

# Rapport d'activité 2022

Équipe PASTIS

## 1 Composition et domaines de recherche du groupe

Le groupe de recherche PASTIS est composé de 12 titulaires, 1 postdoc, 3 doctorants et 4 ATER :

- Adrien Revault d'Allonnes (MCF)
- Alice Millour (MCF - installée en septembre 2022)
- Anna Pappa (MCF)
- Benjamin Dupont (PRAG - rattachement en cours / installé en septembre 2022)
- Farès Belhadj (MCF)
- Françoise Balmas (MCF-HDR)
- Jean-Jacques Bourdin (PU)
- Nicolas Jouandeau (PU - responsable de l'équipe)
- Pablo Rauzy (MCF)
- Revekka Kyriakoglou (MCF)
- Sylvia Chalençon (MCF - mutation au CReSTIC/URCA/Reims en juin 2022)
- Vincent Boyer (MCF-HDR - en disponibilité)
- Sylvain Chichéry (Postdoc ANR ReComp 2022-2023)
- Jean-Pascal Palus (Doctorant 2021 - contrat doctoral ED CLI)
- Maroua Boudabous (Doctorante 2019 - contrat CIFRE avec la société NOVAGEN)
- Stephen Obonyo (Doctorant 2021 - financement Ministère des Affaires Etrangères et de l'Europe (MEAE) / Strathmore University, Nairobi, Kenya)
- Hanane Zerdoum (ATER 2021-2023)
- Oumaima El Joubari (ATER 2021-2023)
- Emna Chebbi (ATER 2021-2022)
- Syrine Saidi (ATER 2021-2022)

Le groupe possède des compétences en synthèse d'images expressives, en enseignement de l'informatique graphique, en résolution de jeux et apprentissage automatique, en représentation du raisonnement et logique non classique, en langage naturel et apprentissage automatique, en analyse de textes, en dynamique symbolique, en sécurité et privacy et en compréhension de programmes.

Jean-Pascal Palus est sous la direction de Adrien Revault d'Allonnes et Nicolas Jouandeau. Son sujet de thèse est « Dynamique et évolution de la confiance : explication et formalisation de la mésinformation ».

Maroua Boudabous est sous la direction de Anna Pappa et Françoise Balmas. Son sujet de thèse est « Modélisation d'un système intelligent d'analyse prédictive des données textuelles massives pour l'aide à la décision ».

Stephen Obonyo est sous la direction de Nicolas Jouandeau. Son sujet de thèse est « Bioinformatic Computation With Tree Search and Learning ».

## 2 Bilan en chiffres

Pour l'année 2022, le groupe PASTIS compte :

- publications en conférence et revues : 9
- soutenances de thèse et HDR : 0
- organisations/participations à des événements scientifiques : 9
- participations à des comités de programme/review : 19
- projets en cours : 1
- nouveaux projets déposés : 5
- implications dans les sociétés savantes, GdR et centres de recherche : 4

Le groupe est impliqué dans le Conseil d'administration de l'AFIG<sup>1</sup>, dirige un chapitre de IEEE CIS France<sup>2</sup>, est impliqué dans le centre de recherche Geopolitique de la datasphere (GEODE)<sup>3</sup>, est impliqué dans le Conseil enseignement de EUROGRAPHICS<sup>4</sup>.

## 3 Soutenances de thèse

(aucune)

## 4 Organisation et participation à des événements scientifiques

- Anna Pappa : poster, « Litte\_Bot : a theatrical dialogue system using Seq2Seq model », Journée AI4LAM Community en distanciel, 19/04/22.
- Anna Pappa : poster, « Litte\_Bot : Le bot qui donne la réplique dans le style Molière », Journée JAI2022 Accès interactif à l'information (AFIA-THL / ARIA / GDR TAL), Campus Pierre et Marie Curie, Paris, 05/05/22.
- Anna Pappa, Tristan Cazenave, Guillaume Grosjean et Baptiste Rozière : poster, « Un chatbot qui donne la réplique! », Atelier Affects, Compagnons Artificiels et Interactions (ACAI), AFIA, Ecole des Mines de Saint-Etienne, du 27/05/22 au 01/06/22.
- Guillaume Grosjean, Anna Pappa, Baptiste Rozière et Tristan Cazenave : poster, « Dialogue avec Molière », 29e conférence sur le Traitement Automatique des Langues Naturelles, TALN-RECITAL-ATALA, Avignon, du 27/06/22 au 01/07/22.
- Alice Millour et Laurent Kevers : poster, « Réalisations, obstacles et perspectives pour l'outillage du corse », Journées Jointes des Groupements de Recherche Linguistique Informatique, Formelle et de Terrain (LIFT) et Traitement Automatique des Langues (TAL), Marseille, 14-15/11/22.
- Maroua Boudabous et Anna Pappa : poster, « Apprentissage actif pour l'extraction des aspects explicites : application à des avis non annotés en français », Journées Jointes des Groupements de Recherche Linguistique Informatique, Formelle et de Terrain (LIFT) et Traitement Automatique des Langues (TAL), Marseille, 14-15/11/22.
- Jean-Jacques Bourdin : co-organisation de la session Education de la conférence EUROGRAPHICS, Université Reims Champagne-Ardenne, 26-29/04/2022.

---

1. <https://www.asso-afig.fr>

2. <http://ieee-ci.lip6.fr/>

3. <https://geode.science/>

4. <https://www.eg.org/>

- Jean-Jacques Bourdin : Animation de la table ronde sur « les outils (en ligne/hors ligne) utilisés pour illustrer les concepts de l'IG et la pédagogie associée », 35èmes journées de l'AFIG 2022, conférence jFIG, Université de Bordeaux, 23-25/11/22.
- Farès Belhadj : Présentation de GL4Dummies, 35èmes journées de l'AFIG 2022, conférence jFIG, Université de Bordeaux, 23-25/11/2022.

Une journée de formation doctorale intitulée « Que faire avec un doctorat en Informatique ? Valorisation des compétences acquises pendant le doctorat pour une multiplicité de carrières » organisée par la SIF<sup>5</sup> (contacts SIF local : Pablo Rauzy, organisation sur le site de l'IUT de Montreuil : Nédra Mellouli) a été proposée aux doctorants.

## 5 Participations à des comités de programme/review

- 19th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)
- 29e conférence sur le Traitement Automatique des Langues Naturelles (TALN)
- 31ème Rencontres francophones sur la logique floue et ses applications (LFA)
- 37th AAAI Conference on Artificial Intelligence (AAAI)
- 9th Games and Natural Language Processing (Games and NLP)
- Computer and Games conference (CG)
- Elsevier Journal of Chemometrics and Intelligent Laboratory Systems (Elsevier-CHEMOLAB)
- Elsevier Journal of Computers and Electronics in Agriculture (Elsevier-COMPAG)
- Elsevier Journal of Systems & Applications in Computer Graphics (Elsevier-CG)
- Elsevier Journal of Entertainment Computing (Elsevier-ENTCOM)
- Eurographics, the 43rd Annual Conference of the European Association for Computer Graphics (EG)
- European Chapter of the Association for Computational Linguistics (EACL)
- European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD)
- IEEE Conference on Games (IEEE-CoG)
- International Conference on Computational Linguistics (COLING)
- International Workshop on AI compliance mechanism (WAICOM)
- IOS Press Journal of International Computer Games Association (IOS-Press-ICGA)
- Premier Conference & Exhibition on Computer Graphics & Interactive Techniques (ACM-SIGGRAPH)
- Robustesse des systèmes de TAL (ROBUSTAL)

## 6 Projets en cours

L'équipe participe a un projet, le projet ReComp financé par l'ANR.

### 6.1 ANR ReComp « Research on Realtime Compliance Mechanism for AI »

**Porteur (à P8) :** Pablo Rauzy

**Partenaires :** Université Paris 8 (France), Sorbonne Université (France), Institut Fur Angewandte Informatik (Allemagne), National Institute of Informatics (Japon).

---

5. <https://www.societe-informatique-de-france.fr/les-journees-sif/journee-autour-du-doctorat-2022/>

**Date de début** : 03/2021

**Date de fin** : 02/2024

**URL** : <http://research.nii.ac.jp/RECOMP/>

**Travaux réalisés en 2022** : Recrutement de Sylvain Chichéry comme postdoc qui travaille sur l'implémentation de Capacity en programmation logique (Answer Set Programming) et a son intégration dans le cadre d'éthiques formelles du projet ReComp, en étroite collaboration avec l'équipe ACASA du LIP6.

## 7 Nouveaux projets déposés

### 7.1 Nouveaux projets académiques

**AAP Paris 8 « Graphes, Reconnaissabilité, Apprentissage et Symétries »**

**Porteurs** : Revekka Kyriakoglou et Nicolas Jouandeau

**Partenaires** : Université Paris 8 (France), Université Technique Tchèque à Prague (Tchéquie), Université Vrije à Amsterdam (Pays Bas)

**Date de début** : 01/2022

**Date de fin** : 12/2023

**URL** : <https://gras.up8.site/>

**Description** : Le projet propose d'étudier le lien entre jeux théoriques et jeux pratiques, au travers du transfert d'apprentissage. D'un point de vue théorique, le projet vise à étudier des jeux combinatoires, à énoncer des propriétés et à fournir des preuves de ces propriétés. D'un point de vue pratique, le projet vise à résoudre des jeux à information imparfaite, en définissant le meilleur coup, le meilleur choix et une stratégie optimale. L'étude de ces jeux est pertinente pour de nombreux domaines (mathématiques appliquées, économie, biologie, sciences politiques, informatique et ingénierie) où ils sont utiles pour modéliser et résoudre des problèmes.

**Travaux réalisés en 2022** : Ce projet a été accepté en février 2022. Nous avons financé 2 mois d'un stage de M2 sur le sujet « Résolution de jeux à reconnaissance ». Les objectifs de ce stage étaient d'étudier l'état de l'art des solutions existantes pour les jeux à information incomplète et implémenter des combinaisons de recherche arborescente et techniques d'apprentissage par renforcement s'adaptant au degré d'information. Deux groupes de travail ont également été organisés et financés par ce projet. Le premier groupe de travail a eu lieu du 8 au 10 juin à l'Université Paris 8 avec notre collaborateur Francesco Dolce de l'Université Technique Tchèque de Prague. Le deuxième a eu lieu du 29 au 30 novembre à l'Université Paris 8 avec Mark van den Bergh de l'Université Vrije d'Amsterdam.

**AAP Paris 8 « Metamoi »**

**Porteurs** : Sylvia Chalençon et Farès Belhadj

**Partenaires** : Université Paris 8 (France)

**Date de début** : 01/2022

**Date de fin** : 12/2022

**Description** : Le projet MetaMoi propose d'étudier et de concevoir un nouveau modèle d'interactions au travers d'un avatar de soi. Dans le domaine des applications informatiques, l'avatar était, par le passé, un assistant pour de l'initiation, de l'aide à la pratique, du dépannage ou encore des astuces d'utilisation. Dans le domaine des jeux vidéos, les avatars sont devenus des représentations des joueurs dans des méta-univers. Avec les réseaux sociaux, les médias, la communication, les jeux vidéos et les mondes virtuels connectés, les avatars vont devenir des représentations de nous même, pilotés par chacun ou pouvant agir et interagir en notre nom, à

notre place, comme nous le ferions ou selon un modèle de comportement choisi. Il s'agit donc de définir comment nous représenter, comment représenter nos expressions, comment permettre aux programmes d'apprendre à réagir comme nous le ferions. Les objectifs visés par ce projet abordent donc de multiples problématiques pour lesquels la recherche et l'élaboration de nouveaux modèles seront nécessaires. Chaque partie produira ainsi ses propres résultats et publications scientifiques en lien direct avec le domaine sous-jacent. L'agrégation totale ou partielle de ces parties permettra de produire des résultats scientifiques en lien avec d'autres domaines comme la santé, l'éducation ou la psychologie.

**Travaux réalisés en 2022 :** Ce projet a été accepté en février 2022. Les premières discussions et décisions prises portaient principalement sur le modèle de représentation du visage. Concernant les expressions faciales, nous avons décidé de nous donner comme premier objectif le pilotage d'une tête virtuelle répondant aux contraintes du Facial Action Coding System proposé par Paul EKMAN. En terme de format de représentation de données, le rigging étant essentiel dans la réalisation de nos objectifs, notre choix s'est porté sur deux formats ouverts, le glTF proposé par le consortium OpenGL et assez répandu sur le Web, et le format USD proposé par Pixar et de plus en plus utilisé dans l'industrie du cinéma. Ces travaux devaient être réalisés dans le cadre du stage financé par l'AAP 2022, sans entrer dans les détails, nous n'avons malheureusement eu que peu de bonnes candidatures, celle néanmoins retenue, de Mme Chaima BEN SALEM (M2 Informatique, Synthèse d'Images et Conception Graphique de Limoges), a été improductive. Malgré cette situation, nous, porteurs, avons néanmoins avancé seuls sur le module de gestion des bones en bas niveau, sur une base glTF, et des bibliothèques ouvertes. Faute de résultats, le projet est en pause depuis septembre 2022 et aucune production scientifique n'est envisagée ni envisageable en l'état. Enfin, suite à cette expérience infructueuse, et au départ en mutation de Sylvia Chalengon, nous avons décidé de ne pas renouveler le financement sur AAP P8 pourtant accordé sur deux ans.

**Projet de recherche doctorale DIM AI4IDF**

**Porteurs :** Nicolas Jouandeau et Jean-Noël Vittaut

**Partenaires :** Université Paris 8 (LIASD) et Sorbonne Université (LIP6)

**Date de début :** 09/2022

**Date de fin :** 12/2025

**Description :** L'objectif de ce projet de recherche doctorale est d'étudier les algorithmes, les stratégies et les méthodes d'apprentissage par renforcement pour les problèmes à information imparfaite avec reconnaissance, d'étudier comment une machine peut par réflexivité, (1) « conceptualiser » une solution en présence d'incertitude, (2) « anticiper » les stratégies possibles des acteurs du monde dans lequel elle agit, et (3) « s'adapter » à l'imperfection de l'information en capitalisant sur les succès et les échecs de ses expériences passées. Ce projet a également été déposé dans le cadre des bourses doctorales de l'ED CLI. Ce projet n'a pas été sélectionné, ni par DIMAI4IDF, ni pas l'ED CLI.

**PHC PAMOJA « Drug design with DRUG »**

**Porteurs :** Dickson Owuor (Strathmore University, Kenya) et Nicolas Jouandeau (Université Paris 8, France)

**Partenaires :** School of Computing and Engineering Sciences (Strathmore, Kenya), School of computer science and IT (DeKUT, Kenya), LIASD (Paris 8, France)

**Date de début :** 05/2022

**Date de fin :** 12/2023

**Description :** L'objectif de ce projet de recherche est d'étudier la synthèse de médicaments par combinaison de recherche arborescente et d'apprentissage par renforcement. Ce projet n'a pas été sélectionné.

## **Algorithmes parallèles de preuve pour les problèmes d'informations incomplètes sur les GPU**

**Porteurs :** Paolo Ciancarini (Universita di Bologna, Italie) et Nicolas Jouandeau (Université Paris 8, France)

**Partenaires :** Dipartimento di Informatica - Scienza e Ingegneria (Universita di Bologna, Italie), LIASD (Paris 8, France)

**Date de début :** 09/2022

**Date de fin :** 12/2025

**Description :** L'objectif est de proposer et d'implémenter de nouveaux algorithmes de recherche parallèles sur GPU qui combinent la recherche arborescente et d'apprentissage automatique. Ce projet n'a pas été sélectionné.

## **7.2 Nouveaux projets industriels**

Pas de nouveaux projet industriels pour cette année.

## **8 Activité du groupe par domaine**

En synthèse d'images expressives, la majeure partie des efforts faits dans ce domaine l'a été dans le cadre du projet MetaMoi.

En enseignement de l'informatique graphique, nous avons participé au groupe de travail de l'AFIG (Association Française d'Informatique Graphique) sur l'enseignement de l'Informatique Graphique. Lors des jFIG 2022 à Bordeaux (journées de l'AFIG), nous avons animé la table ronde sur le thème : « les outils (en ligne/hors ligne) utilisés pour illustrer les concepts de l'IG et la pédagogie associée » comprenant 5 intervenants (présentations) aux approches variées dont nous avons présenté l'usage de sa bibliothèque OpenSource GL4Dummies pour l'enseignement des concepts de l'IG en Licence. Aussi, ces journées ont été l'occasion de proposer une intégration du concours local API8 ([www.api8.fr](http://www.api8.fr)) au sein des prochaines sessions, cela prendra la forme d'une track 64K spéciale AFIG réservée aux étudiants de premier et deuxième cycle. Le concours API8, qui existe depuis 2014, après trois années d'absence liée à la pandémie de Covid, aura de nouveau lieu en 2023, et nous avons obtenu de quoi subventionner une grande partie des dépenses liées à l'événement auprès de l'UFR STN.

En résolution des jeux et apprentissage automatique, nous avons publié un article [1] sur la parallélisation sur GPU de l'algorithme Random Forest, présentant deux solutions avec des gains en rapidité de 1.5 à 3.9 pour plus de 32 arbres sur des datasets de 4 à 15 attributs et de 150 à 250k enregistrements. Nous avons également étudié la résolution de problèmes à un joueur en utilisant un apprentissage par renforcement avec self-play. Ces techniques ont donné de bons résultats dans différents jeux à deux jours tels que les échecs, le go et Othello. Nous avons réalisé des tests sur le morpion solitaire et la conception d'ARN. Nous avons proposé un nouvel algorithme RNASP qui apprend la fonction d'évaluation des états de l'ARN résultant réseau de neurones profond entraîné par renforcement avec selfplay. Les résultats sont meilleurs que les algorithmes existants sur trois datasets et équivalents sur un dataset. Ces travaux ont obtenu le best student paper award [2] à la conférence Neural Computation Theory and Applications (NCTA). Les résultats obtenus sur le morpion solitaire n'ont pas dépassé les résultats existants. Nous avons également étudié les jeux à reconnaissance avec le projet GRAS, au travers des variantes de des échecs (RBC), Nim et Wythoff.

En représentation du raisonnement et logique non classique, nous avons poursuivi des travaux sur la modélisation de la croyance. Nous avons publié une étude bibliographique pluridisciplinaire sur la question aboutissant à une proposition multimodale originale [3]. Ces travaux, présentés

dans une conférence nationale rassemblant la communauté, ont ouvert des perspectives sur le fond et l’amènent à travailler en parallèle sur un article de revue et un nouvel article à soumettre à IJCAI 2023. Par ailleurs, une étude de l’opérateur d’agrégation en logique multivalente (LMV), proposé dans le cadre de la cotation d’informations, a montré que celui-ci permettait de formaliser l’ensemble des normes triangulaires en LMV [4, 5].

En traitement du langage naturel, la production de ressources pour le développement et l’évaluation d’outils de traitement automatique des langues dans plusieurs cadres linguistiques et applicatifs est principalement étudiée. [6] dresse un état des lieux des réalisations pour la langue corse (notamment la mise à disposition d’un premier corpus de référence annoté en morpho-syntaxe). Ces travaux se poursuivent (i) sur les biais d’évaluation en TAL liés à la nature des ressources et aux métriques utilisées (production de corpus du français annotés en entités nommées (EN) permettant la comparaison d’outils de reconnaissance automatique des EN) et (ii) sur la représentativité des ressources linguistique dans le contexte de l’entraînement de modèles (d’annotation et de langue). Nous avons également été effectué dans la création d’un ensemble de mots-clés spécifiques au domaine (souvent appelé lexicon) à partir d’un corpus existant. Pour ce travail, un corpus a été créé à partir des pages web d’entreprises ayant une activité de recherche importante dans leur domaine. Ensuite, pour obtenir le résultat souhaité, elles ont été utilisées à la fois des techniques d’apprentissage profond et des techniques statistiques [7, 8].

En dynamique symbolique, nous travaillons sur la notion de reconnaissabilité des morphismes. Nous étudions plusieurs questions liées à la notion de reconnaissable morphisme et à la notion de la dendricité. La principale question que nous étudions consiste à savoir si les langages créés par des morphismes reconnaissables sont des ensembles dendriques.

En sécurité et privacy, les travaux de recherche sont essentiellement concentré dans le projet ReComp (cf supra dans la section « Projet en cours »). D’autres travaux sont menés concernant les limites des blockchains, dont un article soumis et accepté dans la revue Terminal (à paraître en janvier 2023) et un rapport technique critiquant la volonté d’utilisation de chaîne de blocs pour mettre en œuvre un système de vote électronique [9].

## 9 Présentations et réunions de travail réalisées dans le cadre du séminaire SOIF

Dans la continuité de l’année 2021, le Séminaire d’Ouverture aux Informatiques et de Formation (SOIF) a été maintenu afin d’échanger sur les travaux en cours des membres de l’équipe.

**Jean-Pascal Palus** (doctorant dans l’équipe PASTIS, sous la supervision d’Adrien Revault d’Allonnes et Nicolas Jouandeau)

**Date** : 14 avril 2022

**Titre** : La confiance comme posture de crédulité

**Résumé** : La « confiance » (« trust »), souvent utilisée de manière interchangeable avec la « conviction » (« confidence »), est une position doxastique décrite de multiples façons par les diverses disciplines qui ont tenté de la formaliser. Parfois décrite comme la position épistémique consistant à faire confiance à un agent ou à une proposition, parfois décrite comme l’état psychologique de celui qui fait confiance ; ce concept est néanmoins presque exclusivement décrit comme un moyen d’évaluer un degré de relation entre deux ou plusieurs entités (truster-trusties, truster-proposition) d’un point de vue objectif, ce qui se traduit par l’amalgame de notions telles que la « vérité absolue » ou la « fiabilité » dans ce qui devrait, selon nous, rester un modèle descriptif et prédictif d’adhésion et de confiance dans la valeur de vérité subjective d’une proposition. Nous présenterons ici un aperçu transdisciplinaire de ce qui se dit sur le sujet dans la

littérature et proposerons une approche basée non pas sur l'évaluation externe d'un degré de « confiance » mais sur la modélisation des processus mentaux internes de celui qui fait confiance.

**Julien Lavauzelle** (maître de conférences en mathématiques au LAGA et à l'Université Paris 8)

**Date** : 31 mai 2022

**Titre** : Des codes correcteurs pour le retrait confidentiel d'information

**Résumé** : Le retrait confidentiel d'information (private information retrieval, PIR) est une primitive cryptographique permettant d'accéder à une information stockée à distance, sans révéler l'identité de l'information au système de stockage. Dans cet exposé, on présentera différentes techniques pour réaliser efficacement cette primitive à l'aide de codes correcteurs. Suivant les contextes d'utilisation (faible bande passante, stockage limité, etc.), on en exposera les limites ainsi que des constructions presque optimales.

**Francesco Dolce** (post-doctorant à la Faculté d'informatique de l'Université technique Tchèque, à Prague)

**Date** : 9 juin 2022

**Titre** : Playing with games and words

**Résumé** : Combinatorial games are two-player games with complete and perfect information (each player knows all the possible rules and see all other's player moves), and without chance involved (no dice). In this talk we present two very simple combinatorial games – namely Nim and Wythoff's game – and discuss possible generalizations and how to use Combinatorics on Word to find winning strategies.

**Journée de fin d'année de l'équipe PASTIS** (Séance spéciale où l'ensemble des non-titulaires de l'équipe sont invités à présenter leurs travaux)

**Date** : 30 juin 2022

**Programme détaillé** :

— **14h-14h30** : Accueil et café

— **14h30-15h30** : Session stagiaires

— **Chaima Ben Salem** (MetaMoi : utilisation des bones pour l'animation de modèles 3D)

Il s'agit ici de présenter le contexte du projet MetaMoi et de donner des éléments sur le fonctionnement des modèles 3D animés par Bones. L'objectif actuel est d'implémenter en C/GLSL la fonctionnalité d'animation par Bones en utilisant les bibliothèques Assimp pour l'importation de modèles et GL4D/OpenGL pour la modélisation et le rendu.

— **Emil Raducanu** (Litte\_Bot, a chatbot to bring Molière's characters back to life)

As we celebrate the quadricentennial anniversary of Molière's birth (1622-1673), we would like to implement an interface where the user will be able to exchange with the characters created by the dramatist. This chatbot should fit each characters' personality, whether in a prose or a verse style. The dataset composed of 32 plays provided by the Bibliothèque nationale de France (BnF) has been preprocessed. We separated prose from verse dialogue in order to train efficiently our models. A dictionary of rhymes coming from Molière's work has been formed as well. We are currently fine tuning a French pretrained multi-layer bidirectional Transformer encoder model, camemBERT, to learn the way of writing of Molière. This knowledge will be then distilled to a Seq2Seq model to finally generate text.



- **Jilin He** (Domain-specific keyword extraction)  
 Keyword extraction is a popular research field in natural language processing, which refers to picking up the terms that are most relevant in a given context. However, most of the datasets are scientific articles which makes the model only good at extracting technological keywords. In this internship, we propose to create a domain-specific keyword extraction model based on BERT to grasp the knowledge from a non-technological field. To do so, we've used thousands of scraped text of companies of different sectors that mention their innovation activities. Also, we used a dictionary of 300 keywords created by Cortext and validated by expert, in order to build our dataset. Finally, the keyword extraction based on BERT model is performed. The results give us new and acceptable discovered keywords.
- **Mouna Benabid** (Reconnaissance Blind Chess)  
 Nous avons pu observer ces dernières années de grandes avancées dans la résolution de jeux complexes tels que le jeu de Go, les échecs, et les jeux Atari, et ce à l'aide de l'apprentissage par renforcement (i.e. Renforcement Learning abrégé RL), qui est un sous-domaine de l'apprentissage automatique (i.e. Machine Learning abrégé ML). Cependant, de telles performances nécessitent des programmes d'avoir une connaissance parfaite de leur environnement. Cette connaissance parfaite est toutefois souvent irréaliste dans les cas réels, où nous n'avons généralement accès qu'à une partie de l'information. Les objectifs de ce travail consistent à étudier l'état de l'art des solutions existantes pour les jeux à information complète et à information incomplète et implémenter des combinaisons de recherche arborescente et de RL en s'adaptant au degré d'information. Les jeux à information incomplète avec reconnaissance sont des jeux à information incomplète où on peut faire des perceptions. Sans perception, il s'agit d'un jeu à information incomplète sans reconnaissance et dans ce cas chaque joueur ne voit que ses pièces. Un exemple de jeu à information incomplète avec reconnaissance est le Reconnaissance Blind Chess (RBC). Dans un premier temps, il s'agit d'établir la liste des algorithmes existants déjà expérimentés sur des jeux à information complète, sur des jeux avec chance, ou sur des jeux à information incomplète. Puis, dans un deuxième temps, d'utiliser certains de ces algorithmes pour proposer une solution pour le jeu RBC.
- **15h30-15h45** : Pause
- **15h45-17h15** : Session doctorant.es et ATER
- **Jean-Pascal Palus** (La confiance comme posture de crédulité)  
 La « confiance » (« trust »), souvent utilisée de manière interchangeable avec la « conviction » (« confidence »), est une position doxastique décrite de multiples façons par les diverses disciplines qui ont tenté de la formaliser. Parfois décrite comme la position épistémique consistant à faire confiance à un agent ou à une proposition, parfois décrite comme l'état psychologique de celui qui fait confiance ; ce concept est néanmoins presque exclusivement décrit comme un moyen d'évaluer un degré de relation entre deux ou plusieurs entités (truster-trusties, truster-proposition) d'un point de vue objectif, ce qui se traduit par l'amalgame de notions telles que la « vérité absolue » ou la « fiabilité » dans ce qui devrait, selon nous, rester un modèle descriptif et prédictif d'adhésion et de confiance dans la valeur de vérité subjective d'une proposition. Nous présenterons ici un aperçu transdisciplinaire de ce qui se dit sur le sujet dans la littérature et proposerons une approche basée non pas sur l'évaluation externe d'un degré de « confiance » mais sur la modélisation des processus mentaux internes de celui qui fait confiance.
- **Maroua Boudabous** (Active Learning for Explicit Aspect Term Extraction for Low-

resource Language : A use case of French online reviews)

We propose to use active learning for an end-to-end process to improve explicit aspect extraction, for low-resource language fine-grained labeling, in opinion and sentiment analysis. With the rapid growth of online commerce, capturing customers' opinions about service and product preferences is essential for business companies to ensure competitiveness and enhance their marketing campaigns. Identifying aspect terms in opinionated texts such as online reviews is crucial to achieving this goal. This task, known as Aspect Term Extraction (ATE) in Natural Language Processing (NLP), remains difficult, especially for low-resource languages where labeled data are lacking. Recent advances in cross-domain learning and active learning make it possible to get sufficiently large datasets and reduce the efforts to annotate them. For the training, we use a web-scraped dataset of French reviews on beauty products and electronic devices. We follow a 3-step process starting from defining a pseudo-labeler CRF using cross-domain learning. Then, we proceed to sequence labelling with BiLSTM-CNN-CRF, deep learning model. Finally, we use active learning to deal with label uncertainty resulting from the first pseudo-labeling step. We adapt the query strategy to fit the sequence labeling setting and handle class imbalance. Results show that active learning significantly improves the learning model's performance achieving over 30% of corrections to the initial labels.

- **Syrine Saidi** (Possibilistic preference networks toolbox)

$\pi$ -pref nets are qualitative models for representing conditional preferences in a possibilistic framework. Preferences of a user are described in terms of degrees of satisfaction. This software permits to visualize  $\pi$ -pref nets generated from CP-nets. It allows to compare orderings induced from  $\pi$ -pref nets, CP-nets, the Pareto semantic and default rules. It also makes refinements between different orderings.

- **Oumaima El Joubari** (A TDMA-based protocol for safety-related applications in VANET)

Vehicular Adhoc Network (VANET) is known as an instantiation of MANET designed to provide ubiquitous connectivity to all road network equipment, in which vehicles use wireless transceivers to communicate with vehicles within their radio communication range. Due to the significance of road safety, one of the stringent requirements of VANET performance is low latency and reliable communication between transport equipment. The frequent changes in the topology, bidirectional mobility, and the continuously varying density can lead to medium access collisions and thus affect the network performance. In a VANET, collisions occur when multiple nodes choose a common physical channel to transmit signals simultaneously specifically due to the hidden terminal problem or the merging collision. To address this issue, a novel channel access mechanism called TA-TDMA was designed to avoid collision issues while minimizing access delays. The protocol aims to mitigate the unfairness in time resources allocation by utilizing a traffic prediction model. Traffic conditions on opposite directions, different lanes, and road segments are forecasted and then each frame is partitioned accordingly. Each direction, lane, and segment is allocated a different set of time slots while considering the vehicle traffic conditions. The protocol is also cluster-based and access time among participating nodes is managed by the cluster head (CH).

- **17h15-18h** : Convivialité

**Pablo Rauzy** (maître de conférences en informatique dans l'équipe PASTIS)

**Date** : 25 octobre 2022

**Titre** : Promesses et (dés)illusions : une introduction technocritique aux blockchains

**Résumé** : « Une blockchain est un registre distribué et immuable dans lequel sont écrites des informations qui font consensus. ». Dans cette présentation, nous commencerons par donner du sens à cette phrase et à l'ensemble des termes qui y sont employés, en nous efforçant quand c'est nécessaire de rendre accessibles les notions informatiques (comme la décentralisation, la distribution, l'immuabilité, ou le consensus) et le fonctionnement technique des outils cryptographiques sous-jacents (comme les condensats, les signatures, ou la preuve de travail ou d'enjeu). L'objectif de cette introduction sera d'atteindre une compréhension réelle de ce qu'est une blockchain. Ainsi équipés, nous discuterons ensuite de ce que les blockchains permettent effectivement d'accomplir, et donc surtout ce qu'elles ne permettent pas. Nous questionnerons alors les utilisations qui en sont proposées en nous concentrant sur des cas d'usage typiques des blockchains que nous étudierons plus en détails : les « cryptomonnaies » bien sûr, la certification de documents (avec l'exemple des diplômes), et nous mentionnerons également le cas des NFT. Cela nous permettra en conclusion de questionner de manière générale le caractère d'« innovation de rupture » que l'on associe souvent à cette technologie.

**Stephen Obonyo** (doctorant dans l'équipe PASTIS, sous la supervision de Nicolas Jouandeau)

**Date** : 8 novembre 2022

**Titre** : Designing RNA Sequences By Self-play

**Résumé** : Self-play (SP) is a method in Reinforcement Learning (RL) where an agent learns from the environment by playing against itself until the policy and value functions converge. The SP-based methods have recorded state-of-the-art results in playing different computer games such as Chess, Go and Othello. In this paper, we show how the RNA sequence design problem where a sequence is designed to match a given target structure can be modelled through the SP while performing the state-value evaluation using a deep value network. Our model dubbed RNASP recorded the best and very competitive results on the benchmark RNA design datasets. This work also motivates the application of the self-play to other Computational Biology problems.

**Mark van den Berg** (lecturer à la Vrije Universiteit d'Amsterdam)

**Date** : 29 novembre 2022

**Titre** : Nim variants

**Résumé** : Combinatorial game theory provides results for the class of two-player, deterministic games with perfect information. With the aim of generalizing this theory to the class of non-perfect information games in mind, we introduce and analyze three variants of the game of Nim. In these variants, the opponent only receives partial information on the move executed by the opponent. We model the variants as games in extensive form and compute Nash equilibria for different starting configurations. For one variant, this provides a full characterization of the game. For the other variants, we prove some partial and structural results, but a full characterization remains elusive.

**Benjamin Dupont** (professeur agrégé au département PIF, et membre associé de l'équipe PASTIS)

**Date** : 13 décembre 2022

**Titre** : Calculs de bases par réécriture

**Résumé** : Dans de nombreux contextes en algèbre ont émergé des présentations définies par des générateurs et relations diagrammatiques. L'une des questions principales dans l'étude de

ces présentations est de calculer des bases de certains espaces vectoriels sous-jacents. Dans cet exposé, nous présenterons une méthode issue de la théorie de la réécriture permettant d’approcher ces problèmes. Nous introduirons les propriétés fondamentales de terminaison et de confluence pour des systèmes de réécriture de mots, et nous expliquerons comment ces deux propriétés permettent de calculer des bases dans un contexte linéaire. Nous illustrerons ces constructions sur une famille d’algèbres apparaissant en théorie des algèbres de Lie.

## 10 Liste des publications pour l’année 2022

- [1] K. Senagi and N. Jouandeau. Parallel construction of Random Forest on GPU. In *Journal of Supercomputing*, pages 1–21, Springer, <https://doi.org/10.1007/s11227-021-04290-6>, (JS-2022).
- [2] S. Obonyo, N. Jouandeau and D. Owuor. Designing RNA Sequences By Self-play. In *14th International Conference on Neural Computation Theory and Applications*, pages 1–8, Malte, Best Student Paper Award Winner, SciTePress, (NCTA-2022).
- [3] J-P. Palus, A. Revault d’Allonnes et N. Jouandeau. La confiance comme posture de crédulité. In *Rencontres Francophones sur la Logique Floue et ses Applications*, pages 1–10, Toulouse, France, hal-03832463, (LFA-2022).
- [4] M-J. Lesot and A. Revault d’Allonnes. T-norms in Many-Valued Logics : a Representation Theorem in the ABOP Framework. In *IEEE World Congress on Computational Intelligence*, pages 1–7, Padoue, Italy, hal-03832483, (WCCI-2022).
- [5] M-J. Lesot and A. Revault d’Allonnes. Les t-normes en logique multivalente : un théorème de représentation de l’ABOP. In *Rencontres Francophones sur la Logique Floue et ses Applications*, pages 1–9, Toulouse, France, hal-03832459, (LFA-2022).
- [6] L. Kevers et A. Millour. Réalisations, obstacles et perspectives pour l’outillage du corse. In *Journées Jointes des Groupements de Recherche Linguistique Informatique, Formelle et de Terrain (LIFT) et Traitement Automatique des Langues (TAL)*, pages 1–9, Marseille, France, hal-03846829, 2022.
- [7] T. Cazenave, G. Grosjean, B. Rozière et A. Pappa. Molière, a theatrical agent with speaks like Molière’s characters. In *25th Generative Art Conference*, pages 281–285, ISBN 978-88-96610-44-2, Rome, Italy, (GA-2022).
- [8] M. Boudabous et A. Pappa. Apprentissage actif pour l’extraction des aspects explicites : application à des avis non annotés en français . In *Journées Jointes des Groupements de Recherche Linguistique Informatique, Formelle et de Terrain (LIFT) et Traitement Automatique des Langues (TAL)*, pages 20–28, Marseille, France, hal-03859310, 2022.
- [9] E. Blanchard, F. Li Vigni et P. Rauzy. Auteur·ices, relecteur·ices : redoublons de prudence face aux effets de modes technologiques. In *Rapport technique*, pages 1–23, hal-03741811v2, 2022.