3iA Côte d'Azur
Interdisciplinary Institute
for Artificial Intelligence

Ph.D. research topic

- Title of the proposed topic: **AI for archaeozoology: learning methods to identify and cluster faunal remains**
- Research axis of the 3IA: Axis 1
- **Supervisor (name, affiliation, email):** Marco Corneli (Universtité Côte d'Azur, CEPAM, Inria, CNRS, LJAD, marco.corneli@univ-cotedazur.fr)
- Co-supervisors (name, affiliation):
    - Emmanuelle Vila: CR-CNRS (HDR) in Archaeozoology, Archéorient, Université Lumière Lyon II, CNRS (co-supervisor)
    - Manon Vuillien: post-doc researcher IDEX in Archaeozoology, CEPAM, CNRS, Université Côte d'Azur (advisor or 'co-encadrante')
- The laboratory and/or research group: CEPAM, LJAD (EP MAASAI), Archéorient (MOM)

Application procedure:
Interested candidates should send an e-mail with their CV, their transcripts and two possible referee contacts to marco.corneli@univ-cotedazur.fr and emmanuelle.vila@mom.fr and manon.vuillien@univ-cotedazur.fr. Selected candidates will be invited for an interview organised through video-conferencing services.

Context and research challenges
Although quite recent, the adoption of Machine Learning (ML) techniques for the analysis of archaeological data sets is rapidly increasing [Mackenzie, 2017, Mesanza-Moraza et al., 2020, Bickler, 2021, Palacios, 2023]. ML applications deal with numerical and/or categorical data [e.g. Klassen et al., 2018], textual data [e.g Assael et al., 2022], images [e.g. Horache et al., 2021] and geospatial data [Ramazzotti, 2020]. Applications on bioarchaeological data, such as plants and animal remains, are still underdeveloped, although some works on agriculture [Baum et al., 2016, Puech et al., 2022], marine resources [Bickler, 2018, Ahedo et al., 2021] and taphonomy of animal bones [Cifuentes-Alcobendas and Dom´ınguez-Rodrigo, 2019] are gradually intensifying. Thus, the present PhD project is an opportunity for the development of original ML solutions to boost the impact of AI in archaeozoology and offer new perspectives to researchers in that field. The inspection of faunal remains, i.e. animal bones found in archaeological contexts, provides researchers with information on past human-non-human mammal relationships, paleoenvironments, past animal populations (biology) and on

the subsistence economy of ancient societies. Such remains mainly correspond to complete and/or incomplete bones that have been altered by humans and environments over a long period of time. Typically, the nature of animal bones with poor preservation, fragmentation and state can affect the success of anatomical and taxonomical identification. Indeed, the identification is directly made by the archaeozoologist mainly based on modern anatomical, morphological and biometric criteria [Barone, 1976, Driesch, 1976] and very little on molecular analysis because of the preservation conditions of the material and the cost of these studies. Moreover, if the expert succeeds in the species determination, he is quickly confronted with a strong polymorphism in some closely related species (such as ruminant ungulates, e.g. deer, ibex) and with the emergence of new morphotypes after the domestication processes, especially in sheep, and the intensification of livestock breeding 8 millennia ago to yield specific products still consumed and used today, such as milk and wool. For these reasons, the identification process made by the archaeozoologists is long, tedious and particularly challenging and the development of original and adapted AI tools to support the experts in the taxonomical identification is of particular interest. From a pure machine learning perspective, however, such development is far from being straightforward for several reasons: i) the data (either 2D images or point clouds obtained via 3D scans) might be globally scarce, even in the training phase, and very few observations (i.e. specimens) might be available for some classes (i.e. species); ii) the data might be fragmented and so incomplete; iii) the classification task is difficult *in re ipsa* meaning that the non-expert human eye cannot usually distinguish bones from different species (unless trivial criteria, like size, are discriminant). In particular, point i) undermines most of the recent deep learning machinery used for shapes classification [e.g. PointNet Qi et al., 2017], even if one wished to adopt them for simple feature extraction. So, from one side, there is a need for parsimonious machine learning approaches to classify, reconstruct and possibly segment 3D shapes. From another point of view, the aim of this PhD is to support the activity of the archaeozoologist and not to replace her. On the contrary, her knowledge should be exploited, and this is why there is a need for a comprehensive set of neuro-symbolic AI methods[Sarker et al., 2021] allowing one to inject the expert's knowledge into the learning routine. Thus, the PhD project is articulated in two main axes, that we now describe in detail.

Inter-active supervised learning to identify morphologically close herbivore animals

This part deals with the development of original learning techniques to assist archaeozoologists and expand their workflow on the identification of morphologically close animals found in archaeological sites. The dataset under investigation will be made of astragalus (physical and 3D model) of modern and archaeological ruminants: roe deer, gazella, ibex, sheep and goat. The supervised learning (i.e. the taxonomical identification) is currently accomplished by the experts based on specific anatomical and biometrical criteria that the PhD candidate will become familiar with. Indeed, one option is to develop hybrid architectures on the model of PointSIFT [Jiang, 2018] in which the first layers of the network are "freezed" on handmade keypoints, corresponding to peculiar anatomical features. However, due to the importance of the morphology in our context, we believe that 'pure' machine learning approaches based
on Topological Data Analysis [Chazal and Michel, 2021] should be carefully examined and evaluated on modern and archaeological collections (for the archaeological specimens the identifications obtained via molecular analysis will be adopted for evaluation). Indeed, such approaches have a solid mathematical foundation, are invariant with respect to isometries

that naturally appear during the 3D acquisition and are extremely more parsimonious with respect to other deep learning techniques whose utilisation, as previously said, is limited by the relatively small size of our datasets. At PhD position offer 1 AI for archaeozoology the same time, we aim at extending existing works in AI [Diligenti et al., 2017, Marra et al., September 16–20, 2019] allowing one to inject the ontological knowledge of the experts from a domain into the learning procedure, either via the adoption of constrained loss functions or via knowledge-based active learning [Ciravegna et al., 2021].

Unsupervised learning strategies to study morphological evolution of caprine bones

During the last decades, archaeozoologists have been interested in the morphological and phenotypical diversity of mammals, particularly domestic and wild caprine (sheep, goats, mouflons, ibexes). Their interest is twofold: i) they wish to detect groups of morphologically similar bones, possibly belonging to a same species and ii) to investigate into possible correlations between the morphological features of some specimens and the ecosystem they lived in [Vila et al., 2021]. The investigation is usually made through morphometric studies such as geometric morphometric methods (GMM). Although GMM studies have produced promising results [e.g. Evin et al., 2013, Cucchi and Evin, 2015, Vuillien, 2020], methodological limitations are still present: lack of interpretative context, fragmented archaeological material, amount of 2D or 3D data to process, specific morphological pattern. The last point is of particular interest, because the shape clustering in GMM is performed by selecting ad-hoc points from a 3D point cloud (morphological patterns), based on the experts' knowledge. Then, tools like Procrustes analysis, linear dimensionality reduction (PCA) and standard clustering algorithms are employed. A first objective of our research will aim at better exploiting the hand-craft morphological patterns, by means of recent advances in Optimal Transport, in terms of mixed Procrustes-Wasserstein distances [Alvarez-Melis et al., 2019] and dictionary learning[Vincent-Cuaz et al., 2022b,a]. Later, wemight deal with a more ambitious aim which is to learn the morphological patterns. We remark the originality of this research project, since works combining ML and GMM are still rare in archaeozoology [Matthews et al., 2018, Courtenay et al., 2019]. The dataset under investigation will be made of hundreds of 3D astragalus bones from archaeological ibexes and sheep acquired during the research work conducted by the supervisors (M. Vuillien and E.Vila).

Expected scientific advances

The scientific advances linked to this PhD offer are twofold: first, we aim at developing new architectures or at least at originally extending existing ones for the analysis of the archaeozoological data. This is of interest for the AI community as dedicated approaches are required (see Section 1); second, we aim at building a bridge between two communities (AI and geometric morphometrics) by proposing new algorithms to integrate existing libraries (e.g. Morpho, CRAN R).

Offer

The 3-years PhD will take place at INRIA Universit´e Cˆote d'Azur, in Sophia Antipolis (team MAASAI). The PhD project is conducted in close collaboration with the archaeology laboratories of Cultures & Environments: Prehistory, Antiquity and Middle Ages (CEPAM, Universit´e Cˆote d'Azur, CNRS) in Nice and Arch´eorient (Universit´e Lyon II Lumi`ere, CNRS) in Lyon. The three institutions offer dynamic research environments, exhaustive training opportunities and institutional collaborations. The PhD candidate will benefit from the computational resources available at CEPAM (GPU servers). He/She will benefit from scientific supervision in archaeozoology and bioarchaeology by following update courses offered in the training Master "Sciences of the Earth and Planets, Environment" filed "Prehistory, Palaeoenvironment, Archaeosciences" of the CEPAM lab and will participate in methodological workshops in archaeozoology supported by the Research network BioArcheodat (supervised by E. Vila & A. Dufraisse).

Valorisation

The PhD candidate and project partners will all participate in the results diffusion through scientific publications in peer-reviewed journals and national and international conferences dealing with the two disciplinary fields (e.g. Pattern Recognition, IEEE Transactions on Neural Networks and Learning Systems, Journal of Archaeological Method and Theory, Journal of Archaeological Science, Scientific Reports). At the beginning of the thesis, the candidate will have the opportunity to discuss with international researchers working on similar issues by participating to the 1st International Conference on artificial Intelligence an applied Mathematics for History and Archaeology (IAMAHA) co-organised by CEPAM, Inria Côte d'Azur and I3A in Nice.

Candidate profile and skills

Degree: The candidate must hold a master's degree in mathematics, physics or informatics with a strong knowledge
in machine learning.

Skills: Coding in Python and/or R is required. Previous knowledge in archaeology and zoo-archaeology would be a
plus.

References

V. Ahedo, D. Zurro, J. Caro, and J. Gal´an. Let's go fishing: A quantitative analysis of subsistence choices with a special focus on mixed economies among small-scale societies. *PLoS ONE*, 16, 8:0254539, 2021.

D. Alvarez-Melis, S. Jegelka, and T. Jaakkola. Towards optimal transport with global invariances. In *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR, 2019.

Y. Assael, T. Sommerschield, B. Shillingford, M. Bordbar, J. Pavlopoulos, M. Chatzipanagiotou, I. Androutsopoulos, J. Prag, and N. Freitas. Restoring and attributing ancient texts using deep neural networks. *Nature*, 603, 7900:280–283, 2022.

R. Barone. *Anatomie comparée des mammifères domestiques. Ostéologie*. Vigot, deuxième édition revue et augmentée, Paris, 1976.

T. Baum, C. Nendel, S. Jacomet, M. Colobran, and R. Ebersbach. Slash and burn" or "weed and manure"? a modelling approach to explore hypotheses of late neolithic crop cultivation in pre-alpine wetland sites. *Vegetation History Archaeobotany*, 25, 6:611–27, 2016.

S. Bickler. Machine learning identification and classification of historic ceramics. In B. Petchey, K. Hil, S. Kinaston, and A. Kelly, editors, *Archaeology in New Zealand, New Zealand Archaeological Association*, volume 61, 2, page 20–32. 2018.

S. Bickler. Machine learning arrives in archaeology. *Advances in Archaeological Practice*, 9, 2:186–191, 2021.

F. Chazal and B. Michel. An introduction to topological data analysis: fundamental and practical aspects for data scientists. *Frontiers in artificial intelligence*, 4:667963, 2021.

G. Cifuentes-Alcobendas and M. Domínguez-Rodrigo. Deep learning and taphonomy: high accuracy in the classification of cut marks made on fleshed and defleshed bones using convolutional neural networks. *Scientific reports*, 9, 1:1–12, 2019.

G. Ciravegna, P. F., and M. Gori. Knowledge-driven active learning, 2021. arXiv preprint arXiv:2110.08265.

L. A. Courtenay, J. Yravedra, R. Huguet, J. Aramendi, M. À. Maté-González, D. González-Aguilera, and M. C. Arriaza. Combining machine learning algorithms and geometric morphometrics: a study of carnivore tooth marks. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 522:28–39, 2019.

T. Cucchi and A. M. J. Evin. Morphométrie géométrique et archéozoologie: Concepts, méthodes et applications. In *Messages d'os Archéométrie du squelette animal et humain*, pages 197–216. Editions des archives contemporaines, 2015.

M. Diligenti, M. Gori, and C. Sacca. Semantic-based regularization for learning and inference. *Artificial Intelligence*, 244:143–165, 2017.

A. Driesch. Guide to the measurement of animal bones from archaeological sites. *Peabody Museum Bulletins*, 1:137, 1976.

A. Evin, T. Cucchi, A. Cardini, U. S. Vidarsdottir, G. Larson, and K. Dobney. The long and winding road: identifying pig domestication through molar size and shape. *Journal of Archaeological Science*, 40(1):735–743, 2013.

S. Horache, J. Deschaud, F. Goulette, K. Gruel, T. Lejars, and O. Masson. *Riedones3D: a celtic coin dataset for registration and fine-grained clustering*. Eurographics Workshop on Graphics and Cultural Heritage, 2021. doi: 10.2312/gch.20211410. URL https://doi.org/10.2312/gch.20211410.