
# Mila x Finance : L'ère des agents, du risque et de la protection des consommateur·rice·s

Résultats synthétisés du secteur financier

Mila

# À propos de ce rapport

Le 31 mars 2026, Mila — Institut québécois d'intelligence artificielle a organisé un événement collaboratif de type Café de conversation (aussi appelé World Café). Le rassemblement a réuni environ 30 professionnel·le·s de l'industrie provenant d'une quinzaine d'institutions financières canadiennes. L'objectif principal était de favoriser le dialogue entre Mila et le secteur financier, en abordant les défis communs et en alignant les objectifs de recherche de Mila avec les réalités du marché. L'accent était mis sur le développement d'une IA financière robuste, particulièrement dans les domaines de la gestion des risques et de la protection des consommateur·rice·s.

Les discussions étaient structurées autour de cinq tables rondes thématiques :

I. GOUVERNANCE DE L'IA : CADRE ET PAYSAGE RÉGLEMENTAIRE

II. ÉVALUATION ET SURVEILLANCE DES RISQUES LIÉS À L'IA

III. SÉCURITÉ TECHNIQUE, GARDE-FOUS ET MESURE

IV. L'IA EN GESTION DES RISQUES : CAS D'UTILISATION DE LA DÉTECTION DE FRAUDE

V. L'AVENIR AGENTIQUE

Chaque table était animée par un·e membre de l'équipe de recherche appliquée de Mila et par un·e chercheur·euse universitaire de l'institut. Ce rapport synthétise les points clés et les défis industriels abordés lors de ces discussions.


## Remerciements

Ce rapport représente un effort collaboratif de Mila – Institut québécois d'intelligence artificielle et de représentant·e·s principaux·ales du secteur financier québécois. Nous remercions les personnes suivantes, membres de Mila, pour leur présence essentielle, leurs conseils d'expert·e·s, l'animation des discussions et leurs précieuses contributions à la rédaction et au développement de ce rapport : **Istabrak Abbes**, candidate à la maîtrise en recherche, Université de Montréal, **Arsène Fansi Tchango**, gestionnaire, recherche appliquée en apprentissage automatique, **Simona Gandrabur**, responsable du Studio de sécurité en intelligence artificielle, **Prakhar Ganesh**, candidat au doctorat, Université McGill, **Gaétan Marceau Caron**, directeur principal, recherche appliquée en apprentissage automatique, **Philippe Martin**, candidat au doctorat, Université de Montréal, **Maryam Molamohammadi**, scientifique séniore, recherche en IA responsable, **Mahta Ramezanian**, candidate au doctorat, Université de Montréal, **Shalaleh Rismani**, chercheuse postdoctorale, Université McGill, **Adam Salvail**, gestionnaire, recherche appliquée en apprentissage automatique, **Elnathan Tiokou**, candidat à la maîtrise en recherche, Polytechnique Montréal, et fondateur de la startup Mila Vraust.ai.

L'événement a été organisé et ce rapport a été rédigé sous la direction de **Rheia Khalaf**, directrice des partenariats chez Mila.

Nous souhaitons exprimer notre profonde gratitude aux participant·e·s de l'industrie qui ont partagé leurs points de vue candides et leurs précieuses expériences pratiques. Les participant·e·s comprenaient des représentant·e·s de banques, d'assureurs, de gestionnaires d'actifs et d'organismes de réglementation, tous et toutes engagé·e·s à équilibrer les risques et les opportunités inhérents à l'avancement continu de l'IA dans les services financiers. Les organisations participantes comprenaient : l'Autorité des marchés financiers (AMF), la Banque Nationale du Canada, BNP Paribas, le Mouvement Desjardins, Finance Montréal, iA Groupe financier, la Caisse de dépôt et placement du Québec (La Caisse), Financière Manuvie, le Bureau du surintendant des institutions financières (BSIF), RBC Borealis, Société Générale et Groupe Banque TD.


**Note :** Le contenu de ce rapport a été affiné et amélioré à l'aide de modèles d'IA générative, avec une supervision humaine rigoureuse et une validation finale.

# Avant-propos

L'adoption des systèmes d'IA évolue des projets pilotes expérimentaux vers une restructuration fondamentale de l'environnement corporatif. Ces systèmes fonctionnent désormais comme des collaborateur·rice·s semi-autonomes, et non pas seulement comme des outils logiciels. Cependant, de nombreuses organisations qui adoptent rapidement ces technologies font face au défi stratégique de créer des mécanismes et des systèmes autour de ces technologies qui génèrent une valeur commerciale cohérente.

Cette intégration rapide crée une tension critique avec la nécessité d'une surveillance humaine. Sans une gestion robuste, les entreprises sont exposées à des risques opérationnels et de réputation importants à mesure que l'IA prend en charge des décisions à enjeux élevés. Pour contrer cela, les organisations intègrent directement la sécurité et la gouvernance dans leurs flux de travail, en utilisant des garde-fous dynamiques qui déclenchent automatiquement une intervention humaine lorsque des seuils de risque spécifiques sont dépassés.

Le secteur financier, comme de nombreuses autres industries hautement réglementées, fait face à des défis communs liés à l'adoption rapide de l'IA. Pour aborder ces enjeux, Mila a récemment organisé une discussion stratégique réunissant des expert·e·s du secteur financier et des organismes de réglementation. Le rythme actuel des changements technologiques et de l'adoption de l'IA nécessite une collaboration renouvelée entre les secteurs public, privé et académique. L'objectif est de transcender le piège des projets pilotes et d'adopter un cadre plus AGILE (Conscience, Garde-fous, Innovation, Apprentissage et Résilience de l'écosystème), tel qu'introduit dans le rapport de la deuxième édition du Forum sur l'intelligence artificielle dans le secteur des services financiers (FIASSF)[1]. L'objectif de la discussion était de dépasser les enjeux théoriques et d'aborder les difficultés pratiques de l'intégration de l'IA dans des environnements réglementés :

> → **Aligner l'innovation avec les demandes du marché :** Se concentrer sur la discussion de solutions qui renforcent la résilience du système financier et améliorent la protection robuste des consommateur·rice·s, tout en répondant directement aux contraintes à fort impact telles que l'auditabilité et la souveraineté des données.
>
> → **Favoriser la collaboration interdisciplinaire :** Transformer la gouvernance d'un goulot d'étranglement de fin de parcours en un accélérateur d'innovation sécurisée, assurant une base pour un déploiement d'IA évolutif, digne de confiance et conforme.

Ce rapport présente les risques et les occasions spécifiques liés à l'IA qui ont été identifiés au cours des discussions. Il est structuré de manière à refléter la séquence nécessaire à une opérationnalisation réussie de l'IA, allant des politiques de haut niveau jusqu'à la mise en œuvre et à la vision d'avenir. Il débute par un examen des règles opérationnelles fondamentales et du contexte réglementaire nécessaires à la **gouvernance de l'IA**. Il aborde ensuite les fondements structurels de l'**évaluation et de la surveillance des risques liés à l'IA**. Cette base structurelle prépare le terrain pour la section sur la **sécurité technique et les garde-fous**, qui explore les mécanismes d'application en temps réel requis pour assurer la conformité aux politiques. Ces principes combinés sont ensuite mis à l'épreuve dans la section consacrée à l'**IA appliquée à la gestion des risques**, qui détaille le principal cas d'usage financier — la détection de fraude — ainsi que ses contraintes particulières. Enfin, le rapport se conclut par une réflexion tournée vers l'**avenir agentique**, en analysant les complexités ainsi que l'approche progressive et encadrée nécessaire à l'adoption à grande échelle de systèmes d'IA autonomes.

[1] Deuxième Forum sur l'intelligence artificielle dans le secteur des services financiers, rapport de mars 2026. https://globalriskinstitute.org/publication/fifai-ii-ai-risks-and-opportunities/

# Sommaire exécutif

La mise à l'échelle de l'IA dans le secteur financier est fondamentalement limitée non pas par la technologie, mais par un manque de maturité opérationnelle en matière de gouvernance, de préparation des données et de définition des responsabilités. Les organisations doivent prioriser la construction d'une infrastructure fondamentale robuste, incluant une gouvernance fondée sur le risque et des systèmes de sécurité technique, pour éviter d'accumuler une dette de gouvernance et dépasser la phase des projets pilotes.

Principales conclusions sur l'opérationnalisation de l'IA et la gestion des risques :

I. GOUVERNANCE DE L'IA : La gouvernance de l'IA financière est un mandat obligatoire fondé sur le risque, nécessitant une gouvernance des données fondamentale et intégrée pour naviguer dans des réglementations mondiales complexes et surmonter le fossé entre la preuve de concept (PDC) et la production.

II. ÉVALUATION DES RISQUES LIÉS À L'IA : La surveillance doit continuellement superviser les quatre piliers du risque (fiabilité, éthique, données et vie privée, sécurité et sûreté), gérer les compromis inhérents (paradoxe de l'équité) ainsi qu'aborder les risques des tiers et les lacunes en matière de maturité organisationnelle.

III. SÉCURITÉ TECHNIQUE : Les garde-fous techniques sont une architecture de sécurité essentielle et non optionnelle dans les environnements à enjeux élevés, exigeant des références spécifiques au domaine et un équilibre entre les améliorations de la sécurité et les coûts de latence, notamment dans les systèmes agentiques complexes.

IV. L'IA DANS LA GESTION DES RISQUES : La détection de fraude est le principal cas d'utilisation de l'IA dans le secteur bancaire, confrontée à la rareté des données, aux taux élevés de faux positifs (compromis sur les hallucinations) et parfois à la plus grande efficacité des règles simples par rapport aux modèles d'IA complexes.

V. L'AVENIR AGENTIQUE : Le déploiement des agents d'IA est actuellement limité aux preuves de concept de productivité interne, bloqué à plus grande échelle par des préoccupations de fiabilité, un manque de gouvernance traçable et l'impératif de maintenir le jugement humain comme autorité finale.

# Opérationnalisation et gestion des risques

Nous évoluons dans un environnement en rapide mutation, alors que de nouveaux modèles, de nouveaux risques et de nouvelles menaces apparaissent chaque jour. Le besoin d'un rapport détaillant les pratiques communes actuelles est essentiel pour s'assurer que les organisations peuvent s'adapter de manière proactive à la vitesse du changement, maintenant à la fois la sécurité et l'avantage concurrentiel. Les sections ci-dessous décrivent les pratiques partagées à la date de rédaction de ce rapport, servant de référence vitale dans ce paysage dynamique.

## I. GOUVERNANCE DE L'IA : CADRE ET PAYSAGE RÉGLEMENTAIRE

Le secteur financier est actuellement engagé dans une course à enjeux élevés pour intégrer l'intelligence artificielle, mais le défi central ne réside pas dans la capacité technologique, mais dans l'établissement des règles d'engagement nécessaires. Ce parcours d'opérationnalisation de la gouvernance de l'IA se déroule dans un paysage défini par une expérimentation rapide, des pressions réglementaires complexes et des demandes internes intenses pour démontrer de la valeur.

### 1. Cadres contextuels et fondés sur le risque

→ **La gouvernance de l'IA est réglementaire :** Les institutions financières, en tant que secteur hautement réglementé, font face à des pressions uniques dans l'établissement de la gouvernance de l'IA. Contrairement aux industries moins réglementées, elles ne peuvent pas traiter la gouvernance de l'IA comme une couche optionnelle; c'est un mandat réglementaire entrelacé avec les régimes de conformité existants (p. ex., lutte contre le blanchiment d'argent, protection des consommateur·rice·s, confidentialité des données). Cet environnement exige une approche plus conservatrice et prudente face aux risques, donnant la priorité à l'auditabilité, à une responsabilité claire et à la surveillance humaine dans la boucle, en particulier pour les cas d'utilisation à fort impact. Le défi ne réside pas seulement dans la conformité, mais dans la traduction de réglementations mondiales fragmentées (comme la loi européenne sur l'IA et les lois locales sur les données) en un cadre interne unifié et actionnable qui permet tout de même l'innovation nécessaire sans accumuler une dette technique ou de gouvernance insurmontable.

→ **Données et gouvernance de l'IA :** Un constat récurrent dans l'industrie est que les organisations manquent souvent d'une préparation complète pour la gouvernance de l'IA parce qu'elles traitent encore des défis fondamentaux en matière de gouvernance des données. Les problèmes de qualité des données ont un impact direct sur les systèmes d'IA en aval, soulignant que la gouvernance de l'IA est

inextricablement liée à et fondamentalement dépendante d'une gouvernance des données robuste, car les mauvaises pratiques de données compromettent la responsabilité. Ceci est intrinsèquement lié au besoin d'une meilleure documentation et d'une tenue de dossiers améliorée, essentielles pour la traçabilité, la conformité réglementaire et la compréhension des défaillances au fil du temps.

→ **Passage aux cadres contextuels et fondés sur le risque :** On observe un mouvement fort dans les institutions, s'éloignant d'une gouvernance rigide, uniforme, vers des cadres contextuels et fondés sur le risque. La surveillance est de plus en plus adaptée au cas d'utilisation spécifique de l'IA, ce qui signifie que la distinction entre risque élevé et risque faible détermine le niveau de contrôle et la rapidité du déploiement. Par exemple, les risques associés à l'IA varient considérablement selon des facteurs tels que la nature du système (orienté client ou interne), l'échelle de son impact potentiel et le fait qu'il soit propriétaire ou repose sur des fournisseurs tiers. Cette approche dépendante du contexte informe directement les stratégies de gestion et d'atténuation des risques liés à l'IA. Puisque la nature du cas d'utilisation de l'IA dicte le type et la gravité du risque, les efforts d'atténuation doivent être très spécifiques. Par exemple, le risque de fuite de données dans un produit d'IA générative orienté client nécessite des contrôles différents du risque de biais dans un modèle interne d'approbation de prêts.

→ **La gouvernance comme accélérateur :** La clé du succès est d'intégrer la gouvernance de l'IA dans les processus existants plutôt que de superposer des cadres entièrement nouveaux. Lorsque la gouvernance est introduite tardivement, elle est souvent perçue comme la police et devient un frein à l'innovation. Inversement, lorsque la gouvernance est intégrée tôt dans les flux de développement, de risque et de conformité, elle peut accélérer l'adoption. Les institutions reconnaissent de plus en plus que la gouvernance doit être pratique, intégrée et alignée sur les systèmes existants pour être efficace.

## 2. Défis de gouvernance dans le passage de la preuve de concept (PDC) à la production

→ **Combler le fossé : de la PDC à la production :** Il existe un écart structurel majeur entre l'expérimentation et la mise à l'échelle. Une grande proportion des initiatives reste au stade de PDC, et la transition de la PDC à la production est souvent lente et exige beaucoup de ressources. Les institutions tentent de gérer cela en assouplissant les contraintes lors de l'expérimentation (p. ex., environnements bac à sable) et en réintroduisant des contrôles stricts lorsque nécessaire. Les garde-fous sont également conçus horizontalement pour différents cas d'utilisation, plutôt que d'être réévalués à partir de zéro à chaque fois. Cependant, même les PDC réussies s'enlisent souvent face aux exigences d'intégration et de mise à l'échelle.

→ **Le dilemme budgétaire :** Le financement de projets d'IA, notamment les PDC, reste structurellement difficile. L'absence de rendement de l'investissement clair pour l'expérimentation soulève des questions quant au financement et à l'unité d'affaires responsable de sa gouvernance et de sa responsabilité à long terme. En conséquence, des idées prometteuses peuvent s'enliser avant d'atteindre la production.

→ **L'indicateur de succès non résolu :** Il n'existe pas de consensus clair sur les métriques de succès standard pour la gouvernance de l'IA. Les institutions utilisent un mélange d'approches, incluant les gains d'efficacité (économies de temps et de coûts), les taux d'erreur (comparaison humain par rapport à l'IA) et les indicateurs de risque tels que les violations de données ou l'impact sur la réputation. Après le déploiement, l'accent est fortement mis sur les plans de surveillance, la détection de dérive et l'ajustement continu des seuils, mais sans cadre unifié.

### 3. Fragmentation réglementaire mondiale et traduction

→ **Fragmentation réglementaire mondiale et traduction :** Le paysage réglementaire est fragmenté et évolue rapidement, ce qui complique les choses pour les organisations naviguant dans plusieurs régimes qui se chevauchent, tels que la Loi européenne sur l'IA, le Règlement général sur la protection des données (RGPD), la loi 25 et les lignes directrices émergentes canadiennes et américaines. Cela nécessite des cadres internes hybrides, souvent guidés par la juridiction la plus stricte, et exige que les institutions surveillent constamment les mises à jour et traduisent les exigences légales complexes en mise en œuvre commerciale et technique. L'environnement reste non standardisé : l'Europe favorise les approches prescriptives, l'Amérique du Nord offre de la flexibilité et l'Asie est encore en phase d'expérimentation. Alors que les organismes de réglementation exigent l'explicabilité, la responsabilité et la traçabilité, ils ne prescrivent souvent pas de méthodes de mise en œuvre spécifiques. Par conséquent, les institutions financières évoluent dans un contexte dynamique où les attentes fondamentales sont établies, mais le chemin précis vers une exécution conforme et évolutive est encore défini par l'évolution des pratiques industrielles.

## II. ÉVALUATION ET SURVEILLANCE DES RISQUES LIÉS À L'IA

Pour mettre efficacement l'IA à l'échelle au-delà des phases pilotes initiales, un audit et une gouvernance robustes sont essentiels. Cela nécessite d'aller au-delà des simples listes de contrôle de conformité et d'adopter des cadres contextuels et fondés sur le risque pour l'évaluation des risques liés à l'IA.

### 1. Cadres d'évaluation des risques liés à l'IA

→ **Piliers principaux de l'évaluation des risques :** Ces cadres déplacent l'attention des préoccupations de sécurité informatique conventionnelles vers les impacts sociotechniques plus larges de l'IA, catégorisant couramment les risques en quatre piliers principaux.

- **Fiabilité technique :** L'exactitude et la robustesse sont primordiales. Cela inclut la prévention des hallucinations et l'assurance que l'IA fonctionne de manière fiable même lorsqu'elle est exposée à des données réelles qui s'écartent de son ensemble d'entraînement. Étant donné que l'IA générative repose sur des ensembles de données vastes et non structurés (textes, images et codes collectés sur le web), le risque de « mauvaises données en entrée, mauvais résultats en sortie » est important. La **traçabilité des données** est le cadre utilisé pour s'assurer que la « connaissance » du modèle est traçable, vérifiable et protégée contre les injections malveillantes ou la dégradation significative.
- **Éthique et équité :** Ce domaine se concentre sur l'identification des biais algorithmiques qui pourraient mener à la discrimination. Il priorise également l'explicabilité, en s'assurant que les décisions de l'IA ne sont pas des boîtes noires, mais peuvent être comprises par les humains.
- **Données et vie privée :** Cela concerne le carburant de l'IA. Les risques clés incluent les fuites de données (où le modèle peut révéler des informations sensibles), le manque de consentement des utilisateur·rice·s et l'utilisation de matériel protégé par des droits d'auteur, ce qui peut créer une responsabilité juridique.
- **Sécurité et sûreté :** Cela va au-delà des menaces cybersécuritaires traditionnelles pour inclure les attaques adversariales, telles que l'injection de requêtes, ainsi que les risques pour la sécurité physique, notamment dans des contextes comme les systèmes de fabrication automatisés ou les véhicules autonomes.

→ **Complexité dans la gestion et l'atténuation des risques liés à l'IA :** La gestion des risques liés à l'IA est intrinsèquement complexe parce qu'elle nécessite d'équilibrer des objectifs profondément interconnectés et souvent contradictoires, où l'optimisation pour une mesure de protection peut involontairement affaiblir une autre. Cela est plus visible dans le paradoxe de l'équité, où l'adhésion stricte aux principes de **confidentialité des données**, comme la suppression des informations personnelles identifiables (IPI) pour protéger les identités, compromet directement l'**équité éthique** en supprimant les données démographiques nécessaires pour auditer le système à la recherche de biais. Au-delà de ces compromis internes, le paysage est encore compliqué par la nature de « boîte noire » de nombreux modèles, où la **transparence** est souvent sacrifiée pour une plus grande **précision** prédictive. Parce que les domaines de risque de la vie privée, de l'équité, de la sécurité et de la performance n'existent pas de manière isolée, l'atténuation n'est pas une simple liste de contrôle, mais un acte continu de calibrage; une correction pour une vulnérabilité peut créer un effet d'entraînement qui introduit des responsabilités juridiques, sociales ou techniques ailleurs.

→ **Cycle de gestion continue des risques :** Un cadre de gestion des risques robuste nécessite un cycle continu de surveillance, de mesure et d'adaptation. Cela garantit que les contrôles restent proportionnels aux risques dynamiques et spécifiques au contexte posés par le système d'IA tout au long de son cycle de vie complet. L'IA générative est sujette à la dérive, où les sorties du modèle se dégradent ou évoluent au fil du temps à mesure qu'il interagit avec de nouvelles données. Cela renforce la nécessité pour les organisations de disposer de systèmes de surveillance continue pour valider que l'IA fonctionne toujours dans ses paramètres de sécurité et d'exactitude d'origine.

## 2. Souveraineté des données et risques liés aux tiers

→ **Préoccupations relatives à la souveraineté des données :** Le recours à des outils d'IA externes soulève des préoccupations importantes concernant la souveraineté des données. Lorsque des fournisseurs tiers traitent des données, les institutions font face à des risques liés à l'endroit où les données sont stockées, à la juridiction légale applicable et au potentiel d'accès non autorisé ou d'utilisation d'informations sensibles par le fournisseur externe. Cette perte de contrôle sur l'emplacement et l'accès aux données est un facteur clé pour le développement de solutions d'IA internes, malgré les compromis en termes de performance.

→ **Compromis liés aux tiers :** Une préoccupation majeure est le compromis entre les outils d'IA tiers et la souveraineté des données. Les outils externes sont souvent plus puissants et conviviaux, mais introduisent des risques liés aux fuites de données, à la perte de contrôle et au manque de visibilité. En conséquence, les institutions poussent à développer des outils internes, même s'ils sont moins performants. En pratique, cela crée des comportements fantômes, où les employé·e·s utilisent des comptes personnels ou des canaux non officiels pour accéder à de meilleurs outils.

→ **Responsabilité en matière de risques des modèles de fournisseurs :** Un défi spécifique et croissant est la gestion du risque des modèles (GRM) des outils d'IA génératives intégrés de tiers. Bien que les institutions financières demeurent entièrement responsables de la GRM de ces modèles, de nombreux fournisseurs refusent de partager suffisamment de détails sur leur évaluation des risques ou leur fonctionnement interne. Ce manque de transparence nécessite que les politiques de gouvernance de l'IA établissent des cadres clairs et des exigences contractuelles pour mandater l'accès à l'information et le partage des données de risque avec les fournisseurs.

→ **Réglementation et solutions non techniques :** Une gouvernance efficace de l'IA reconnaît les limites des solutions purement techniques. La gouvernance doit plutôt intégrer des mesures non techniques, telles que des contrats plus solides, une meilleure documentation (p. ex., fiches de modèles[2]),

[2] https://huggingface.co/docs/hub/model-cards

et des attentes plus claires envers les fournisseurs tiers. Étant donné que les petites entreprises manquent souvent de levier face aux grands fournisseurs d'IA monopolistiques, les réglementations jouent un rôle essentiel dans l'établissement des normes industrielles. L'incertitude persiste autour des cadres réglementaires existants et de la division évolutive des rôles entre développeur·euse·s et déployeur·euse·s. Dans l'ensemble, une gouvernance efficace de l'IA nécessite une coordination entre les pratiques internes et les structures réglementaires externes.

### 3. Structure organisationnelle et les lignes de défense

→ **Le modèle des trois lignes de défense (3LD) :** La mise en œuvre d'une gouvernance efficace de l'IA implique souvent **trois lignes de défense**, comprenant généralement les développeur·euse·s, une équipe d'audit interne et une équipe d'audit indépendante séparée. Un défi institutionnel courant est le manque d'expertise suffisante en IA dans ces trois lignes de défense, entraînant des lacunes dans la surveillance. Dans certains cas, les équipes d'audit impliquent les développeur·euse·s dans l'audit de leurs propres systèmes, ce qui soulève des préoccupations quant à l'indépendance et aux biais ou contaminations potentiel·le·s.

→ **Le problème de la multiplicité des responsables :** Cette complexité contribue au **problème de la multiplicité des responsables**, où la diffusion des responsabilités entre plusieurs équipes rend la responsabilité peu claire. Une gouvernance efficace nécessite une collaboration interdisciplinaire et un changement culturel où les échecs sont acceptés comme des opportunités d'améliorer les structures, plutôt que de blâmer des individus. La gestion proactive des risques, plutôt que les réponses réactives de type « salle d'urgence », requiert l'engagement de la haute direction.

→ **La tension entre efficacité et surveillance :** La proposition de valeur centrale de l'efficacité de l'IA est continuellement tempérée par la demande non négociable de surveillance humaine. Dans des domaines à risque élevé comme la lutte contre le blanchiment d'argent, le jugement humain est essentiel pour la responsabilité, le respect des exigences réglementaires et l'atténuation des risques. Une atténuation efficace des risques nécessite une approche proactive, intégrant une analyse juridique approfondie et une surveillance humaine tout au long du cycle de développement de l'IA. Cependant, l'efficacité de la surveillance humaine reste discutable, et cette exigence introduit des frictions opérationnelles. Ces frictions peuvent diminuer les gains d'efficacité attendus de l'IA et entraîner des retards opérationnels, administratifs et financiers. À l'inverse, le déploiement rapide de nouvelles fonctionnalités d'IA sans examen juridique adéquat nécessite souvent un retrait ou une révision coûteux par la suite.

→ **Naviguer dans les dynamiques internes : Littératie, alignement et gestion du changement :** En interne, les organisations font face à une pression importante de gestion du changement due à des désalignements. Alors que les dirigeant·e·s sont impatient·e·s de rentabiliser leur investissement et poussé·e·s par l'urgence d'adopter l'IA avant leurs concurrents, les employé·e·s sont partagé·e·s entre la curiosité et l'inquiétude face à la transformation de leur travail. Les cadres intermédiaires ont pour mission de mettre en œuvre l'IA et de concilier ces attentes contradictoires, ce qui met en évidence un besoin crucial de renforcer les connaissances en matière d'IA, de proposer des formations et d'assurer une communication claire, tout en reconnaissant que la capacité d'adaptation devient une compétence essentielle.

## III. SÉCURITÉ TECHNIQUE, GARDE-FOUS ET MESURE

La clarté des attentes réglementaires (explicabilité, responsabilité et traçabilité) contraste fortement avec les méthodes de mise en œuvre non prescriptives. Cette dynamique déplace l'attention de la gouvernance générale vers les mécanismes techniques critiques et en temps réel requis pour faire respecter les exigences

organisationnelles et réglementaires. Ce mécanisme est le **garde-fou**, un système de sécurité indépendant qui filtre les entrées et les sorties de l'IA en temps réel par rapport à des politiques spécifiques pour prévenir les violations de politique.

## 1. Conception et application de garde-fous en temps réel

→ **Le rôle évolutif des garde-fous dans la finance :** Dans des secteurs à enjeux élevés comme la finance, les garde-fous ne peuvent plus être traités comme des ajouts optionnels aux modèles de langage. Ils doivent être conçus dans le cadre de l'architecture du système, surveillés au fil du temps et évalués par rapport à des modes de défaillance réalistes. De plus, le même modèle peut nécessiter différentes configurations de sécurité selon le cas d'utilisation, le niveau d'autonomie et la population d'utilisateur·rice·s. Cela fait évoluer le rôle des garde-fous au-delà des simples systèmes de refus[3]. En pratique, les garde-fous doivent aider à faire respecter les limites organisationnelles, détecter les violations de politique et soutenir les exigences de conformité tout en restant compatibles avec les flux de travail d'entreprise. Les contrôles administratifs et techniques doivent être considérés ensemble plutôt que séparément.

→ **Gestion de la complexité des systèmes agentiques :** Avec les systèmes agentiques, la préoccupation ne se limite plus à savoir si une réponse est correcte. Le système peut naviguer, déclencher des outils, écrire des données, extraire du contenu ou opérer sur une infrastructure interne, ce qui crée une plus grande surface d'attaque. L'injection de requêtes, l'injection web, les permissions excessives et les boucles de validation humaine mal définies présentent des préoccupations concrètes. La complexité elle-même devient un multiplicateur de risques : à mesure que les systèmes deviennent plus autonomes, l'analyse des défaillances, l'attribution des responsabilités et la gestion du changement deviennent plus difficiles.

→ **Intégration de garde-fous spécifiques au secteur :** Le défi ne concerne pas seulement la génération de contenu nuisible, mais aussi l'accès, la propagation ou la transformation non autorisé·e·s d'informations sensibles. Il existe un besoin fort de références de garde-fous adaptées au domaine pour la finance, traitant spécifiquement la fuite d'informations privilégiées, sensibles ou réglementées, la gestion de la vie privée et des informations personnelles identifiables (IPI), ainsi que les conseils d'investissement non sécuritaires. Ces références servent à évaluer l'efficacité, la fiabilité et la rapidité des garde-fous. La voie à suivre implique des architectures de garde-fous modulaires combinant des classificateurs spécialisés rapides, des bibliothèques de politiques configurables et l'utilisation sélective de juges par modèle de langage ou de révision humaine. L'évaluation continue et la robustesse multilingue sont des domaines majeurs d'amélioration, ainsi que la conception de flux de travail axés sur la sécurité qui estiment le risque avant et après une action, et pas seulement après un incident.

## 2. Défis d'évaluation et compromis de latence

→ **Adapter les références aux besoins du domaine :** L'évaluation est l'un des besoins les plus urgents. Un garde-fou n'est significatif que si sa performance peut être mesurée d'une manière qui correspond au contexte de déploiement. Un virage utile s'oriente vers des références adaptées, une évaluation dynamique en temps réel, des tableaux de bord de surveillance et des métriques spécifiques aux cas d'utilisation plutôt que vers des comparaisons génériques de type classement. Le secteur

[3] Un système de refus simple en apprentissage automatique est conçu pour identifier et décliner les requêtes nuisibles, non sécuritaires ou hors de portée. Ces systèmes sont essentiels pour la sécurité, afin de s'assurer que les modèles, tels que les agents conversationnels, ne génèrent pas de contenu dangereux, illégal ou contraire à l'éthique.

financier manque de références publiques largement partagées parce que les ensembles de données réalistes sont difficiles et coûteux à construire, notamment lorsque des contraintes de vie privée et de confidentialité s'appliquent. En conséquence, des pipelines d'évaluation personnalisés, souvent demandés par les clients ou construits en interne, semblent être la voie pratique à suivre.

→ **Tension entre sécurité et latence :** Les approches de type modèle de langage comme juge (*LLM-as-a-judge*), une technique d'évaluation de l'IA où un grand modèle de langage hautement performant est sollicité pour évaluer les sorties d'un autre modèle ou système, sont perçues comme utiles pour la vérification et l'évaluation post hoc, mais leur coût et leur surcharge de latence, dans certains cas, doublent ou triplent approximativement le temps de réponse. Cela les rend intéressantes pour les décisions à risque élevé, les audits ou l'évaluation asynchrone, mais moins adaptées comme solution d'exécution universelle. Cela soulève une question architecturale plus large : quels risques doivent être traités par des détecteurs spécialisés légers, lesquels par adjudication basée sur un modèle de langage et lesquels par escalade humaine ? On peut soutenir que l'exactitude importe souvent plus que la rapidité dans les scénarios critiques, mais seulement jusqu'à un certain point; les institutions auront encore besoin de piles de sécurité graduées qui allouent les budgets de calcul selon la gravité du risque.

### 3. Meilleures pratiques de mesure et de positionnement

→ **Stratégie de positionnement :** Certains préjudices sont mieux traités à l'étape d'entrée, par exemple lorsqu'un sujet ne doit pas être abordé du tout, tandis que d'autres sont mieux traités à l'étape de sortie, où le système doit décider comment répondre de manière sécuritaire dans un domaine autorisé. Pour les assistants d'entreprise ou les systèmes orientés client, le risque peut s'accumuler au fil des tours plutôt que d'apparaître dans un seul énoncé. Dans ce cas, le filtrage statique à tour unique est insuffisant; les paramètres de diffusion en continu et de tours multiples (maintien du contexte conversationnel sur plusieurs interactions) sont plus utiles. En général, le positionnement peut varier selon le domaine : pour certaines applications sensibles, la vérification après génération peut être plus efficace que le filtrage de requêtes avant génération seul.

→ **Orientation de la mesure :** Les pratiques clés pour la mesure des garde-fous comprennent la concentration sur les taux de refus de réponse, le développement de taxonomies d'attaques spécifiques au domaine, la conduite de tests de requêtes adversariales et la surveillance des changements comportementaux dans les systèmes d'IA au fil du temps.

→ **Protocoles robustes :** Étant donné que l'IA est non déterministe, les cadres de mesure doivent reconnaître qu'une réponse différente n'est pas toujours un bogue. Les efforts d'évaluation doivent donc se concentrer sur la distinction entre la variation inoffensive et la dérive de sécurité significative. Cela nécessite l'adoption de protocoles d'évaluation robustes et longitudinaux plutôt que de s'appuyer sur des exécutions de références ponctuelles.

## IV. L'IA DANS LA GESTION DES RISQUES : CAS D'UTILISATION POUR LA DÉTECTION DE LA FRAUDE

Une intégration pratique de l'IA se situe dans la fonction centrale du secteur financier : la gestion des risques. Ce domaine englobe l'intégration de l'IA dans les processus existants pour gérer des risques tels que les menaces préidentifiées, comme la fraude.

## 1. Cas d'utilisation central : détection de fraude et contraintes opérationnelles

Dans le contexte de l'IA pour la gestion des risques, la détection de fraude représente le cas d'utilisation le plus important dans le secteur bancaire. Les budgets des départements de lutte contre la fraude ont considérablement augmenté au cours des dernières décennies dans de nombreuses banques mondiales. Avec le rythme de développement de l'IA, les activités frauduleuses sont devenues beaucoup plus efficaces, incluant la falsification de documents d'identité, les courriels d'hameçonnage crédibles et les atteintes à la cybersécurité. Les institutions financières doivent intégrer avec soin des outils d'IA pour l'atténuation de la fraude tout en restant conscientes des risques qui en découlent.

→ **Rareté des données et techniques de détection d'anomalies :** La détection de fraude est souvent entravée par un problème de supervision faible, où les données frauduleuses étiquetées (véritables étiquettes de fraude positives) sont rares par rapport aux transactions légitimes, et les faux négatifs sont dominants. Plusieurs solutions pour la détection de fraude sont utilisées :

- **Données synthétiques :** Créer artificiellement des scénarios de fraude pour entraîner des modèles lorsque les exemples réels sont insuffisants. La génération synthétique de données de comportement financier avec des scénarios correctement définis bénéficie aux algorithmes de détection, car ces données nouvellement générées peuvent être utilisées de manière supervisée dans des algorithmes de score ou de classification qui manquent actuellement de données correctement étiquetées.

- **Détection d'anomalies :** L'utilisation de techniques non supervisées comme l'Isolation Forest[4], qui sont des méthodes populaires pour la détection et l'évaluation de comportements anormaux. Cela aide à identifier les valeurs aberrantes, comme les schémas de transactions inhabituels dans les comptes appartenant à des populations vulnérables (p. ex., transactions par des personnes âgées de plus de 80 ans dans des boîtes de nuit), ou des activités suspectes liées aux transferts d'argent et aux paiements vers des pays et organisations à risque.

- **Avantages opérationnels des nouveaux agents :** Les nouveaux agents d'IA peuvent offrir des avantages en identifiant des scénarios de fraude connus (généralement issus du contenu web inclus dans les données d'entraînement du modèle de langage) et en les testant automatiquement, ainsi qu'en appliquant différents algorithmes pour générer des scores de risque client.

→ **Le compromis des faux positifs et les interventions simples :** Dans la détection de fraude, augmenter la sensibilité d'un modèle pour détecter de nouveaux schémas de fraude entraîne souvent un volume élevé de faux positifs (alertes signalant incorrectement des transactions légitimes). Ce problème est distinct des hallucinations d'un modèle (génération de sorties factuellement incorrectes ou insensées), car les faux positifs font référence à des erreurs de classification dans un domaine défini. Un taux élevé de faux positifs, comme un bond de 10 % à 35 %, est économiquement insoutenable, car il submerge les analystes humains de bruit. Pour rester en avance sur les fraudeur·euse·s sophistiqué·e·s qui utilisent l'IA pour éliminer les signaux d'hameçonnage traditionnels (p. ex., les fautes d'orthographe), une logique simple s'avère souvent plus efficace que l'IA complexe seule. Par exemple, exiger une correspondance exacte du nom pour les virements IBAN peut éliminer une vaste catégorie de fraude aux paiements.

[4] L'Isolation Forest est un algorithme d'apprentissage automatique non supervisé spécialement conçu pour la détection d'anomalies. Contrairement aux méthodes traditionnelles qui établissent un profil de données « normales » pour trouver des déviations, il se concentre explicitement sur l'isolation des observations individuelles.

# V. L'AVENIR AGENTIQUE

Bien que les agents d'IA détiennent un potentiel significatif pour transformer les opérations financières, l'adoption de ces systèmes autonomes en est encore à ses débuts. La plupart des organisations mènent des preuves de concept plutôt que de déployer des agents en production, avec des efforts actuellement concentrés sur les outils internes et les actions contrôlées.

## 1. L'avenir agentique : déploiement et mise à l'échelle

→ **Les cas d'utilisation se concentrent sur les gains de productivité :** L'automatisation des tâches répétitives de niveau junior peut économiser environ 30 à 40 % du temps. Les applications orientées client se limitent aux agents conversationnels et aux flux de travail guidés restreints aux données non sensibles. Les agents sont également utilisés pour le soutien décisionnel par l'analyse qualitative et les rapports structurés, ainsi que pour l'automatisation des flux de travail tels que la surveillance sectorielle et la génération de rapports standardisés.

→ **Gestion des risques pour le commerce agentique :** À terme, des achats et des transactions financières se dérouleront de bout en bout entre des agents d'IA autorisés sans intervention humaine. Cet environnement autonome crée un besoin aigu de nouveaux protocoles de gestion des risques, de traçabilité et de responsabilité définie, notamment en ce qui concerne la validation des transactions, la fraude et la protection des consommateur·rice·s lorsque des décisions financières sont exécutées exclusivement par des agents.

→ **Plusieurs obstacles empêchent un déploiement plus large :** Les entreprises manquent d'environnements contrôlés avec une gouvernance, une surveillance et des capacités de retour arrière suffisantes. Des préoccupations de fiabilité persistent, notamment des hallucinations et des sorties instables dans des contextes à enjeux élevés. L'expertise interne est limitée, ce qui rend difficile de suivre le rythme de l'innovation rapide des fournisseurs. Les organisations font face à un compromis entre construire en interne, ce qui offre du contrôle mais est lent, et travailler avec des fournisseurs, ce qui accélère la livraison mais introduit des risques d'auditabilité et de dépendance, comme vu précédemment. Les contraintes réglementaires exigent une validation, une traçabilité et une responsabilité claire, tandis que la rotation rapide des fournisseurs augmente la complexité opérationnelle. Les architectures de données existantes ne sont souvent pas prêtes pour les systèmes basés sur des agents.

## 2. Meilleures pratiques de gouvernance et techniques pour les systèmes agentiques

→ **Surveillance continue des entrées, sorties et lignée des modèles :** La responsabilité est actuellement attribuée aux utilisateur·rice·s, bien que la réglementation future puisse déplacer la responsabilité vers les fournisseurs de modèles. La mauvaise compréhension des licences de code source ouvert et de la provenance des modèles introduit également un risque juridique.

→ **Meilleures pratiques techniques :** Celles-ci incluent la limitation des actions des agents aux opérations en lecture seule ou étroitement contraintes, la mise en œuvre d'une observabilité détaillée avec des journaux immuables[5], et le déploiement d'outils d'évaluation automatisés pour détecter les hallucinations et assurer la conformité. Les données doivent être segmentées, en utilisant des ensembles de données synthétiques ou dé-identifiés dans les environnements de test et des contrôles d'accès stricts en production.

[5] https://trainingcamp.com/glossary/immutable-logs/

→ **Indicateurs clés de performance et de sécurité :** Ceux-ci comprennent le taux d'hallucination, le taux d'erreur d'action, le temps de détection et de correction des problèmes, les taux de faux positifs et négatifs pour la détection des risques, et l'effort humain requis pour vérifier les sorties.

### 3. Principaux risques et voie à suivre

→ **Mise à l'échelle contrôlée :** Les principaux risques concernent les dommages à la réputation causés par des erreurs orientées client, l'escalade réglementaire due à une auditabilité insuffisante, et la dette technique provenant des intégrations fragmentées de fournisseurs. Les progrès nécessitent une approche contrôlée et progressive : commencer par des environnements bac à sable, se concentrer sur des flux de travail étroits avec des métriques claires, mettre en œuvre une évaluation automatisée et appliquer des contrôles d'audit des fournisseurs. Cela permet aux organisations de gérer les risques tout en préservant la possibilité de mettre à l'échelle les projets pilotes réussis.

→ **La frontière humaine :** Bien que les agents d'IA serviront de plus en plus d'amplificateurs de productivité pour les expert·e·s financier·ère·s, le jugement humain demeure l'arbitre final pour les décisions à enjeux élevés. Les employé·e·s doivent donc suivre une formation importante pour mesurer le potentiel, les limites et les risques associés aux nouveaux agents d'IA. L'avenir de la finance réside dans l'équilibre entre la détection algorithmique de pointe et la préservation des normes juridiques et éthiques centrées sur l'humain.


# Conclusion

L'événement « Mila x Finance : L'ère des agents, du risque et de la protection des consommateur·rice·s » a souligné un consensus critique au sein du secteur financier canadien : la mise à l'échelle réussie de l'IA dépend de la maturité des cadres de gouvernance, de risque et de sécurité, et non pas seulement de la capacité technologique. La transition des preuves de concept (PDC) isolées vers une production à l'échelle de l'entreprise nécessite un changement systémique, passant du développement en silos à une opérationnalisation intégrée et fondée sur le risque.

Les discussions ont mis en évidence que l'adoption efficace de l'IA exige :

→ **Une intégration en amont de la gouvernance :** La gouvernance de l'IA doit être traitée comme un mandat réglementaire non optionnel, intégrée tôt dans le cycle de développement et construite sur une base solide de gouvernance des données pour assurer la responsabilité et l'auditabilité.

→ **Une gestion continue et contextuelle des risques :** Les cadres de surveillance doivent superviser continuellement les risques de fiabilité, de biais éthique, de vie privée et de sécurité, en reconnaissant les compromis inhérents (comme le paradoxe de l'équité) et la complexité introduite par les fournisseurs tiers et le problème de la multiplicité des responsables.

→ **Une sécurité technique non négociable :** Les garde-fous techniques sont des composants architecturaux essentiels pour l'application de la conformité en temps réel dans les environnements à enjeux élevés. Ils nécessitent des références spécifiques au domaine et des stratégies sophistiquées pour gérer l'effet multiplicateur de risques de la complexité agentique sans introduire une latence excessive.

→ **Une mise à l'échelle contrôlée et progressive :** Bien que l'avenir agentique promette des gains d'efficacité significatifs, son déploiement actuel est limité par des préoccupations de fiabilité et d'auditabilité. La voie à suivre nécessite une approche contrôlée, en commençant par des environnements bac à sable, en se concentrant sur des flux de travail étroits et en s'assurant que le jugement humain reste l'autorité décisionnelle finale dans les contextes à enjeux élevés.

En fin de compte, l'avenir de l'IA dans la finance réside dans l'atteinte d'un équilibre robuste. Il s'agit de transcender le piège des projets pilotes en adoptant un cadre AGILE qui priorise la conscience, les garde-fous protecteurs et la résilience de l'écosystème. Cet effort collaboratif entre le monde académique, Mila, l'industrie et les organismes de réglementation est essentiel pour s'assurer qu'à mesure que l'IA évolue pour devenir un collègue de travail, elle reste digne de confiance, conforme et fondamentalement protectrice des consommateur·rice·s.

**« L'événement « Mila x Finance : L'ère des agents, du risque et de la protection des consommateur·rice·s » a mis en lumière l'évolution fulgurante de l'IA et les nouveaux risques qui l'accompagnent dans le domaine financier. Si les outils de détection de fraude progressent, les fraudeurs utilisent désormais eux-mêmes l'IA pour contourner les garde-fous bancaires; un défi quasi inexistant il y a seulement deux ans. Grâce aux tables rondes, nous avons pu échanger entre expert·e·s sur les applications concrètes en interne et partager ce qui fonctionne réellement sur le terrain de la détection. Face à une technologie qui se transforme si vite, des rendez-vous annuels comme celui-ci sont essentiels. Ils nous permettent de rester à l'affût et d'anticiper des risques que nous n'avions pas encore prévus. »**

**- Philippe Martin, doctorant à l'Université de Montréal, animateur de la table ronde « L'IA dans la gestion des risques : cas d'utilisation pour la détection de la fraude »**

## Devenez partenaire de Mila

Que vous souhaitiez lancer une collaboration de recherche, explorer des applications de l'IA en finance et en gestion des risques, accélérer l'adoption responsable de l'IA ou vous connecter avec les meilleurs talents en IA, nous voulons vous entendre.

En devenant partenaire de Mila, votre organisation a accès à :

**Une expertise de calibre mondial** — Collaborez avec des chercheur·euse·s de premier plan en apprentissage automatique, en IA générative, en optimisation, en prévision et en IA de confiance.

**Une innovation appliquée** — Explorez des cas d'utilisation à fort impact dans les services financiers, notamment la gestion des risques, la détection des fraudes, la finance climatique, l'optimisation de portefeuille et la conformité réglementaire.

**Un écosystème unique** — Tirez parti des relations de Mila avec les leaders du secteur, les start-ups, les organismes de réglementation et l'ensemble de la communauté de l'IA.

**Développement des talents et des compétences** — Engagez-vous auprès des meilleurs talents en IA grâce à des projets collaboratifs, des ateliers et des programmes d'innovation.

Communiquez avec l'équipe des partenariats de Mila pour explorer comment votre organisation peut contribuer à façonner l'avenir de l'IA en finance.