured data such as texts and images are also being analysed, which can make it difficult to assess quality.

FINMA assessed whether the supervised institutions have defined requirements in their internal rules and directives to ensure that data is complete, correct and of integrity and that the availability of and access to data is secured.

2.4 Tests and ongoing monitoring

FINMA observed weaknesses in the selection of performance indicators, tests and ongoing monitoring at some of the supervised institutions.

FINMA assessed whether the supervised institutions schedule tests to ensure the data quality and functionality of the AI applications, which include checks for accuracy, robustness and stability and, if necessary, bias.[8] It assessed whether experts in the respective area of application provided questions and predefined expectations and whether performance indicators were defined in advance in order to assess how well an AI application

[7] AI is not a high-risk application per se. The risk associated with it depends on the complexity, adaptivity and autonomy of the respective application, its area of application and its integration into processes.

[8] There are a variety of tests to assess the performance and results of an application. These include tests in which the user knows the correct result and checks whether the application delivers it (e.g. backtesting, out-of-sample testing), constructed tests to understand how the application behaves in certain borderline cases (e.g. sensitivity analyses or stress testing), tests with incorrect input data (e.g. adversarial testing), or tests against additional, possibly simpler benchmark models. Tests can also be used to assess potential application limits and to check results for “repeatability”.

achieves the set goals.[9] With regard to regular checks to be carried out, FINMA assessed, for example, whether the supervised institutions had defined thresholds or other validation methods to ensure the correctness and ongoing quality of the outputs.[10] It also assessed whether the supervised institutions monitor changes in input data to ensure that models remain applicable in a changing environment (recognition and treatment of data drift). Monitoring also includes analysing cases in which the output has been ignored or changed by users, as such manual corrections can provide information about weaknesses. Finally, FINMA assessed whether the supervised institutions give prior consideration to recognising and handling exceptions.

2.5 Documentation

FINMA observed that some supervised institutions do not have centralised documentation requirements and that some of the existing documentation is not sufficiently detailed and recipient-oriented.

For material applications, FINMA assessed whether the supervised institutions address the purpose of the application, data selection and preparation, model selection, performance measures, assumptions, limitations, testing and controls as well as fallback solutions in the documentation. Regarding the selection of data, FINMA considered whether the supervised institutions presented data sources and data quality checks including integrity, correctness, appropriateness, relevance, bias and stability. It also considered how the supervised institutions ensure the robustness, reliability and traceability of the application and whether they carry out an appropriate categorisation into a risk category and the associated justification and review.

2.6 Explainability

FINMA observed that results often cannot be understood, explained or reproduced and therefore cannot be critically assessed.

Where decisions had to be justified to investors, clients, employees, the supervisory authority or the audit firm, FINMA assessed the explainability of the applications in greater depth. This includes understanding the drivers of the applications or the behaviour under different conditions in order to be able to assess the plausibility and robustness of the results.

[9] The more essential and complex the application and the less is known about how the system works or the underlying data, the more important it is to assess whether the application is working according to its purpose before productive use, in the event of changes and – especially due to the adaptive nature of today's applications – on an ongoing basis. It is also important to consider fallback mechanisms in order to be prepared if the AI develops in an undesirable direction and no longer fulfils the originally defined objectives.

[10] Sampling, backtesting, predefined test cases or benchmarking, for example, can contribute to this.

2.7 Independent review

FINMA did not observe a clear distinction between the development of AI applications and the independent review in all cases.

It also observed that only a few supervised institutions carry out an independent review of the entire model development process by qualified personnel in order to consistently identify and reduce model risks.

For material applications, FINMA assessed whether the independent review included the submission of an objective, informed and unbiased opinion on the appropriateness and reliability of a process for a particular application and whether the results of the independent review were taken into account in the development of the application.

3 Outlook

The understanding of risks associated with the use of AI by supervised institutions is still developing. Based on its supervisory experience, and in line with relevant international developments, FINMA will also refine its expectations of appropriate governance and risk management by supervised institutions in connection with AI and, where necessary, make them transparent in the market. As with other relevant risk drivers, FINMA strives for a technology-neutral, proportional and standardised approach across all sectors, taking into account significant differences between the sectors and international standards.