



HAL
open science

ArchéoBot : Vers une pédagogie interactive et inclusive en archéologie grâce à l'IA générative

Guillaume Simiand, Vincenzo Capozzoli, Alain Duploux

► To cite this version:

Guillaume Simiand, Vincenzo Capozzoli, Alain Duploux. ArchéoBot : Vers une pédagogie interactive et inclusive en archéologie grâce à l'IA générative. Journées d'Études sur l'Intelligence Artificielle JEDDAME, Université Jean Jaurès Toulouse, Mar 2024, Toulouse, France. hal-04503459

HAL Id: hal-04503459

<https://hal.science/hal-04503459v1>

Submitted on 13 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

ArchéoBot : Vers une pédagogie interactive et inclusive en archéologie grâce à l'IA générative

Guillaume Simiand^{1,2}, Vincenzo Capozzoli^{1,3}, Alain Duplouy^{1,3}

¹ Université Paris 1 Panthéon-Sorbonne

² UR 4150 - IRJS : Institut de recherche juridique de la Sorbonne

³ UMR 8215 – Trajectoires. De la sédentarisation à l'état

L'avènement des technologies de l'intelligence artificielle, et plus particulièrement des grands modèles de langage, a ouvert de nouvelles perspectives dans le domaine de l'éducation. Le projet ArchéoBot, développé au sein de l'École d'histoire de l'art et de l'archéologie de la Sorbonne (Université Paris 1 Panthéon-Sorbonne) et destiné à accompagner les enseignants et les étudiants, de la salle de cours jusqu'à l'évaluation finale, illustre parfaitement cette révolution pédagogique.

ArchéoBot déploie des architectures récemment proposées pour surmonter les défis bien identifiés des modèles de langage à grande échelle (LLMs) dans leurs versions initiales, tels que les hallucinations et la concision excessive des réponses. En enrichissant un LLM (potentiellement open source) avec une architecture RAG (Retrieval Augmented Generation), soit une base de données vectorielles permettant d'augmenter le prompt initial avec des documents issus d'une bibliographie constituée par des experts, le dispositif assure une meilleure traçabilité des sources sollicitées. Ce processus remet ainsi l'importance de la recherche documentaire au cœur de l'interrogation du LLM. En outre, pour des réponses plus complètes, ArchéoBot intègre une étape de décomposition de la question de type *skeleton of thought*, combinaison à notre connaissance inédite. Le souci de mettre en œuvre à chaque étape des logiciels open source, pour une adoption durable et inclusive, appuie une démarche rigoureuse vis-à-vis de la gestion des données pédagogiques.

Spécifiquement conçu pour l'enseignement de l'archéologie et l'exploration de ses méthodes, ArchéoBot se distingue par sa capacité à fournir des réponses précises et contextuelles, grâce à une base de données vectorielle robuste alimentée par des contenus pédagogiques vérifiés, incluant des supports multimédias et textuels et garantissant la fiabilité des réponses fournies. Des fonctionnalités de gestion des erreurs sont également implémentées pour minimiser les hallucinations.

Une des innovations majeures d'ArchéoBot réside dans sa capacité à soutenir les étudiants dans la prise de notes. Reconnaisant les défis associés à la documentation des cours, en particulier pour ceux qui éprouvent des difficultés à capturer les points clés en temps réel, ArchéoBot offre une solution interactive pour vérifier et approfondir les notes prises, facilitant ainsi une meilleure compréhension et une correcte transcription de l'information. En interrogeant ArchéoBot, les étudiants pourront ainsi construire et étoffer leur propre corpus de connaissances, mais aussi développer l'apprentissage autonome et la réflexion critique, en s'offrant une expérience interactive proche d'une conversation avec un enseignant.

Cette contribution explore les potentialités de l'intelligence artificielle générative pour renforcer et enrichir l'enseignement et l'apprentissage en archéologie. On présentera les étapes clés du projet : rédaction des contenus, vectorisation et stockage dans une base de données, développement

du chatbot, intégration dans l'environnement numérique de cours. Nous aborderons également la complémentarité assumée entre ArchéoBot et l'enseignant ainsi que les précautions prises pour protéger les ressources pédagogiques.

En particulier, après avoir présenté les origines et les raisons principales du projet, nous aborderons la conception et l'implémentation d'ArchéoBot, mettant en lumière les défis techniques et éthiques surmontés pour réaliser cette plateforme d'apprentissage innovante et nous focalisant sur la construction et l'organisation d'une base de connaissances dédiée, les méthodologies d'interrogation et de vectorisation des informations textuelles, et les défis spécifiques tels que la gestion des images en archéologie, la fiabilité des sources et la réduction des hallucinations de l'IA. Cette présentation détaillera par ailleurs les raisons derrière le choix de LangChain, mettant en avant sa capacité à faciliter des adaptations rapides aux modèles de langage, essentielles pour répondre aux besoins évolutifs en termes de contenu pédagogique. Nous soulignerons également l'importance de la transition vers l'utilisation de modèles de langage open source, reflétant notre engagement envers une éducation accessible et éthiquement responsable, tout en garantissant la pérennité du projet.

En outre, nous aborderons la gestion et la protection des données au sein d'ArchéoBot. L'utilisation de LangChain permet une exécution locale et contrôlée des processus d'IA, offrant une garantie supplémentaire contre la réutilisation non autorisée des données pour le réentraînement des modèles de langage. Cette approche souligne notre détermination à préserver l'intégrité académique et la confidentialité des ressources pédagogiques, tout en naviguant dans le paysage complexe des considérations éthiques et des coûts associés à l'IA générative.

La présentation envisagera ensuite les défis et les perspectives futures d'ArchéoBot dans le contexte de l'enseignement en général et en l'occurrence de l'archéologie, mettant en lumière comment cette initiative s'inscrit dans une révolution intellectuelle et technologique plus large. Elle invitera à une réflexion sur le potentiel transformatif de l'IA dans l'éducation, en abordant les implications pratiques, éthiques et sociétales de son intégration dans les processus d'enseignement, d'apprentissage et de recherche. En offrant une aide personnalisée à la prise de notes et donc un apprentissage personnalisé et immersif, ArchéoBot représente un pas en avant vers une pédagogie plus interactive et inclusive, reflétant notre engagement envers l'excellence éducative et l'accessibilité.

En outre, cette communication interrogera plus fondamentalement la manière dont l'IA générative redessine les modalités de la formation et de l'évaluation universitaires. Plutôt que de se cantonner à un rôle d'assistance à l'enseignement traditionnel, ArchéoBot nous permet d'esquisser de nouveaux paradigmes pédagogiques. A terme, il permettra par exemple la génération automatique de questions de quiz et de QCM, mais aussi de sujets et de commentaires pour les devoirs ouverts des étudiants. L'IA générative ouvre ainsi la voie à une implication active des chatbots dans les processus de formation et d'évaluation.

Enfin, l'exemple d'ArchéoBot constitue un tremplin vers une réflexion élargie sur le rôle transformateur des grands modèles de langage dans l'enrichissement de l'expérience éducative au sein de l'enseignement supérieur. Cette analyse invite à explorer les stratégies d'intégration de ces technologies avancées au cœur des SHS, engageant un dialogue entre les « sciences molles » et les « sciences dures », inéluctablement destiné à redéfinir profondément nos approches pédagogiques dans les années à venir.