

**Le présent défi est publié en vertu de l'appel de propositions lié au défi du PICC (EN578-DB1800).  
Veuillez consulter [les documents de l'invitation à soumissionner](#).**

**Nom du défi :** Apprentissage machine et traitement du langage naturel pour l'automatisation du développement de la revue systématique et de la méta-analyse

**Commanditaire du défi :** Agence de la santé publique du Canada (ASPC) / ministère de la Défense nationale (MDN)

**Avis du défi :** EN578-DB1801

**Zone de priorité :** Volet militaire

### **Financement**

Ce défi est entièrement financé par l'ASPC.

Phase I – 2018-2019 : Le financement disponible pour la phase I est d'environ 225 000 \$ par contrat. Jusqu'à deux contrats pourraient être attribués pour cette phase.

Phase II – 2019-2020 : Le financement disponible pour la phase II peut aller jusqu'à 450 000 \$. La phase II peut être exercée pour l'un des entrepreneurs à qui un contrat de phase I a été attribué au départ.

Le financement est en dollars canadiens. Le fait de divulguer l'estimation du financement disponible n'engage aucunement le Canada à payer les sommes mentionnées ci-dessus.

### **Contexte/résumé**

En tant qu'organisme fédéral responsable de la prévention et de la gestion des maladies, l'Agence de la santé publique du Canada (ASPC) a pour mandat de promouvoir et de protéger la santé des Canadiens. Afin d'aider les professionnels de la santé, les décideurs et les citoyens à prendre des décisions éclairées, l'ASPC doit procéder en temps opportun à une analyse précise et complète des données probantes. Pour s'acquitter de son mandat, l'ASPC a besoin d'une capacité accrue d'effectuer des examens rapides (en temps réel) et complets des données probantes.

Une plateforme d'apprentissage machine (AM) et de traitement du langage naturel (TLN) est requise pour effectuer des examens systématiques des preuves qui s'accumulent en lien avec le processus décisionnel concernant les politiques et programmes cliniques de santé publique.

Dans ce contexte, une innovation AM-TLN est recherchée afin d'entreprendre une approche systématique, structurée et transparente pour la recherche, le tri et le classement de la documentation scientifique publiée et examinée par les pairs en fonction d'une question de recherche prédéfinie. L'innovation doit avoir la capacité d'extraire et d'exporter les principaux résultats d'études publiées dans des tableaux de données et de preuves prédéfinis.

### **Détails du défi**

Le défi vise à déterminer « l'innovation » (une plateforme d'apprentissage machine [AM] et de traitement du langage naturel [TLN] pour améliorer les examens systématiques des données sur la santé) et à mettre à l'essai la performance de l'innovation aux bureaux de l'ASPC et du MDN à Ottawa.

**Note :** Les éléments de défi suivants seront évalués conformément au formulaire de soumission de propositions et à la grille d'évaluation.

L'innovation doit posséder toutes les caractéristiques suivantes :

- La plateforme doit pouvoir traiter un éventail de questions de santé publique en fonction de différentes populations et interventions et de différents groupes de comparaison et de résultats attendus.
- L'innovation doit être modulaire (p. ex., modules « de récupération de données et de publications », « d'examen préalable et de sélection », « d'extraction de données » et « de compte rendu sommaire ») et devrait contenir les éléments particuliers plus bas, qui seront évalués conformément au formulaire de soumission de propositions et à la grille d'évaluation.
- La recherche et la sélection de publications et de données issues de recherches pertinentes (« récupération de données et de publications »), filtrées en fonction du cadre population, intervention, facteurs de comparaison et résultats (PICO) ou de variables du cadre PICO (p. ex., population, résultats) à l'aide d'une combinaison des stratégies suivantes :
  - Traiter des articles pertinents en libre accès et récupérables en texte intégral HTML au moyen d'interfaces de programmation d'applications (API) d'éditeurs.
  - Traiter des articles pertinents en texte intégral en format PDF qui peuvent être versés manuellement dans la plateforme par l'utilisateur final.
- L'extraction de données d'études sélectionnées ou pertinentes dans des tableaux de données organisés étude par étude en fonction des renseignements clés pertinents pour la question de recherche (p. ex., caractéristiques de la population de chaque étude, résultats rapportés de chaque étude).
- Documentation transparente du processus d'examen systématique (p. ex., combien d'études triées aux fins d'inclusion, combien d'études sélectionnées, etc.), affichée de façon graphique sous la forme d'un organigramme généré automatiquement.

L'innovation devrait posséder autant des caractéristiques suivantes que possible :

- Pour la récupération initiale de publications et de données (module de tri et de sélection), il est préférable que ce module comporte une interface directe avec des bases de données de publications (p. ex., Scopus, OvidMEDLINE) pour accéder à tous les articles et résumés aux fins de tri en fonction de la pertinence. S'il n'est pas possible d'avoir une interface directe avec des bases de données de publications pour trier des catalogues entiers, il serait acceptable de commencer le tri selon la pertinence à partir d'une liste prédéfinie d'études potentiellement pertinentes établie à l'aide d'une recherche par mot clé dans les bases de données des publications selon les directives du bibliothécaire du domaine de la santé du gouvernement du Canada (GC). Il est aussi préférable que ce module ait une capacité d'exécuter d'autres tris des références citées dans les études sélectionnées et d'inclure ces études citées dans l'analyse finale lorsqu'elles sont pertinentes (c.-à-d., « recherche manuelle » ou « recherche à la chaîne »);
- Il est fortement encouragé que les modules soient développés « à la chaîne »;
- Il est fortement recommandé que les modules soient conçus selon le format d'annotation normalisé (p. ex., le format d'annotation TLN ou un format similaire) afin d'assurer l'interopérabilité avec des applications actuelles et à venir de tierces parties et du GC, selon les normes numériques du GC (<https://ouvert.canada.ca/fr/blog/amelioration-des-services-du-gouvernement-lere-numerique>);
- Après l'extraction de données des études sélectionnées, la plateforme devrait comprendre un module permettant de générer automatiquement du texte pour la production d'un compte rendu sommaire de l'estimation des résultats;
- Il est préférable que la plateforme soit dotée d'une capacité permettant de classer automatiquement les études en fonction du type et du format ainsi que de la pertinence avec la question du cadre PICO, conformément aux normes établies (p. ex., évaluation du risque de partialité de Cochrane pour les essais aléatoires).

### **Proposition**

Les soumissionnaires doivent fournir dans leurs propositions les détails techniques de l'innovation et expliquer comment elle satisfait aux objectifs ci-dessus. Vous pouvez accéder au formulaire électronique de soumission de la proposition dans [l'avis du défi](#).