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

Communication FINMA sur la surveillance 08/2024

Gouvernance et gestion des risques en lien avec l'utilisation de l'intelligence artificielle

18 décembre 2024

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

Table des matières

1 Introduction

L'intelligence artificielle (IA) est toujours plus utilisée sur les marchés financiers.[1] Pour les assujettis, cela implique des opportunités, mais aussi des risques. Par la présente communication sur la surveillance, la FINMA attire l'attention sur les risques en question et sur la nécessité de les identifier, de les limiter et de les contrôler de manière adéquate.

Il n'existe pas, jusqu'à présent, de législation spécifique à l'IA en Suisse. Dans le droit des marchés financiers, les exigences prudentielles – neutres à l'égard de la technologie et fondées sur des principes – relatives à une gouvernance et une gestion des risques efficaces couvrent les risques liés à l'utilisation de l'IA. Conformément à la pratique internationale, la FINMA attend des assujettis utilisant l'IA qu'ils s'intéressent activement aux répercussions de cette utilisation sur leur profil de risque et qu'ils adaptent en conséquence leur gouvernance, leur gestion des risques et leurs systèmes de contrôle. Outre la taille, la complexité, la structure et le profil de risque des assujettis, il faut notamment tenir compte de l'importance des applications d'IA ainsi que de la probabilité que les risques découlant de l'utilisation de ces applications se réalisent. [2]

2 Enseignements tirés de la surveillance

Les risques liés à l'utilisation de l'IA se trouvent principalement dans le domaine des risques opérationnels[3], notamment les risques liés aux modèles (par ex. le manque de robustesse, d'exactitude ou d'explicabilité ainsi que les biais) ainsi que les risques informatiques et les cyberrisques. Ils résultent également d'une dépendance croissante envers des tiers, tels que les fournisseurs de solutions matérielles, de modèles ou de services *cloud*, sur un marché toujours plus concentré.[4] Enfin, il existe des risques juridiques et de

[1] Voir le rapport du CSF « The Financial Stability Implications of Artificial Intelligence » du 14.11.2024, p. 3 ss, concernant l'utilisation de l'IA sur les marchés financiers (ci-après : rapport CSF).

[2] Parmi les facteurs potentiellement déterminants pour le caractère important d'une application, on compte (liste non exhaustive) : l'importance pour le respect de la législation sur les marchés financiers, les conséquences financières pour l'entreprise, les risques juridiques et de réputation, l'importance du produit pour l'entreprise, le nombre de clients ou d'investisseurs concernés, le type de clients ou d'investisseurs (détail/institutionnel), l'importance du produit pour les clients ou les investisseurs, les conséquences en cas d'erreur ou de défaillance. Parmi les facteurs potentiellement déterminants pour la probabilité de survenance des événements liés aux risques, on compte (liste non exhaustive) : la complexité (par ex. explicabilité, prévisibilité), le type et la quantité de données utilisées (par ex. données non structurées, intégrité, pertinence, données personnelles), le caractère inadapté des processus de développement ou de surveillance, le degré d'autonomie et d'intégration des processus, la dynamique (par ex. cycles de calibrage courts), l'interconnexion de plusieurs modèles, le potentiel d'attaques ou de défaillances (par ex. accru en raison d'une externalisation).

[3] Cf. art. 89 OFR : Par risque opérationnel, on entend le risque de perte lié à l'inadéquation ou à la défaillance de processus ou de systèmes internes, à la défaillance de personnes ou encore à des facteurs externes.

[4] Cf. rapport CSF, p. 16 ss.

réputation ainsi que des difficultés concernant l'attribution des responsabilités en raison du fonctionnement autonome et difficilement explicable de ces systèmes et de la dispersion des compétences relatives aux applications d'IA chez les assujettis.

Dans le cadre de la surveillance courante, notamment lors d'entretiens de surveillance et des premiers contrôles spécifiques sur place, la FINMA a observé des mesures visant à traiter les risques résultant spécifiquement des applications d'IA. Des exemples de telles mesures sont présentés ci-après dans le but d'aider les assujettis à identifier, évaluer, gérer et contrôler les risques liés aux applications d'IA internes et externes.

2.1 Gouvernance

La FINMA a constaté que les assujettis se focalisent en premier lieu sur les risques liés à la protection des données, mais moins sur les risques liés aux modèles tels que le manque de robustesse et d'exactitude, les biais ou le manque de stabilité et d'explicabilité. Le développement d'applications d'IA est par ailleurs souvent décentralisé ; il est de ce fait difficile de mettre en œuvre de manière systématique des normes, d'attribuer des responsabilités claires à des collaborateurs disposant des compétences et de l'expérience nécessaires et de gérer tous les risques pertinents. En ce qui concerne les applications et les services acquis auprès de tiers, il est parfois difficile pour les assujettis de déterminer si elles impliquent de l'IA, quelles données et méthodes sont utilisées et si une diligence raisonnable suffisante existe.

La FINMA a évalué si les assujettis ayant recours à de nombreuses applications ou à des applications importantes disposaient d'une gouvernance en matière d'IA comprenant notamment un inventaire centralisé, y compris une classification des risques et les mesures qui en découlent, la détermination des compétences et des responsabilités pour le développement, la mise en œuvre, la surveillance et l'utilisation de l'IA, des prescriptions relatives aux tests des modèles et aux contrôles auxiliaires du système, des normes de documentation ainsi que des formations étendues. En cas d'externalisation, elle a évalué si les assujettis avaient mis en place des tests, des contrôles et des clauses contractuelles supplémentaires régissant les compétences et les questions de responsabilité et s'ils s'étaient assurés que les tiers auxquels ils faisaient appel disposaient des compétences et de l'expérience nécessaires.

2.2 Inventaire et classification des risques

La FINMA a constaté que les assujettis définissaient parfois l'IA de manière étroite pour se concentrer sur des risques considérés comme plus importants ou nouveaux. Pour de nombreux assujettis, il est difficile de garantir l'exhaustivité des inventaires, car le développement et l'utilisation de l'IA

sont souvent largement répandus dans l'entreprise. De plus, des applications sont accessibles à tous depuis l'avènement de l'IA générative. Tous les assujettis n'ont par ailleurs pas établi de critères systématiques pour identifier les applications qui, en raison de leur importance, des risques spécifiques qu'elles impliquent et de la probabilité de réalisation desdits risques, nécessitent une attention particulière dans la gestion des risques.[5]

La FINMA a évalué si les assujettis avaient une définition suffisamment large de l'IA ;[6] les applications classiques peuvent en effet présenter des risques similaires et les mêmes risques doivent être traités de la même manière.[7] Elle a ensuite évalué l'existence et l'exhaustivité des inventaires d'IA ainsi que la classification des risques des applications d'IA.

2.3 Qualité des données

La FINMA a constaté que tous les assujettis n'avaient pas défini des prescriptions et des contrôles pour garantir la qualité des données dans les applications d'IA.

Les applications d'IA apprennent souvent de manière automatisée à partir des données, sans intervention humaine. La qualité des données est donc souvent plus importante que le choix du modèle concret. Les données peuvent toutefois être erronées, incohérentes, incomplètes, non représentatives ou obsolètes et donc de mauvaise qualité. Les données historiques peuvent contenir un biais qui se répercute sur les prévisions futures ou peuvent ne plus être représentatives pour les prévisions en raison d'évolutions de l'environnement. Dans le cas de solutions acquises auprès de tiers, les assujettis n'ont souvent aucune influence sur les données sous-jacentes, voire ne les connaissent pas. Il est donc possible que celles-ci ne soient pas appropriées pour les assujettis ou l'utilisation concrète et le risque d'utiliser inconsciemment des données sciemment manipulées augmente. L'utilisation accrue de l'IA s'accompagne en outre d'une augmentation de l'exploitation de données non structurées, telles que du texte et des images, pour lesquelles la qualité est difficile à évaluer.

[5] Les risques sont généralement plus importants lorsque l'IA est utilisée pour respecter le droit de la surveillance ou pour exécuter des fonctions critiques, ou lorsque la clientèle ou les collaborateurs sont fortement concernés par ses résultats. Les assujettis devraient définir les critères de classification.

[6] Voir la définition de l'OCDE : OCDE, « Explanatory memorandum on the Updated OECD Definition of an AI System », OECD Artificial Intelligence Papers, mars 2024 (n° 8).

[7] L'IA n'est pas en soi une application à haut risque. Le risque associé à l'IA dépend de la complexité, de l'adaptabilité et de l'autonomie de l'application en question, de son champ d'application et de son intégration dans les processus.

La FINMA a évalué si les assujettis ont défini dans leurs instructions et directives internes des prescriptions visant à garantir l'exhaustivité, la correction et l'intégrité des données et à s'assurer que celles-ci sont disponibles et accessibles.

2.4 Tests et surveillance constante

La FINMA a constaté des faiblesses chez certains assujettis concernant le choix des indicateurs de performance, les tests et la surveillance courante.

La FINMA a évalué si les assujettis prévoient des tests pour garantir la qualité des données et le bon fonctionnement des applications d'IA, comprenant des contrôles de l'exactitude, de la robustesse et de la stabilité ainsi que, le cas échéant, des biais.[8] Elle a évalué si les spécialistes du domaine d'application concerné avaient posé des questions et défini des attentes à ce sujet et si des indicateurs de performance avaient été fixés à l'avance pour évaluer dans quelle mesure une application d'IA atteint les objectifs fixés.[9] Concernant les contrôles réguliers, la FINMA a notamment évalué si les assujettis avaient défini des seuils ou d'autres méthodes de validation pour garantir l'exactitude et la qualité continue des résultats.[10] Elle a aussi évalué si les assujettis surveillaient les changements dans les données d'entrée afin de s'assurer que les modèles restent applicables même face à une évolution de l'environnement (détection et traitement des dérives des données). La surveillance comprend aussi l'analyse des cas où le résultat a été ignoré ou modifié par les utilisateurs, puisque de telles corrections manuelles peuvent indiquer des points faibles. Enfin, la FINMA a évalué si les assujettis menaient préalablement des réflexions sur l'identification et le traitement des exceptions.

[8] Il existe une multitude de tests permettant d'évaluer les performances et les résultats d'une application, notamment les tests dans lesquels les utilisateurs connaissent le résultat correct et vérifient si l'application le fournit (par ex. *backtesting*, *out-of-sample testing*), des tests conçus pour comprendre comment l'application se comporte dans certains cas limites (par ex. analyses de sensibilité ou *stress testing*), des tests avec des données d'entrée erronées (par ex. *adversarial testing*), ou encore des tests par rapport à d'autres modèles de référence, le cas échéant plus simples. Les tests permettent aussi d'évaluer les éventuelles limites de l'application et de vérifier la reproductibilité des résultats.

[9] Plus l'application est essentielle et complexe, et moins on en sait sur le fonctionnement du système ou sur les données sous-jacentes, plus il est important d'évaluer en permanence, avant l'utilisation productive, si l'application fonctionne conformément à son objectif en cas d'évolutions, notamment en raison de la capacité d'adaptation des applications actuelles. Il est aussi important de réfléchir à des solutions de repli afin d'être préparé si l'IA évolue dans une direction non souhaitée et ne remplit plus les objectifs initialement définis.

[10] Les échantillons, le *backtesting*, des cas de test prédéfinis ou le *benchmarking* peuvent par exemple être pertinents à cet égard.

2.5 Documentation

La FINMA a constaté que certains assujettis ne disposaient d'aucune prescription centrale en matière de documentation et que la documentation existante n'était pas toujours suffisamment détaillée et orientée vers le destinataire.

Pour les applications importantes, la FINMA a évalué si les assujettis abordaient dans la documentation le but de l'application, la sélection et la préparation des données, le choix du modèle, les mesures de performance, les hypothèses, les limitations, les tests et les contrôles ainsi que les solutions de repli. Concernant la sélection des données, la FINMA a examiné si les assujettis présentaient des sources de données et des contrôles de la qualité des données, y compris l'intégrité, l'exactitude, l'adéquation, la pertinence, les biais et la stabilité. Elle a aussi examiné la manière dont les assujettis assurent la robustesse et la fiabilité ainsi que la compréhensibilité de l'application et s'ils procèdent à une classification appropriée dans une catégorie de risque ainsi qu'à la justification et à l'examen correspondants.

2.6 Explicabilité

La FINMA a constaté que les résultats ne sont souvent pas compris, expliqués ou reproduits et qu'ils ne peuvent donc pas être évalués de manière critique.

La FINMA a évalué de manière approfondie l'explicabilité des applications lorsque des décisions devaient être motivées envers des investisseurs, la clientèle, des collaborateurs, la surveillance ou la société d'audit. Il s'agit notamment de comprendre les vecteurs des applications ou le comportement dans différentes conditions afin de pouvoir évaluer la plausibilité et la robustesse des résultats.

2.7 Vérification indépendante

La FINMA n'a pas toujours constaté de délimitation nette entre le développement d'applications d'IA et la vérification indépendante.

Elle a en outre observé que peu d'assujettis soumettent l'ensemble du processus de développement des modèles à une vérification indépendante par du personnel qualifié en la matière dans le but d'identifier et de réduire systématiquement les risques liés aux modèles.

Pour les applications importantes, la FINMA a évalué si la vérification indépendante inclut l'émission d'un avis objectif, averti et impartial sur l'adéquation et la fiabilité d'une procédure pour un cas d'application donné, et si les conclusions de la vérification indépendante est pris en compte lors du développement de l'application.

3 Perspectives

La compréhension des risques liés à l'utilisation de l'IA chez les assujettis en est encore à ses débuts. Sur la base des expériences acquises dans le cadre de la surveillance et en référence aux évolutions internationales pertinentes, la FINMA va aussi développer ses attentes envers les assujettis en matière de gouvernance et de gestion des risques adéquates en lien avec l'utilisation de l'IA. Au besoin, elle communiquera ces attentes de manière transparente sur le marché. À cet égard, la FINMA vise, comme pour d'autres facteurs de risque pertinents, une approche neutre à l'égard de la technologie, proportionnelle et uniforme dans tous les secteurs, tout en tenant compte des différences importantes qui existent entre les secteurs ainsi que des normes internationales.