



À l'attention de l'**Ecole Universitaire de Recherche ArTec**

Le Kremlin-Bicêtre, le 25 septembre 2024

Objet : Lettre d'engagement pour le projet ArTec 2025 "Les femmes dans les métiers du cinéma français"

Madame, Monsieur,

Le Laboratoire de Recherche de l'EPITA (LRE) s'engage pleinement dans la réalisation des tâches qui lui incombent dans le cadre du projet "Les femmes dans les métiers du cinéma français" candidat à l'appel à projet recherche 2025 d'ArTeC. Notre laboratoire, fort de son expertise acquise au travers de divers projets de recherche, notamment le projet ANR SODUCO (<https://anr.fr/Projet-ANR-18-CE38-0013>), honorera les engagements suivants pour l'année 1 du projet :

1. Le LRE assurera la mise en œuvre de la fouille de données sur des annuaires professionnels du cinéma numérisés disponibles dans Gallica. Notre travail portera notamment sur l'analyse de mise en page, l'océrisation (OCR) et l'extraction d'informations. L'objectif est de produire un flux de données structurées pour les phases ultérieures du projet, comportant les entrées de ces annuaires. La boîte à outils du projet SODUCO sera adaptée et mise en œuvre, avec des ajustements nécessaires au niveau des modèles de reconnaissance d'entités : noms, professions, organisations, lieux, etc.
2. Le LRE prendra en charge le traitement d'un volume de 15 000 à 50 000 pages, issues de documents numérisés, représentant entre 10 et 33 annuaires. Ces données seront ensuite diffusées librement via des plateformes ouvertes comme api.bnf.fr et Zenodo, dans le respect des dispositions de la propriété intellectuelle et du RGPD concernant les données personnelles.
3. Deux livraisons de données seront réalisées, permettant un travail itératif avec les partenaires du projet.
4. Les données publiques seront traitées à l'aide des ressources de calcul internes du LRE, tandis que celles relevant du domaine privé seront traitées via les ressources du BnF Datalab.
5. Deux enseignants-chercheurs permanents superviseront et contribueront à la réalisation des travaux, selon un effort estimé à 3 semaines.personne au total.

ÉCOLE POUR L'INFORMATIQUE ET LES TECHNIQUES AVANCÉES

EPITA PARIS | 14-16 RUE VOLTAIRE | 94270 LE KREMLIN-BICÊTRE | MÉTRO : PORTE D'ITALIE

Tél. 01 44 08 01 01 | Fax. 01 44 08 01 99 | www.epita.fr | e-mail : contact@epita.fr

SA EPITA – RCS Créteil N 898 192 075 | Siret 898 192 075 00015 | Établissement d'Enseignement Supérieur Privé



En contrepartie de ces engagements, le LRE recevra :

1. Une liste exhaustive des pages à traiter, contenant leur typologie, année, provenance et URL permettant leur téléchargement via Gallica.
2. Une somme d'un montant de 10 000 € au titre de la prestation de traitement des données des annuaires, visant à couvrir partiellement les frais de personnels non-permanents, de matériel et d'infrastructure de calcul.
3. L'accès aux ressources de calcul du BnF Datalab pour le traitement des données relevant du domaine privé.

Il est à noter que le LRE restera libre de publier les résultats de ses travaux exclusifs sur la base des données libres du projet.

Nous vous prions d'agréer, Madame, Monsieur, l'expression de nos salutations distinguées.

Thierry Géraud
Directeur de la Recherche de l'EPITA



14-16 rue Voltaire
94276 LE KREMLIN-BICÊTRE CEDEX