
De #Cicero à IAristote : l'analyse du discours politique assistée par chatbot

Julien Longhi^{*1,2,3} and Boris Borzic⁴

¹Institut des humanités numériques – Université de Cergy Pontoise – France

²Laboratoire AGORA – Université de Cergy Pontoise : EA7392 – France

³CY Cergy Paris Université – Université de Cergy Pontoise – France

⁴Equipes Traitement de l'Information et Systèmes – UCP, ENSEA, CNRS – France

Résumé

Cette démonstration montrera comment l'intégration d'un chatbot vidéo (qui s'inscrit dans le domaine du Retrieval-augmented generation – RAG - for Videos (1)) couplé à une IA générative a permis de déployer de nouvelles fonctionnalités et de poser de nouvelles questions de valorisation de recherche en lien avec la plateforme #Cicero (2). En effet, lancé en 2019, le projet #Cicero répondait au besoin sociétal d'une meilleure compréhension des événements politiques, grâce à la production d'une plateforme en ligne d'analyse des messages produits par des comptes sélectionnés (expérimenté dans le cadre des élections européennes 2019). L'originalité de cette plateforme est que l'analyse des messages était complétée par une prise en compte des contenus vidéo, peu analysés par les outils traditionnels. En nous appuyant sur l'analyse automatisée des sous-titres, nous proposons un couplage inédit entre analyse textuelle et analyse vidéo/image. Ce projet avait été présenté au Salon Innovatives 2019 du CNRS (3) et avait notamment permis de développer un moteur de recherche vidéo capable de traiter les prises de parole des candidats aux élections européennes dans les vidéos partagées en ligne. L'idée était de procéder à des requêtes dans cet ensemble, et de produire, grâce à des outils numériques, des résultats sur ce corpus (fréquences de termes, réseaux de mots, usages spécifiques selon les personnalités, etc.).

Grâce à l'évolution de l'intelligence artificielle, une nouvelle version de #Cicero, appelée IAristote, intègre un LLM, qui va décomposer la structure des contenus, en s'appuyant notamment sur les connecteurs logiques, et générer ensuite des questions automatiques. Ainsi, l'utilisateur peut utiliser les suggestions du chatbot, du type : Qu'est-ce que (concept, terme, sujet) ? / Pouvez-vous expliquer (concept, processus, principe) ? / Comment définiriez-vous (terme) ? / Pouvez-vous donner une définition de (concept) ? / Pouvez-vous citer un exemple de (concept, principe) ? Il accèdera ainsi aux réponses fournies par les candidats dans le corpus, et pourra mesurer les proximités, les écarts, et les potentiels enjeux électoraux. Mais la demande peut être plus complexe, avec des questions comme : Quelle est la différence entre (concept A) et (concept B) ? / Quels sont les facteurs qui contribuent à (phénomène) ?

L'objectif de la démonstration de la plateforme sera alors de présenter l'outil développé (à la suite notamment de (4) pour une première version en ligne, et (5) pour son utilisation dans l'analyse linguistique de communautés), et illustrer le fait que l'intelligence artificielle permet de faire évoluer l'usage d'une plateforme interactive de navigation dans du contenu,

*Intervenant

en facilitant l'accès aux informations, et en proposant, via des échanges avec l'assistant conversationnel, l'identification de sujets d'intérêts, et l'accès à des informations complexes et pertinentes (esquissés dans (6) dans une moindre mesure).

Références & liens :

(1) Arefeen, M. A., Debnath, B., Uddin, M. Y. S., & Chakradhar, S. (2024). iRAG: An Incremental Retrieval Augmented Generation System for Videos. arXiv preprint arXiv:2404.12309. <https://arxiv.org/abs/2404.12309>

(2) #Cicero : une plateforme veut outiller les citoyens pour décrypter les discours politiques. <https://usbeketrica.com/fr/article/cicero-outiller-citoyen-decrypter-discours-politique>

(3) Innovatives SHS 2019. Projet CICERO. https://25images.msh-lse.fr/innovatives-shs_2019/video/projet-cicero/

(4) Marinica, C., Longhi, J., Hassine, N., Alkhouli, A., & Borzic, B. (2018, January). #Idéo2017: une plateforme citoyenne dédiée à l'analyse des tweets lors des événements politiques. In *Extraction et Gestion des Connaissances (EGC)*.

(5) Saigh, D., Borzic, B., Alkhouli, A., & Longhi, J. (2017). Contribution linguistique à une classification automatique des communautés de sens et à leur analyse. La controverse sur le statut des intermittents du spectacle. *Questions de communication*, (31), 161-182.

(6) Régionales 2021 : Franciliens, testez vos connaissances ! <https://theconversation.com/regionales-2021-franciliens-testez-vos-connaissances-162301>