

## Master Humanités numériques

Semaine de compétitions et de hackathons | Janvier 2026



## Projet 1 : Plonger au cœur des réseaux à la cour de France (XVII<sup>e</sup>–XVIII<sup>e</sup> siècles)

### Hackathon @ENC x Prosocour @CRCV

Explorer, à partir de la base Prosocour, les réseaux de sociabilité de la cour de France, de Louis XIV à Louis XVI, le tout à partir de données fiabilisées.





## Projet 2 : Entités nommées dans la BiblioBase

La BiblioBase est une édition numérique de la Bibliographie de la France, contenant les références d'ouvrages littéraires dont la publication est annoncée au XIX<sup>e</sup> siècle. Dans la BiblioBase, les noms d'auteurs, de traducteurs et de para-auteurs sont en transcription diplomatique.

L'objectif est d'aligner ces entités nommées avec un format de référence comme l'ISNI.





## Projet 3 : Abbreviations and algorithms - The icelandic challenge

Icelandic manuscripts are known for being dark, damaged and carrying highly abbreviated texts, posing challenges for ATR technology. In this project, we will test how far we can squeeze out of these manuscripts by leveraging computer vision and NLP approaches.

The aim is to build a flexible pipeline applicable well beyond the Icelandic case, which will allow editors to automate routine transcription steps.



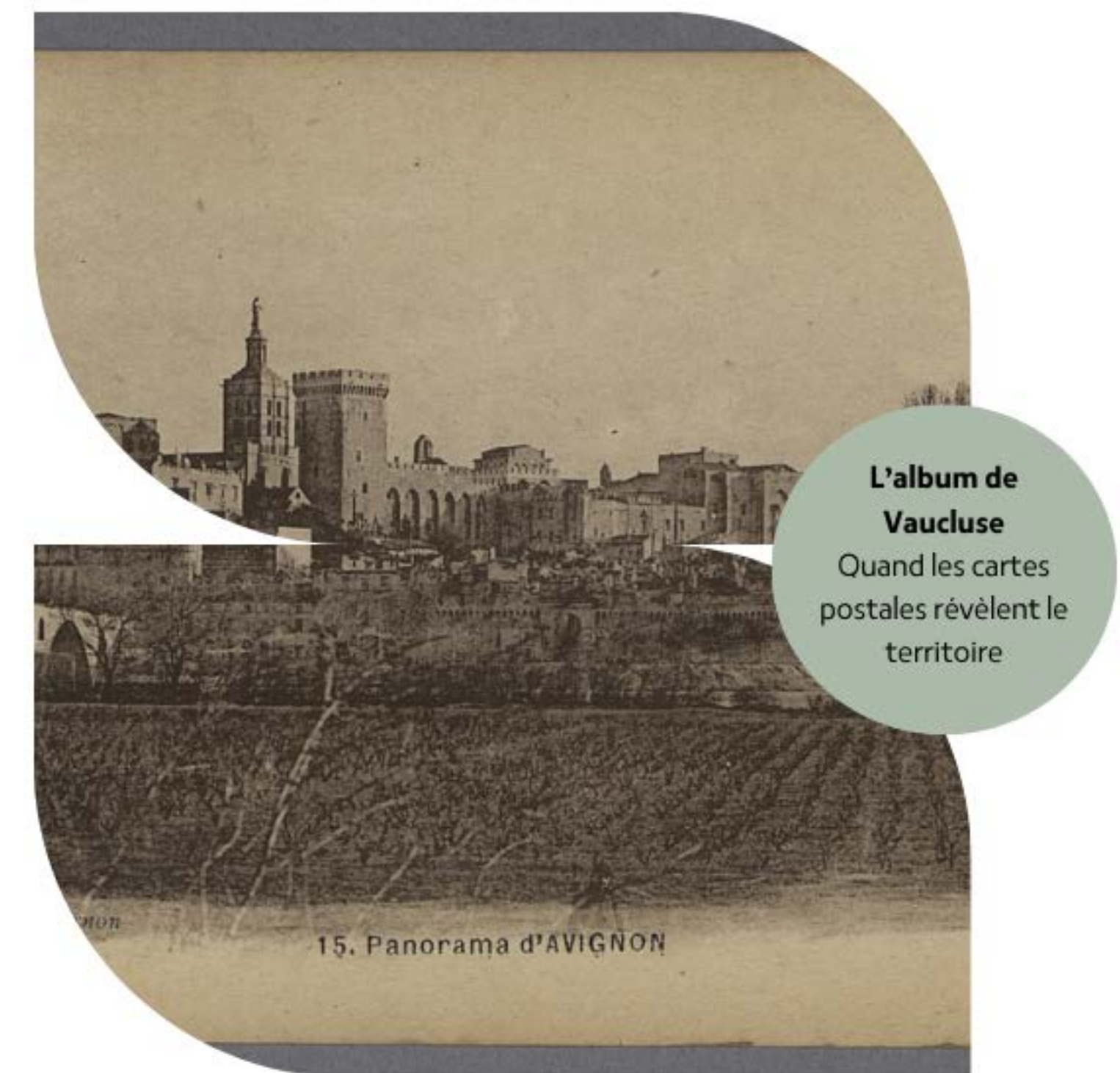
AM02 0132-is, MG 5998



## Projet 4 : Album de Vaucluse : Quand les cartes postales révèlent le territoire

L'objectif final est de réaliser une cartographie du territoire de Vaucluse, une cartographie interactive sur laquelle les lieux et les images s'affichent.

Pour procéder, nous demandons aux étudiants d'analyser les images pour en extraire la légende et pour identifier les vues similaires afin de les regrouper. Les étudiants seront amenés, en fonction des éléments analysés, à associer les coordonnées géographiques aux images.





## Projet 5 : Explorer les voies de classement et de description des photographies du fonds Thierry

Contribuez au projet ThierryNum (INHA / DIM PAMIR / UMR 8167 Orient et Méditerranée) en imaginant des méthodes d'indexation automatique et autres outils pour explorer un vaste fonds photographique relatif au patrimoine de l'Arménie et de la Géorgie médiévales.



institut  
national  
d'histoire  
de l'art



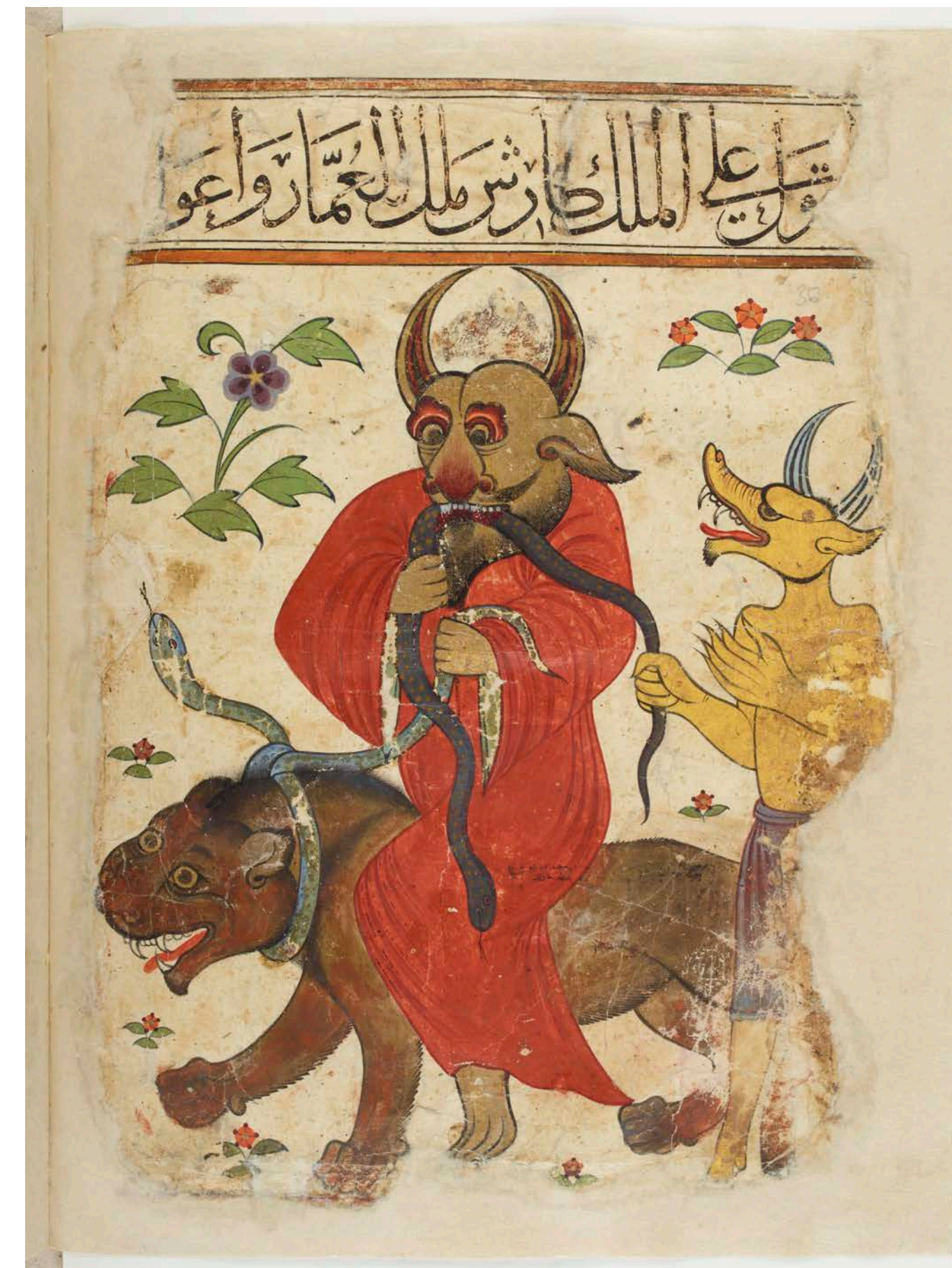
Fonds Thierry, INHA



## Projet 6 : Naturalis historia : Cartographier l'imaginaire naturel médiéval dans Mandragore

La base Mandragore offre un cadre privilégié pour l'analyse automatisée de l'enluminure médiévale. En s'appuyant sur les métadonnées existantes, le défi consiste à concevoir un pipeline associant vision par ordinateur et modélisation des données patrimoniales.

L'objectif est de repérer, annoter et mettre en relation les éléments iconographiques des miniatures, tout en expérimentant différents modèles de computer vision.



Paris, BNF, ARABE 2583, fol. 2v



## Projet 7 : L'avortement à la télévision française (1960-2024)

Modélisation des discours, des représentations et des thématiques associés à l'avortement

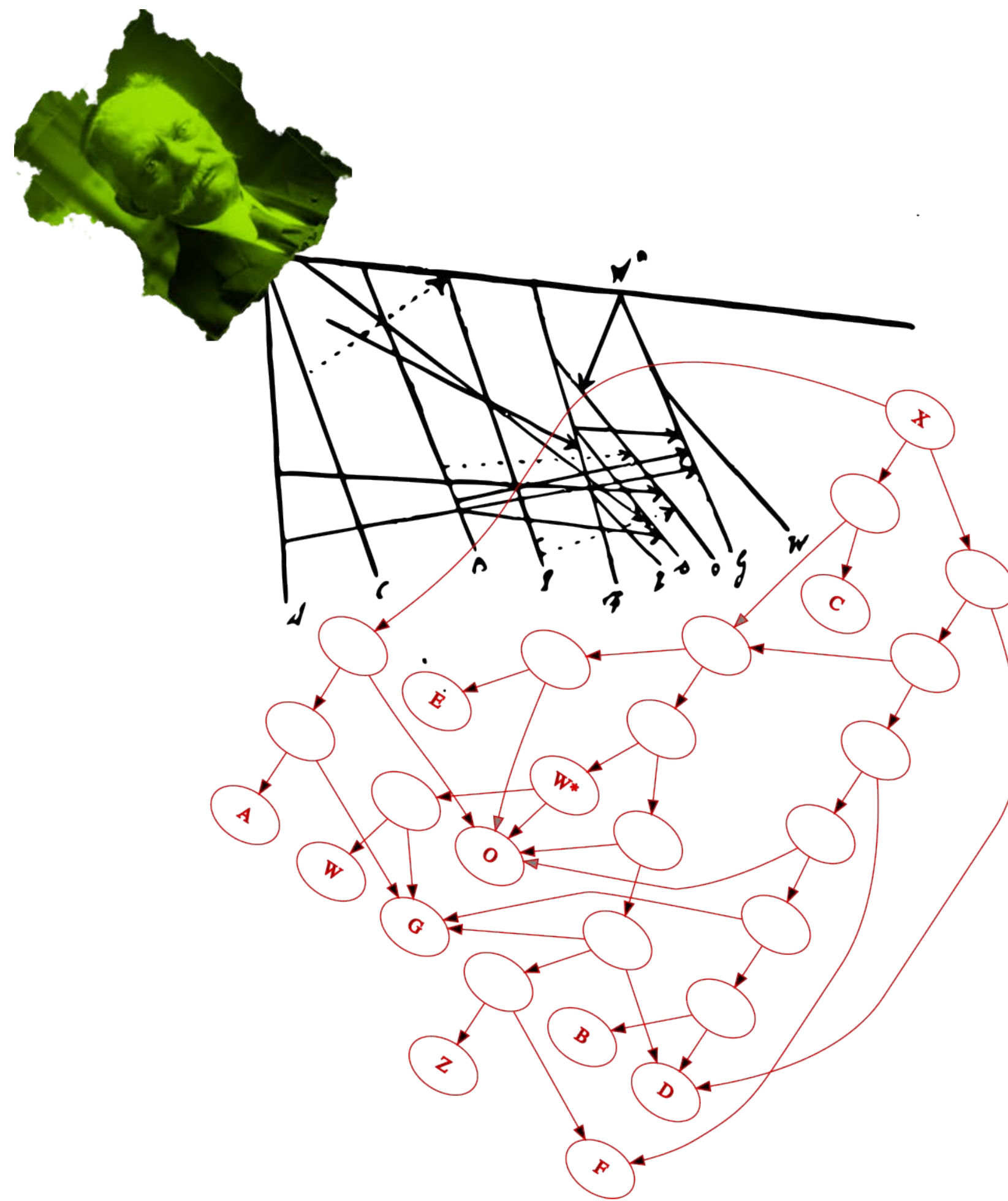


©Josée Lorenzo, INA, 1972



## Projet 8 : Trouver les arbres cachés...

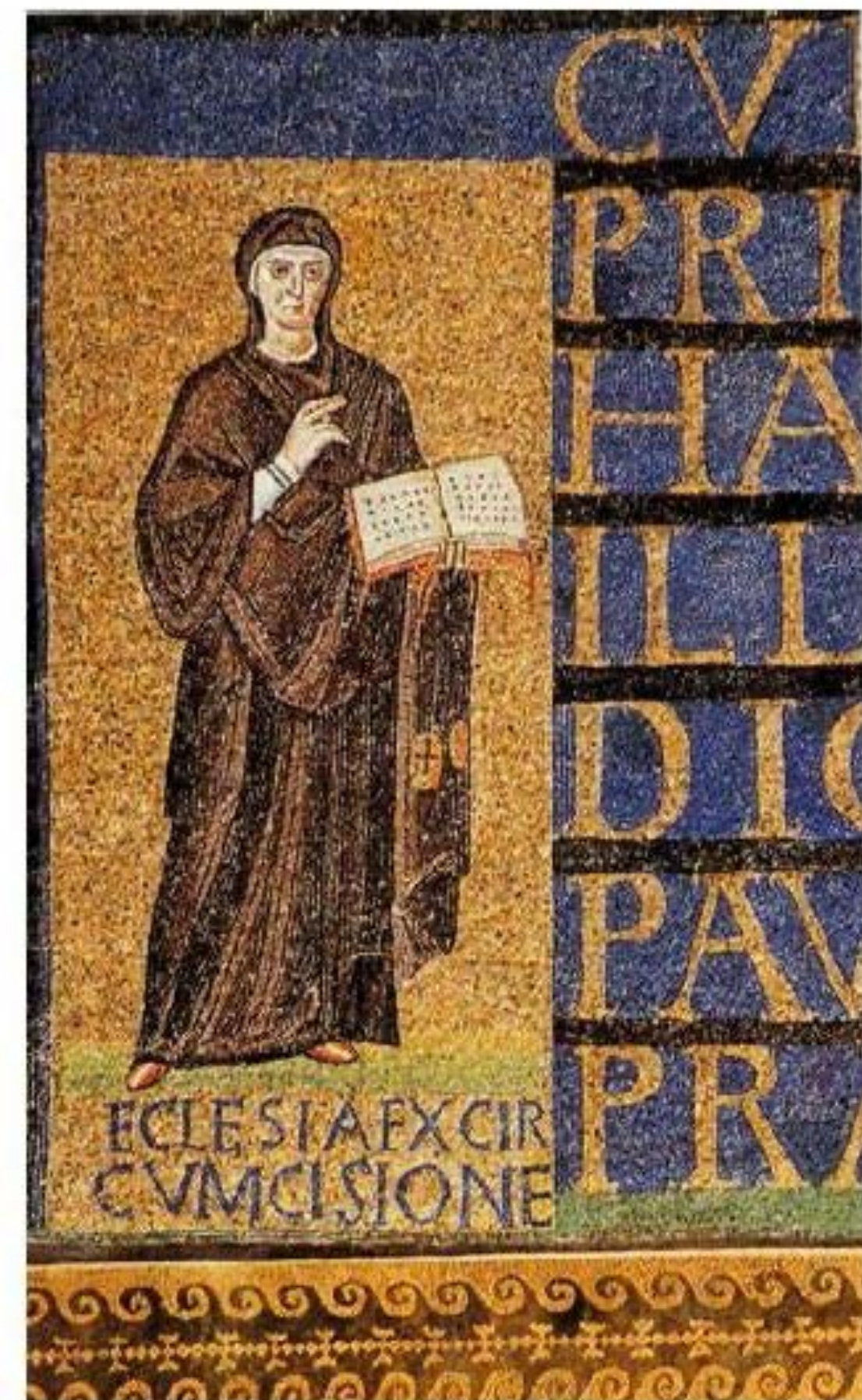
Dans le cadre des projets LostMA (The Lost Manuscripts of Medieval Europe: Modelling the Transmission of Texts) et du projet collaboratif OpenStemmata, nous organisons une compétition dédiée à l'identification, la transcription et la contextualisation de généalogies de manuscrits (stemmata) à partir de corpus de travaux scientifiques numérisés.





## Projet 9 : Des corpus en relief : identification, analyse et sérialisation de corpus épigraphiques

Construire une chaîne reproductible pour filtrer, explorer et encoder les éléments qui composent le discours architectural observé dans des corpus épigraphiques latins publiés en ligne, dans le cadre du projet ANR E-cclesia (ANR-24-CE54-6743).



Eglise de S. Sabina, Aventin, Rome



## Projet 10 : O.D.I.L : Indexer automatiquement les images du Moyen-Âge par l'IA

Dans le cadre du projet O.D.I.L (ENC-PSL, EHESS-ALHoMA, DHLab-EPFL et Biblissima), nous invitons les participants à relever le défi d'une tâche de classification multi-label extrême en vision par ordinateur pour indexer automatiquement les images médiévales à partir des étiquettes du thésaurus TIMEL (plus de 1 000 concepts).





## Projet 11 : Aide à la réunion de microfaciès de silicites par la reconnaissance de formes dans des images microscopiques

Dans le cadre du GDR SILEX, l'identification des vues au stéréomicroscope repose aujourd'hui sur une analyse visuelle experte, longue et coûteuse en temps.

Le hackathon vise à automatiser ce processus en exploitant les catalogues de référence : extraire des données d'images, définir des critères pertinents, structurer des microfaciès cohérents et attribuer les pièces archéologiques aux types du référentiel.

L'objectif est d'évaluer la faisabilité et les conditions d'application des algorithmes de vision par ordinateur en pétroarchéologie.

**GDR** Groupement  
de recherche  
**SILEX**  
Caractérisation et provenance





## Projet 12 : One prompt to rule them all? Compétition de transcription automatique de manuscrits médiévaux français

Dans le cadre du projet LostMa, nous proposons de développer une solution d'HTR pour des manuscrits français de chansons de gestes et de romans en prose. Les participants entraîneront des modèles capables de transcrire automatiquement des pages manuscrites. Les prédictions seront évaluées sur un jeu de test confidentiel à l'aide du CER.





## Projet 13 : Quand Archelec rencontre l'IA

Génération automatique de métadonnées à partir  
de professions de foi électorales françaises de  
2017 à 2024

# SciencesPo



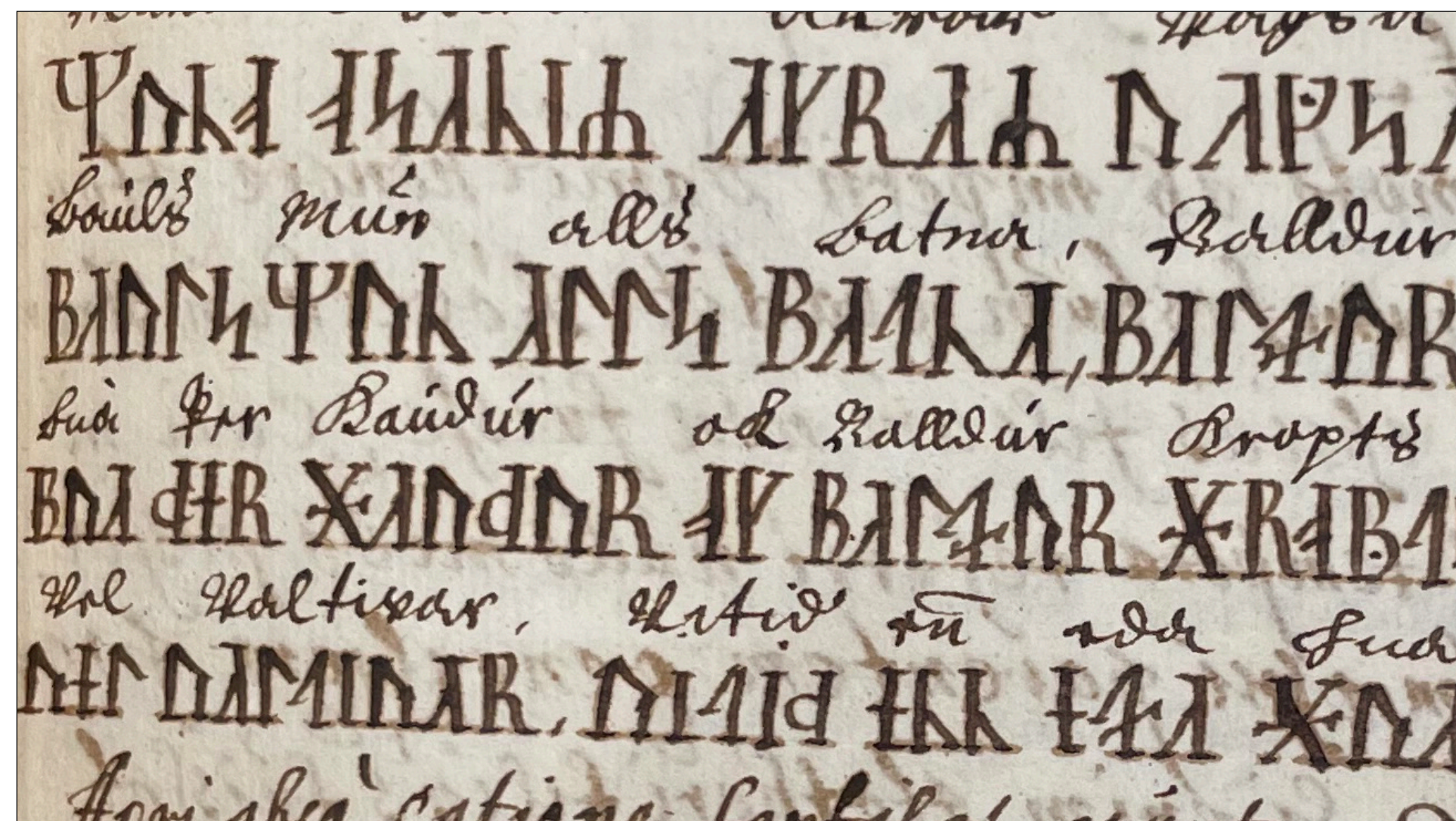
Professions de foi des élections législatives 2022. Mention des droits :  
Caroline Maufruid/Sciences Po.



## Projet 14 : Text Acquisition Pipeline for Early-Modern Latin Manuscripts

Libraries and archives around the world hold vast collections of undigitised manuscripts that remain largely inaccessible for computational research.

This project aims to develop a reusable, end-to-end pipeline leveraging computer vision and vision-language models to convert photographs taken by researchers into machine-readable transcriptions, adaptable across diverse manuscript types.





## Projet 15 : Analyser en direct les flux de chaînes d'information en continu

InfoVoxTracker est une base de données compilant les prises de paroles des invités ou des personnalités publiques sur les chaînes françaises d'information en continu.

En s'appuyant sur les données existantes, le défi consiste à expérimenter différents modèles afin de réaliser un décompte précis des temps de paroles politiques par analyse des flux sonores, ou encore d'effectuer une analyse sémantique des discours à l'antenne.



# InfoVox Tracker



# Partenaires et soutiens

