



GOUVERNEMENT

*Liberté
Égalité
Fraternité*

bpifrance



Appel à projets

« Communs numériques pour l'intelligence artificielle générative »

L'appel à projets est ouvert¹ jusqu'au 24 octobre 2023 à 12h00 (midi, heure de Paris).
Les projets ne seront étudiés qu'à partir de cette date.

En cas d'épuisement des moyens financiers affectés à cet appel à projets, il peut être arrêté de manière anticipée par arrêté de la Première ministre pris sur avis du Secrétariat général pour l'investissement (SGPI).

Les porteurs de projets sont invités à déposer leur dossier de candidature en ligne sur la plateforme de Bpifrance :
<https://www.picxel.bpifrance.fr/projets>

APPEL À PROJETS
Juin 2023



¹ sous réserve de la publication au Journal Officiel de l'arrêté relatif à l'approbation du cahier des charges

Sommaire

2_ Sommaire

3_ Contexte et objectifs de l'AAP

- _ Le plan d'investissement France 2030
- _ La stratégie d'accélération en intelligence artificielle
- _ Objectif de l'appel à projets

6_ Projets attendus

- _ Nature des projets
- _ Porteurs de projets
- _ Travaux et dépenses éligibles
- _ Modalités de financement

10_ Processus de sélection

- _ Critères d'éligibilité
- _ Critères de sélection
- _ Processus de sélection

12_ Mise en œuvre, suivi des projets et allocation des fonds

- _ Conventionnement
- _ Suivi des projets et étapes d'allocation des fonds
- _ Communication
- _ Conditions de reporting

Contexte et objectifs de l'appel à projets

● Le plan d'investissement France 2030

- **Traduit une double ambition** : transformer durablement des secteurs clefs de notre économie (énergie, automobile, aéronautique, espace, formation tout au long de la vie, etc.) par l'innovation technologique et positionner la France non pas seulement en acteur mais en leader du monde de demain. De la recherche fondamentale à la production d'un produit ou service nouveau, France 2030 soutient tout le cycle de vie de l'innovation jusqu'à son industrialisation.
- **Est inédit par son ampleur** : 54 Md€ seront investis pour que nos entreprises, nos universités, nos organismes de recherche, réussissent pleinement leurs transitions dans ces filières stratégiques. L'enjeu est de leur permettre de répondre de manière compétitive aux défis écologiques et d'attractivité du monde qui vient, et de faire émerger les futurs champions de nos filières d'excellence. France 2030 est défini par deux objectifs transversaux consistant à consacrer 50 % de ses dépenses à la décarbonation de l'économie et 50% à des acteurs émergents porteurs d'innovation, et à intervenir sans engager de dépenses défavorables à l'environnement (au sens du principe Do No Significant Harm).
- **Sera mis en œuvre collectivement** : pensé et déployé en concertation avec les acteurs économiques, académiques, locaux et européens pour en déterminer les orientations stratégiques et les actions phares. Les porteurs de projets sont invités à déposer leur dossier via des procédures ouvertes, exigeantes et sélectives pour bénéficier de l'accompagnement de l'Etat.
- **Est piloté par le Secrétariat général pour l'investissement** pour le compte de la Première ministre et mis en œuvre par l'Agence de la transition écologique (ADEME), l'Agence nationale de la recherche (ANR), la Banque publique d'investissement (Bpifrance) et la Caisse des Dépôts et Consignations (CDC).

● Un nouvel axe pour la stratégie d'accélération en intelligence artificielle

La Stratégie nationale pour l'intelligence artificielle² (SNIA) a été lancée en 2018 et renouvelée fin 2021³. La priorité de la phase 2 de la SNIA, qui s'inscrit dans le cadre du plan France 2030, est de favoriser la diffusion de l'IA dans notre économie. Trois leviers sont activés pour atteindre cet objectif : la formation, le soutien à l'offre « deep tech » et la mobilisation des intermédiaires contribuant au rapprochement entre l'offre et la demande en IA.

Les développements d'IA génératives s'accroissent ces derniers mois à la suite de premières expérimentations grand public réussies, offrant de nouvelles perspectives de progrès économique et social. Il a donc été décidé d'associer un nouvel axe dédié à l'IA générative aux trois axes identifiés comme prioritaires pour le soutien à l'offre « deep tech » (l'IA embarquée, l'IA de confiance, l'IA frugale) dans le cadre de la phase 2 de la SNIA.

Une IA est dite générative lorsqu'elle est capable de générer automatiquement du contenu de différents types (texte, image, son, etc.) à partir de requêtes qui peuvent être exprimées en langage naturel ou de façon programmatique. Ces IA font généralement l'objet d'entraînement sur de très larges bases de données, mobilisant d'importantes capacités de calcul⁴. Depuis 2017, elles reposent principalement sur l'architecture *transformer* et le concept d'*attention*⁵, qui ont constitué une rupture dans le domaine de l'apprentissage automatique. D'abord appliquées à plusieurs tâches du traitement automatique des langues (TAL) pour la génération de contenus textuels (question-réponse, synthèse de document, traduction, etc.), ces IA génératives ont trouvé des applications pour d'autres modalités (images, vidéos, etc.) et même pour des tâches de reconnaissance (reconnaissance d'entités nommées, modération de contenu, etc.) pour lesquelles elles n'étaient pas initialement destinées.

Au vu des performances actuelles des systèmes d'IA générative, des perspectives de **gains importants sont entrevus pour de nombreuses tâches** comme :

- la recherche et la synthèse documentaire pour la réalisation d'expertises et l'automatisation de tâches administratives (instruction et décision judiciaires, traitement automatisé de contributions aux débats démocratiques, etc.) ;
- les assistants conversationnels pour l'entreprise (éventuellement personnalisables) ;
- le soutien à la conception et à l'innovation dans l'industrie (nouveaux médicaments, matériaux, circuits électroniques, etc.) ;
- la génération automatique de code informatique ;
- l'assistance aux artistes dans les industries culturelles et créatives (environnements 3D et scripts dans les jeux vidéo,

² L'intelligence artificielle est entendue ici comme l'ensemble des techniques conférant à une machine des capacités d'analyse et de décision lui permettant de s'adapter aux situations rencontrées en faisant des prédictions à partir de données et de connaissances acquises. Les définitions proposées dans le projet de règlement européen sur l'IA par la Commission européenne et dans « Principes sur l'Intelligence Artificielle de l'OCDE » sont retenues dans le présent contexte.

³ https://ec.europa.eu/commission/presscorner/detail/fr/IP_21_1682

⁴ Le giga-modèle de langage BLOOM développé dans le cadre du projet BigScience comporte par exemple 176 milliards de paramètres et a été entraîné sur le supercalculateur Jean Zay financé dans le cadre de la SNIA

⁵ <https://arxiv.org/abs/1706.03762>. Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).

design de meubles et de vêtements, etc.) ;

- le soutien à la formation (conception de ressources et parcours pédagogiques, accompagnement personnalisé, etc.).

Si ces approches algorithmiques sont bien connues de la communauté scientifique depuis plusieurs années, la démocratisation de l'accès à ces outils est récente. De nombreux **autres cas d'usage devraient émerger**. En particulier, **l'adoption par les entreprises de ces outils** représente une importante opportunité de progrès économique.

Les modèles de « fondation » (*foundation models*)⁶ correspondent à des modèles pré-entraînés pouvant faire l'objet d'une adaptation (*fine-tuning*) à une tâche spécifique ou pour un domaine spécifique, c'est-à-dire d'une poursuite de leur entraînement sur un ensemble de données de taille réduite mais ciblé sur l'application finale visée. Les IA génératives reposant aujourd'hui sur de très larges modèles longs et coûteux à entraîner, elles reposent principalement sur des modèles de fondation ensuite déclinés pour automatiser différentes tâches métier. De la même manière que des progrès sur un modèle de fondation peuvent se traduire en progrès sur l'ensemble des tâches dérivées sur lesquelles le modèle est adapté par *fine-tuning*, toute fragilité ou biais négatif⁷ de ce modèle de fondation peut se répercuter sur les applications sous-jacentes. Dans ce contexte, il y a un risque de dépendance vis-à-vis des développeurs de ces modèles et d'enfermement propriétaire dans le cadre d'une utilisation commerciale de ces outils. Cela pourra pénaliser la diffusion de ces outils au service de notre économie. Pour remédier à cette situation, garantir notre autonomie stratégique et s'assurer d'une maîtrise complète des outils, il est essentiel qu'une diversité d'acteurs se positionne sur les différentes briques de la chaîne de valeur de l'IA générative :

- collecte, nettoyage, qualification et annotation des données d'apprentissage et de test,
- développement de modèles pré-entraînés « de fondation »,
- production de modèles « adaptés » pour des applications spécifiques,
- intégration des IA génératives au sein de systèmes plus larges (processus industriels ou administratifs, robots, etc.),
- mise à disposition de plateformes d'accès et de services d'IA générative,
- codéveloppement des outils génératifs par les utilisateurs,
- location et fourniture d'infrastructures de calcul et de stockage des données, etc.

L'objet de cet appel à projets (AAP) est de participer à la constitution d'un tel écosystème en favorisant la mise en place de « communs numériques pour l'IA générative » et l'émergence des meilleurs cas d'usage, en lien avec les objectifs de France 2030.

Le présent cahier des charges décrit les modalités de l'AAP « Communs numériques pour l'intelligence artificielle générative » pour les interventions en aides d'Etat. Il est opéré pour le compte de l'Etat par Bpifrance.

Un projet ayant été déposé à cet AAP pourra être soumis au processus d'instruction d'un autre AAP ou AMI de France 2030 jugé plus adapté, sans besoin de re-dépôt par le(s) porteur(s) de projet.

● Objectif de l'appel à projets

Cet AAP a vocation à soutenir des projets collaboratifs avec une double ambition :

- a) accélérer la **création et la mise en accessibilité de communs numériques sur l'ensemble de la chaîne de valeur de l'IA générative** à condition que le caractère incitatif de l'aide publique soit justifié et avéré ;
- b) assurer une utilisation par le plus grand nombre du ou des communs numériques développés pour **développer des produits ou services innovants**.

Il est entendu ici par commun numérique une ressource produite et/ou entretenue collectivement par une communauté d'acteurs hétérogènes, et gouvernée par des règles qui lui assurent son caractère collectif et partagé. Il est dit numérique lorsque la ressource est dématérialisée : logiciel, base de données, contenu numérique (texte, image, vidéo et/ou son), etc. Les caractéristiques des communs numériques sont les suivantes :

- l'usage de la ressource par les uns ne limite pas les possibilités d'usage par les autres (la ressource est non rivale) ;
- sa préservation ne passe pas par la réservation du droit d'usage à une communauté restreinte (la ressource est non-exclusive).

Les communs numériques pourront notamment concerner :

- les bases de données d'apprentissage et de test valorisant le patrimoine national de données (textes, sons, images vidéos, etc.) ;
- les giga-modèles génératifs pré-entraînés ;
- les adaptations de modèles pré-entraînés génératifs à des cas d'usages spécifiques (*fine-tuning*, distillation, etc.) ;
- les interfaces de programmation applicatives pour utiliser ces modèles (*prompting*, etc.) ;
- les outils d'évaluation des IA génératives.

⁶<https://arxiv.org/abs/2108.07258>. Bommasani, Rishi, et al. "On the opportunities and risks of foundation models." arXiv preprint arXiv:2108.07258 (2021).

⁷ <https://dl.acm.org/doi/pdf/10.1145/3442188.3445922>. Bender, Emily M., et al. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?." *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 2021.

Les projets soumis à ce dispositif pourront couvrir tout ou partie de cette chaîne de valeur de l'IA générative. Une attention particulière sera accordée à la capacité du projet à mettre à disposition des briques (corpus, modèles, etc.) en accès libre afin de **conférer à des développeurs de solutions un avantage concurrentiel dans la mise sur le marché d'un produit ou service**. Concernant les bases de données, l'un des principaux objectifs du présent AAP est de s'assurer de la **qualité et de la légalité des bases de données** constituées pour valoriser le patrimoine national français, francophone et européen dans l'entraînement de modèles d'intelligence artificielle.

Il est complémentaire avec :

- l' AAP « Maturation technologique et démonstration de solutions d'intelligence artificielle embarquée »⁸ dont l'objectif est de favoriser la maturation d'architectures et plateformes matérielles dédiées à l'IA embarquée et la démonstration de leur efficacité en situation réelle ;
- l' AAP « Maturation technologique et démonstration de systèmes de confiance intégrant de l'intelligence artificielle »⁹ dont l'objectif est de valider à l'échelle d'un système industriel complet le fonctionnement de plusieurs technologies dites « de confiance », notamment via une évaluation du système au regard de critères techniques et réglementaires ;
- l' AAP « Démonstrateurs d'IA frugale au service de la transition écologique dans les territoires »¹⁰ dont l'objectif est de développer et mettre au service de la transition écologique dans les territoires des systèmes d'intelligence artificielle frugaux ;
- l' AAP « Thématiques Spécifiques en Intelligence Artificielle » (TSIA) 2023, qui comprend un volet d'encouragement des collaborations de recherche sur les giga-modèles pour le traitement automatique du langage naturel et des données multimodales¹¹.

⁸<https://www.bpifrance.fr/nos-appels-a-projets-concours/appele-a-projets-maturation-technologique-et-demonstration-de-solutions-dintelligence-artificielle-embarquee>

⁹ Insérer ici le lien

¹⁰ Insérer ici le lien

¹¹ <https://anr.fr/fr/detail/call/appele-a-projets-thematiques-specifiques-en-intelligence-artificielle-tsia-2023/>

Projets attendus

● Nature des projets

Les projets devront être orientés vers la démonstration d'un besoin fonctionnel clairement identifié et permettant à terme de renforcer la position des acteurs sur un marché existant ou d'ouvrir sur un nouveau marché. Le développement de solutions sur des marchés innovants émergents en lien avec les objectifs de France 2030 sera particulièrement encouragé.

Le présent appel à projets vise à soutenir les types de projets suivants :

1. la mise en accessibilité ouverte ou partagée d'une ou plusieurs **base(s) de données valorisant de patrimoine français, francophone et européen** dans l'entraînement d'IA génératives, qu'il s'agisse de texte, de fréquences, d'images, de signaux ou d'image 3D,
2. la mise en accessibilité ouverte ou partagée de **giga-modèles génératifs pré-entraînés offrant une plus-value significative** par rapport aux modèles pré-entraînés déjà disponibles en accès ouvert et/ou libre de droits,
3. la mise en accessibilité ouverte ou partagée de **giga-modèles génératifs « adaptés » à des cas d'usage métier spécifiques**, qui permettront à certains partenaires du projet de commercialiser un ou plusieurs produits ou services innovants,
4. la mise en accessibilité ouverte ou partagée **d'API permettant des utilisations ciblées** de modèles d'intelligence artificielle générative,
5. la mise en accessibilité ouverte ou partagée **d'outils d'évaluation des IA génératives** pour différentes tâches,
6. d'autres communs de la chaîne de valeur de l'IA générative, à condition que le caractère incitatif de l'aide publique soit justifié et avéré.

1) Concernant les **bases de données**, des actions de collecte, regroupement, nettoyage (y compris mise en conformité au RGPD et à la réglementation sur les droits d'auteur), annotation, qualification (distribution des facteurs d'influence, réalisme, etc.) et mise en accessibilité de tous types de données pertinents pour l'entraînement d'IA génératives sont notamment attendues. Les formats de données utilisés doivent permettre une exploitation ultérieure par le plus grand nombre. Le caractère inédit et le grand volume des bases de données mises à disposition seront particulièrement appréciés pour la sélection du projet. En particulier, il est encouragé la mise à disposition de données constitutives du patrimoine national français (données géographiques, archives, données relatives aux œuvres d'art, etc.), en associant le détenteur de ces données au projet. Les bases de données vectorielles (où les mots sont représentés par des vecteurs), particulièrement utiles pour l'IA générative (pour mettre en place des approches de type vector search), sont également concernées.

Des exemples d'initiatives encouragées dans le cadre du présent AAP sont présentés ci-dessous :

- des bases de données valorisant le patrimoine français, francophone et européen et dont la qualité est démontrée supérieure à celles issues du *crawling* et *scraping* du web (Common Crawl, etc.),
- des bases de données respectant le droit d'auteur (attribution systématique d'une œuvre incluse dans la base à son auteur, possibilité pour l'auteur de ne pas autoriser l'utilisation de ses œuvres, etc.),
- des bases de données dont les contenus toxiques sont filtrés en conformité avec les normes de l'organisation internationale du travail,
- les bases de données multimodales avec des associations de modalités originales,
- des bases de données intégrant des paires image-texte à forte corrélation (par exemple une image et sa légende),
- des bases de données images/vidéos de haute résolution.

A l'inverse, il faudra éviter toute intégration de données générées par une IA générative dans ces bases de données (pour limiter le risque d'hallucination qui résulte de ce genre de pratique).

Ces bases de données seront constituées dans l'objectif d'une **utilisation ultérieure** par les membres du consortium

Base(s) de données mutualisées :

Une deuxième typologie de projet se concentrera exclusivement sur la **mutualisation des données intra-filière ou inter-filières**, dans l'objectif de les mettre à **disposition** de l'ensemble d'une communauté d'acteurs ou d'une filière ; ces projets devront présenter un certain **nombre de caractéristiques** décrites en **annexe 2**.

2) Concernant les **giga-modèles génératifs pré-entraînés** (de fondation), il s'agira de justifier l'intérêt d'entraîner un ou plusieurs nouveaux modèles au regard de ceux déjà existants et accessibles. Une justification des choix de conception envisagés (taille du modèle, taille et composition de la base de données d'entraînement, etc.) sera également attendue. Des exemples d'axes d'amélioration par rapport à l'état de l'art sont fournis ci-dessous :

- Performance sur des langues spécifiques ou pour des données spécifiques,
- Robustesse et généralisabilité (maîtrise des biais, taille de l'environnement de fonctionnement, nombre de tâches dérivées couvertes par le modèle de fondation, etc.),
- Résilience (capacité à fonctionner en dehors de l'environnement de fonctionnement prévu),

- Frugalité (en données, en consommation électrique pour l'entraînement et l'inférence, etc.) et embarquabilité, etc.

Sans que cela constitue une obligation, les projets visant à réconcilier l'IA générative avec les autres axes du levier « soutien à l'offre deep tech » de la SNIA (confiance, frugalité, embarquabilité) seront appréciés.

3) Concernant les **giga-modèles adaptés**, il s'agira de se positionner sur des tâches métier à forte valeur ajoutée (gain de compétitivité, innovation, etc.) ou pour lesquelles les modèles génératifs actuellement disponibles sont moins performants dans un contexte national que pour d'autres pays (couverture des spécificités du droit français, des références culturelles et sociales françaises, etc.). Il est notamment encouragé le réentraînement et la mise en accessibilité de modèles :

- linguistiques pour améliorer substantiellement les performances en langue française ou dans les langues nationales ou régionales de l'Union européenne (vocabulaire métier ou représentatif d'une variété d'usages, d'une variété de locuteurs de la Francophonie, etc.) lorsqu'ils sont associés à une opportunité commerciale ou sociétale justifiée,
- linguistiques sur des bases de données de langage informatique conférant un avantage concurrentiel pour un ensemble d'entreprises d'un secteur économique ou industriel, ou ayant un intérêt sociétal avéré
- génératifs non linguistiques sur des bases de données de tous types (images, vidéos, fréquences, etc.),
- génératifs dédiés à la constitution ou amélioration de bases de données (transcription, légende d'images, etc.).

Les actions d'optimisation (distillation, couplage à une base pour faire de la recherche documentaire, etc.) et d'alignement de ces modèles (*Reinforcement learning from human feedback*, etc.) sont également concernées.

4) Les API (ou interfaces de programmation applicative) désignent ici tout outillage informatique permettant d'accéder de manière facilitée et logicielle à un ensemble de méthodes, fonctions ou données. En particulier, les projets d'API seront appréciés au regard de l'aspect innovant en termes de nouveaux services mis à disposition.

5) Les outils d'évaluation désignent ici tout outillage informatique permettant de rendre compte de la performance et autres caractéristiques (robustesse, résilience, frugalité, etc.) des IA génératives, notamment des outils d'application automatique de métriques d'évaluation, d'analyse statistique des résultats, etc.

Seront sélectionnés les projets qui démontreront leur capacité :

- a) à rapidement **positionner les acteurs nationaux** sur les applications de l'IA générative présentant les meilleures opportunités de marché,
- b) à garantir dès la soumission du projet la **réutilisation par un large ensemble d'acteurs** (entreprises ou administration publiques) les communs numériques mis à disposition, par exemple car ils sont utilisables pour différents cas d'usages (les communs destinés à l'ensemble d'une filière industrielle ou d'un secteur économique seront encouragés et les lettres de soutien seront alors appréciées),
- c) à garantir que la stratégie de mise en accessibilité adoptée (choix des licences des bases de données et des modèles, choix des modèles pré-entraînés pour le *fine-tuning*, plateformes de partage, etc.) répond aux **enjeux de souveraineté** (limitation des dépendances, sécurisation et confidentialité des données), **de confiance et de conformité réglementaire** (choix des stratégies et méthodologies d'annotation pour assurer la traçabilité de la propriété intellectuelle et de la responsabilité, etc.) associés à l'IA générative.

Le projet devra présenter une assiette de travaux comprise entre 1 M€ et 20 M€ et viser un niveau de TRL indicatif entre 6 et 8 en fin de projet. La durée des projets devra être comprise entre 12 et 24 mois.

Les projets concernant les **Bases de données mutualisées** se référeront à l'annexe 2.

Le dossier précisera la méthodologie utilisée pour parvenir aux résultats présentés dans le cadre de l'auto-évaluation d'impact environnemental. La méthodologie de calcul de l'impact de l'entraînement du système d'IA s'appuiera par exemple sur l'outil Green Algorithms (GT), Lannelongue et al, (<https://www.green-algorithms.org/>) et précisera le pays de localisation des serveurs.

Les efforts des porteurs de projets en matière d'écoconception, de maîtrise des consommations énergétiques et de ressources ainsi que de recyclabilité seront valorisés.

● **Porteurs de projets**

Le chef de file des projets peut être assuré par tout type de porteur (entreprise, organisme de recherche, administration publique, etc.). Il est encouragé que l'initiative implique (en tant que partenaires ou via des lettres d'engagement) le plus grand nombre d'acteurs ayant un intérêt propre dans l'utilisation des communs numériques de l'IA générative mis en place (pour la mise sur le marché d'un produit ou d'un service, pour une utilisation interne, etc.).

Une collaboration effective existe :

- entre des entreprises parmi lesquelles figure au moins une PME ou une start-up et où aucune entreprise unique ne

- supporte seule plus de 70% des dépenses éligibles ou,
 - entre une entreprise et un ou plusieurs organismes de recherche et de diffusion de connaissances et où ces derniers supportent au moins 10% des dépenses éligibles et ont le droit de publier les résultats de leurs propres recherches.

Les projets concernant les **Bases de données mutualisées** se référeront à l'annexe 2.

● Travaux et dépenses éligibles

Les dépenses éligibles sont directement affectées au projet (hormis les frais connexes qui sont calculés par un forfait). Dans le cas général (régime RDI), la nature des dépenses éligibles est précisée ci-dessous :

| Type de dépenses | Principes |
|---------------------------------|--|
| Salaires et charges | Salaires chargés du personnel du projet (non environnés) appartenant aux catégories suivantes : chercheurs (post-doc inclus), ingénieurs, techniciens. |
| Frais connexes | Montant forfaitaire des dépenses de personnel (20% des salaires chargés non environnés). |
| Coûts de sous-traitance | Coûts de prestations liés aux activités de R&D (cible : 30% max des coûts projet). |
| Contribution aux amortissements | Coûts d'amortissements comptables des instruments et du matériel de R&D au prorata de leur utilisation dans le projet. Exemple : pour un équipement amorti de façon linéaire sur une durée de 10 ans, et utilisé à 100% durant 2 ans pour le projet, le montant éligible à une aide sera égal à 2/10 ^e du montant total de l'investissement dans cet équipement. |
| Frais de mission | Frais réels des déplacements liés à la réalisation du projet. |
| Autres coûts | Autres frais d'exploitation directement liés à l'activité du projet (consommables non amortis dans les comptes). |

La contribution d'un partenaire représentant moins de 5% ou moins de 200 k€ de l'assiette de dépenses totales du projet a vocation à être prise en charge en sous-traitance.

Les dépenses sont éligibles à compter du lendemain de la constatation par Bpifrance du caractère complet du dossier, à l'issue du délai de soumission du dossier (se référer à la date de clôture de l'AAP en page de couverture).

Les projets concernant les **Bases de données mutualisées** se référeront à l'annexe 2.

● Modalités de financement

L'intervention publique s'effectue dans le respect de la réglementation de l'Union européenne applicable en matière d'aides d'État ([articles 107 à 109 du TFUE](#)).

Il peut être fait application de l'un des régimes d'aides d'Etat visés ci-dessous :

- régime cadre exempté de notification n° SA.58995 relatif aux aides à la recherche, au développement et à l'innovation (RDI) pour la période 2014-2023 ;
- régime cadre exempté de notification n° SA.100189 relatif aux aides en faveur des PME pour la période 2014-2023 ;
- régime cadre exempté de notification n° SA.59107 relatif aux aides en faveur de l'accès des PME au financement pour la période 2014-2023 ;
- régime cadre exempté de notification n° SA.59108 relatif aux aides à la protection de l'environnement pour la période 2014-2023.

Aides proposées pour les activités économiques

Sont considérées comme « économiques » les activités des entités, généralement des entreprises, consistant à offrir des biens ou services sur un marché potentiel, avec l'espérance de retours financiers basés sur les résultats du projet.

Dans le cas général, le taux de l'aide s'applique sur les dépenses éligibles et dans la limite des intensités maximales permises par les régimes d'aides évoqués ci-dessus, identifiées lors de l'instruction approfondie des projets.

L'aide apportée aux activités économiques sera constituée de subvention selon les modalités suivantes :

Pour les projets pour lesquels les dépenses relèvent de recherche et de développement, le régime d'aides applicable est le régime cadre exempté de notification n° SA.58995 relatif aux aides à la recherche, au développement et à l'innovation (RDI) pour la période 2014-2023.

| Type de recherche | Type d'entreprise | | |
|----------------------------|------------------------|-------------------------|-------------------------------|
| | Petite Entreprise (PE) | Entreprise Moyenne (ME) | Grande Entreprise (GE ou ETI) |
| Recherche industrielle | 70 % | 60 % | 50 % |
| Développement expérimental | 45 % | 35 % | 25 % |

Des taux différents s'appliquent pour les projets concernant les **Bases de données mutualisées** ; se référer à l'annexe 2.

L'aide apportée aux activités économiques sera constituée **de subventions**.

Les projets mobilisant une offre Cloud française ou européenne pour les entrainements de modèles bénéficieront d'une attention particulière.

Aides proposées pour les activités non économiques

Sont considérées comme « non économiques », les activités des entités, généralement des organismes de recherche, quel que soit leur statut, remplissant une mission d'intérêt général en consacrant une part prépondérante de leur activité à la R&D. Les activités relevant de prérogatives de puissance publique lorsque les entités publiques agissent en leur qualité d'autorité publique sont également considérées comme « non économiques ».

Pour les activités non économiques, l'aide sera apportée sous forme de subventions selon les modalités suivantes :

| Type d'acteur | Intensité de l'aide |
|---|--------------------------|
| Organismes de recherche et assimilés (type de coûts au choix de l'entité) | 100% des coûts marginaux |
| | 50% des coûts complets |

Toute dépense d'un organisme de recherche et assimilé liée à des travaux applicatifs pour le développement d'une solution portée par un des membres du consortium est à considérer en sous-traitance de ce dernier.

Les organismes de recherche et assimilés ne peuvent porter plus de 20% des dépenses totales du projet.

Processus de sélection

● Critères d'éligibilité

Pour être éligible, un projet doit :

- être complet au sens administratif et être soumis dans les délais, au format imposé, sous forme électronique via la plateforme de Bpifrance ;
- répondre aux objectifs et attendus indiqués au paragraphe 2 et satisfaire les contraintes indiquées, notamment relatives au montant d'assiette de dépenses ;
- porter sur des travaux réalisés en France et non-engagés avant le dépôt de la demande d'aide (la date d'éligibilité des dépenses correspond au lendemain de la date de réception du dossier complet par Bpifrance, conformément au calendrier des relèves indiqué en page de couverture) ;
- être porté par un consortium composé uniquement de partenaires éligibles à recevoir des aides publiques (en particulier, les partenaires doivent être à jour de leurs obligations fiscales et sociales, ne pas être sous le coup de la récupération d'aides déclarées illégales ou incompatibles par la Commission européenne, et ne pas avoir le statut d'« entreprise en difficulté » au sens de la réglementation européenne des aides d'Etat). Les projets concernant les **Bases de données mutualisées** se référeront à l'annexe 2.

● Critères de sélection

Les dossiers seront notamment évalués selon les critères ci-dessous.

| Critères | Précisions |
|--|--|
| Pertinence | <p>Pertinence au regard du texte de l'appel à projets et aux exigences clés d'accessibilité, de souveraineté et de performance associés aux communs numériques pour l'IA générative.</p> <p>Justification des opportunités économiques générées par la mise en accessibilité du commun, notamment grâce à une analyse de marché.</p> <p>Contribution du projet à la valorisation du patrimoine national de données.</p> |
| Montage du projet | <p>Gouvernance, planning et jalons décisionnels, gestion des risques notamment en matière de délais et de surcoûts, description des coûts projet, clarté de la rédaction.</p> |
| Structure du projet | <p>Existence d'une collaboration structurée ou d'un effet diffusant au sein d'une filière ou d'un écosystème.</p> <p>Pertinence du consortium, implication d'au moins un acteur émergent de l'innovation selon la définition France 2030.</p> <p>Diversité des acteurs associés à l'initiative (partenaires et soutiens) pour l'utilisation du commun.</p> <p>Les projets concernant les Bases de données mutualisées se référeront à l'annexe 2.</p> |
| Indicateur clé de performances | <p>Proposition d'indicateurs clé de performance permettant de mesurer la mise en accessibilité ouverte et partagée du projet présenté.</p> |
| Plan de financement | <p>Description des modalités de financement du projet.</p> <p>Incitativité de l'aide.</p> <p>Capacité à mener à terme le projet.</p> |
| Innovation | <p>Capacité du projet à lever des verrous technologiques et d'innovation, à combler des failles de marché, afin de conférer des avantages concurrentiels à un ensemble d'utilisateurs ou à l'échelle d'une filière industrielle ou économique.</p> |
| Infrastructures | <p>Justification des choix quant aux infrastructures utilisées pour la constitution du commun (enjeux de souveraineté, coût, etc.), en particulier concernant les ressources de calcul utilisées (supercalculateur, cloud, etc.).</p> |
| Accessibilité | <p>Justification des choix effectués pour favoriser l'accès au commun, en termes de licence, de plateforme d'accès, de communication et diffusion au public et vers des filières ciblées, etc.</p> |
| Aspects de confiance/ données personnelles / robustesse / compliance | <p>Capacité à assurer la confiance et la légalité du commun mis en place (débiaisage des données, alignement des modèles, respect de la réglementation sur les données personnelles et sur les droits d'auteur, etc.)</p> |

| | |
|---------------------------------|--|
| AI Act etc. | |
| Impacts environnementaux | Démonstration qualitative et quantitative des éléments annoncés dans le dossier de candidature (éléments pertinents pour apprécier les impacts, positifs ou négatifs, sur les 6 axes de la taxonomie européenne). Les projets causant un préjudice important du point de vue de l'environnement seront exclus (application du principe DNSH – Do No Significant Harm ou « absence de préjudice important ») au sens de l'article 17 du règlement européen sur la taxonomie. ¹² |
| Répliquabilité de la solution | Caractère généralisable de la solution, implémentation des résultats dans les référentiels européens et internationaux. |
| Pertinence du modèle d'affaires | Justification des opportunités économiques engendrées par le commun : description du modèle d'affaires (produits et services envisagés / segments de marchés) et du plan d'affaires, en s'appuyant sur des études prévisionnelles des marchés correspondants, à l'échelle nationale, européenne, et, le cas échéant, internationale, tenant compte des évolutions technologiques et des développements d'autres solutions, et présentation des facteurs-clés de succès de la solution proposée. |
| Impacts socio-économiques | Démonstration du caractère structurant du projet pour la filière concernée (capacité d'entraînement des sous-traitants, perspectives d'investissement, etc.). Justification de la pertinence du projet par rapport aux enjeux socio-économiques et sociétaux, notamment concernant les opportunités d'automatisation des tâches industrielles et administratives, et la mise en conformité aux futures exigences de la réglementation européenne sur l'IA. Démonstration de la pérennité du commun après la fin du projet et du soutien public, anticipation du besoin de maintenabilité du commun sur le long terme, identification des évolutions possibles. |
| Cybersécurité | Prise en compte de la cybersécurité dans les travaux et les investissements réalisés, identification des actions réalisées et/ou planifiées pour assurer un bon niveau de cybersécurité au dispositif. |

La labellisation par un ou plusieurs pôles de compétitivité, bien que non obligatoire, sera prise en compte pour juger de la pertinence des projets, notamment quant à leur caractère innovant, leur solidité technique et leur caractère stratégique pour la performance de l'écosystème ou de la filière.

● Processus de sélection

La procédure de sélection relève de la gouvernance mise en œuvre dans le cadre du Plan d'investissement France 2030.

Présélection et sélection

A la suite de chaque relèvement de l'AAP, Bpifrance conduit une première analyse en termes d'éligibilité et présélectionne les meilleurs projets pour audition, sur la base des critères de sélection, en lien, en tant que de besoin avec les représentants des ministères sectoriels concernés.

Les porteurs des projets ainsi présélectionnés sont auditionnés par un jury composé de Bpifrance, d'experts externes à l'administration et, le cas échéant de représentants des ministères concernés.

Sur la base de l'avis du jury d'audition, le comité de présélection décide, en accord avec l'Etat, des projets qui entrent en phase d'instruction.

Instruction

Bpifrance envoie au porteur du projet ou au chef de file du consortium une notification de la décision d'entrée en instruction approfondie, accompagnée des compléments de dossier détaillés que le porteur devra déposer sur la plateforme de dépôt de Bpifrance dans un délai maximum de 21 jours.

L'instruction est conduite sous la responsabilité de Bpifrance, qui pourra s'appuyer sur l'expertise de deux experts externes à l'administration. Dans ce cadre, le porteur peut être invité à détailler de façon approfondie son projet lors d'une réunion d'expertise pouvant aller jusqu'à une journée.

A l'issue de cette phase d'instruction, Bpifrance présente au comité interministériel compétent les conclusions de l'instruction qui comprennent les recommandations et propositions d'un éventuel soutien.

A l'issue de cette dernière phase, la Première ministre prend les décisions finales d'octroi de l'aide après avis du SGPI.

¹² cf. annexe 1 « critères de performance environnementale » du présent cahier des charges

Mise en œuvre, suivi des projets et allocation des fonds

● Conventonnement

Chaque bénéficiaire signe une convention avec Bpifrance. Cette convention précise notamment l'utilisation des crédits, le contenu du projet, le calendrier de réalisation, les modalités de pilotage du projet, le montant des tranches et les critères de déclenchement des tranches successives, les prévisions de cofinancement des projets, les modalités de restitution des données nécessaires au suivi et à l'évaluation des investissements, et les modalités de communication.

La convention d'aide est signée dans le cas général dans un délai de 4 mois à compter de la décision de la Première ministre, sous peine de perte du bénéfice de la décision d'aide.

Ces conditions sont indépendantes des modalités de conventionnement définies par la Commission Européenne pour le financement du projet au niveau européen.

● Suivi des projets et étapes d'allocation des fonds

Le bénéficiaire met en place un tableau de bord comportant des indicateurs de suivi technique, industriel et financier de l'avancement des projets. Il le transmet régulièrement à Bpifrance ainsi que l'ensemble des documents demandés à chaque versement d'aide (rapport d'avancement, ERDA certifiés, ...) selon les modalités prévues par la convention.

La mise en accessibilité ouverte et partagée du projet financé sera contrôlée à chaque étape du projet et conditionnera le versement de chaque tranche d'aide.

Pour chaque projet soutenu, des réunions d'avancement peuvent être organisées en tant que de besoin. Demandée par Bpifrance et organisée par le chef de file ou le porteur de projet, elle associe les membres du comité de sélection ou leur représentant. Cette réunion a pour objet de suivre la mise en œuvre du projet et notamment le niveau d'exécution budgétaire, l'avancement des opérations financées et le respect du planning¹³.

● Communication

Bpifrance s'assure que les documents transmis sont soumis à la plus stricte confidentialité et ne sont communiqués que dans le cadre de l'expertise et de la gouvernance de France 2030. L'ensemble des personnes ayant accès aux dossiers de candidature est tenu à la plus stricte confidentialité.

Une fois le projet sélectionné, chaque bénéficiaire soutenu par France 2030 est tenu de mentionner ce soutien dans ses actions de communication, ou la publication des résultats du projet, avec la mention unique : « **Ce projet a été soutenu par le plan France 2030** », accompagnée du logo de France 2030. L'État se réserve le droit de communiquer sur les objectifs généraux de l'action, ses enjeux et ses résultats, le cas échéant à base d'exemples anonymisés et dans le respect du secret des affaires. Toute autre communication est soumise à l'accord préalable du bénéficiaire.

● Conditions de reporting

Le bénéficiaire est tenu de communiquer régulièrement à Bpifrance et à l'Etat les éléments d'informations nécessaires à l'évaluation de l'avancement du projet ainsi que de l'impact environnemental (performance commerciale, emplois créés, brevets déposés, effets environnementaux et énergétiques), Cette évaluation pourra se poursuivre après réalisation du projet. Ces éléments, et leurs évolutions, sont précisés dans conditions générales de la convention d'aide entre Bpifrance et le bénéficiaire.

¹³ Ces revues techniques font l'objet de rapport des experts de la Commission Européenne.



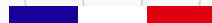
Contacts

Les renseignements concernant le processus administratif (constitution du dossier, démarches en ligne, taux d'aide) pourront être obtenus auprès de Bpifrance par courriel en indiquant dans l'objet du message « IA générative » pour un traitement plus rapide de la demande :

strategies-acceleration@bpifrance.fr

Les renseignements concernant les projets (cohérence avec les finalités de l'AAP) pourront être obtenus auprès de la Direction Générale des Entreprises (DGE) par courriel en indiquant dans l'objet du message le nom de l'AAP pour un traitement plus rapide de la demande :

la.dge@finances.gouv.fr



Annexe 1 : Critères de performance environnementale

Les projets causant un préjudice important du point de vue de l'environnement seront exclus (application du principe DNSH – Do No Significant Harm ou « absence de préjudice important ») au sens de l'article 17 du règlement européen sur la taxonomie. En créant un langage commun et une définition claire de ce qui est « durable », la taxonomie est destinée à limiter les risques d'écoblanchiment (ou "greenwashing") et de distorsion de concurrence, et à faciliter la transformation de l'économie vers une durabilité environnementale accrue. Ainsi, la taxonomie définit la durabilité au regard des six objectifs environnementaux suivants :

- l'atténuation du changement climatique ;
- l'adaptation au changement climatique ;
- l'utilisation durable et la protection des ressources aquatiques et marines ;
- la transition vers une économie circulaire ;
- la prévention et la réduction de la pollution ;
- la protection et la restauration de la biodiversité et des écosystèmes.

Pour l'évaluation technique de l'impact du projet vis-à-vis de chaque objectif environnemental, le déposant doit renseigner le document dédié disponible sur le site de l'appel à projet (dossier de candidature) et le joindre au dossier de candidature.

Il s'agira d'autoévaluer les impacts prévisibles de la solution proposée (faisant l'objet de l'aide) par rapport à une solution de référence pertinente, explicite et argumentée. Cette analyse tient compte du cycle de vie des process et du ou des produits ou livrables du projet, suivant les usages qui en sont faits. En tant que de besoin, ces estimations pourront être étayées par des analyses en cycle de vie plus complètes.

Annexe 2 : Caractéristiques spécifiques attendues sur les projets de type « Bases de données mutualisées »

Les porteurs de ces projets de bases de données mutualisées sont invités à prendre contact avec Bpifrance en amont du dépôt de leur dossier (contact : strategies-acceleration@bpifrance.fr)

● Porteurs de projets

Le projet est porté préférentiellement par une société de projet ou éventuellement par une structure fédérant plusieurs entreprises, voire une entité représentative des entreprises de la filière (telle une fédération professionnelle, un GIE, une association, un pôle de compétitivité...).

● Gouvernance

Les projets sont dotés d'une gouvernance décisionnaire sur la stratégie et les décisions d'investissements. La gouvernance intègre des représentants de la ou des filières concernées, qui détiennent au moins la moitié des droits de vote. Les représentants de la ou des filières peuvent être une fédération, une association professionnelle ou un ensemble d'entreprises.

● Dimension innovante

Les projets présentés doivent proposer des cas d'usage innovants autour de la donnée, et reposer sur des briques technologiques innovantes.

● Transparence et accessibilité

Les outils et plateformes de partage de données ainsi créés doivent être ouverts aux tiers, dans le cadre de conditions d'accès transparentes, non discriminatoires et clairement établies. Le projet doit se conformer aux principes, règles et labels établis par l'initiative européenne Gaia-X en la matière, notamment en termes de transparence vis-à-vis du recours à des services de traitement de données exposés à des législations extraterritoriales ainsi que leur conformité aux législations européennes (i.e. : RGPD, règlement sur la gouvernance européenne des données, etc.).

● Interopérabilité

Les projets attendus doivent aboutir à une interopérabilité des outils numériques employés par les parties prenantes, au moyen par exemple du développement d'API, et/ou à une interopérabilité sémantique. Notamment, le projet doit se conformer aux principes, règles et labels établis par l'initiative européenne Gaia-X en la matière.

● Travaux et dépenses éligibles

| Type de dépenses | Principes |
|------------------|---|
| Investissements | Investissements dans des actifs et incorporels actifs n'ayant aucune forme physique ni financière tels que les brevets, les licences, le savoir-faire ou d'autres types de propriété intellectuelle |
| Fonctionnement | Frais de personnel et frais administratifs (frais généraux compris) liés aux activités suivantes : <ul style="list-style-type: none">● l'animation du projet en vue de favoriser la collaboration, le partage d'informations et la prestation ou la mise à disposition de services de soutien aux entreprises |

| | |
|--|--|
| | <p>spécialisés et adaptés aux besoins de ces dernières ;</p> <ul style="list-style-type: none">• les opérations de marketing du projet visant à renforcer la participation de nouvelles entreprises ou organisations ainsi qu'à accroître la visibilité du projet ;• la gestion des installations du projet ; l'organisation de programmes de formation, d'ateliers et de conférences pour faciliter le transfert de connaissances et le travail en réseau entre les participants ainsi que la coopération transnationale |
|--|--|

● Taux d'aide

Pour les projets relevant des Bases de données mutualisées, pour lesquels les dépenses relèvent de structuration de la filière, la section « pôle d'innovation » du régime cadre exempté de notification n° SA.58995 est applicable. Pour ces projets, le taux d'aide peut s'élever jusqu'à 50 % maximum des dépenses éligibles dans la limite de la réglementation européenne sur les dépenses de fonctionnement et d'investissement.

Les typologies de projets suivants, quels que soient leurs mérites propres, ne sont pas éligibles au présent appel à projets :

- un projet de développement dont le bénéfice économique est limité à un nombre restreint d'acteurs au regard de la ou des filières concernées dans leur ensemble ;
- un projet de R&D mutualisé entre plusieurs entreprises ;
- un projet de communication ou de sensibilisation en faveur d'une filière ou d'une thématique portée par une filière.