

PROGRAMME DE RECHERCHE DE L'ICSIA AU CIFAR

Appel à manifestations d'intérêt

Réseaux de solutions pour la sécurité de l'IA

Table des matières

Contexte	2
Portée et objectif de l'appel	3
Volet A : Atténuer les risques pour la sécurité des contenus synthétiques ..	4
Volet B : La sécurité de l'IA dans les pays du Sud	5
Financement et durée	6
Critères d'admissibilité	7
Procédure de dépôt et échéancier	8
Exigences pour les manifestations d'intérêt	9
Processus d'examen et critères d'évaluation de la manifestation d'intérêt	12
Foire aux questions	13

Contexte

Le programme de recherche de l'Institut canadien de la sécurité de l'intelligence artificielle (ICSIA) au CIFAR a le plaisir de lancer un **appel à manifestations d'intérêt pour des réseaux de solutions pour la sécurité de l'IA.**

Depuis l'avènement des modèles fondateurs, leur commercialisation et leur utilisation généralisée, les aspects liés à la sécurité de l'IA ont fait l'objet d'un examen et d'un débat publics approfondis. À la suite du lancement de ChatGPT en novembre 2022, la communauté de recherche et des spécialistes ont commencé à soulever des préoccupations au sujet des capacités croissantes de ces systèmes de pointe, comme la prolifération de contenus synthétiques générés par l'IA — les hypertrucages —, les enjeux de confidentialité et de sécurité, l'incidence sur le marché du travail et sur les droits de la personne, etc. Face à l'augmentation de l'utilisation et des capacités des modèles d'IA avancée à usage général au cours des dernières années, un mouvement international de plus en plus important s'est formé pour assurer la sécurité des systèmes d'IA avancée.

Pour relever ces défis, le CIFAR a reçu 27 millions de dollars du gouvernement fédéral en novembre 2024 pour lancer le programme de recherche de l'ICSIA au CIFAR. Ce programme est le volet de recherche appliquée et dirigée par des chercheurs et chercheuses de l'[ICSIA](#), qui relève d'Innovation, Sciences et Développement économique Canada. **L'un des axes centraux du Programme de recherche de l'ICSIA au CIFAR est de cerner et d'atténuer les risques pour la sécurité associés aux systèmes d'IA avancée.**

Le Programme de recherche de l'ICSIA au CIFAR tire parti des forces actuelles de la Stratégie pancanadienne en matière d'IA qui, depuis sa création en 2017, a entretenu une communauté de recherche de premier plan à l'échelle mondiale qui crée des solutions d'IA transformatrices et responsables pour les gens et la planète. Le programme IA et société du CIFAR, conformément à son mandat visant à faire progresser les travaux portant sur les implications sociétales de l'IA, a piloté le modèle des réseaux de solutions entre 2020 et 2024.

Le Programme de recherche de l'ICSIA au CIFAR considère la sécurité de l'IA comme un domaine vaste et multidisciplinaire qui englobe à la fois les perspectives techniques de l'informatique et les perspectives sociotechniques des sciences sociales et humaines.

Portée et objectif de l'appel

Les réseaux de solutions du CIFAR constituent des occasions de collaboration pour des équipes multidisciplinaires dirigées par un noyau central de six personnes au maximum. Ces équipes travaillent ensemble pour développer des approches et des applications (solutions) visant à relever les défis sociotechniques actuels et à saisir les occasions en matière de sécurité de l'IA. Les solutions doivent mettre l'accent sur l'atténuation des risques associés aux systèmes d'IA avancée et de pointe.

Les solutions pourraient s'appliquer à tous les secteurs de la société, y compris les milieux universitaires, les pouvoirs publics, l'industrie et la société civile. Elles devraient se concentrer sur la mise au point de prototypes, de pratiques et d'outils définis — sur les plans techniques et de la gouvernance — dans le domaine de la sécurité de l'IA.

L'objectif de cet appel à manifestations d'intérêt est de recenser jusqu'à trois réseaux multidisciplinaires de spécialistes, qui collaboreront à l'élaboration de solutions susceptibles d'avoir une incidence transformatrice sur la sécurité de l'IA. Le réseau de solutions financé dans le cadre de cet appel devrait :

- appliquer des connaissances de pointe issues de la recherche et de la pratique au développement de nouvelles interventions, applications ou approches pour répondre aux défis sociotechniques actuels liés à la sécurité de l'IA;
- mettre à l'essai des prototypes pour favoriser leur adoption, ce qui devrait comprendre une consultation de la communauté ou des parties prenantes au cours des 18 premiers mois du projet (si elle n'est pas déjà réalisée au moment où le réseau est financé);
- éclairer les politiques ou les pratiques qui renforcent la sécurité de l'IA et atténuent les préjudices sociaux, éthiques et économiques potentiels de l'IA avancée;
- entreprendre des efforts visant à développer l'intervention, au-delà de la mise au point et de l'essai des prototypes.

Les propositions qui ne sont pas axées sur l'atténuation des risques pour la sécurité associés aux systèmes d'IA avancée à grande échelle ne seront pas prises en considération. Les propositions dont l'objectif principal est de faire progresser les connaissances ou la compréhension d'un domaine d'étude ou qui se concentrent uniquement sur des solutions techniques (p. ex., l'amélioration des algorithmes d'apprentissage automatique) ne seront pas prises en considération.

Cet appel comporte deux volets, chacun axé sur un thème ou une zone géographique spécifique.

Volet A : Atténuer les risques pour la sécurité des contenus synthétiques

La popularité et l'accessibilité croissantes des modèles d'IA ont facilité plus que jamais la création et la diffusion de contenus synthétiques. Ceux-ci peuvent avoir des effets positifs dans de nombreux domaines. Toutefois, lorsqu'ils sont déployés dans des contextes particuliers, ils peuvent présenter des risques sur les plans technique — p. ex., provoquer l'effondrement du modèle — et social — p. ex., faciliter la fraude ou la tromperie par le biais d'hypertrucages, miner la confiance ou diffuser de la désinformation et de la désinformation. Les effets néfastes des contenus synthétiques recourent également d'autres effets néfastes, comme les biais et la discrimination dans les modèles d'IA et leur déploiement. Comprendre les contextes dans lesquels les contenus synthétiques sont nuisibles et atténuer les risques pour la sécurité constituent une préoccupation centrale du Programme de recherche de l'ICSIA au CIFAR.

Les réseaux de solutions bénéficiant d'un financement dans le cadre de ce volet mettront au point des solutions sociotechniques pour la détection, l'atténuation et la régulation des effets néfastes des contenus synthétiques, les testeront et présenteront comment les déployer.

Nous accueillons favorablement les propositions qui abordent les risques potentiels et les préoccupations liées à la sécurité des intrants ou des extrants de l'IA avancée. Plus précisément, les sujets peuvent comprendre à la fois les intrants et les extrants des systèmes d'apprentissage automatique :

- **Extrants des contenus synthétiques issus des modèles d'apprentissage automatique** : incidence sociétale des extrants des modèles; perception erronée des capacités des modèles; éducation et alphabétisation; filigrane, droits d'auteur; désinformation et confiance;
- **Données synthétiques traitées en tant qu'intrants d'apprentissage automatique** : incidence des données d'entraînement synthétiques sur les modèles, p. ex., sur les biais; cybersécurité; pollution des ensembles d'entraînement et changements de distribution induits par le modèle; réglementation internationale des données;
- **Risques pour la sécurité des systèmes d'IA générative multiagent.**

Cette liste n'est pas exhaustive. Les propositions portant sur d'autres aspects des contenus synthétiques seront prises en considération, à condition qu'elles se concentrent sur l'atténuation des risques liés aux systèmes d'IA avancée à grande échelle.

Deux réseaux de solutions recevront du financement dans le cadre de ce volet.

Volet B : La sécurité de l'IA dans les pays du Sud

Les propositions de ce volet peuvent porter sur n'importe quel sujet lié à l'atténuation des risques pour la sécurité associés aux systèmes d'IA avancée. Dans le cadre de ce volet, les propositions doivent tenir compte des défis et des occasions uniques dans la ou les régions où la solution sera mise en œuvre. Nous nous intéressons particulièrement aux propositions qui abordent les domaines prioritaires de la version finale du [Rapport international sur la sécurité de l'IA](#) (en anglais), en mettant l'accent sur la gestion des risques et les stratégies d'atténuation.

Les équipes doivent constituer des collaborations internationales dont la composition est équilibrée entre le Canada et les pays du Sud (voir les critères d'admissibilité pour plus de détails). La solution financée par l'intermédiaire de cet appel doit être axée sur les pays du Sud classés comme étant à revenu faible ou moyen selon les [définitions de la Banque mondiale](#) (en anglais).

Le réseau de solutions retenu dans ce volet devrait promouvoir des applications et des politiques d'IA sûres et durables qui tiennent compte à la fois des avantages et des risques de l'IA — en mettant à l'échelle des innovations responsables en matière d'IA, en soutenant la recherche sur les politiques, le renforcement des capacités et la création d'ensembles de données locaux.

Un réseau de solutions recevra du financement dans le cadre de ce volet.

Financement et durée

Dans le cadre de cet appel, le CIFAR fournira, pendant **une période de plus de deux ans**, un soutien financier spécifique et flexible pour les activités de recherche et développement (R-D) admissibles de trois réseaux de solutions sélectionnés, qui devront concevoir, développer et mettre en œuvre une solution novatrice à un défi lié à la sécurité de l'IA. Le réseau retenu recevra jusqu'à 700 000 \$ CA en tout, répartis comme suit :

- 60 000 \$ CA par année pour chaque coresponsable (jusqu'à concurrence de 2) et 25 000 \$ CA par année pour chaque membre du noyau central (jusqu'à concurrence de 4) afin de soutenir les activités de recherche et de développement liées au réseau;
- jusqu'à 60 000 \$ CA par année pour couvrir les coûts directs de recherche et de développement du réseau;
- un soutien financier, logistique et de gestion de projet de la part du CIFAR pour permettre deux rencontres en personne par année entre les membres du réseau et des spécialistes externes.

Critères d'admissibilité

Chaque réseau de solutions devrait être composé d'une équipe multidisciplinaire et :

- être dirigé par deux coresponsables et comprendre au plus quatre membres additionnels;
- inclure au moins une personne affiliée à une faculté ou membre du personnel d'un institut national d'intelligence artificielle (Amii, Mila ou Institut Vecteur);
- tenir compte de l'équité, de la diversité et de l'inclusion dans la composition de l'équipe (y compris la représentation des genres).

Les membres de l'équipe peuvent représenter n'importe quel secteur (universitaire, gouvernemental, industriel), mais ils doivent participer au projet en raison de leur expertise, et non en tant que mandataires d'une organisation. Le CIFAR ne fournira pas d'allocations aux membres de l'équipe provenant du secteur public ou de l'industrie.

- **Volet A : Atténuer les risques pour la sécurité des contenus synthétiques**

- L'équipe doit comprendre au moins une personne spécialisée dans les sciences sociales, les sciences humaines ou le droit.
- Elle doit également inclure au moins une personne possédant une expertise technique en matière d'IA.
- Des collaborateurs ou collaboratrices de l'étranger peuvent en faire partie, mais la plupart des membres de l'équipe doivent être établis au Canada.

- **Volet B : La sécurité de l'IA dans les pays du Sud**

- L'équipe doit compter au moins une personne établie au Canada et au moins une personne établie dans un pays du Sud (le lieu où la solution sera appliquée).
- La préférence est accordée aux équipes qui comprennent des membres provenant d'au moins deux pays du Sud.

Procédure de dépôt et échéancier

- La période de dépôt des manifestations d'intérêt s'étend **du 4 mars au 15 avril 2025**. **Toutes les manifestations d'intérêt** doivent avoir été soumises par l'entremise du portail en ligne du CIFAR au plus tard à 23 h 59, peu importe le fuseau horaire.
- Un accusé de réception sera envoyé à tous les réseaux candidats dont la manifestation d'intérêt aura été reçue avant la date limite.
- Les propositions (manifestations d'intérêt et propositions complètes) peuvent être soumises en français ou en anglais.
- Les réseaux candidats invités à soumettre une proposition complète seront avisés au plus tard à la mi-mai 2025. Les critères d'évaluation et d'examen des propositions complètes seront fournis à ce moment.
- Les propositions complètes doivent être soumises au moyen du système de proposition en ligne au plus tard le 15 juillet 2025 à 23 h 59, peu importe le fuseau horaire. Les propositions reçues après cette date limite ne seront pas prises en compte.
- Le comité d'évaluation composé de spécialistes invitera les finalistes sélectionnés à présenter leur proposition complète en personne ou de façon virtuelle lors d'une réunion du comité d'évaluation.
- La réunion du comité d'évaluation se tiendra en septembre 2025. Le réseau de solutions sera choisi à la suite de cette réunion.
- Le CIFAR s'efforcera de notifier le réseau retenu au début d'octobre 2025; une annonce publique suivra.

Exigences pour les manifestations d'intérêt

Les exigences exactes relatives à la soumission, les questions et la longueur limite peuvent être consultées sur le portail du CIFAR. Les soumissions doivent comprendre l'information suivante :

- Renseignements administratifs :
 - Titre du projet
 - Noms, affiliations et adresses complètes des membres de l'équipe du réseau
- Solution
 - Justification
 - Le CIFAR rassemble les plus brillants cerveaux du monde pour se pencher sur les enjeux les plus complexes de notre temps. Les propositions doivent décrire le contexte actuel dans lequel s'inscrit la solution ainsi que les lacunes et les défis auxquels elle remédiera. Les propositions doivent préciser comment cette solution diffère des autres solutions qui ont été tentées.
 - Retombées
 - Les solutions proposées doivent transcender l'avancement des connaissances et l'élaboration de principes et de cadres pour la gouvernance de l'IA. Les propositions retenues devront avoir des retombées importantes dans un secteur spécifique de la société et devront préciser comment la solution proposée entraînera un changement positif, à quoi elle ressemblera et quelles sont les répercussions sociétales escomptées, y compris les personnes qui en bénéficieront.
 - Méthodologie
 - Les propositions doivent présenter la méthodologie utilisée pour mettre la solution en œuvre, y compris les spécialistes que le réseau prévoit embaucher pendant le développement et la mise en œuvre du projet.
 - Pour le volet A, il sera important d'impliquer les utilisateurs et utilisatrices d'outils ou les communautés concernées par les risques pour la sécurité des contenus synthétiques.

- Pour le volet B, la participation du public ou de parties prenantes régionales est importante afin que la solution tienne compte des limites et des possibilités propres à la région concernée et y soit adaptée.
- Les propositions doivent inclure un plan pour favoriser l'intégration des principes d'équité, de diversité, d'inclusion et de durabilité lors de la conception et de la mise en œuvre du projet. Les réseaux candidats sont encouragés à faire évaluer leur proposition pour qu'elle respecte les exigences du processus d'évaluation éthique de leur établissement.
- Alignement avec la mission et les priorités du Programme de recherche de l'ICSIA au CIFAR
 - Les réseaux candidats doivent décrire comment le modèle du CIFAR, qui soutient un engagement international ou multidisciplinaire et une étroite collaboration, ajoutera une valeur particulière au projet et contribuera à positionner l'équipe de manière unique afin qu'elle génère des retombées.
 - Le dossier de candidature doit décrire comment son réseau de solutions est en phase avec la description du volet applicable.
 - Le projet doit s'inscrire dans le contexte du [Rapport international sur la sécurité de l'IA 2025](#) (en anglais) et expliquer comment il abordera directement les risques précis recensés dans le rapport.
- Évaluation des risques du projet
 - Les propositions doivent décrire tout risque potentiel prévu lors de la mise en œuvre de la solution. Elles doivent inclure tout facteur culturel, politique, technique, informatique ou sociétal qui pourrait nuire au succès de la solution et décrire comment le réseau prévoit d'atténuer les risques envisagés.
- Réseau de solutions
 - Leadership potentiel
 - Selon l'expérience du CIFAR, le leadership joue un rôle crucial dans la réussite d'un projet donné. Dans leur manifestation d'intérêt, les réseaux candidats doivent décrire le leadership potentiel pour le projet proposé. La sélection des deux coresponsables potentiels doit se faire non seulement en fonction de leur influence dans leur secteur d'activité, mais aussi de leur ouverture à la pensée de rupture, leur capacité à favoriser une solide collaboration entre les disciplines, les zones géographiques et les secteurs d'activité, ainsi que leur engagement à générer des retombées importantes. La proposition

doit indiquer clairement pourquoi ces deux personnes sont les plus aptes à diriger cette équipe et cette solution.

- Excellence de l'équipe
 - Le CIFAR crée des communautés mondiales regroupant de remarquables spécialistes qui s'engagent à collaborer de façon ouverte et soutenue, à explorer des solutions et à produire des résultats. Les membres de l'équipe doivent être des spécialistes au sein de leur communauté. La proposition doit démontrer l'excellence des membres de l'équipe et indiquer clairement qu'ils ont l'expertise multidisciplinaire nécessaire pour soutenir la solution proposée.

Processus d'examen et critères d'évaluation de la manifestation d'intérêt

Les **examens** seront effectués par les membres du Conseil de recherche de l'ICSIA au CIFAR et par d'autres spécialistes. Les manifestations d'intérêt seront évaluées en fonction des critères suivants :

- Pertinence et retombées de la solution (30 %)
 - Originalité et pertinence de la solution, notamment l'alignement avec l'appel (15 %)
 - Importance sociétale et technique des retombées attendues (15 %)
- Méthodologie (20 %)
 - Pertinence de la portée et de l'importance des activités et de la méthodologie pour atteindre les résultats escomptés, y compris la faisabilité du calendrier et les ressources (10 %)
 - Prise en compte adéquate de l'engagement du public et des principes d'équité, de diversité, d'inclusion et de durabilité dans la conception et la mise en œuvre du projet (10 %)
- Évaluation des risques du projet (10 %)
 - Description des risques associés à la mise en œuvre et des stratégies d'atténuation appropriées
- Leadership (20 %)
 - Démonstration de l'excellence des coresponsables désignés et de leur sens du leadership; preuve de leur capacité à diriger le projet, à encourager la collaboration et à favoriser la pensée de rupture
- Équipe (20 %)
 - Démonstration de l'excellence des membres de l'équipe et de leur capacité à atteindre les résultats escomptés; mise à profit de différents types de spécialistes et intégration appropriée au projet de leurs connaissances et points de vue particuliers
 - Équilibre entre les femmes et les hommes au sein de l'équipe
 - Représentation de l'expertise, soit par discipline, soit par zone géographique, en fonction des critères d'admissibilité du volet en question

Foire aux questions

1. Qu'est-ce qu'un réseau de solutions?

Un réseau de solutions est une équipe de spécialistes intersectoriels et interdisciplinaires qui conçoivent, développent et mettent en œuvre des solutions d'IA responsables, sécuritaires et bénéfiques.

2. Une organisation peut-elle faire partie d'un réseau de solutions?

Non, les membres proposés par un réseau de solutions doivent être des personnes. Une organisation dans son ensemble ne peut pas faire partie du réseau de solutions. Une personne peut représenter son organisation, mais elle doit contribuer activement au réseau de solutions en tant que membre assidu.

3. Une personne issue d'une organisation à but lucratif peut-elle faire partie d'un réseau de solutions?

Oui, mais elle ne bénéficiera pas du soutien financier accordé aux membres du réseau. Le CIFAR lui fournira une aide financière uniquement pour les frais de déplacement et d'hébergement.

4. Une personne peut-elle présenter sa candidature dans plus d'un réseau de solutions?

Non, une personne ne peut présenter sa candidature que dans le cadre d'un seul réseau de solutions.

5. Les manifestations d'intérêt non retenues recevront-elles une rétroaction?

Non. En raison du nombre de manifestations d'intérêt attendues, nous ne fournirons pas de rétroaction aux candidatures non retenues.

6. Est-il possible de proposer une solution qui a déjà été conçue?

Les manifestations d'intérêt peuvent présenter une solution qui a déjà été conçue, mais qui doit encore être développée et mise en œuvre.

7. Le CIFAR peut-il m'aider à entrer en contact avec d'autres personnes intéressées par cet appel?

Le personnel du CIFAR peut, à sa discrétion, vous aider à établir des liens avec des personnes d'intérêt pour la communauté du CIFAR (comme les titulaires de bourses de recherche du CIFAR ou les titulaires de chaires en IA Canada-CIFAR).

8. Comment le financement peut-il être utilisé par le réseau de solutions?

Le financement du réseau de solutions est flexible et peut servir à payer les frais liés au dégageement de fonctions (enseignement ou cours), au soutien à la rémunération, à l'aide aux stagiaires, aux frais de voyage et d'hébergement, aux services professionnels et techniques, en plus d'autres coûts associés aux activités de R-D.

9. Le réseau de solutions peut-il recevoir du financement d'autres sources pour la solution proposée?

Oui, le réseau de solutions est autorisé à recevoir du financement d'autres sources pour la solution proposée.

10. D'autres dispositions financières peuvent-elles être prises dans le cas où un membre du réseau de solutions ne pourrait pas recevoir de soutien financier?

Non, nous ne pouvons pas prendre d'autres dispositions si une personne ne peut pas recevoir de financement dans le cadre de ce programme.

11. Combien de réseaux de solutions recevront du financement?

Pour le volet A, deux réseaux de solutions retenus seront financés. En ce qui concerne le volet B, un réseau de solutions retenu recevra du financement. Les prochains appels du programme Réseaux de solutions seront annoncés sur le site Web cifar.ca.

12. Quels sont les différents éléments de la manifestation d'intérêt?

Les questions de l'appel à manifestation d'intérêt se trouvent dans le [portail de candidature du CIFAR](#). Il suffit de créer un compte pour les consulter. La personne qui soumet la candidature doit être un membre du réseau. Nous encourageons les équipes à travailler à partir d'un document partagé avant de soumettre leurs réponses dans le portail de candidature.

13. Quelles qualifications en matière d'expérience et d'expertise recherchez-vous chez les membres du réseau de solutions proposé?

Outre les critères d'admissibilité définis pour les membres du réseau, il n'y a pas de critères particuliers en fait d'expérience professionnelle et d'expertise. Le CIFAR crée des communautés mondiales regroupant de remarquables spécialistes qui s'engagent à collaborer de façon ouverte et soutenue, à explorer des solutions et à produire des résultats. Les membres de l'équipe doivent être des spécialistes au sein de leur communauté. La proposition doit démontrer l'excellence des membres de l'équipe et indiquer clairement qu'ils ont l'expertise nécessaire pour soutenir la solution proposée.

14. Les propositions doivent-elles être davantage axées sur la recherche ou sur la mise en œuvre de solutions à des problèmes concrets avec de vrais utilisateurs et utilisatrices?

Le réseau de solutions financé dans le cadre de cet appel devrait :

- appliquer des connaissances de pointe issues de la recherche et de la pratique au développement de nouvelles interventions, applications ou approches pour répondre aux défis sociétaux actuels liés à la mise en œuvre sécuritaire de l'IA;
- mettre à l'essai des prototypes pour favoriser leur adoption, ce qui devrait comprendre une consultation de la communauté au cours des 18 premiers mois du projet (si elle n'est pas déjà réalisée au moment où le réseau est financé);
- éclairer les politiques ou les pratiques qui soutiennent et renforcent la confiance envers l'IA et qui atténuent les préjudices sociaux, éthiques et économiques potentiels;
- entreprendre des efforts pour la mise à l'échelle de l'intervention.

Les propositions dont l'objectif principal est de faire progresser les connaissances ou la compréhension d'un domaine d'étude ou qui se concentrent uniquement sur des solutions techniques (p. ex., l'amélioration des algorithmes d'apprentissage automatique) ne seront pas prises en considération.

Pour toute question concernant cet appel, veuillez écrire à ai@cifar.ca