



HAL
open science

ArchéoBot : une IA conversationnelle pour l'enseignement de l'archéologie

Vincenzo Capozzoli, Alain Duploux, Guillaume Simiand

► **To cite this version:**

Vincenzo Capozzoli, Alain Duploux, Guillaume Simiand. ArchéoBot : une IA conversationnelle pour l'enseignement de l'archéologie. Culture et recherche, 2024, 147, pp.106-109. hal-04794139

HAL Id: hal-04794139

<https://hal.science/hal-04794139v1>

Submitted on 20 Nov 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.

Copyright



Recherche et intelligence artificielle

ArchéoBot :

une IA conversationnelle pour l'enseignement de l'archéologie

Le contexte de transformation numérique accélérée, de croissance et de diversification des effectifs étudiants à l'université, d'augmentation significative des publications scientifiques (le plus souvent en format numérique) et d'émergence de *chatbots* d'intelligence artificielle générative depuis novembre 2022 incite à repenser la pédagogie universitaire, notamment dans le domaine des sciences humaines et sociales.

VINCENZO CAPOZZOLI

Maître de conférences, archéologue,
Université Paris 1 Panthéon-Sorbonne –
UMR CNRS 8215 Trajectoires.
De la sédentarisation à l'État

ALAIN DUPLOUY

Maître de conférences, HDR,
archéologue, Université Paris 1
Panthéon-Sorbonne – UMR CNRS 8215
Trajectoires. De la sédentarisation
à l'État

GUILLAUME SIMIAND

Chercheur en humanités numériques,
Université Paris 1 Panthéon-Sorbonne,
Institut de recherche juridique
de la Sorbonne (IRJS)

Dans cette optique, en septembre 2023, répondant à un appel à projets pédagogiques numériques de l'Université Paris 1 Panthéon-Sorbonne, une équipe interdisciplinaire composée de deux archéologues et d'un expert en humanités numériques a lancé le projet ArchéoBot. Nous avons ainsi décidé de tirer parti des dernières avancées des IA conversationnelles tout en jugulant leurs limites, en particulier les « hallucinations¹ », la brièveté parfois excessive et décontextualisée des textes générés et, surtout, l'absence de traçabilité des sources et de transparence dans l'utilisation en aval des données disponibles dans la base de connaissance, deux problèmes rédhibitoires pour un usage pédagogique. Spécialement conçu pour le domaine des méthodes de l'archéologie, ArchéoBot a d'abord été pensé comme un moyen d'enrichir l'offre de formation en salle de classe et d'accompagner les étudiants, tant en licence qu'en master, dans leurs processus d'apprentissage et de révision.

Méthodes de l'archéologie et pédagogie universitaire au XXI^e siècle : la contribution d'ArchéoBot

L'enseignement des méthodes de l'archéologie fait face à plusieurs défis majeurs. D'une part, le domaine repose sur un vaste corpus de connaissances et de données empiriques, souvent dispersé dans de nombreuses publications spécialisées en plusieurs langues, ce qui peut rendre l'accès à l'information difficile pour les étudiants, surtout s'ils sont peu familiers avec le travail de recherche bibliographique. D'autre part, la nature même de la discipline, qui combine des aspects théoriques et pratiques, nécessite une approche pédagogique capable de transmettre non seulement des connaissances, mais aussi des méthodes d'analyse et d'interprétation. La diffusion de plus en plus large d'outils d'IA générative, mal adaptés au travail scientifique, peut en outre induire les étudiants en erreur et les pousser à acquérir de mauvaises habitudes du point de vue de la fiabilité des sources à un moment

où celle-ci est plus importante que jamais. ArchéoBot se penche ainsi sur les questions cruciales du contexte et de la pratique discursive des LLM disponibles, qui véhiculent trop souvent l'illusion d'une pensée structurée et objective. Face à une science contextuelle comme l'archéologie, ces limites sont particulièrement problématiques.

Avec notre outil, si un étudiant pose la question suivante : « Quelles sont les principales méthodes de prospection en archéologie ? », ArchéoBot fournit une réponse structurée, expliquant les différentes méthodes et technologies employées, mais surtout, il cite des sources précises issues de sa base de connaissances vectorielle (manuels de référence ou articles scientifiques récents).

La structure de l'outil

ArchéoBot s'appuie sur la combinaison innovante de trois éléments clés, à savoir :

1. Les architectures RAG (*Retrieval-augmented generation*), qui associent à un modèle de langue avancé une base de données vectorielle. Cette dernière permet de représenter sous forme numérique une approximation du contenu sémantique des documents, ce qui améliore grandement les capacités de recherche du système en allant au-delà des approches basées uniquement sur des mots-clés.
2. Une décomposition de la question initiale (*skeleton of thought*), qui vise à pallier certains défauts des grands modèles de langue, notamment l'imprécision et la brièveté des réponses, en segmentant la tâche initiale et en parallélisant le traitement des fragments.
3. L'intégration de Zotero², outil libre de gestion bibliographique largement adopté dans le milieu académique, qui permet aux utilisateurs de se concentrer sur la construction d'une bibliographie qualitative, sans se soucier de la gestion de la base de données.

1. L'habitude a été prise de nommer ainsi des productions de texte vraisemblables, mais entièrement erronées. Leur présence ne peut être entièrement éradiquée des grands modèles de langue actuels (*large language models*, LLM), mais différents dispositifs permettent de les limiter.

2. Pour Zotero, voir les articles : « Ouvrir les données de la Bibliothèque nationale de France à la recherche », *Culture et Recherche*, n° 144, printemps-été 2023, p. 30-33 ; Laurent Jonchère et Olivier Troccaz, « HALUR : des outils connectés à HAL au service des chercheurs », *Culture et Recherche*, n° 144, printemps-été 2023, p. 40-41.

×

Pour contribuer à l'amélioration du modèle, merci de laisser un feedback à la fin de la session!

Nouveau chat

Modèle

gpt-4-turbo

Nombre de documents convoqués par étape

10

Collections interrogées

Toutes collections

Archéobot

Agent conversationnel | méthodes en archéologie

Peux-tu me fournir une explication détaillée des différentes méthodes de prospection en archéologie ? J'aimerais avoir une définition du terme "prospection" et par la suite comprendre les différentes techniques adoptées par les archéologues (pédestres, aériennes et géophysiques).

Plan de la réponse:

1. Définir la "prospection" en archéologie et son objectif principal.
2. Expliquer les méthodes de prospection par observation sans contact direct.
3. Détailler les techniques de prospection pédestre et leur mise en œuvre.
4. Présenter les principes et outils de la prospection aérienne.
5. Exposer les méthodes de prospection géophysique et leurs applications.
6. Mentionner les technologies spécifiques comme le LIDAR et leur utilité.
7. Discuter des avantages et des limites de chaque méthode de prospection.
8. Illustrer l'importance de l'intégration des données dans un SIG pour l'analyse.

Sources consultées

- Dabas 2018 = M. Dabas, *L'archéo-géophysique ou la détection et la cartographie des sites archéologiques*, in *Annales des Rencontres archéologiques de Saint-Céré*, 25, 2018, p. 117-124, <https://hal.science/hal-02188753>. [\[lien\]](#)
- Djindjian 2011 = F. Djindjian, *Chapitre 2 - Prévention et prospection archéologiques*, *Manuel d'archéologie*, Paris, 2011 (Collection U), p. 123-139. [\[lien\]](#)
- Djindjian 2017 = F. Djindjian, *Chapitre 2. Prévention et prospection archéologiques*, *L'archéologie*, Paris, 2017 (Hors collection), p. 131-147. [\[lien\]](#)
- Lehoërf - Giligny 2020 = A. Lehoërf, F. Giligny, *2 / Le travail de terrain*, Paris, 2020 (*Repères*), p. 40-98. [\[lien\]](#)
- Tabbagh 2018 = A. Tabbagh, *La prospection : évolution de la sous-discipline, évolution du métier*, in *ArcheoSciences. Revue d'archéométrie*, 42, 2018, p. 103-108, <https://journals.openedition.org/archeosciences/5440>. [\[lien\]](#)

« Des réponses structurées... »
(capture d'écran des auteurs).

Une interface versatile et transparente
Initialement conçu comme un simple agent conversationnel, ArchéoBot a rapidement transcendé cette fonction de base. Aujourd'hui, il intègre une base de données spécialisée dans plusieurs domaines de l'archéologie, contenant 900 documents

soigneusement sélectionnés pour leur qualité et pertinence, cumulant environ 90 millions de caractères. Cette base inclut des cours sous forme textuelle ou vidéo, des articles scientifiques, des ouvrages de référence, des *podcasts*, ainsi que des archives numérisées. Chaque texte généré par ArchéoBot est accompagné

« ... appuyées sur des données aux sources clairement identifiées »
(capture d'écran des auteurs).

1. Définir la "prospection" en archéologie et son objectif principal.

La prospection archéologique, telle que définie dans les travaux de Djindjian (2017), englobe l'ensemble des méthodes et des techniques utilisées pour acquérir des informations sur l'occupation humaine d'un espace donné sans perturber le sol ou les structures présentes. Cette démarche se distingue de la fouille archéologique par son caractère non invasif et par l'étendue souvent vaste du terrain investigué (Tabbagh, 2018). L'objectif principal de la prospection est de localiser et de documenter les sites archéologiques, de cartographier les différentes occupations de l'espace à travers le temps, et d'évaluer des paramètres tels que la date, la surface, l'environnement et la démographie des sites archéologiques (Djindjian, 2011).

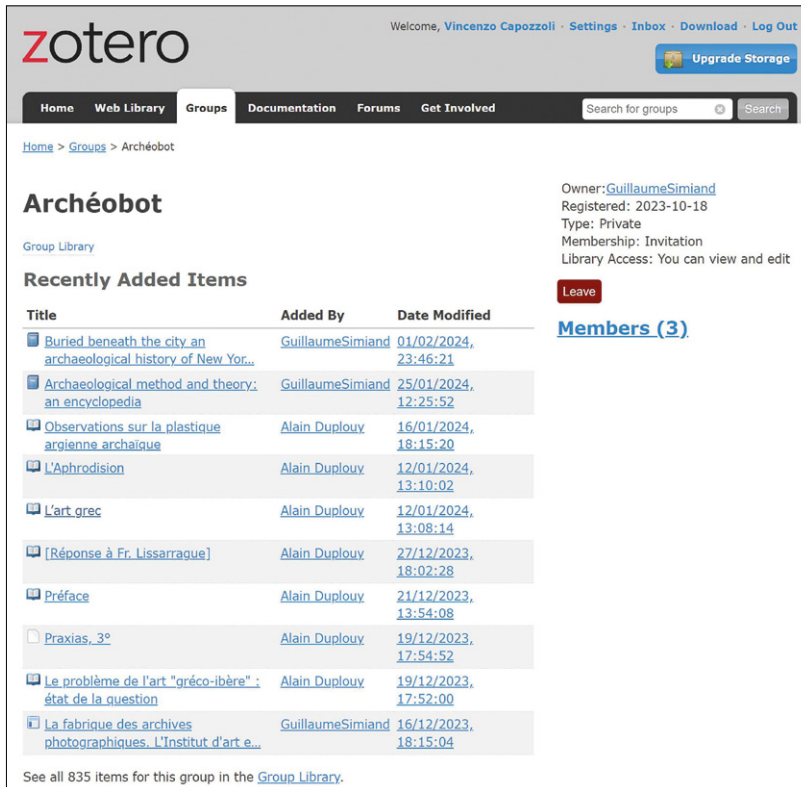
Les techniques employées varient grandement et peuvent inclure la prospection pédestre, où les chercheurs parcourent méthodiquement le terrain à la recherche de vestiges visibles à la surface, la prospection aérienne qui utilise des photographies prises depuis le ciel pour identifier des anomalies indiquant la présence de structures enfouies, et la prospection géophysique qui fait appel à des instruments mesurant les propriétés physiques du sous-sol pour détecter des anomalies (Djindjian, 2017). Ces méthodes peuvent être complétées par l'analyse de documents historiques, l'examen de toponymes, et l'utilisation de technologies avancées comme le LIDAR, qui permet de détecter des variations minimales dans la topographie d'un terrain, souvent masquées par la végétation ou d'autres éléments (Lehoërf, 2022).

La prospection archéologique est donc une étape cruciale dans la recherche archéologique, permettant de maximiser l'efficacité des fouilles tout en minimisant les perturbations des sites, et de fournir une vue d'ensemble des patterns d'occupation humaine sur de vastes zones, ce qui est essentiel pour la planification de recherches futures et la conservation du patrimoine archéologique.

Sources consultées pour cette section

- Djindjian 2011 = F. Djindjian, *Chapitre 1 - Les conditions de conservation et de découverte des sites et des objets archéologiques*, *Manuel d'archéologie*, Paris, 2011 (Collection U), p. 104-122. [\[lien\]](#)
- Djindjian 2011 = F. Djindjian, *Chapitre 2 - Prévention et prospection archéologiques*, *Manuel d'archéologie*, Paris, 2011 (Collection U), p. 123-139. [\[lien\]](#)
- Djindjian 2017 = F. Djindjian, *Chapitre 1. Les conditions de conservation et de découverte des sites et des objets archéologiques*, *L'archéologie*, Paris, 2017 (Hors collection), p. 111-130. [\[lien\]](#)
- Djindjian 2017 = F. Djindjian, *Chapitre 2. Prévention et prospection archéologiques*, *L'archéologie*, Paris, 2017 (Hors collection), p. 131-147. [\[lien\]](#)
- Lehoërf 2022 = A. Lehoërf, *Chapitre IV. Une minutieuse et ambitieuse enquête*, Paris cedex 14, 2022 (*Que sais-je ?*), p. 76-119. [\[lien\]](#)
- Tabbagh 2018 = A. Tabbagh, *La prospection : évolution de la sous-discipline, évolution du métier*, in *ArcheoSciences. Revue d'archéométrie*, 42, 2018, p. 103-108, <https://journals.openedition.org/archeosciences/5440>. [\[lien\]](#)

CULTURE ET RECHERCHE n° 147 automne-hiver 2024 Recherche et intelligence artificielle 107



« Le recours à Zotero pour gérer la base de données bibliographiques allège le travail des experts de la discipline » (capture d'écran des auteurs).

de sa bibliographie. Cette dernière, qui pointe vers les documents source *via* des liens directs vers des bases externes lorsqu'elles sont disponibles, permet aux étudiants et aux enseignants de vérifier et d'approfondir les informations fournies. Cette « auditabilité » est au cœur de la logique de développement d'ArchéoBot. Enfin, ArchéoBot offre à ses utilisateurs plusieurs paramètres personnalisables, incluant le choix du LLM (avec des options telles que les GPT d'OpenAI, les modèles de MistralAI, et bientôt Llama 3 et Claude) ainsi que le nombre de documents à intégrer dans le système RAG. Grâce à cette approche, ArchéoBot place autant que possible son utilisateur aux commandes et tourne le dos à la logique de « boîte noire » trop souvent présente dans les interfaces d'IA générative.

Perspectives

ArchéoBot est actuellement en phase de test et son développement se poursuit activement. Outre l'enrichissement de sa base bibliographique, il supportera de nouveaux modèles de langage, comme Llama 3,

dont le caractère partiellement ouvert assurera une mise en œuvre sans que les données ne sortent vers des systèmes externes. Une autre perspective de développement pour ArchéoBot sera sa mise en compatibilité avec les modèles multimodaux, notamment ceux consacrés à l'image, laquelle joue un rôle essentiel dans la démarche scientifique en archéologie : schémas, plans topographiques et dessins de mobilier. Cette intégration permettra également de mieux gérer les sources documentaires anciennes grâce à une gestion plus fine des opérations de numérisation. L'amélioration des techniques de segmentation des documents facilitera par ailleurs une meilleure préservation du contexte dans les publications scientifiques longues, repoussant ainsi les limites du système. Enfin, l'optimisation de l'architecture contribuera à limiter les coûts de fonctionnement et la charge de calcul, en réduisant l'empreinte environnementale du projet et en le rendant accessible à un plus grand nombre d'étudiants et de chercheurs.

Au-delà des aspects techniques, l'éventuelle intégration d'ArchéoBot dans l'enseignement soulève des questions plus larges sur la transformation de la pédagogie et des pratiques universitaires à l'ère des IA génératives. Dans cet univers à l'évolution extrêmement rapide, l'utilisation de tels systèmes soulève de nombreuses questions éthiques et épistémologiques. Comment un tel outil modifie-t-il la relation entre l'enseignant et l'étudiant, selon que l'utilisateur soit l'un ou l'autre ? Comment garantir que les étudiants développent une pensée critique et ne deviennent pas dépendants de l'outil ? Comment ArchéoBot pourra-t-il maximiser le signal dans les textes qu'il génère et minimiser le bruit à faible valeur ajoutée ?

La construction de nouveaux savoirs n'est pas un processus anodin. Cependant, nous croyons qu'ArchéoBot, en s'appuyant sur une base de données scientifique vérifiable et en se préparant à intégrer des modèles multimodaux, participera inévitablement à une rupture dans la production des savoirs déjà amorcée par les IA génératives. En établissant des connexions inattendues entre ses sources et en obligeant enseignants et étudiants à une confrontation constante, ArchéoBot finira par entraîner des changements significatifs tant dans l'enseignement de l'archéologie que dans la manière dont les étudiants perçoivent cette discipline, ouvrant ainsi la voie à des nouvelles formes de pédagogie et, *in fine*, à un changement épistémologique majeur de cette science éminemment humaine. ■

Bibliographie et sitographie

Vincenzo Capozzoli, Alain Duploux, François Giligny et Guillaume Simiand, « ArchéoBot, le chatbot », *La collection numérique*, n° 30, 2023, p. 48-49.

« ArchéoBot : un compagnon intelligent et vérifié pour explorer l'archéologie et ses méthodes avec LangChain », *Observatoire IA*, 3 décembre 2023 (<https://observatoire-ia.pantheonsorbonne.fr/actualite/archeobot-compagnon-intelligent-et-verifie-explorer-larcheologie-et-ses-methodes>, site consulté le 8 juillet 2024).

« Peut-on réviser les cours avec ArchéoBot ? », *SUN Paris 1*, 4 mars 2024 (<https://pedagogie-numerique.pantheonsorbonne.fr/actualite/peut-reviser-cours-archeo-bot>, site consulté le 8 juillet 2024).

Guillaume Simiand, Vincenzo Capozzoli et Alain Duploux, « ArchéoBot : vers une pédagogie interactive et inclusive en archéologie grâce à l'IA générative », conférence tenue lors des Journées d'études sur l'intelligence artificielle (JEDDAME), Toulouse, 14 et 15 mars 2024 (<https://hal.science/hal-04503459>, site consulté le 12 juillet 2024).

Page de droite : « L'interface est compatible avec Moodle, le LMS le plus utilisé dans l'enseignement supérieur en France » (capture d'écran des auteurs).

✕

Pour contribuer à l'amélioration du modèle, merci de laisser un feedback à la fin de la session!

Nouveau chat

Modèle

gpt-4-turbo

Nombre de documents convoqués par étape

10

Collections interrogées

Toutes collections ✕

Archéobot

Agent conversationnel | méthodes en archéologie

Posez-moi une question relative à l'archéologie. ➤

✕

Pour contribuer à l'amélioration du modèle, merci de laisser un feedback à la fin de la session!

Nouveau chat

Modèle

gpt-4-turbo

Nombre de documents convoqués par étape

10

Collections interrogées

Toutes collections ✕

Archéobot

Agent conversationnel | méthodes en archéologie

Modèle

- gpt-4-turbo
- mistral-medium-latest
- mixtral8x7b
- mistral-large-latest

Nombre de documents convoqués par étape

- 10
- 5
- 8
- 10
- 15
- 20
- 30
- 40
- 50