

[\[French version below\]](#)

## Thesis Proposal: Deep Learning and Knowledge Integration for Temporal Relations Extraction

### **SUPERVISORS:**

Anaïs Lefeuvre-Halftermeyer ([anaïs.halftermeyer@univ-orleans.fr](mailto:anaïs.halftermeyer@univ-orleans.fr)) LIFO, U. Orléans

Thi Bich Hanh Dao ([thi-bich-hanh.dao@univ-orleans.fr](mailto:thi-bich-hanh.dao@univ-orleans.fr)) LIFO, U. Orléans

### **PROFILE:**

Ideally, the recruited person will hold a **Master's degree in computer science** and have theoretical and practical knowledge in deep learning. An interest in language and its automatic processing would be appreciated but is not a prerequisite for recruitment. The candidate must demonstrate:

- Programming skills, such as proficiency in Python, for example
- Experience in Machine Learning, data mining, or applied mathematics
- Synthesis and writing skills allowing for clear and effective reporting of work done
- Ability to communicate in French or English, both orally and in writing

The recruited person will work at LIFO, University of Orléans (Campus de la Source, Orléans). They will be integrated into the Contraintes et Apprentissage team of LIFO(<https://www.univ-orleans.fr/lifo/equipes/CA/>).

The thesis will start in October 2024, and funding will last for three years.

An audition will take place before the MIPTIS doctoral school jury on June 12 to finalize the selection process.

### **REMUNERATION:**

Remuneration follows current legislation (2100 euros for gross salary), see <https://www.enseignementsup-recherche.gouv.fr/fr/le-financement-doctoral-46472>

### **APPLICATION SUBMISSION:**

Applications should be submitted by email until May 4, 2024, to Anaïs Halftermeyer ([anaïs.halftermeyer@univ-orleans.fr](mailto:anaïs.halftermeyer@univ-orleans.fr)) and Thi Bich Hanh Dao ([thi-bich-hanh.dao@univ-orleans.fr](mailto:thi-bich-hanh.dao@univ-orleans.fr)). Please submit in a single PDF:

- A detailed CV of your past activities (with possibly a list of publications)
- A letter presenting your motivations, qualifications, and experiences related to the topic
- University certificates and transcripts (both BSc and MSc degrees marks)
- Contact details or letters of recommendation (a minimum of 2 appreciated)
- Possibly an English language certificate
- Please note that all documents should be in English or French.

If necessary, a critical review of a scientific article may be requested for selection.

## SUBJECT DESCRIPTION:

We propose to work within the framework of temporal information extraction, which associates a synthetic representation of the events described in natural language text. A classical representation of such data is a graph of temporal relations between the events described and/or between temporal expressions [1].

Recent advances in deep learning in terms of language skills lead us to question human mastery over natural language processing tasks. These models have increasingly complex architectures and are increasingly demanding in terms of computing power and training data. However, they remain insufficient since general knowledge about temporal relations is not exploited to better guide and explain the results. In the context of this thesis topic, we propose to explore the integration of knowledge into a deep learning system, based on a language model, to solve temporal reasoning tasks.

A preliminary system [3] proposed to construct a temporal graph from medical texts by leveraging BERT, using rules in probabilistic logic during the model learning phase, as well as during the global inference phase. This hybrid work opened research avenues on the considerable contribution that temporal knowledge could represent through rule-based work. In order to make the systems more efficient, another study [4] proposed to successfully utilize syntactic analysis of inputs. In line with [2], we propose to leverage temporal knowledge representation to enhance system performance and explainability.

We are interested in integrating knowledge into these models to best solve temporal reasoning tasks, and this via constraint expression to:

- Leverage the best of both worlds, constraints, and language models acquired by deep learning
- Propose partly explainable hybrid models
- Base our systems on controlled computing power combined with a reproducible methodology of knowledge injection

Concretely, given a deep learning system based on a language model trained to translate text into a temporal graph representing the events narrated in the input text, injecting knowledge via constraint expression will modify the system's outputs. We aim to incrementally inject knowledge to guide our system while controlling:

- The size of our model
- The size of our training data
- The complexity of our constraints

## RÉFÉRENCES

- [1] T. Knez and S. Žitnik. *Event-centric temporal knowledge graph construction: A survey*. Mathematics, 11(23), 2023.
- [2] B. Zhang and L. Li. *Piper: A logic-driven deep contrastive optimization pipeline for event temporal reasoning*. Neural Networks, 164:186–202, 2023.
- [3] Y. Zhou, Y. Yan, R. Han, J. H. Caufield, K.-W. Chang, Y. Sun, P. Ping, and W. Wang. *Clinical temporal relation extraction with probabilistic soft logic regularization and global inference*. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 35, pages 14647–14655, 2021.
- [4] L. Zhuang, H. Fei, and P. Hu. *Syntax-based dynamic latent graph for event relation extraction*. Information Processing Management, 60(5):103469, 2023

## Proposition de thèse : Apprentissage profond et intégration de connaissances pour l'extraction de relations temporelles

### RESPONSABLES :

Anaïs Lefevre-Halftermeyer ([anaïs.halftermeyer@univ-orleans.fr](mailto:anaïs.halftermeyer@univ-orleans.fr)) LIFO, U. Orléans

Thi Bich Hanh Dao ([thi-bich-hanh.dao@univ-orleans.fr](mailto:thi-bich-hanh.dao@univ-orleans.fr)) LIFO, U. Orléans

### PROFIL RECHERCHÉ :

Idéalement, la personne recrutée sera titulaire d'un **Master en informatique**, et disposera de connaissances théoriques et pratiques sur les techniques par apprentissage profond. Un intérêt pour la langue et son traitement automatique serait apprécié, sans être un pré-requis à recrutement. Le candidat doit démontrer :

- Des compétences en programmation, telles que la maîtrise de Python, par exemple
- Une expérience en apprentissage automatique, en fouille de données ou en mathématiques appliquées
- Des compétences en synthèse et rédaction permettant de rendre compte clairement et efficacement du travail réalisé
- La capacité de communiquer en français ou en anglais, à l'oral et à l'écrit.

La personne recrutée travaillera au LIFO, Université d'Orléans (Campus de la Source, Orléans). Elle s'intégrera dans l'équipe Contraintes et Apprentissage (<https://www.univ-orleans.fr/lifo/equipes/CA/>) du LIFO.

Une audition aura lieu devant le jury de l'école doctorale MIPTIS le 12 juin pour finaliser le processus de sélection.

### RÉMUNÉRATION :

La rémunération suit la législation en vigueur (2100 euros brut) voir : <https://www.enseignementsup-recherche.gouv.fr/fr/le-financement-doctoral-46472>

### CONTACT - DÉPÔTS DE CANDIDATURE :

Dépôt des candidatures par courrier électronique auprès de mail : Thi Bich Hanh Dao ([thi-bich-hanh.dao@univ-orleans.fr](mailto:thi-bich-hanh.dao@univ-orleans.fr)) et Anaïs Lefevre-Halftermeyer ([anaïs.halftermeyer@univ-orleans.fr](mailto:anaïs.halftermeyer@univ-orleans.fr)). Veuillez soumettre en un seul PDF :

- Un CV détaillé de vos activités passées (éventuellement avec une liste de publications)
- Une lettre présentant vos motivations, qualifications et expériences en lien avec le sujet
- Certificats universitaires et relevés de notes (notes des diplômes de licence et de master)
- Références ou lettres de recommandation (un minimum de 2 serait apprécié)
- Éventuellement un certificat de langue anglaise
- Veuillez noter que tous les documents doivent être en anglais ou en français.

Le cas échéant, une lecture critique d'article scientifique pourra être demandée pour la sélection.

## CONTEXTE SCIENTIFIQUE :

Nous proposons de travailler dans le cadre de l'extraction d'informations temporelles qui associe à un texte en langue naturelle une représentation synthétique des événements qui y sont relatés. Une représentation classique de ce type de données est un graphe de relations temporelles entre les événements relatés et/ou entre expressions temporelles [1].

Les récentes avancées du deep learning en matière de compétences langagières nous amène à nous interroger sur la maîtrise de l'humain sur les processus de résolution de tâche reposant sur la langue naturelle. Ces modèles ont des architectures de plus en plus complexes et sont de plus en plus gourmands en puissance de calcul et en données d'entraînement. Cependant ils restent insuffisants puisque les connaissances générales sur des relations temporelles ne sont pas exploitées pour mieux orienter les résultats, ainsi que pour expliquer ces résultats.

Nous proposons, dans le cadre de ce sujet de thèse, d'explorer l'intégration de connaissances dans un système par apprentissage profond, reposant sur un modèle de langue, à résoudre des tâches de raisonnement temporel.

Un premier système [3] a proposé de construire un graphe temporel à partir de textes médicaux en s'appuyant de BERT, de la mise à profit de règles en logique probabiliste dans la phase d'apprentissage du modèle, ainsi qu'à la phase d'inférence globale. Ce travail hybride a ouvert des pistes de recherche sur l'apport considérable que pouvait représenter la connaissance propre à la temporalité via le travail sur les règles. Afin de rendre plus performant les systèmes, [4] a proposé de tirer partie de l'analyse syntaxique des entrées avec succès. Dans la même lignée de [2], nous proposons de tirer partie de la représentation de la connaissance temporelle comme levier de performance et d'explicabilité du système dans sa totalité.

Nous nous intéressons à intégrer de la connaissance dans ces modèles afin de résoudre au mieux les tâches de raisonnement temporel, et ce via l'expression de contraintes afin de :

- tirer profit du meilleur des deux mondes que sont les contraintes et les modèles de langues acquis par deep learning
- proposer des modèles hybrides en partie explicables
- faire reposer nos systèmes sur une puissance de calcul maîtrisée combinée à une méthodologie reproductible d'injection de connaissances

Concrètement, étant donné un système d'apprentissage profond reposant sur un modèle de langue entraîné à traduire un texte en un graphe temporel représentant les événements narrés dans le texte en entrée, l'injection de connaissances via l'expression de contraintes modifiera les sorties du système.

Nous cherchons à injecter des connaissances de manière incrémentale afin de guider notre système tout en maîtrisant :

- la taille de notre modèle
- la taille de nos données d'entraînement
- la complexité de nos contraintes

## RÉFÉRENCES :

- [1] T. Knez and S. Žitnik. *Event-centric temporal knowledge graph construction: A survey*. Mathematics, 11(23), 2023.
- [2] B. Zhang and L. Li. *Piper: A logic-driven deep contrastive optimization pipeline for event temporal reasoning*. Neural Networks, 164:186–202, 2023.
- [3] Y. Zhou, Y. Yan, R. Han, J. H. Caufield, K.-W. Chang, Y. Sun, P. Ping, and W. Wang. *Clinical temporal relation extraction with probabilistic soft logic regularization and global inference*. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 35, pages 14647–14655, 2021.
- [4] L. Zhuang, H. Fei, and P. Hu. *Syntax-based dynamic latent graph for event relation extraction*. Information Processing Management, 60(5):103469, 2023