

Thématique : Santé / Intelligence Artificielle

Challenges Intelligence Artificielle

Règlement du challenge IA

Sponsor : CHU de Toulouse



Fiche d'identité du Challenge

ELEMENTS CLES DU CHALLENGE

Nom du sponsor : Centre Hospitalier
Universitaire de Toulouse

Statut juridique¹ : Etablissement public de santé

Type de structure (PME / ETI /GE etc.) :
établissement de recherche

CONTACT POUR LE CHALLENGE

Prosper BURQ

Fonction : Data scientist

Numéro de téléphone : 05 61 77 75 68

RESUME DU CHALLENGE

Ce challenge consiste à apporter une solution novatrice permettant d'intégrer les dernières avancées en Intelligence Artificielle pour faciliter l'anonymisation des données de santé et dynamiser l'écosystème IA en santé. Le CHU dispose de nombreuses bases de données à fort potentiel. Le lauréat pourra ensuite commercialiser la solution auprès d'autres producteurs de données de santé. Il leur permettra ainsi de mettre en œuvre un certain nombre de partenariats avec le tissu économique et de valoriser leurs données.

¹ Si le sponsor est un consortium, précisez à chaque fois les différents membres du consortium.

Sommaire

1. CONTEXTE ET OBJECTIFS DU CHALLENGE.....	4
2. ETAT DE L'ART DE LA SOLUTION ACTUELLE ET CHALLENGE A RESOUDRE	4
3. RESULTATS ATTENDUS.....	5
4. JEUX DE DONNEES A DISPOSITION ET EQUIPE MOBILISEE PAR LE SPONSOR.	5
5. COMPETENCES ET ENGAGEMENTS DU SPONSOR	6
6. DEROULEMENT DU CHALLENGE	7
7. EXIGENCE ET CRITERES DE SELECTION.....	7
8. RESTITUTION DU CHALLENGE.....	7
9. PERSPECTIVES ET RETOMBEEES POSSIBLES DU CHALLENGE POUR LE LAUREAT	8

1. CONTEXTE ET OBJECTIFS DU CHALLENGE

L'enjeu principal du moment dans la santé est l'exploitation des données de vie réelle (hospitalière), qu'elles soient rétrospectives ou prospectives. Le CHU de Toulouse dispose d'une très grande quantité de données dans différentes spécialités médicales.

L'Intelligence Artificielle est un levier de l'utilisation et de l'exploitation de ces données. Elle permet d'atteindre des objectifs d'amélioration de prise en charge des patients. Ces nouveaux outils faciliteront le développement de nouveaux usages autour de la donnée et des pratiques médicales (aide au diagnostic, médecine personnalisée, e-santé, analyse automatique d'image...).

Afin d'exploiter ces données, plusieurs défis techniques restent à résoudre et notamment le partage et le transfert des données dans un cadre réglementaire protecteur de la vie privée.

2. ETAT DE L'ART DE LA SOLUTION ACTUELLE ET CHALLENGE A RESOUDRE

Afin d'exploiter les données de santé dans le cadre réglementaire actuel et sans dénaturer la donnée, plusieurs défis techniques restent à résoudre :

- L'extraction des données et leur restitution à des formats intelligibles, normés et interopérables ;
- L'exploitation mathématique de ces données massives par des outils adaptés au monde médical ;
- Le partage et le transfert des données dans un cadre réglementaire respectueux de la vie privée.

C'est ce troisième défi que nous souhaitons aborder car il nous paraît actuellement le plus limitant dans l'ensemble des projets de collaboration entre les établissements détenteurs de données de santé et le tissu économique local.

Pour cela, il faudrait développer une solution d'anonymisation des jeux de données et de ciblage des informations identifiantes sur des données de santé.

La plupart des solutions d'anonymisation actuelle se contentent de supprimer ou flouter certains champs de la donnée (en suivant les préconisations de la HIPAA qui datent de 1996) ou de supprimer des métadonnées dans les DICOMs. Cela n'est pas suffisant car de nombreuses données directement ou indirectement identifiantes peuvent toujours se trouver dans les données.

D'autres méthodes plus abouties ajoutent tellement de bruit dans les données qu'elles en deviennent inutilisables par des algorithmes. Le groupe de travail « Article 29 » de la Commission Européenne a publié un avis avec des critères à respecter. Ceux-ci pourraient être mis en place en bénéficiant des techniques d'IA pour faciliter l'automatisation et la généralisation de l'anonymisation.

Le seul moyen de travailler sur des données de santé avec un risque zéro vis-à-vis du patient est de les anonymiser. Il faut donc trouver une solution technique permettant de respecter la réglementation et la vie privée du patient, tout en permettant l'innovation.

3. RESULTATS ATTENDUS

L'objectif est de fournir une solution viable d'anonymisation de jeu de données en exploitant les mécanismes de représentation de la donnée utilisée en intelligence artificielle (par exemple en utilisant des mécanismes d' « embeddings »). Ces représentations abstraites de la donnée pourraient permettre d'anonymiser efficacement nos données de santé sans endommager leurs propriétés statistiques. Elles resteront donc utilisables facilement par des entreprises pour développer des produits innovants.

4. JEUX DE DONNEES A DISPOSITION ET EQUIPE MOBILISEE PAR LE SPONSOR

Dans le cadre d'un projet de recherche en cours, les données suivantes portant sur 100 patients seront disponibles :

- Données d'imagerie ;
- Texte ;
- Données biologiques ;
- Données génomiques ;
- Données observationnelles.

Ce projet couvre le périmètre de données le plus large mais ne concerne qu'une petite cohorte de patients. Pour entraîner des algorithmes spécifiques, nous pourrions aussi fournir des jeux de données plus restreints (par exemple uniquement des textes) mais sur des cohortes beaucoup plus importantes. Enfin, s'il est trop complexe de mettre en place un outil couvrant tout ce périmètre dans le délai imparti, nous pouvons prioriser ces différents secteurs (voire utiliser un autre jeu de donnée plus approprié si nécessaire).

Dans le cadre de ce projet, nous pourrions solliciter différents utilisateurs de la donnée finale afin d'avoir un retour sur la qualité de l'anonymisation.

Pour ce challenge, le sponsor s'engage à mettre à disposition un environnement datascience complet. Celui-ci sera composé à minima d'espaces de travail sécurisés (Python, R) et des ressources de calculs nécessaires (par exemple serveur de calcul avec GPU).

Si le lauréat a des besoins supplémentaires, nous pourrions faire évoluer cet environnement de façon adéquate tant sur la partie hardware que software.

Les personnes suivantes participeront régulièrement à la mise en œuvre du projet :

- Directeur des affaires financières du CHU de Toulouse : Arnaud Joan-Grangé ;
- Data scientist : Prosper Burq ;
- Data manager : Yan Zhao ;
- Chef de projet : Beatrice Appiah ;
- ARC Règlementaire : Isabelle Olivier.

D'autres intervenants pourront être mobilisés au fil du projet en fonction des besoins d'expertises.

S'agissant de données de santé, le lauréat devra respecter la confidentialité des informations reçues et les démarches relatives au RGPD devront être réalisées.

5. COMPETENCES ET ENGAGEMENTS DU SPONSOR

Le CHU de Toulouse est le premier employeur de l'ex région Midi-Pyrénées, accomplissant ses missions de soins, de prévention, d'enseignement et de recherche grâce à près de 4 000 médecins et 12 000 personnels hospitaliers (soit près de 16 000 personnes au service de la santé).

Il prend en charge environ 280 000 malades et 800 000 consultants chaque année, ce qui le situe au quatrième rang des hôpitaux français en terme d'activité. Seul CHU de la région Occitanie Ouest, les soins de recours représentent plus de 50 % de son activité, grâce à des plateaux techniques et des équipes de très grande qualité. Le CHU de Toulouse coordonne au niveau national ou est associé à 13 centres de références des maladies rares et 68 centres de compétences.

Dans le domaine des soins, le CHU de Toulouse se distingue, depuis 20 ans, en tête des palmarès réalisés par la presse nationale.

Le CHU de Toulouse est un acteur majeur de la cancérologie en Occitanie Ouest : le traitement du cancer représente environ 25 % de son activité. En 2014, il s'est associé avec ses partenaires publics et privés, en particulier l'Institut Claudius Regaud pour former l'Institut universitaire du cancer de Toulouse (IUCT), nouveau modèle français d'organisation des soins, de la recherche et de l'enseignement en cancérologie.

La mission de recherche et d'innovation du CHU fait partie intégrante à la fois de son activité quotidienne et de sa stratégie pour l'avenir. Elle est menée en collaboration étroite avec les facultés et les organismes de recherche que sont notamment l'Institut national de la santé et de la recherche médicale (Inserm) et le Centre national de la recherche scientifique (CNRS). Les équipes du CHU mènent également de nombreux projets avec les entreprises du médicament et du dispositif médical, les associations et les autres centres hospitaliers.

Le CHU de Toulouse dispose d'une énorme quantité de données. À ce jour, l'exploitation d'une petite partie de ces données dans le cadre réglementaire actuel est extrêmement compliquée et requiert beaucoup de temps et de personnes.

Une solution d'anonymisation complète faciliterait et accélérerait donc considérablement l'exploitation de l'ensemble des données du CHU de Toulouse. Ainsi, nous serions en mesure de :

- Valoriser nos données dans le cadre de partenariats tiers ;
- Accélérer la recherche sur données et ainsi faire plus de publications.

6. DEROULEMENT DU CHALLENGE

- M0-M2 : Démarches règlementaires et Signature du contrat de partenariat entre le sponsor et le lauréat (ou sponsor et Bpifrance) assurant la confidentialité des informations transmises ;
- M0-M1 : Réunion(s) de préparation du projet ;
- M2 : Mise à disposition sécurisée des données ;
- M3-M5 : Première version v0 du prototype ;
- M5-M11 : Itération du lauréat et sponsor pour ajouter des données ;
- M5-M11 : Evaluation de la solution à chaque itération ;
- M11 : Finalisation de la solution ;
- M9-M11 : Validation de la solution par le sponsor.

7. EXIGENCE ET CRITERES DE SELECTION

- Expérience dans le traitement des données massives ;
- Expérience dans la diversité des données de santé (standards et formats) ;
- Expertise dans l'IA ;
- Culture du développement agile ;
- Culture de la sécurité des données ;
- Expertise sur les problématiques liées à l'anonymisation.

Ouverture pour une proposition partenariale et commerciale entre le lauréat et le CHU pour l'accès à la solution développée, son utilisation et sa promotion.

8. RESTITUTION DU CHALLENGE

Outre les restitutions régulières pendant le développement, une restitution plus élargie est attendue au plus tard 1 an après la sélection du lauréat sous forme d'une réunion avec l'ensemble des utilisateurs du CHU. Cette présentation sera accompagnée d'un atelier.

9. PERSPECTIVES ET RETOMBÉES POSSIBLES DU CHALLENGE POUR LE LAUREAT

La solution du challenge pourra cibler le marché de la donnée de santé c'est-à-dire l'ensemble des centres producteurs de soin mais également les TPE, PME et industries qui souhaitent exploiter les données de santé en respectant le cadre légal.

Un tel produit pourrait en premier lieu adresser le marché médical public, gros producteur de données de santé, et aboutir à terme à la vente d'un service français qui sera commercialisé par le lauréat du challenge.

L'ambition pour ce produit est qu'il soit le plus générique possible. On peut donc très bien envisager qu'il puisse être commercialisé ensuite à l'international, en particulier dans les pays où la réglementation sur les données de santé est particulièrement protectrice.