



Appel à Manifestation d'Intérêt

« Mutualisation de données pour l'intelligence artificielle »

Date de clôture : 16 novembre 2018

Cet appel à manifestation d'intérêt s'inscrit dans le cadre de de la stratégie française en matière d'intelligence artificielle présentée le 29 mars 2018 par le président de la République.

Les réponses¹ doivent être adressées exclusivement sous forme électronique à l'adresse ami-mutualisation.dge@finances.gouv.fr avant la clôture de l'appel à manifestation d'intérêt, la date et l'heure de réception faisant foi :

16 novembre 2018 à 18h00

Pendant toute la durée de cet appel à manifestation d'intérêt, des questions pourront être soumises à l'adresse ami-mutualisation.dge@finances.gouv.fr. Les réponses à ces questions seront publiées au fil de l'eau sur le site de la Direction Générale des Entreprises, à l'adresse <https://www.entreprises.gouv.fr/numerique/mutualisation-de-donnees-pour-intelligence-artificielle>.

¹ La taille des courriers électroniques contenant les réponses ne pourra pas excéder 10 Mo. Si nécessaire, pour les réponses dépassant les 10Mo, un mode alternatif de réception pourra être proposé sur demande à l'adresse ami-mutualisation.dge@finances.gouv.fr

Table des matières

1 -	Contexte	4
2 -	Objectif.....	4
3 -	Cadrage.....	5
a)	Approche générale.....	5
b)	Principales caractéristiques des initiatives que l'Etat pourrait soutenir	5
c)	Types de dépenses susceptibles de faire l'objet d'un soutien financier de l'Etat	5
d)	Propositions de modalités d'intervention de l'Etat	5
4 -	Calendrier	6
	ANNEXE - Modèle de format de réponse	7

1 - Contexte

Partant du constat que les données sont la matière première de l'intelligence artificielle (IA) contemporaine et que de leur disponibilité dépend l'émergence de nombreux usages et applications, **la mission Villani² a souligné la nécessité de mener « une politique offensive de la donnée visant à favoriser son accès, sa circulation et son partage » et recommandé de « fédérer les différents écosystèmes autour de plateformes de mutualisation sectorielles ».**

L'apprentissage par données est en effet actuellement une méthode massivement utilisée pour construire des algorithmes d'IA. Par exemple, l'apprentissage profond ou *deep learning* nécessite d'entraîner les algorithmes associés à ces méthodes sur des jeux massifs de données. Au fur et à mesure de leurs itérations sur ces jeux de données préalablement préparés, les algorithmes auto-adaptent leurs paramètres pour fournir la solution attendue. De manière générale, plus les données représentatives des problématiques à traiter sont nombreuses, plus les processus d'ajustement des paramètres sont pertinents. **En outre, au-delà des méthodes d'apprentissage, la constitution de jeux de données est aussi nécessaire pour évaluer les algorithmes d'IA, que ce soit pour mesurer leur performance, contrôler la conformité de leur fonctionnement par rapport à leur finalité ou évaluer leur sécurité.**

Un enjeu fondamental pour ce qui est des applications industrielles de l'IA, est donc la collecte et l'annotation de larges corpus de données. **Or de tels jeux de données sont aujourd'hui des ressources rares pour les acteurs français.** En fonction des secteurs d'activité, les raisons de cette rareté sont souvent différentes : volumes insuffisants de données réelles existantes, capacités de collecte limitées, etc. Les pratiques de mutualisation de données sont régulièrement avancées comme une solution pertinente car les opérations de collecte et d'annotation sont d'une grande complexité, et peuvent être très consommatrice en temps, en ressources humaines et financières. Il est en outre fréquent que celui qui collecte la donnée ne soit pas le seul à pouvoir en tirer un bénéfice, ou le mieux placé pour l'exploiter.

Ces constats ont conduit le Président de la République à annoncer, dans le cadre de la stratégie nationale en matière d'IA, le soutien à la création de plateformes de partage de données³.

2 - Objectif

Cet appel à manifestation d'intérêt (AMI) vise à recueillir l'intérêt d'acteurs privés et publics pour un futur appel à projets soutenant des initiatives de mutualisation de données au sein de plateformes sectorielles ou cross-sectorielles, **leur avis sur les modalités de cofinancements privé-public les plus adaptées** qui pourraient être mises en oeuvre et **leurs propositions de modèles économiques** permettant aux initiatives soutenues de générer une activité viable et pérenne, sans nouveaux financements publics, à un horizon moyen de 3 ans.

Le présent appel à manifestation d'intérêt s'inscrit par ailleurs dans une démarche de co-construction afin de s'assurer de l'adéquation du futur appel à projets avec les besoins réels des acteurs. De nombreux obstacles à l'émergence de pratiques de mutualisation de données ont en effet d'ores et déjà été identifiés par la DGE (caractère stratégique de ces données pour les entreprises, coût d'identification pour chaque acteur des données non sensibles mais à valeur ajoutée pour l'entraînement d'algorithmes, complexité de la gouvernance pour garantir la pérennité et l'enrichissement des jeux de données, modèle économique restant à inventer...). **Il est ainsi attendu que les réponses reçues contribuent à définir de manière plus précise les caractéristiques des initiatives ciblées** par le futur appel à projets (assiette financière, nombre de partenaires, type de partenaires, type de structures porteuses, activités de ces structures ...).

Cet AMI s'adresse aussi bien aux potentiels futurs porteurs de projets, qu'aux structures susceptibles de les accompagner ou de structurer leur réflexion, en particulier les pôles de compétitivité et les fédérations professionnelles. En s'appuyant notamment sur les informations recueillies au travers des

² https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf

³ <http://www.elysee.fr/declarations/article/transcription-du-discours-du-president-de-la-republique-emmanuel-macron-sur-l-intelligence-artificielle/>

réponses à cet AMI, le cahier des charges d'un appel à projets sera rédigé puis publié début 2019. Une contribution à cet AMI ne sera pas un préalable nécessaire à la soumission d'un dossier en réponse au futur appel à projets.

3 - Cadrage

a) Approche générale

L'objet principal du soutien de l'Etat doit être une initiative sectorielle ou cross-sectorielle visant **la constitution et/ou l'enrichissement et/ou l'exploitation de très grandes bases de données mutualisées répondant aux attentes avérées d'un nombre significatif d'acteurs** pour le développement de solutions d'IA. Il est demandé aux acteurs de s'exprimer sur les projets qu'ils souhaiteraient voir soutenus, *a priori* par des dispositifs existants mais sans exclure des idées de nouveaux dispositifs qu'ils trouveraient pertinents.

b) Principales caractéristiques des initiatives que l'Etat pourrait soutenir

Le soutien de l'Etat sera a minima conditionné à :

- la démonstration d'une initiative économiquement viable et pérenne sans aide financière de l'Etat à un horizon moyen de 3 ans ;
- la démonstration d'un fort impact sur l'économie nationale ;
- l'utilisation de formats de données exclusivement ouverts et d'interfaces applicatives de programmation (ou *API*) documentées ;
- la démonstration du respect du règlement relatif à la protection des données personnelles pour les initiatives concernées et la formalisation de principe de gouvernance des données.

c) Types de dépenses susceptibles de faire l'objet d'un soutien financier de l'Etat

Il s'agira, a minima, des dépenses pour les activités suivantes :

- la constitution d'une base de données massive (répartie ou non) à partir de bases tierces ;
- l'intégration de nouvelles données dans une base de données massive déjà constituée ;
- l'annotation des données avec, si nécessaire, le développement d'outils et de méthodologies associés ;
- le développement des interfaces techniques permettant la mise à disposition des données ;
- le développement d'outils de gestion technique et de facturation.

Si la démonstration de leur opportunité est apportée, notamment pour l'équilibre et la viabilité du modèle économique de l'initiative, d'autres activités pourraient faire l'objet d'un soutien de l'Etat, en particulier l'utilisation ou l'achat d'une infrastructure matérielle et le développement d'outils et de méthodologies d'évaluation d'algorithmes, notamment à des fins de délivrance de labels

En aucun cas ni l'immobilier ni la reconstruction d'outils existants ne fera l'objet d'un soutien financier de l'Etat.

d) Propositions de modalités d'intervention de l'Etat

Deux types d'interventions distincts de l'Etat sont envisagés :

- interventions en subventions et avances remboursables, dans une logique d'amorçage, en attendant un modèle de financement pérenne ;
- prises de participations minoritaires.

Concernant les interventions en subventions et avances remboursables, le modèle de l'appel à projets « structuration de filières »⁴ pourra être considéré. D'autres modalités d'intervention peuvent être envisagées et suggérées en s'appuyant sur les régimes d'aides proposés à la Commission européenne sous forme d'exemption (cf. règlements UE n° 651/2014 et n° 2017/1084). Aucune procédure de notification à la Commission ne sera toutefois entreprise dans le cadre du soutien aux initiatives de mutualisation de données au sein de plateformes sectorielles.

Les prises de participations concerneront des structures, dotées d'une personnalité juridique propre, dont l'activité principale sera le portage d'une initiative de mutualisation de données au sein d'une plateforme sectorielle. Les prises de participation seront toujours minoritaires. La participation des partenaires privés, majoritaire, pourra être constituée d'apports en nature et d'apports en numéraire. Les apports en numéraire des partenaires privés devront être dimensionnés pour répondre aux besoins de financement, tant de l'investissement que du fonds de roulement.

4 - Calendrier

L'AMI se déroulera de la manière suivante :

- le 10 septembre 2018, lancement de l'AMI sur la base du présent document ;
- pendant toute la durée pendant laquelle cet AMI sera ouvert, des questions pourront être soumises par écrit à l'adresse ami-mutualisation.dge@finances.gouv.fr - des réponses seront fournies au fil de l'eau sur le site de la Direction Générale des Entreprises, à l'adresse <https://www.entreprises.gouv.fr/numerique/mutualisation-de-donnees-pour-intelligence-artificielle> ;
- le 16 novembre 2018 à 18h00, clôture de la réception des réponses à l'AMI, les répondants sont invités à déposer leur réponse⁵ exclusivement sous forme électronique à l'adresse ami-mutualisation.dge@finances.gouv.fr ;
- de mi-novembre à fin décembre, des auditions ponctuelles de porteurs de réponses pourront être organisées en tant que de besoin afin d'approfondir certains points d'intérêt.

Précisions :

Un modèle de format de réponse est proposé en annexe de ce document. Il précise les informations requises et facultatives. L'utilisation de ce modèle est conseillé mais non obligatoire. Les répondants utilisant leur propre format de réponse devront veiller à y intégrer toutes les informations recensées comme requises dans le modèle proposé en annexe.

Des réponses communes à des groupements d'acteurs, préfigurateurs de consortia qui pourraient soumettre un dossier de candidature au futur appel à projets, sont souhaitées. Toutefois, des réponses faites par des contributeurs uniques seront également considérées s'ils sont en mesure de susciter des initiatives à fort impact sur l'économie nationale et économiquement viables et pérennes.

Pour chaque réponse à l'AMI, un référent doit être identifié. Il sera l'interlocuteur de l'administration pour l'organisation d'éventuelles auditions après la clôture de l'AMI.

⁴ <http://www.bpifrance.fr/A-la-une/Appels-a-projets-concours/Appel-a-projets-Structuration-de-Filiere-38393>

⁵ La taille des courriers électroniques contenant les réponses ne pourra pas excéder 10 Mo. Si nécessaire, pour les réponses dépassant les 10Mo, un mode alternatif de réception pourra être proposé sur demande à l'adresse ami-mutualisation.dge@finances.gouv.fr

ANNEXE

MODELE DE FORMAT DE REPONSE

Les informations relatives aux sections A, C.1., D et F sont requises a minima.
Les autres informations sont facultatives et renseignées autant que possible.

A. Identification du référent

Nom :	
Prénom	
Fonction :	
Organisation :	
Secteur d'activité :	
Mail :	
Téléphone :	
Adresse postale :	

B. Identification du groupement d'acteurs

(Il peut s'agir d'un groupement d'acteurs constitué pour répondre à l'AMI et éventuellement pressenti pour répondre au futur appel à projets.)

Nom	Statut	Informations clés
<nom de l'acteur>	<PME, grande entreprise (GE), association (AS), établissement de recherche (ER)>	<Nature de l'activité, marchés adressés, chiffres d'affaires, effectifs>
Groupement pressenti pour répondre au futur AAP : Oui/Non – Commentaires.		

C. Description de l'initiative envisagée**C.1. Principales finalités de l'initiative et justification de son intérêt**

<Description des besoins du secteur, du marché, d'usages qui justifient l'initiative>

<Description de la réponse apportée par l'initiative>

C.2. Description détaillée de l'initiative

<Services proposés - Fonctionnement de ceux-ci – Publics visés>

<Acteurs nécessaires - Rôle et valeur ajoutée de chacun>

<Gouvernance de l'initiative>

<Structure juridique>

C.3. Mise en œuvre du RGPD et gestion de la gouvernance des données

<Explications concernant la prise en compte de la réglementation en vigueur concernant la protection des données personnelles pour les initiatives concernées>

<Description des principes de gouvernance des données>

D. Description du modèle économique

<Présentation du modèle économique et de la stratégie de valorisation économique ; arguments appuyant la capacité à bénéficier de retombées économiques>

<Présentation d'éventuelles initiatives concurrentes (au niveau national, communautaire voire international) – Critères de différenciation>

E. Description des moyens et investissements nécessaires

E.1. Moyens à mettre en œuvre

<Description des moyens nécessaires et de leur montée en charge jusqu'à l'atteinte d'un « régime de croisière »: bâtiments éventuels, équipements, équipes, etc.>

<Précisions sur les moyens apportés par les partenaires et les moyens supplémentaires requis>

E.2. Investissements

<Quantification des besoins d'investissement>

<Précisions sur les montants des investissements réalisés par les partenaires et le montant de l'aide financière publique éventuelle>

F. Description des modalités de soutien public qui semblent les plus adaptées

<Description des formes de soutien public souhaitées et justifications>

<Description des modalités de mise en œuvre du soutien public qui semblent les plus adaptées>

G. Financement / Plan d'affaires prévisionnel

<Présentation d'un tableau de financement sommaire distinguant les financements publics et les financements apportés par les partenaires>

<Présentation d'un plan d'affaires prévisionnel synthétique jusqu'à l'atteinte de l'autofinancement précisant les investissements, leur nature, les frais de personnel et de fonctionnement, les recettes attendues – ce plan d'affaires pourra inclure le tableau de financement>