



# APPEL À MANIFESTATION D'INTÉRÊT

## ► DATA CHALLENGE

Tu es un acteur local azuréen ? Tu possèdes un patrimoine de données spécifiques à ton domaine et tu penses que l'usage du Machine Learning et de l'Intelligence Artificielle appliqués à ces données accélérerait ton business ? Ce data challenge est là pour toi. Deviens sponsor data du premier data challenge industriel de la Côte d'Azur !



## QU'EST CE QU'UN DATA CHALLENGE ?

► Un **data challenge** est une compétition visant à résoudre des problématiques spécifiques en utilisant l'analyse des données. Ce challenge est organisé par un comité de pilotage constitué par la MIA, le ClusterIA, le 3IA Côte d'Azur et le Pôle SCS.

► Le **sponsor data** fournit une base de données appropriée aux **compétiteurs** pour servir de source d'information à l'analyse. Les compétiteurs participants accèdent à ces données via une plateforme dédiée. Ils développent ensuite des algorithmes d'apprentissage automatique pour répondre à la question posée.

► À la fin de la compétition, les meilleurs compétiteurs sont récompensés par un prix en reconnaissance de leur travail et des résultats obtenus par l'algorithme qu'ils ont développé. Les data challenges permettent ainsi de stimuler l'innovation et la résolution de problèmes dans différents domaines.



## OBJECTIFS

Le présent appel à projet vise à sélectionner un sujet de data challenge qui bénéficiera d'un accompagnement de 4 mois de bout en bout par les membres du comité de pilotage.

**Cette démarche poursuit les intérêts et objectifs suivants :**

- favoriser le développement d'outils précis d'analyse et d'assistance par l'usage de l'intelligence artificielle,
- promouvoir l'innovation,
- pour les compétiteurs, l'occasion de se faire repérer par des acteurs majeurs du domaine et dénicher des talents,
- démontrer par des résultats rapides le bénéfice concret de l'utilisation de la donnée du sponsor data,
- enrichir le catalogue d'algorithmes partagés à la communauté data du territoire azuréen.



## ACCOMPAGNEMENT PROPOSÉ

L'organisation du data challenge nécessitera une forte mobilisation de votre organisation. Un data challenge réussi passe par une forte implication du sponsor data !

Le comité de pilotage fournira un soutien complet au projet sélectionné tout au long du développement de la mise en place de la problématique data challenge, en mettant à disposition des compétiteurs les ressources techniques et humaines nécessaires.

### ► Définition et cadrage du data challenge

- aide au cadrage scientifique et organisationnel du projet,
- assistance juridique sur la réglementation applicable au traitement des données,
- coordination avec des prestataires spécialisés dans l'hébergement de data challenges,
- partage des bonnes pratiques.

### ► Gestion de l'infrastructure d'hébergement des données

- mise à disposition de serveurs sécurisés propriétaires souverains et opérés par le comité de pilotage : 16 GPU,
- pas de financement pour le sponsor data pour l'hébergement et la mise à disposition des données.

### ► Mise en valeur des résultats

- communication pré et post challenge,
- accompagnement dans la publication d'articles scientifiques,
- support dans les démarches Open Data et Open Source,
- mise en relation des acteurs de l'écosystème,
- aide restreinte à la collecte et à l'anonymisation des données.



## ORGANISATION DE LA COMPÉTITION

### ÉTAPE 1

JUIN

#### ► Manifestation d'intérêt

- constitution du comité de pilotage, mise en place des suivis,
- recrutement CDP,
- définition des thématiques d'intérêts,
- rédaction d'un AMI,
- campagne de prospection,
- jury de sélection,
- contractualisation ou cadrage avec le sponsor data.

### ÉTAPE 2

JUILLET À SEPTEMBRE

#### ► Cadrage challenge

- définition de la question et des aspects scientifiques,
- constitution des comités,
- identification des coûts et financement projet,
- définition du règlement intérieur de la compétition,
- cadrage juridique, définition protocole de collecte.

### ÉTAPE 3

JUIN JUILLET OCTOBRE

#### ► Mise en place plateforme

- définition des besoins, sélection, contractualisation avec le prestataire éventuel,
- rédaction des instructions et infos à destination des compétiteurs (FR et EN) à publier sur la plateforme,
- mise à disposition des jeux de données.

### ÉTAPE 4

JUIN À NOVEMBRE

#### ► Valorisation et communication

- définition de la stratégie de communication,
- diffusion auprès de l'écosystème, des communautés,
- analyse des résultats et publications scientifiques éventuelles.

### ÉTAPE 5

NOVEMBRE

#### ► Présentation lauréat lors du Sophia Summit

- 20 novembre 2023





## DÉROULEMENT DE LA COMPÉTITION

► **Début de la compétition** : lundi 25 septembre 2023

► **Fin de la compétition** : lundi 6 novembre 2023

► Les participants à la compétition passent par plusieurs phases :

→ **Phase d'inscription**

Les compétiteurs s'inscrivent au data challenge.

→ **Phase d'apprentissage**

Les compétiteurs entraînent leurs algorithmes en utilisant les données annotées mises à disposition sur la plateforme du challenge.

→ **Phase de soumission**

Les compétiteurs évaluent leurs algorithmes en les testant sur un jeu de données intermédiaire fourni par l'organisation. Les compétiteurs sont limités en nombre de soumission et puissance de calcul à un volume hebdomadaire afin de maintenir un engagement tout au long de la compétition. Un LeaderBoard est à disposition afin de suivre la progression des meilleurs compétiteurs.

→ **Phase de promotion des lauréats**

► Les 3 meilleurs compétiteurs ayant obtenu le plus haut score lors des soumissions présentent leurs travaux et l'approche proposée lors du Sophia Summit 2023 organisé par le 3IA Côte d'Azur le 20 Novembre 2023 à Nice.



## CANDIDATURE SPONSOR DATA

Les sponsor data candidats à l'appel à manifestation d'intérêts devront constituer un support unique (max 10 pages) au format pdf afin de décrire les différents aspects de leur projet à fournir au comité de pilotage : [contact@clusteria.fr](mailto:contact@clusteria.fr)

► **Date limite de soumission** : vendredi 23 juin 2023

► **Le support de candidature devra présenter les éléments suivants :**

→ présentation du porteur de projet unique (point de contact unique) et description de l'équipe projet qui sera mobilisée tout au long des étapes de préparation et pilotage du data challenge.

→ domaine et problématique envisagée, intérêt et caractère innovant,

→ description et justification des données fournies par le candidat, source, présence, qualité et traitement des données RGPD si applicable, capacité à ré-identifier la base fournie à l'organisation,

→ présentation du problème métier à résoudre lors de la compétition : approche IA, métrique d'évaluation et critères de mesure du succès,

→ modalités de partage des données souhaité avec l'organisation / capacité à collecter les données par l'organisation, → principales difficultés identifiées (ressources, matériel, qualité des données) par le candidat sur sa problématique.

► **Les thématiques éligibles à candidature sont les suivantes :**

→ IA fondamentale,

→ IA pour la médecine computationnelle intégrative,

→ IA pour la biologie computationnelle,

→ Territoire intelligent et sécurisé,

→ En complément de ces 4 axes les sujets relatifs à l'IA pour la cybersécurité, l'IA pour l'observation de l'espace et la terre et l'IA pour l'environnement seront considérés.

► **Jalonnements section sponsor :**

