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	• Les assujettis se concentrent principalement sur les risques liés à la protection des données, et non sur les risques liés aux modèles d'IA. • Le développement de l'IA est souvent décentralisé. • En cas d'utilisation d'une application acquise auprès d'un tiers, il est parfois difficile de savoir si elle implique de l'IA.	• Définition trop étroite de l'IA par les assujettis • Difficulté pour certains assujettis d'assurer une exhaustivité de l'inventaire. • Absence de critère permettant d'identifier les applications d'IA qui présentent un risque
Attentes de la FINMA	• Inventaire centralisé et complet des applications d'IA • Les compétences et les responsabilités (pour les phases de développement et d'utilisation) doivent être clairement définies. • Mise en place de règles relatives aux tests des modèles d'IA, à leur documentation et à des formations • En cas d'externalisation, des clauses contractuelles précises concernant les responsabilités du prestataire	• Définition suffisamment large et uniforme de l'IA • Mise en place de critères pour identifier les applications d'IA importantes et les risques spécifiques nécessitant une attention particulière
Mise en œuvre concrète (exemples découlant de la pratique d'OA)	Mise en place d'une directive interne qui alloue des responsabilités (à éviter: projets gérés uniquement par la 1ère ligne ou qui émergent de manière décentralisée) *mais important de souligner* 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 (art. 21 LPD). Compréhension des prestations offerts par des tiers + engagements contractuels à obtenir de tiers (même si marché concentré avec un nombre réduit de prestataires) → enjeux similaires à l'outsourcing	*Checklist* et *heat map* pour documenter, pour chaque cas d'utilisation, (i) les *model risks* (*robustness*, caractère correct, *bias*, stabilité et *explainability*) d'IA et (ii), selon les cas, les engagements contractuels pris par des prestataires tiers. Inventaire avec une classification en fonction des risques

Communication FINMA 08/2024 (lien)

Gouvernance et gestion des risques en lien avec l'utilisation de l'intelligence artificielle (2/3)

Thématiques	Pilier 3: Qualité des données	Pilier 4: Tests / surveillance constante	Pilier 5: Documentation
Constats de la FINMA	• Tous les assujettis n'ont pas défini des règles et des processus pour garantir la qualité des données dans les applications d'IA.	• Faiblesses relevées dans la planification et la mise en œuvre de tests et de contrôle • Peu d'indicateurs de performance spécifiques sont définis à l'avance.	• Certains assujettis ne disposent pas de directives pour documenter le recours à l'IA. • Documentation incomplète, peu détaillée et pas calibrée en fonction des utilisateurs de chaque application
Attentes de la FINMA	• Établir des instructions/directives internes pour garantir l'exhaustivité, la correction, l'intégrité et l'accessibilité des données utilisées	• Mise en place de processus de test pour vérifier les modèles d'IA et s'assurer que les applications atteignent les objectifs prévus. • Réalisation de contrôle réguliers des réponses d'IA	• Fournir une documentation détaillée pour les applications importantes couvrant: les objectifs de l'application, la fiabilité, les risques, la sélection des données et leur qualité.
Mise en œuvre concrète (exemples découlant de la pratique d'OA)	Processus interne de contrôle de la qualité des données d'entrée (*input*) Scepticisme implicite du régulateur à l'égard des LLM (car contrôle de la qualité des données très difficile en pratique) → risque que le déploiement de LLM soit soumis à des limites règlementaires à l'avenir?	Définition de KPI Contrôles *ex post* pour lutter contre le phénomène de *model/data drift* 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 *fallback*

Communication FINMA 08/2024 (lien)

Gouvernance et gestion des risques en lien avec l'utilisation de l'intelligence artificielle (3/3)

Thématiques abordées	Pilier 6: Explicabilité	Pilier 7: Vérification indépendante
Constats de la FINMA	• Les résultats des modèles d'IA sont souvent peu explicables et reproductibles, ce qui limite la possibilité d'en faire une évaluation critique.	• Des processus de vérification indépendants du développement des modèles d'IA sont rarement mis en œuvre.
Attentes de la FINMA	• Assurer que les résultats des modèles sont compréhensibles pour les parties prenantes, qu'il s'agisse par exemple d'investisseurs, de clients ou de collaborateurs. • Comprendre les mécanismes de fonctionnement des modèles pour garantir leur plausibilité et robustesse.	• Mettre en place, pour les applications importantes, une vérification indépendante couvrant tout le cycle de son développement, afin d'obtenir des avis objectifs et d'identifier les risques.
Mise en œuvre concrète (exemples découlant de la pratique d'OA)	*Due diligence* des applications d'IA → objectif (difficile!) qui devrait être visé: reproductibilité de l'*output* afin de pouvoir comprendre son origine / analyse de sensibilité / indication des sources dans le cadre des projets de *Retrieval Augmented Generation (RAG)* Également important pour permettre de se défendre en cas de prétention en responsabilité civile d'un tiers → jurisprudence "prémonitoire": ATF 4A_301/2023 (en cas de liquidation des positions du client avec un solde négatif, la banque de prouver ces pertes).	Séparation fonctionnelle (*chinese walls*?) entre les *developers* 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'*output* et les audits