



HAL
open science

Répondre aux requêtes des étudiants avec un agent conversationnel à mémoire supervisée

Florian Baud, Alex Aussem

► **To cite this version:**

Florian Baud, Alex Aussem. Répondre aux requêtes des étudiants avec un agent conversationnel à mémoire supervisée. Extraction et Gestion des Connaissances, Jan 2023, Lyon, France. hal-04249159

HAL Id: hal-04249159

<https://hal.science/hal-04249159>

Submitted on 20 Oct 2023

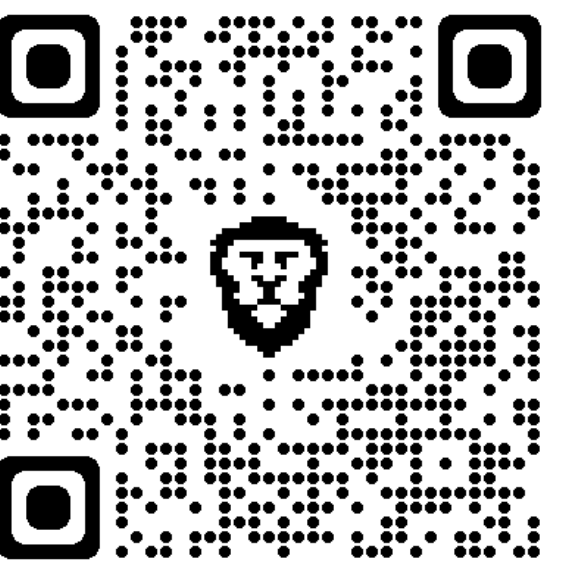
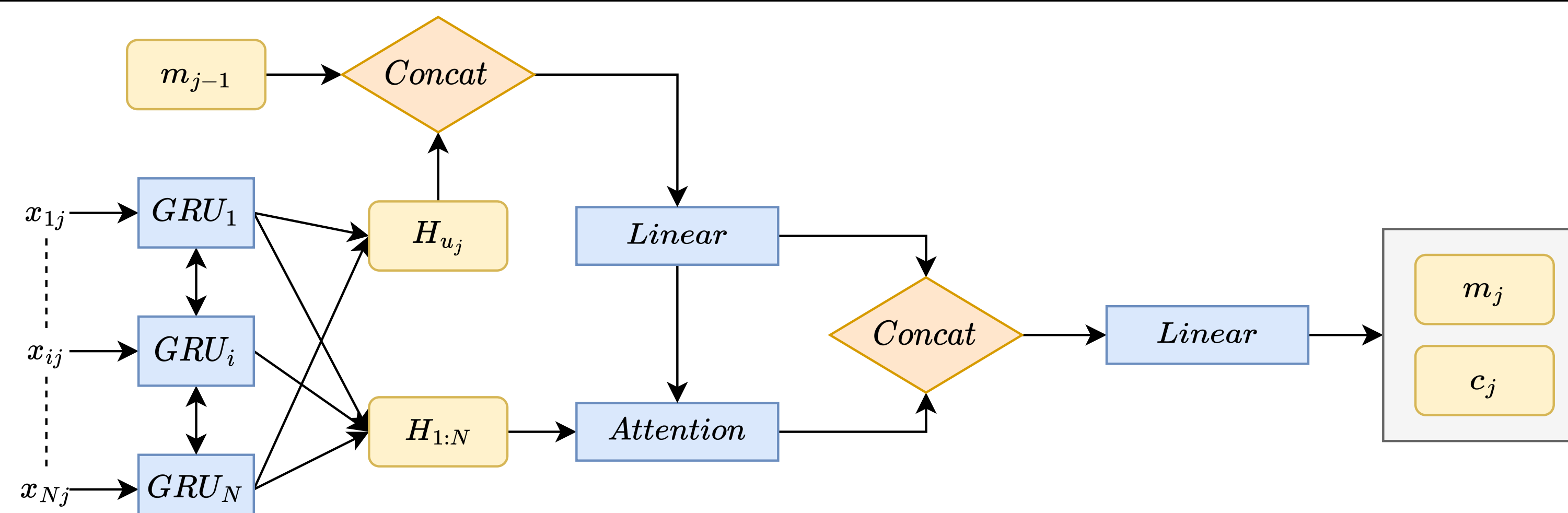
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

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.

Résumé

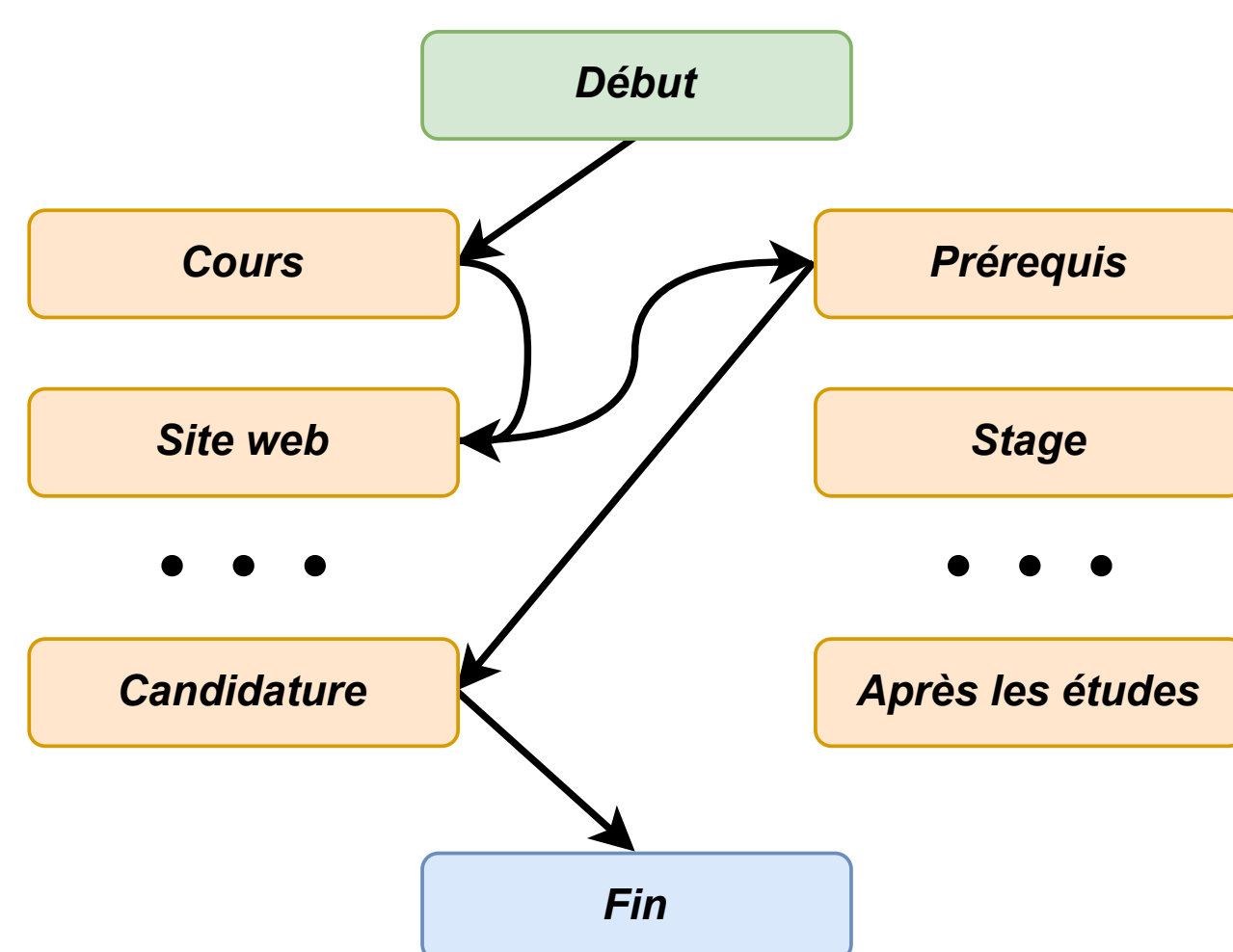
Nous présentons l'agent conversationnel (neuronal) qui a été développé à l'Université Lyon 1 pour répondre aux questions des candidats au Master Data Science. Basé sur une architecture Seq2Seq [3] combinée avec une mémoire supervisée, l'agent est capable d'identifier l'intention de l'utilisateur et d'encoder les informations pertinentes des échanges passés afin de produire une réponse personnalisée. Un générateur de dialogue a permis de constituer une base d'entraînement de 7500 dialogues synthétiques pour l'apprentissage des paramètres du modèle. L'agent conversationnel a été déployé et est accessible en ligne à l'adresse suivante : <http://chatbotinfo.univ-lyon1.fr/>.

L'architecture du modèle



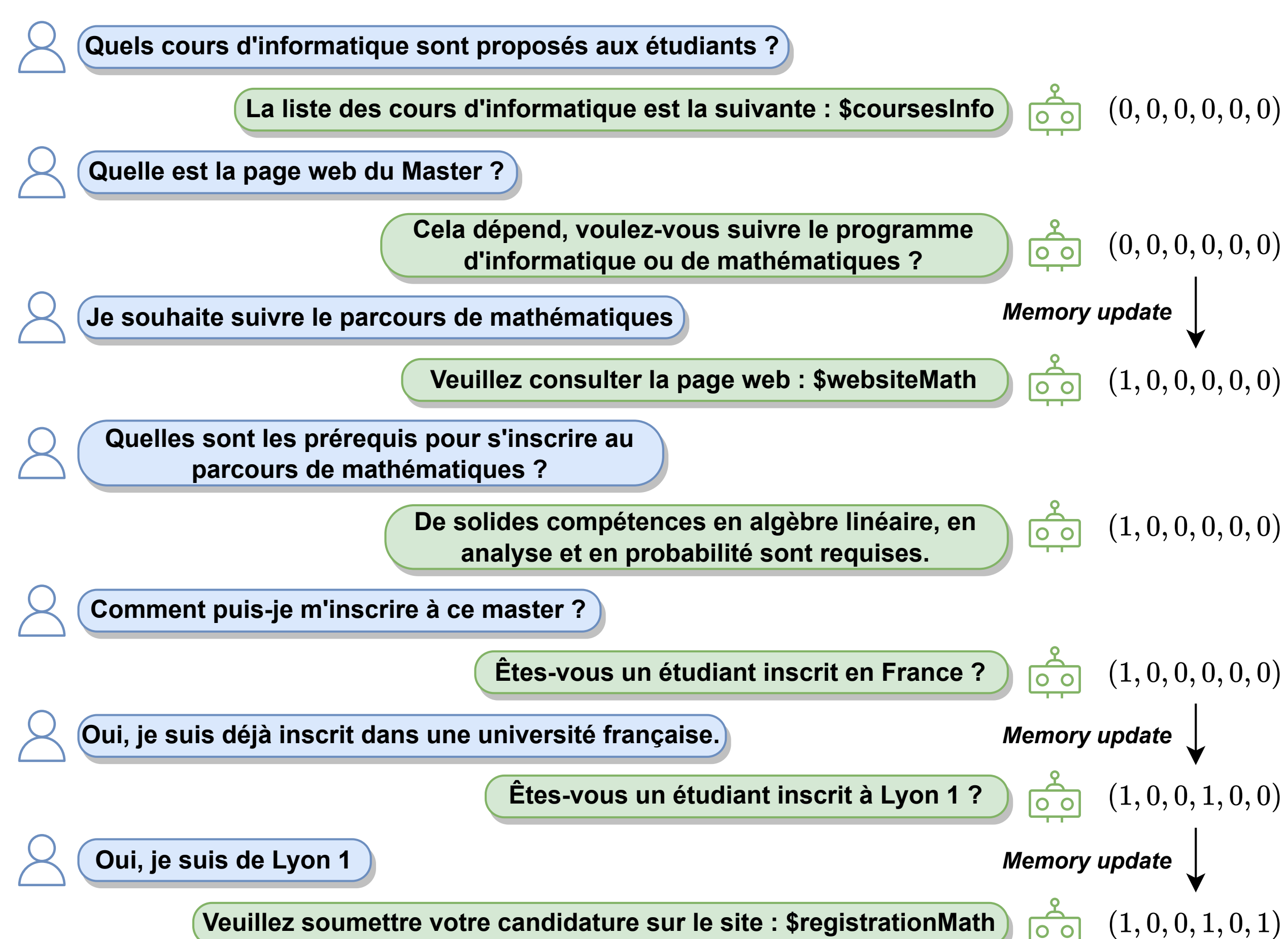
Inspiré d'un précédent travail [1], cet agent conversationnel est constitué d'un bi-GRU permettant d'encoder la phrase de l'utilisateur. Puis la mémoire m_{j-1} est incorporée avec un mécanisme d'attention [2] produisant une réponse c_j et un nouvel état de la mémoire m_j . Comparé à d'autres modèles, la mémoire est contrôlable et explicable. De plus, le modèle a très peu de paramètres surtout vis à vis des modèles de langage.

La génération de dialogue



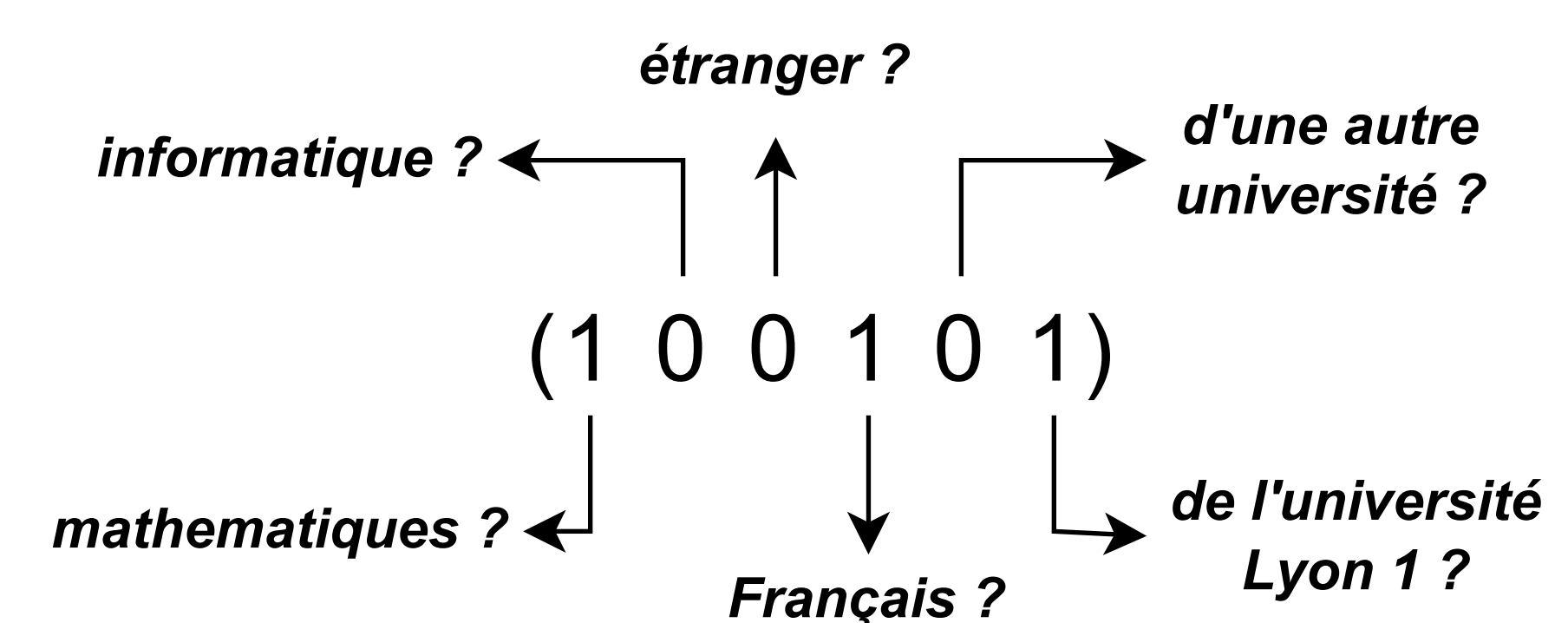
La génération d'un dialogue est une marche aléatoire sur différents thèmes préalablement définis. La mémoire est générée en même temps.

Exemple d'un dialogue généré



Ce dialogue est généré à partir d'une base de connaissances construite avec des données collectées auprès des étudiants du Master *Data Science*. Des variables sont présentes afin de modifier les informations qui changent au fil des années.

Représentation de la mémoire



La mémoire est un vecteur binaire qui encode le profil de l'étudiant. Grâce à ce profil extrait au fil de la conversation, l'agent répond de façon personnalisée aux questions des étudiants.

Limitations

Cet agent est "simple" et parcimonieux en nombre de paramètres (comparé à un modèle de langage), il présente toutefois quelques limites. Tout d'abord, l'encodeur est une architecture bi-GRU qui doit être entraînée à partir de zéro. Un modèle linguistique plus sophistiqué, couplé à la mémoire supervisée et adapté à notre tâche, permettrait certainement d'améliorer la compréhension des questions des étudiants. En outre, nos résultats sont plutôt mitigés : seulement 36,11% des réponses sont jugées correctes, 6,22% sont fausses, et dans 57,67% des cas, l'agent a demandé à l'utilisateur de reformuler la question.

Références

- [1] Jean-Baptiste Aujogue and Alex Aussem. Hierarchical recurrent attention networks for context-aware education chatbots. In *International Joint Conference on Neural Networks, IJCNN 2019 Budapest, Hungary, July 14-19, 2019*, pages 1-8. IEEE, 2019.
- [2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015. Conference Track Proceedings*, 2015.
- [3] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724-1734, Doha, Qatar, October 2014. Association for Computational Linguistics.