

## Offre de stage de Master 2 - IA & Humanités numériques

### Évaluation et ajustement de modèles d'ATR/HTR pour le florentin ancien

6 mois (de janvier à juin 2027)

#### Informations générales

Type de stage : obligatoire, niveau Master 2

Durée : 6 mois (4 janvier 2027 – 30 juin 2027)

Date limite des candidatures : 20 octobre 2026

Gratification légale de stage

Établissement d'accueil : Université Grenoble Alpes (SIRET 13002608100013)

Laboratoire d'accueil : LUHCIE (EA n°7421), Bâtiment ARSH, 1281 rue des Universités, 38610 Gières

Projet : FlorA – Archive numérique des textes mémoriels florentins en langue vulgaire

*Ce stage est financé par des fonds France 2030, dans le cadre de l'appel à manifestation d'intérêt Compétences et Métiers d'Avenir, et est rattaché au projet AMI CMA EFELIA-MIAI.*

#### Le stage

Comment adapter des modèles d'intelligence artificielle à des documents historiques qui n'ont jamais été conçus pour être lus par une machine ? Ce stage propose d'explorer cette question à partir d'un corpus de textes florentins des XIV<sup>e</sup>-XVI<sup>e</sup> siècles, imprimés et manuscrits. Il s'inscrit dans le cadre d'un projet d'archive numérique consacré aux écrits mémoriels florentins : chroniques, livres de famille, *ricordi* et *ricordanze*, récits d'ambassade, etc. Une part importante de ces sources est encore aujourd'hui accessible uniquement sous forme d'éditions anciennes ou de manuscrits. **L'objectif du stage est de tester, évaluer et adapter des modèles de reconnaissance automatique de textes (ATR/HTR), notamment avec Kraken HTR (via la plateforme eScriptorium) afin de les rendre plus performants sur le florentin ancien.** Les modèles disponibles sont encore peu adaptés à l'italien ancien. Le problème est particulièrement épineux pour les manuscrits, dont l'écriture est généralement une cursive marchande (*mercantesca*), mais également pour les imprimés anciens, qui présentent d'importantes variations graphiques et orthographiques.

#### Missions

Le stage combine expérimentation en IA, *computer vision*, traitement automatique de documents, préparation de données et humanités numériques. Le/la stagiaire sera notamment amené-e à :

- prendre en main Kraken HTR (via la plateforme eScriptorium) ;
- tester et évaluer plusieurs modèles existants sur des échantillons de textes florentins ;
- constituer et préparer des données d'entraînement et d'évaluation ;
- analyser les erreurs et mesurer les performances des modèles ;
- entraîner et/ou ajuster un ou plusieurs modèles pour le florentin ancien ;
- comparer les performances sur des documents imprimés et manuscrits ;
- documenter les expérimentations et les résultats ;
- préparer les données produites pour leur intégration dans une chaîne éditoriale (XML-TEI) ;
- contribuer à la mise en place d'un pipeline reproductible de traitement des données textuelles et de leurs métadonnées.

Le stage doit notamment aboutir à un ou plusieurs modèles adaptés au florentin ancien et à une chaîne de traitement susceptible d'être réutilisée à plus grande échelle dans FlorA.

#### Profil recherché

Le stage s'adresse prioritairement à des étudiant-es de Master 2 en informatique, informatique appliquée aux SHS, humanités numériques ou en formations interdisciplinaires proches. Nous recherchons en particulier :

- un intérêt pour l'IA appliquée aux documents et corpus patrimoniaux ;

- un intérêt pour le traitement d'images / la *computer vision*, notamment les problématiques d'OCR/HTR et de *machine learning*
- de bonnes bases en programmation (Python) ;
- une familiarité avec le traitement et la structuration de données ;
- une bonne compréhension des enjeux d'évaluation (séparation des jeux d'entraînement et de test, métriques d'évaluation CER, WER)
- une appétence pour le travail expérimental et l'analyse de résultats.
- un niveau d'anglais B2 minimum ; un niveau de français B1 ou supérieur (facultatif mais apprécié)

Des connaissances en XML-TEI, en édition numérique, en italien ou dans les domaines historiques et littéraires concernés sont un plus, mais ne sont pas indispensables. L'objectif est précisément de travailler à l'interface entre compétences techniques et problématiques de recherche en SHS.

## Environnement de recherche

Le stage permettra de travailler dans un environnement interdisciplinaire associant IA, ingénierie des données, humanités numériques et recherche en sciences humaines. Le/la stagiaire sera intégré-e au projet FlorA et bénéficiera d'un double encadrement :

- Elise Leclerc (LUHCIE) pour le volet scientifique et les problématiques liées au corpus ;
- Karolina Boczoń (GRICAD) pour le volet technique, en lien avec les référents de la plateforme *eScriptorium* au sein de l'équipe ELAN de Litt&Arts.

Le stage, d'une durée de 6 mois à temps plein, à partir du 4 janvier 2027, sera basé à Grenoble, au LUHCIE. Un bureau sera mis à disposition du ou de la stagiaire. Un ou plusieurs jours de travail à distance par semaine pourront être organisés, en fonction des exigences du projet et des contraintes du candidat.

## Comment postuler ?

Les candidat-es intéressé-es doivent soumettre les documents suivants (en français ou en anglais) :

- **curriculum vitae (CV)**
- **relevé de notes de master**
- **lettre de motivation** présentant votre intérêt pour le stage et vos éventuelles expériences en IA, computer vision, programmation, traitement de données ou humanités numériques.
- **lettre de recommandation ou de références académiques** : facultative.

Les candidatures sont à adresser à [elise.leclerc@univ-grenoble-alpes.fr](mailto:elise.leclerc@univ-grenoble-alpes.fr) avant le **20 octobre 2026 à 18h** en indiquant en objet « Candidature pour le stage FlorA - MIAI-EFELIA ».

## Calendrier

En raison de délais contraints par un changement du logiciel comptable de l'Université, les candidatures seront examinées dès réception ; les candidat-es présélectionné-es seront contacté-es pour un entretien (en visio) qui se déroulera au plus tard le vendredi 23 octobre 2026.

**20 octobre 2026 à 18h** – date limite d'envoi des candidatures.

**26 octobre 2026 à 9h** – communication des résultats

**29 octobre 2026 à 12h** – date limite d'envoi des informations administratives et financières nécessaires à la mise en place de la gratification.

**4 janvier 2027 à 9h** – début du stage.

**30 juin 2027 à 17h** – fin du stage.

**Rémunération** : 4,50 €/h (tarif légal), soit environ 630 €/mois pour 35h/semaine.

N.B. Les congés et les jours fériés non gratifiés.

**Travail à fournir après le stage** : rapport de stage.

Pour toute demande d'information : [elise.leclerc@univ-grenoble-alpes.fr](mailto:elise.leclerc@univ-grenoble-alpes.fr)

## **Pour en savoir plus sur l'environnement du stage**

Grenoble, nichée au cœur des Alpes françaises, abrite plusieurs universités et instituts de recherche renommés, dont l'Université Grenoble Alpes. En plus de son offre universitaire, Grenoble est également entourée de magnifiques sites naturels avec un accès facile aux montagnes pour le ski, la randonnée et d'autres activités de plein air. La population de Grenoble est diversifiée et internationale. L'Université Grenoble Alpes compte 60 000 étudiant-es, 6 000 chercheur-es et 3 000 doctorant-es. Elle est classée chaque année parmi le groupe 100-250 des meilleures universités mondiales dans les grands classements internationaux et figure parmi les 10 premiers établissements d'enseignement supérieur français.

L'UGA compte 71 unités de recherche, dont le LUHCIE (Laboratoire Universitaire Histoire Cultures Italie Europe, EA n°7421). Le laboratoire réunit des historiens, des historiens de l'art, des musicologues et des spécialistes de la langue, de la littérature et de la civilisation italiennes. Il rassemble 34 enseignant-chercheur-es et 50 doctorant-es rattaché-es aux Écoles doctorales SHPT (Sciences de l'Homme, du Politique et du Territoire) et LLSH (Langues, Littératures et Sciences Humaines). Il est intégré au Pôle de recherche « Sciences humaines et sociales » (SHS) de l'Université Grenoble Alpes. Le laboratoire est engagé dans de nombreux programmes de recherche concernant l'histoire, l'histoire de l'art, la musicologie, la littérature, l'image, le patrimoine, le politique, la culture, la traduction et les humanités numériques. Ces programmes sont structurés selon quatre axes : Circulations, transferts, frontières – Écritures, objets, formes de communication – Sciences et savoirs – Territoires, pratiques et discours politiques (<https://luhcie.univ-grenoble-alpes.fr/>).

Le Cluster IA MIAI (Institut multidisciplinaire en intelligence artificielle) vise à mener des recherches en intelligence artificielle au plus haut niveau, à proposer des cours attractifs pour les étudiant-es et professionnel·les de tous niveaux, à soutenir l'innovation dans les grandes entreprises, PME et startups, ainsi qu'à informer et interagir avec les citoyen·nes sur tous les aspects de l'IA. Il réunit des partenaires des secteurs public et privé et vise à développer une IA pour les humains et l'environnement (<https://miai-cluster.univ-grenoble-alpes.fr/miai-institute>)

## Master 2 Internship Offer - AI & Digital Humanities

### Evaluation and fine-tuning of ATR/HTR models for Old Florentine

6 months (January to June 2027)

#### General information

Type of internship: compulsory, Master 2 level

Duration: 6 months (4 January 2027 – 30 June 2027)

Application deadline: 20 October 2026

Statutory internship allowance

Host institution: Université Grenoble Alpes (SIRET 13002608100013)

Host laboratory: LUHCIE (EA no. 7421), Bâtiment ARSH, 1281 rue des Universités, 38610 Gières

Project: FlorA – Digital archive of Florentine memorial texts in the vernacular

*This internship is funded by France 2030, as part of the call for expressions of interest Compétences et Métiers d'Avenir, and is attached to the AMI CMA EFELIA-MIAI project.*

#### The internship

How can artificial intelligence models be adapted to historical documents that were never designed to be read by a machine? This internship sets out to explore this question on the basis of a corpus of Florentine texts from the 14th-16th centuries, both printed and handwritten. It is part of a digital archive project devoted to Florentine memorial writings: chronicles, family books, *ricordi* and *ricordanze*, embassy accounts, etc. A significant proportion of these sources is still accessible today only in the form of early editions or manuscripts.

**The aim of the internship is to test, evaluate and adapt automatic text recognition (ATR/HTR) models, in particular with Kraken HTR (via the eScriptorium platform), in order to improve their performance on Old Florentine.** The models available are still not well-suited to Old Italian. The problem is particularly thorny for manuscripts, whose script is generally a merchant cursive (*mercantesca*), but also for early printed books, which display considerable graphic and orthographic variation.

#### Tasks

The internship combines experimentation in AI, computer vision, automatic document processing, data preparation and digital humanities. In particular, the intern will be expected to:

- get to grips with Kraken HTR (via the eScriptorium platform);
- test and evaluate several existing models on samples of Florentine texts;
- build and prepare training and evaluation data;
- analyse errors and measure the performance of the models;
- train and/or fine-tune one or more models for Old Florentine;
- compare performance on printed and handwritten documents;
- document the experiments and the results;
- prepare the data produced for integration into an editorial pipeline (XML-TEI);
- contribute to setting up a reproducible pipeline for processing textual data and their metadata.

In particular, the internship is expected to result in one or more models adapted to Old Florentine and in a processing pipeline that can be reused on a larger scale within FlorA.

#### Candidate profile

The internship is aimed primarily at Master 2 students in computer science, computer science applied to the humanities and social sciences (SHS), digital humanities or closely related interdisciplinary programmes. We are looking in particular for:

- an interest in AI applied to heritage documents and corpora;
- an interest in image processing / computer vision, in particular OCR/HTR and machine learning issues
- a solid grounding in programming (Python);

- familiarity with data processing and data structuring;
- a good understanding of evaluation issues (separation of training and test sets, CER and WER evaluation metrics)
- a taste for experimental work and the analysis of results.
- English at B2 level or above; French at B1 level or above (optional but appreciated)

Knowledge of XML-TEI, digital publishing, Italian or the relevant historical and literary fields is a plus, but is not essential. The aim is precisely to work at the interface between technical skills and research questions in the humanities and social sciences.

## Research environment

The internship will provide an opportunity to work in an interdisciplinary environment combining AI, data engineering, digital humanities and research in the humanities. The intern will be integrated into the FlorA project and will benefit from joint supervision:

- Elise Leclerc (LUHCIE) for the scientific side and the issues relating to the corpus;
- Karolina Boczoń (GRICAD) for the technical side, in liaison with the contact persons for the *eScriptorium* platform within the ELAN team of Litt&Arts.

The internship, lasting 6 months full-time and starting on 4 January 2027, will be based in Grenoble, at LUHCIE. An office will be made available to the intern. One or more days of remote work per week may be arranged, depending on the requirements of the project and the candidate's constraints (geographical distance).

## How to apply?

Interested candidates must submit the following documents (in French or in English):

- **curriculum vitae (CV)**
- **Master's transcript of records**
- **cover letter** setting out your interest in the internship and any experience you may have in AI, computer vision, programming, data processing or digital humanities.
- **letter of recommendation or academic references:** optional.

Applications should be sent to [elise.leclerc@univ-grenoble-alpes.fr](mailto:elise.leclerc@univ-grenoble-alpes.fr) before **20 October 2026 at 6:00 pm**, with the subject line "Candidature pour le stage FlorA - MIAI-EFELIA".

## Timeline

Owing to tight deadlines imposed by a change in the University's accounting software, applications will be reviewed as soon as they are received; shortlisted candidates will be contacted for an interview (by video call), which will take place no later than Friday 23 October 2026.

**20 October 2026 at 6:00 pm** – deadline for submitting applications.

**26 October 2026 at 9:00 am** – notification of results

**29 October 2026 at 12:00 noon** – deadline for sending the administrative and financial information required to set up the internship allowance.

**4 January 2027 at 9:00 am** – start of the internship.

**30 June 2027 at 5:00 pm** – end of the internship.

**Remuneration:** €4.50/h (statutory rate), i.e. approximately €630/month for 35h/week.

N.B. Leave and public holidays are not remunerated.

**Work to be submitted after the internship:** internship report.

For any enquiries: [elise.leclerc@univ-grenoble-alpes.fr](mailto:elise.leclerc@univ-grenoble-alpes.fr)

## **Find out more about the internship environment**

Grenoble, nestled in the heart of the French Alps, is home to several renowned universities and research institutes, including Université Grenoble Alpes. In addition to its academic offering, Grenoble is also surrounded by magnificent natural sites, with easy access to the mountains for skiing, hiking and other outdoor activities. Grenoble's population is diverse and international.

Université Grenoble Alpes has 60,000 students, 6,000 researchers and 3,000 doctoral students.

It is ranked every year in the 100-250 group of the world's best universities in the major international rankings and is among the top 10 French higher education institutions.

UGA has 71 research units, including LUHCIE (Laboratoire Universitaire Histoire Cultures Italie Europe, EA no. 7421). The laboratory brings together historians, art historians, musicologists and specialists in Italian language, literature and civilisation. It comprises 34 academic staff and 50 doctoral students attached to the doctoral schools SHPT (Sciences de l'Homme, du Politique et du Territoire) and LLSH (Langues, Littératures et Sciences Humaines). It is part of the "Humanities and Social Sciences" (SHS) Research Department of Université Grenoble Alpes. The laboratory is involved in numerous research programmes concerning history, art history, musicology, literature, the image, heritage, politics, culture, translation and digital humanities. These programmes are structured around four research axes: Circulations, transfers, borders – Writings, objects, forms of communication – Sciences and knowledge – Territories, political practices and discourses (<https://luhcie.univ-grenoble-alpes.fr/>).

The MIAI AI Cluster (Multidisciplinary Institute in Artificial Intelligence) aims to conduct research in artificial intelligence at the highest level, to offer attractive courses for students and professionals at all levels, to support innovation in large companies, SMEs and startups, and to inform and interact with citizens on all aspects of AI. It brings together partners from the public and private sectors and aims to develop AI for people and the environment (<https://miai-cluster.univ-grenoble-alpes.fr/miai-institute>)