

## COMMUNICATION FINMA 08/2024

Philipp Fischer &amp; Antoine Amiguet—OBERSON ABELS SA—www.obersonabels.com

Communication FINMA 08/2024 (lien)  
Gouvernance et gestion des risques en lien avec  
l'utilisation de l'intelligence artificielle (1/3).

Thématiques	Pilier 1: Gouvernance	Pilier 2: Inventaire / classification des risques
Constats de la FINMA	<ul style="list-style-type: none"><li>Les assujettis se concentrent principalement sur les risques liés à la protection des données, et <b>non sur les risques liés aux modèles d'IA</b>.</li><li>Le développement de l'IA est souvent <b>décentralisé</b>.</li><li>En cas d'utilisation d'une application acquise auprès d'un tiers, il est parfois <b>difficile de savoir si elle implique de l'IA</b>.</li></ul>	<ul style="list-style-type: none"><li><b>Définition trop étroite de l'IA</b> par les assujettis</li><li>Difficulté pour certains assujettis d'assurer une <b>exhaustivité de l'inventaire</b>.</li><li><b>Absence de critère</b> permettant d'identifier les applications d'IA qui présentent un risque</li></ul>
Attentes de la FINMA	<ul style="list-style-type: none"><li><b>Inventaire</b> centralisé et complet des applications d'IA</li><li>Les <b>compétences</b> et les <b>responsabilités</b> doivent être clairement définies.</li><li>Mise en place de règles relatives aux <b>tests</b> des modèles d'IA, à leur documentation et à des formations</li><li>En cas d'externalisation, des clauses contractuelles précises concernant les <b>responsabilités du prestataire</b></li></ul>	<ul style="list-style-type: none"><li><b>Définition large et uniforme</b> de l'IA</li><li>Mise en place de <b>critères</b> pour identifier les <b>applications d'IA importantes</b> et les <b>risques spécifiques</b> nécessitant une attention particulière</li></ul>
Mise en œuvre concrète (exemples découlant de la pratique d'OA)	Mise en place d'une <b>directive interne</b> qui alloue des responsabilités (à éviter: projets gérés uniquement par la 1 <sup>ère</sup> ligne ou qui émergent de manière décentralisée) <b>mais important de souligner</b> ce que la FINMA ne dit pas: l'autorité n'interdit pas la prise de décision par le biais de l'IA (mais un "humain" doit en assumer la responsabilité ultime) → autorisation des décisions individuelles automatisées ( <a href="#">art. 21 LPD</a> ).  Compréhension des prestations offerts par des tiers + <b>engagements contractuels</b> à obtenir de tiers (même si marché concentré avec un nombre réduit de prestataires) → enjeux similaires à l'outsourcing	<b>Checklist et heat map</b> pour documenter, pour chaque cas d'utilisation, (i) les <b>model risks</b> ( <b>robustness</b> , caractère correct, <b>bias</b> , stabilité et <b>explainability</b> ) d'IA et (ii), selon les cas, les engagements contractuels pris par des prestataires tiers.  Inventaire avec une classification en fonction des risques

La FINMA a récemment publié sa Communication 08/2024 concernant la gouvernance et la gestion des risques liés à l'utilisation de l'intelligence artificielle (IA) dans le secteur financier.

Ce document reflète les constats de la FINMA à l'égard des pratiques actuelles et ses attentes envers les assujettis pour identifier, limiter et contrôler les risques spécifiques liés à l'IA. En résumé, la FINMA indique avoir observé ce qui suit:

1. Une **compréhension approfondie des risques** et une **structure de gouvernance adéquate** font parfois défaut. La pratique montre que ces deux aspects évoluent plus lentement que la technologie elle-même.

- L'adoption d'une nouvelle technologie comme l'IA s'effectue souvent **de manière décentralisée** et en-dehors des cadres de gouvernance établis.
- Une confiance parfois excessive est accordée aux **prestataires tiers**, alors que la concentration du marché limite le choix des prestataires et augmente le risque de **'vendor lock-in effect'**.
- Les **coûts liés à l'évaluation**, à la **supervision** et au **contrôle** des prestations externalisées est parfois sous-estimé.

Sur cette base, la FINMA formule des attentes, qui sont résumées dans la Legal Update en annexe (aussi disponible en anglais). Nous présentons également des idées de mise en œuvre que nous avons développées et observées dans le cadre de nos échanges avec des clients et d'autres praticiens.

Thématiques	Pilier 3: Qualité des données	Pilier 4: Tests / surveillance constante	Pilier 5: Documentation
Constats de la FINMA	<ul style="list-style-type: none"><li>Tous les assujettis n'ont pas défini des règles et des processus pour garantir la <b>qualité des données</b> dans les applications d'IA.</li></ul>	<ul style="list-style-type: none"><li>Faiblesses relevées dans la planification et la mise en œuvre de <b>tests</b> et de <b>contrôle</b></li><li>Peu d'<b>indicateurs de performance spécifiques</b> sont définis à l'avance.</li></ul>	<ul style="list-style-type: none"><li>Certains assujettis ne disposent pas de <b>directives pour documenter</b> le recours à l'IA.</li><li>Documentation incomplète, peu détaillée, et mal adaptée aux enjeux et risques de l'IA</li></ul>
Attentes de la FINMA	<ul style="list-style-type: none"><li>Établir des <b>instructions/directives internes</b> pour garantir l'exhaustivité, la correction, l'intégrité et l'accessibilité des données utilisées</li></ul>	<ul style="list-style-type: none"><li>Mise en place de <b>processus de test</b> pour vérifier les modèles d'IA et s'assurer que les applications atteignent les objectifs prévus.</li><li><b>Réalisation de contrôle réguliers</b> des réponses d'IA</li></ul>	<ul style="list-style-type: none"><li>Fournir une <b>documentation détaillée pour les applications importantes</b> couvrant: les objectifs de l'application, la fiabilité, les risques, la sélection des données et leur qualité.</li></ul>
Mise en œuvre concrète (exemples découlant de la pratique d'OA)	Processus interne de contrôle de la qualité des données  <b>Scepticisme</b> implicite du régulateur à l'égard de l'utilisation de <b>LLM</b> (car contrôle de la qualité des données très difficile en pratique) → risque que le déploiement de LLM soit soumis à des limites réglementaires à l'avenir?	Définition de <b>KPI</b>  Contrôles <b>ex post</b> pour lutter contre le phénomène de <b>model/data drift</b>  Processus d'audit (le cas échéant par un expert tiers)	Documentation des applications utilisées: (i) objectif des applications, (ii) sélection et préparation des données, (iii) sélection des modèles, (iv) KPIs, (v) les tests et les contrôles et (vi) les solutions de <b>fallback</b>

Thématiques abordées	Pilier 6: Explicabilité	Pilier 7: Vérification indépendante
Constats de la FINMA	<ul style="list-style-type: none"><li>Les résultats des modèles d'IA sont souvent <b>peu reproductibles</b>, ce qui limite la possibilité d'en faire une évaluation critique.</li></ul>	<ul style="list-style-type: none"><li>Des processus de <b>vérification indépendants</b> du développement des modèles d'IA sont rarement mis en œuvre.</li></ul>
Attentes de la FINMA	<ul style="list-style-type: none"><li>Assurer que les résultats des modèles sont <b>compréhensibles</b> pour les parties prenantes, qu'il s'agisse par exemple d'investisseurs, de clients ou de collaborateurs.</li><li>Comprendre les mécanismes de fonctionnement des modèles pour garantir leur plausibilité et robustesse.</li></ul>	<ul style="list-style-type: none"><li>Mettre en place, pour les applications importantes, une <b>vérification indépendante</b> couvrant tout le cycle de son développement, afin d'obtenir des avis objectifs et d'identifier les risques.</li></ul>
Mise en œuvre concrète (exemples découlant de la pratique d'OA)	<b>Due diligence</b> des applications d'IA → objectif (difficile!) qui devrait être visé: reproductibilité de l' <b>output</b> afin de pouvoir comprendre son origine / analyse de sensibilité / indication des sources dans le cadre des projets de <b>Retrieval Augmented Generation (RAG)</b>  Également important pour permettre de se défendre en cas de prétenition en responsabilité civile d'un tiers → jurisprudence "prémonitoire": ATF <a href="#">4A_301/2023</a> (en cas de liquidation des positions du client avec un solde négatif, la banque de prouver ces pertes).	<b>Séparation fonctionnelle</b> ( <i>chinese walls?</i> ) entre les <b>developers</b> et les personnes en charge de la revue  Implications d'experts tiers au niveau technique et juridique, notamment pour la définition des KPIs, le contrôle de l' <b>output</b> et les audits

En termes de **next steps concrets**, notre suggestion est, en sus de la mise en place d'un cadre de gouvernance interne, d'effectuer:

- un **mapping** des applications existantes qui impliquent un recours à l'IA, avec,
- un **classement sur une échelle de risques** (en tenant compte des 'model risks' (exemples: 'robustness', caractère correct, 'bias', stabilité, 'explainability', recours à des prestataires tiers) en sus des risques liés à la protection des données qui sont souvent déjà appréhendés), puis,
- une **évaluation en fonction du 'risk appetite'** de l'établissement.