

Appel à projets 2023-2024

REGLEMENT

Inauguré en décembre 2022, le lab (<https://inalelab.hypotheses.org/>) de l'Institut national de l'audiovisuel est une structure d'appui à la recherche orientée « données ». Il a pour mission d'encourager et faciliter l'exploitation scientifique des collections de l'INA à grande échelle, à l'aide d'outils d'intelligence artificielle et d'analyse automatisée.

Dans le cadre du présent appel à chercheurs 2023-2024 (ci-après « l'Appel »), le lab propose de soutenir en 2023 jusqu'à quatre projets de recherche, appuyés sur ces technologies et portant sur des jeux de données issus des collections de l'Institut (ci-après « le(s) Projet(s) de Recherche »).

Cet appel s'inscrit dans le prolongement des [dispositifs de soutien à la recherche de l'INA](#).

L'INA offrira aux porteurs de projets sélectionnés dans le cadre de cet l'Appel (ci-après « le(s) Lauréat(s) ») un accompagnement personnalisé de la part de professionnel(le)s et expert(e)s de l'INA, assorti d'un soutien à la fois technique et méthodologique.

La participation à l'Appel implique l'acceptation complète et sans réserve des termes du présent règlement (ci-après « le Règlement ») par les doctorant(e)s ou chercheurs qui y déposeront leur candidature (ci-après « le(s) Candidat(s) »).

1. Organisation et calendrier

1.1. Organisateur

Institut national de l'audiovisuel (INA) – 4 avenue de l'Europe 94366 Bry-sur-Marne, France.

Les locaux du lab sont situés à l'INA thèque : Bibliothèque nationale de France. Site François Mitterrand (rez-de-jardin). Quai François-Mauriac 75706 Paris cedex 13.

Contact : Arthur Lezer, lelab@ina.fr

1.2. Calendrier

Lancement de l'Appel :	13 avril 2023
Date limite de dépôt des dossiers de candidature ¹ :	10 juin 2023
Publication de la liste des Lauréat(e)s :	17 juillet 2023
Débuts des travaux des Lauréat(e)s :	1er octobre 2023

1.3. Modalités de dépôt des candidatures

Les Candidat(e)s devront adresser leur dossier de candidature au plus tard **le 10 juin 2023 à 23h59**, par e-mail, à lelab@ina.fr.

Ce dossier devra comprendre impérativement les pièces suivantes (en fichiers séparés) :

1° Le Bulletin de candidature, au format Word (à télécharger, remplir et signer) ;

2° Un CV, mentionnant la liste des publications du ou de la Candidat(e) ;

3° Le Projet de recherche décrivant en 15 000 signes maximum :

- Le ou les objectifs du Projet de recherche, faisant apparaître la bonne adéquation de ces travaux à la mission du lab, à savoir l'étude des médias par la donnée, le développement de synergies et du dialogue entre les sciences informatiques et les humanités.

¹ Comprenant notamment la fiche de candidature complétée et signée (voir annexe 1 au Règlement).

- Une ou plusieurs question(s) de recherche(s) et/ou hypothèses de recherche,
- Le(s) jeu(x) de données INA envisagé(s), et le cas échéant, les autres jeux de données mobilisés, en indiquant pour les données INA une évaluation la plus précise possible du volume de données concerné, ou du périmètre du corpus (en nombre de documents, ou en heures, etc.),
- Les verrous technologiques éventuels,
- La méthodologie envisagée, et le calendrier de travail associé,
- Les modalités envisagées pour la valorisation des résultats.

Un accusé de réception de son dossier de candidature sera adressé à chaque Candidat(e) par retour de mail.

Aucune remise en mains propres de dossier de candidature ne pourra être acceptée.

Les dossiers parvenus incomplets ou hors délai ne seront pas examinés.

Pour toute question relative à la soumission des candidatures, à l'élaboration du Projet de recherche et aux données disponibles, il est possible de contacter l'INA pendant la période d'ouverture des candidatures et de solliciter un rendez-vous en écrivant à lelab@ina.fr.

2. Candidat(e)s éligibles

L'Appel s'adresse aux chercheuses et chercheurs de toutes disciplines scientifiques, sans condition d'âge, de résidence ou de nationalité.

Les Candidat(e)s doivent être :

- inscrit(e)s dans un établissement d'enseignement supérieur, français ou non, pour un cursus de recherche niveau doctorat ;
- ou titulaires d'un doctorat, et être membre titulaire ou associé d'un laboratoire ou d'une unité de recherche.

Les travaux de recherche appliquée, les projets menés au sein de laboratoires ou d'entreprises privés, le développement de prototypes, pilotes ou preuves de concept à vocation commerciale ne sont pas éligibles.

Les candidat(e)s répondant à l'appel à projets du lab ne peuvent candidater en parallèle à l'appel à chercheurs de l'INA. Les deux dispositifs de soutien ne sont pas cumulables.

3. Projets de Recherche éligibles

Les dossiers de candidature devront être rédigés en langue française ou en langue anglaise.

L'Appel vise à soutenir les Projets de recherche :

- dont les questions et problématiques s'appuient principalement sur les collections et bases de données de l'INA. Ces collections représentent des archives uniques qui couvre l'ensemble des médias français, TV et Radio, depuis le début de leur existence ; 34 000 titres de cinéma ; une captation 24h/24 des flux Radio et TV depuis 2001, ainsi qu'un échantillon des sites Internet en lien avec l'écosystème des médias audiovisuels depuis 1996, de comptes Twitter, et des chaînes hébergées sur les plateformes vidéo représentatifs de l'actualité médiatique française depuis 2014.

Et

- comportant une dimension quantitative et/ou un défi technologique dans l'exploitation de gros jeux de données, en adéquation avec la mission du lab.

Le Projet de recherche pourra être en démarrage ou déjà en cours.

Il pourra s'agir d'un travail individuel ou collectif, impliquant plusieurs chercheurs et institutions. Dans tous les cas, les candidatures sont déposées par un(e) chercheur(se), considéré(e) comme porteur et responsable du projet.

Il pourra s'agir d'une recherche menée spécifiquement dans le cadre de cet Appel, ou bien inscrite dans un projet plus large (thèse, ouvrage...).

4. Modalités de sélection

Jusqu'à quatre Candidat(e)s pourront être sélectionnés et se voir octroyer le statut de porteurs de projet du lab.

Les Lauréats sont désignés par un comité de sélection, composé de l'équipe de pilotage du lab, ainsi que de membres de son Conseil Scientifique ([liste](#)), selon leurs expertises.

Chaque projet sera évalué sur la base des critères suivants :

- Clarté et pertinence des objectifs ;
- Clarté, originalité et pertinence de la ou des questions de recherche, des hypothèses ;
- Clarté, originalité et cohérence de la méthodologie et du protocole expérimental proposés pour y répondre ;
- Faisabilité technique et coût estimatif, évalués par l'INA des besoins en accompagnement humain nécessaire pour le Projet, des performances des outils, de la capacité de calcul et de la capacité de stockage proposés par l'INA ;
- Perspectives de valorisation des résultats.

Une attention particulière sera portée à la bonne adéquation de ces travaux à la mission du lab, à savoir l'étude des médias par la donnée, le développement de synergies et du dialogue entre les sciences informatiques et les humanités.

Une attention particulière sera portée aux Projets de recherche exploitant des jeux de données et/ou des technologies à fort enjeu pour le lab, et plus généralement pour l'INA, **dans sa mission d'éclairage du patrimoine et de l'actualité médiatique.**

Une liste en est proposée ci-dessous à titre indicatif ; les Candidat(e)s sont également libres de proposer leurs propres corpus constitués à partir des données de l'INA, méthodologies, et outils.

La liste des Lauréat(e)s sera publiée sur le site <https://inalelab.hypotheses.org/> à partir du 17 juillet 2023.

Les Candidat(e)s non retenu(e)s ne feront l'objet d'aucune notification individuelle.

5. Octroi du statut de porteur de projet du lab au(x) Lauréat(s)

Le Porteur de projet du lab bénéficiera d'un accompagnement personnalisé de la part de professionnel(le)s et expert(e)s de l'INA, qui apporteront leur soutien à la fois technique et méthodologique.

Les séances de travail en présentiel avec les Lauréat(e)s, si elles sont nécessaires, seront organisés à l'INA thèque, à Paris (adresse : Bibliothèque nationale de France. Site François Mitterrand (rez-de-jardin). Quai François-Mauriac 75706 Paris cedex 13). Les frais de déplacement et d'hébergement ne seront pas pris en charge.

Le Projet de recherche sera soutenu pour une durée d'1 an à partir de la date de début des travaux (1^{er} octobre 2023), reconductible une fois pour la même durée après demande expresse du ou de la Lauréat(e) et approbation du comité de sélection.

L'accompagnement pourra porter sur :

- l'aide à la recherche documentaire et à la constitution d'un corpus de médias et/ou de données, l'accès aux collections et le partage d'une expertise sur le patrimoine audiovisuel et l'actualité médiatique ;
- la mise en œuvre de technologies d'intelligence artificielle développées à l'INA pour l'analyse et la fouille de corpus audiovisuels : Vision par ordinateur, Extraction d'images, Traitement automatique du langage, transcription des dialogues, détection de copies, etc. (voir la dernière section pour plus de détails) ;
- l'accès à une infrastructure de calcul, avec le temps de serveur et la puissance adaptée à la mise en œuvre de ces traitements. Cette infrastructure pourrait également être mobilisée pour exécuter des algorithmes ou programmes autres que ceux développés à l'INA, après évaluation de ces besoins par l'équipe du lab et validation des critères nécessaires pour assurer la sécurité informatique des systèmes d'information de l'INA.

Le ou la Lauréat(e) et son équipe, disposeront d'un accès privilégié aux locaux du lab et à son équipement, pour l'organisation de séances de travail et la consultation des données et des collections de l'INA. Cela comprend la réservation d'une salle de groupe située en mezzanine de l'INAthèque qui peut accueillir jusqu'à 10 personnes, sous réserve qu'elles disposent d'un [titre d'accès au rez-de-jardin de la BnF](#). Le ou la Lauréat(e) et son équipe bénéficieront en outre d'un accès facilité aux jeux de données et aux fichiers de visualisation des médias d'archive, y compris un accès distant dans la limite des capacités techniques et des obligations légales de l'INA. Cet accès sera

proposé au nom seul du ou de la Lauréat(e) et pourra être fourni à d'autres membres de son équipe sur examen du besoin.

Enfin, le ou la lauréat(e), porteur de projet, sera invité(e) à participer au programme d'animation scientifique du lab, notamment à son séminaire, et bénéficiera du soutien de son réseau, et d'actions de valorisation en France et à l'international.

6. Engagements des Lauréat(e)s

Il sera demandé aux Lauréat(e)s, au plus tard dans les 6 mois après la fin de la période de soutien apportée par l'INA :

- Une présentation de leurs travaux, en commun avec les autres chercheur(e)s soutenu(e)s par l'INA ;
- Et/ou une présentation de leurs travaux et résultats dans le cadre du séminaire du lab, organisé à Paris (<https://inalelab.hypotheses.org/seminaires>) ;
- Et un rapport écrit présentant les résultats de leur recherche. Un article soumis à évaluation en vue de sa publication dans une revue scientifique en lien avec les travaux menés dans le cadre du Projet de recherche soutenu pourra faire office de rapport. Le ou la lauréat(e) pourra également proposer un "billet de recherche" pour publication sur le "carnet de recherches" en ligne du lab : <https://inalelab.hypotheses.org/>

Dans le cas où les résultats de recherche ne pourraient faire l'objet d'une licence ouverte (opendata) pour des raisons juridiques, le ou la Lauréat(e) s'engage à autoriser l'INA et toute personne autorisée par l'INA à utiliser (et le cas échéant à exploiter) les résultats de son Projet de Recherche, à des fins de recherche pour une durée de 10 ans.

Les publications ou communications du/de la Lauréat(e) sur les résultats de recherche devront mentionner le soutien du lab et de l'INA, et dans la mesure du possible le logo de l'INA. Afin que l'INA puisse valider l'apposition de ladite mention et de son logo, les supports de communication, projets de publications liés aux travaux de recherche, papier et en ligne, seront fournis préalablement à l'INA.

Ces publications et/ou communications ne pourront en aucun reproduire/représenter les médias et données, code informatique, interfaces de logiciels, sans autorisation préalable et écrite de l'INA et/ou de tous titulaires de droits et devront respecter les réglementations en vigueur, notamment la réglementation en matière de données personnelles (RGPD).

7. Garanties

Chaque Candidat(e) déclare et garantit :

- qu'il détient tous les droits nécessaires à la présentation de son Projet de Recherche dans le cadre de l'Appel, que les contenus attachés à ce Projet de Recherche sont originaux et n'enfreignent aucun droit, y compris les droits de propriété intellectuelle, de quiconque ;
- que son Projet de Recherche ne portera pas atteinte au droit à la vie privée, au droit à l'image, au droit de propriété ni à quelque autre droit de quelque individu ou personne morale, ni ne contiennent d'éléments diffamatoires, obscènes ou injurieux ;
- qu'il s'engage à indemniser et à défendre l'INA contre tous recours, action, réclamation, éviction de tiers (y compris dommages, honoraires, frais et débours en résultant) découlant d'une violation par lui du Règlement et/ou des droits desdits tiers.

8. Responsabilités

Le Règlement sera mis en ligne sur le site internet du lab <https://inalelab.hypotheses.org/> . Il pourra être modifié et complété, sans préavis ni formalité préalables, par l'INA, toute modification éventuelle étant également mise en ligne sur le site <https://inalelab.hypotheses.org/>. En cas de modification(s) portant sur le calendrier indicatif prévu à l'article 1.2, l'INA fera en outre mes meilleurs efforts pour informer les Candidats par courriel.

La participation à l'Appel implique de la part des Candidat(e)s (ainsi que, le cas échéant, des Lauréat(e)s) l'adhésion complète et sans réserve au Règlement.

L'INA se réserve à tout moment le droit de suspendre ou d'interrompre l'Appel, momentanément ou définitivement, sans préavis ni formalités. Il en informera les Candidat(e)s par courriel ainsi que, le cas échéant, par une publication *ad hoc* sur le site <https://inalelab.hypotheses.org/> . L'INA ne pourra voir sa responsabilité engagée de ce fait par quiconque et aucun remboursement, ni aucune indemnité d'aucune sorte, ne seront dus aux Candidat(e)s.

La participation des Candidat(e)s à l'Appel s'effectue sous leur entière responsabilité. En particulier, l'INA ne saurait en aucune circonstance être tenu responsable de l'impossibilité pour un(e) ou des Candidat(es) de faire parvenir son dossier à l'adresse électronique susmentionnée.

L'INA ne pourra en aucun cas être tenu responsable des dommages directs ou indirects subis par le Lauréat du fait de l'utilisation par ledit Lauréat des outils, technologies, collections et jeux de données, de difficultés survenues à l'occasion de leur utilisation ou de l'impossibilité de les utiliser.

9. Données personnelles

Les données personnelles du(de la) candidat(e) recueillies font l'objet d'un traitement informatique de la part de l'INA dans le cadre du présent règlement, dans les conditions ci-après définies :

7.1 Finalités de la collecte

L'INA, en sa qualité de responsable de traitement, traite les données à caractère personnel du Candidat pour les besoins du présent Appel (du dépôt des candidatures à la publication des Lauréats).

La candidature étant soumise à l'acceptation du Règlement, la base juridique de ce traitement est l'exécution d'un contrat.

Les catégories de données collectées sont : les données d'état civil et d'identité (nom, prénom), numéro de téléphone, adresse e-mail et l'adresse postale des candidats et Lauréats. Ces données sont indispensables au traitement et sont destinées aux services concernés de l'INA.

7.2- Durée de conservation

La durée de conservation des données personnelles sera limitée au temps nécessaire à l'accomplissement de l'objectif poursuivi lors de leur collecte. A l'issue de l'Appel, les données seront conservées en archivage intermédiaire pendant 3 ans. Après cette période d'archivage, les données seront définitivement supprimées.

7.3 - Sécurité et accès aux données informatisées

L'accès aux données du Candidat est limité aux personnes habilitées pour effectuer les opérations de traitements de données, de sécurité et de maintenance informatique.

7.4 - Droit d'accès, mise à jour et droit de rectification des données individuelles

Conformément et dans les limites de la réglementation relative à la protection des données personnelles, le Candidat dispose d'un droit d'interrogation, d'accès, de limitation du traitement de ses données, de rectification ou d'effacement des données le concernant, ainsi que du droit à la portabilité de ses données.

7.5 - Sort des données après le décès

Le Participant dispose aussi du droit de définir des directives relatives au sort de ses données après son décès.

7.6 - Contact

Pour exercer ses droits, le Candidat doit adresser sa demande par courrier électronique à dpdina@ina.fr ou par courrier postal auprès de l'Institut national de l'audiovisuel, à l'attention du Délégué à la Protection des Données, Direction Juridique, situé 4 avenue de l'Europe 94366 Bry-sur-Marne Cedex, accompagné d'une copie d'un titre d'identité.

Le Candidat dispose enfin du droit d'introduire une réclamation auprès de la Commission nationale de l'Informatique et des libertés (Cnil).

10. Loi applicable

Le Règlement est soumis au droit français.

Toute contestation portant sur l'interprétation ou l'exécution du Règlement sera soumise aux Tribunaux compétents.

Appel à projets du lab/ 2023-2024

Bulletin de candidature

1. Candidat(e)

NOM :

Prénom :

Adresse personnelle :

Téléphone (mobile) :

Courriel :

Dernier diplôme obtenu par le Candidat :

Date d'obtention du dernier diplôme :

Diplôme préparé ou statut du ou de la Candidat(e) au 02/10/2023 (doctorant, post-doctorant,...) (et date d'inscription) :

Université ou École (et département), laboratoire ou organisme d'inscription du ou de la Candidat(e) à la rentrée 2023 :

2. Autres membres de l'équipe / partenaires (le cas échéant)

Pour chaque personne, donner Prénom, Nom, Laboratoire / Unité de recherche / Institution

3. Projet de Recherche proposé au lab de l'INA

Titre court du Projet de Recherche proposé :

Cadre de mise en œuvre du projet (thèse de doctorat, HDR, projet de recherche financé par une agence régionale, nationale ou européenne, projet d'ouvrage...) :

J'accepte de recevoir les informations liées à l'activité du lab et d'être contacté(e) par l'INA pour intervenir lors de manifestations (tables rondes, séminaires...) en lien avec mon Projet de Recherche

Date :

Signature :

Les données personnelles recueillies par l'INA dans le cadre de votre candidature pourront être utilisées par l'INA et par le Comité pendant la durée nécessaire à la gestion de votre dossier. Si vous l'avez accepté, elles pourront également être utilisées par l'INA pour vous inviter à intervenir lors de manifestations en lien notamment avec votre Projet de Recherche, et ce pendant une durée de 10 ans à compter de la date du présent formulaire. Conformément à la loi « informatique et libertés » du 6 janvier 1978, vous bénéficiez d'un droit d'accès, de rectification voire d'opposition aux données vous concernant, que vous pouvez exercer en vous adressant à dpdina@ina.fr.

Le ou la Candidat(e) doit impérativement joindre à son dossier de candidature (envoi en fichiers séparés) : un CV, et la présentation de son Projet de Recherche (en 15 000 signes espaces compris maximum).

Jeux de données et outils disponibles

Jeux de données proposés à titre indicatif

- La transcription textuelle complète des JT des chaînes "hertziennes" (TF1, F2, F3, Canal+, M6, Arte), ainsi que des principales tranches horaires des chaînes d'information en continu depuis janvier 2019, effectuée par des outils de reconnaissance automatique de la voix (ASR).
- Toutes les occurrences à la télévision, par chaîne et par heure, des portions de flux répétées durant les 15 dernières années ; autrement dit, l'identification des rediffusions ou des extraits repris d'une chaîne par les autres chaînes, depuis leur première occurrence.
- Une base de données comprenant toutes les publicités répertoriées par l'Autorité de Régulation Professionnelle de la Publicité (ARPP) de 1968 à nos jours, cataloguées par annonceur, secteur publicitaire, etc. ; et indexée via le moteur de recherche visuel Snoop (voir ci-dessous).
- Une archive de plusieurs centaines de millions de tweets, correspondant à un échantillon glissant de 4500 hashtags et comptes twitter suivant l'actualité médiatique française depuis 2015, ainsi que les médias (vidéo, images, liens hypertexte) attachés à ces tweets ; y compris ceux ayant été depuis supprimés, modifiés, etc.

Outils et technologies pouvant être mobilisés

- La détection des visages et des voix pour l'extraction des caractéristiques des individus : âge et sexe ([inaFaceanalyzer](#), [inaSpeechsegmenter](#)), identité ([Trombinos](#)), pour alimenter des recherches sur la représentativité, la répartition du temps de parole, etc.
- La transcription automatique des voix et dialogues (ASR), ouvrant la possibilité d'appliquer l'analyse textuelle aux contenus audiovisuels : la fouille « *full-text* » dans un corpus audiovisuel ; la textométrie, le « *text as data* » et toute méthode de traitement automatique du langage (TAL) y compris celles à base d'apprentissage profond. Enfin, les bibliothèques / frameworks / boîtes à outils d'application de ces traitements aux humanités numériques ([Gargantext](#), [Cortext](#)...).
- Les moteurs d'indexation et de recherche visuels comme [Snoop](#), qui permet d'identifier au sein d'un corpus les images contenant un visage, un objet, un bâtiment, une posture, etc. ; et d'effectuer des mesures de similarité et des classifications sur l'ensemble du corpus, à partir de ces objets.
- L'étude des traces numériques via les outils d'analyse et de curation de corpus, de contenus web-natifs (curation de corpus de pages web avec [hyphe](#)), de tweets, etc.
- L'étude des données sous forme de graphes et de réseaux ; les outils d'analyse et de visualisation dans ce domaine (tels que [ipysigma](#), [gephi](#), etc.)
- Potentiellement, toute technologie ou bibliothèque open-source, libre, gratuite ou développée pour les besoins du projet