

# Rapport d'activité 2024

Équipe PASTIS

## 1 Composition et domaines de recherche du groupe

Le groupe de recherche PASTIS est composé de 13 titulaires, 5 doctorants :

- Adrien Revault d'Allonnes (MCF)
- Alexandros Singh (MCF)
- Alice Millour (MCF)
- Anna Pappa (MCF-HDR)
- Benjamin Dupont (PRAG)
- Farès Belhadj (MCF)
- Françoise Balmas (MCF-HDR)
- Jean-Jacques Bourdin (PU)
- Louis Falissard (MCF)
- Nicolas Jouandeau (PU - responsable de l'équipe)
- Pablo Rauzy (MCF)
- Revekka Kyriakoglou (MCF)
- Vincent Boyer (MCF-HDR - en disponibilité)
- Daniel Ruiru (Doctorant 2024 - financement Ministère des Affaires Etrangères et de l'Europe (MEAE) / Strathmore University, Nairobi, Kenya)
- Florent Giauna (Doctorant 2024 - contrat doctoral ED CLI)
- Marina Seghier (Doctorante 2023 - 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)

Le groupe possède des compétences 1) Synthèse d'image expressive, 2) Enseignement de l'informatique graphique, 3) Combinatoire et jeux, 4) Analyse de textes, ressources et modèles de TALN, 5) Modélisation de la confiance avec les logiques non classiques et 6) Sécurité informatique émancipatrice et privacy.

Daniel Ruiru est sous la direction de Dickson Owuor (Strathmore University, Kenya) et Nicolas Jouandeau. Son sujet de thèse est « Solving Resource Allocation Problems with Deep Reinforcement Learning ».

Florent Giauna est sous la direction de Louis Falissard et Nicolas Jouandeau. Son sujet de thèse est « Interprétation géométrique du transfert d'apprentissage et applications en télédétections ».

Marina Seghier est sous la direction d'Alice Millour et Nicolas Jouandeau. Son sujet de thèse est « Explicabilité des modèles de TAL par les ressources : caractérisation de la variation textuelle et paradigme d'évaluation systématique ».

---

\*version 250122-1318

Maroua Boudabous est sous la direction d’Anna Pappa et Françoise Balmas. Son sujet de thèse est « Extraction intelligente de données web et reconnaissance d’aspects dans les corpus multilingues par apprentissage actif et par transfert. ».

Stephen Obonyo est sous la direction de Nicolas Jouandeau et de Dickson Owuor (Strathmore University, Kenya). Son sujet de thèse est « Bioinformatic Computation With Tree Search and Learning ». Sa soutenance est prévue pour le 9 décembre 2024.

## 2 Bilan en chiffres

Pour l’année 2024, le groupe PASTIS compte :

- publications en conférence et revues : 12
- interventions académiques sans publication : 4
- publications grand public/médiation/valorisation : 1
- interventions grand public/médiation/valorisation : 7
- soutenances de thèse et HDR : 2
- organisations d’évènements scientifiques : 2
- participations à des comités de programme/review : 15
- 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 HDR

**Anna Pappa**

**Sujet :** *Contributions à la création de corpus et de modèles d’apprentissage profond pour les données textuelles multilingues.*

**Jury :** Christian Boitet (LIG, Rapporteur), Benoit Crabbe (LLF, Rapporteur), Max Silberstein (ELLIADD, Examineur), Tita Kyriacopoulou (LIGM, Examinatrice), Tristan Cazenave (LAMSADE, Examineur), Jean-Jacques Bourdin (LIASD, Examineur), Nicolas Jouandeau (LIASD, Garant/Rapporteur).

**Date :** 22 avril 2024.

## 4 Soutenances de thèse

**Stephen Obonyo**

**Sujet :** *Bioinformatic Computation With Tree Search and Learning.*

**Jury :** Frédéric Saubion (LERIA, Rapporteur), Tristan Cazenave (LAMSADE, Rapporteur), Jean-Noël Vittaut (LIP6, Examineur), Anna Pappa (LIASD, Examinatrice), Louis Falissard (LIASD, Examineur), Nicolas Jouandeau (LIASD, Directeur de thèse).

---

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

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

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

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

**Date** : 9 décembre 2024.

## 5 Organisations d'évènements scientifiques

1. 7ème édition du concours national API8 ([api8.fr](http://api8.fr)). **F. Belhadj**, **B. Dupont** et **A. Revault d'Allonnes**. journée de présentation des travaux. Université Paris 8, 29 mai 2024
2. Unblocked & Unchained Symposium ([unblockchained.sciencesconf.org](http://unblockchained.sciencesconf.org)). **P. Rauzy**. Journée d'étude et de médiation interdisciplinaires "Analyses critiques autour des blockchains". Université Paris 8, 13 décembre 2024.

## 6 Participations à des comités de programme/review

- IEEE Conference on Games (IEEE-CoG)
- IEEE Transactions on Games (IEEE-ToG)
- International Computer Games Association Journal (IOSPress-ICGA)
- ICGA Computer and Games Conference (CG)
- 23rd IFIP International Conference on Entertainment Computing (IFIP-ICEC)
- International Conference on Agents and Artificial Intelligence (ICAART)
- International Conference on Informatics in Control, Automation and Robotics (ICINCO)
- Empirical Methods in Natural Language Processing (EMNLP, conférence principale et comité d'éthique)
- Association of Computational Linguistics Rolling-Review (ARR, conférence principale et comité d'éthique)
- Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-Coling)
- Special Interest Group on Under-resourced Languages (SIGUL)
- Games for NLP (Games4NLP)
- PremierConference&Exhibition onComputer Graphics& Interactive Techniques(ACM-SIGGRAPH)
- Eurographics, the 45th Annual Conference of the European Association for Computer Graphics (EG)
- Computer & Graphics Journal

### 6.1 Projets en cours

### 6.2 CONDOR

**Porteurs** : Jean-Jacques Bourdin

**Partenaires** : Archives de l'Académie des sciences de Paris ; Archives nationales de France ; Bibliothèque de l'Institut de France ; Bibliothèque nationale de France ; Campus Condorcet, Paris-Aubervilliers ; Institut d'histoire de la Révolution française, Paris.

**Date de début** : 2016

**Date de fin** : 2024

**URL** : <https://www.inventaire-condorcet.com/>

## 7 Nouveaux projets

### 7.1 Nouveaux projets académiques

#### 7.1.1 AAP Paris 8 « AbSynth »

**Porteurs :** Farès Belhadj

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

**Date de début :** 01/2024

**Date de fin :** 12/2024

**URL :** <https://expreg.org/amsi/C/projets/AbSynth/>

**Description :** AbSynth vise à aborder le rendu expressif, d'une part, avec une perspective fondée sur un processus d'abstraction menant à des degrés de figuration du vide, et, d'autre part, en exploitant l'expression fluidifiée du mouvement perçu pour synthétiser la scène observée ou modélisée.

Le concept de « non peint » est central dans notre approche, il compose le paysage et donne la place des éléments ostensiblement présents dans la scène. Aussi, il préfigure le vide qui incarne les constituants caractérisés par une « absence de forme », comme l'est évidemment le vide mais aussi tout ce qui serait considéré « fluide » à une échelle donnée : air / vent, nuages / fumées, eau, etc. Ainsi, pour pouvoir correctement caractériser ces vides, nous proposons de porter une attention particulière au mouvement et à la manière avec laquelle « il a été perçu » (ou, dans le cas d'une scène 3D conçu/élaboré).

Concernant l'exécution du rendu, nous proposons de maintenir un point de vue dirigé par le mouvement, ce dernier étant persistant quelque soit la scène, a minima provoqué par l'action de « rendre », que le medium soit fluide ou non. Évidemment, il ne s'agit pas de minimiser la place de la lumière ni même son traitement (perception/interprétation), mais plutôt de transférer son exploitation au sein du processus de rendu pour en faire l'expression tangible du mouvement. À ce stade, il est important de préciser que l'ensemble de notre processus est nécessairement déroulé dans le temps, l'image produite étant qu'une expression ponctuelle du rendu.

**Travaux réalisés en 2024 :** Démarrage de la première phase du projet avec la captation des flux vidéos de plans fixes qui ont été effectuées dans les Jardins de Giverny en mai 2024. Premiers tests et développements sont réalisés sur la partie abstraction, ils permettent au moins de confirmer expérimentalement, et sur des échantillons bien choisis, la validité de l'approche sur cette partie. Il reste évidemment nécessaire de généraliser l'approche pour l'automatiser sur la détection et l'étiquetage « des vides » (préfiguration du vide) avant de commencer à s'intéresser à la partie synthèse.

#### 7.1.2 AAP Paris 8 « CoMTE »

**Porteurs :** Alexandros Singh et Revekka Kyriakoglou

**Partenaires :** Benjamin Dupont

**Date de début :** 01/2024

**Date de fin :** 12/2024

**URL :** <https://wintershammer.github.io/comte/>

**Description :** Le projet CoMTE porte sur les arbres, leurs polynômes de Shabat et leurs groupes de monodromie. Dans le cadre de ce projet, notre objectif principal est de déterminer les polynômes de Shabat associés à des passeports bien précis. Un passeport est une donnée combinatoire qui encode les informations nécessaires pour reconstruire la structure d'un arbre.

**Travaux réalisés en 2024 :**

- Une série de six séminaires en ligne qui ont eu lieu entre le 06/03 et le 04/07, donnés par des membres du projet en France et en Grèce, dont le but était de fournir une introduction aux dessins des enfants de Grothendieck, aux fonctions de Belyi, et à des notions mathématiques apparentées. Ces sujets constituent la base de notre recherche dans le cadre de ce projet (Détails sur la page : <https://wintershammer.github.io/comte/>).
- Une série de réunions de recherche régulières (hebdomadaires) d’une heure, qui a débuté le 20/07. Ces réunions servent à coordonner nos efforts dans le cadre de ce projet. Outre les mises à jour régulières sur les derniers résultats de nos recherches, nous discutons au cours de ces réunions de nos efforts pour rédiger les appels à projets, les ordres de mission et les tâches administratives liées au projet.
- Une visite de 3 membres du projet basés en France (Benjamin Dupont, Revekka Kyriakoglou, Alexandros Singh) à l’Université de l’Égée à Karlovasi, Samos, accueillis par 2 membres du projet basés en Grèce (Metaftsis Vasileios, Prasadis Efstratios). La visite a eu lieu au cours de la première semaine de septembre dans le but de travailler sur un article scientifique (actuellement en cours de rédaction) rassemblant nos résultats jusqu’à présent.
- Une visite de 2 membres du projet basés en Grèce (Metaftsis Vasileios, Prasadis Efstratios) à l’Université de Paris 8, accueillis par 3 membres du projet basés en France (Benjamin Dupont, Revekka Kyriakoglou, Alexandros Singh). La visite s’est déroulée du 11 au 15 novembre dans le but de poursuivre notre collaboration sur l’article susmentionné, en présentant nos résultats au séminaire de l’équipe PASTIS du laboratoire LIASD de l’UP8 (via une conférence donnée par le Prof. Prasadis le 14).
- Un effort continu pour organiser une école d’été (qui aura lieu fin août 2025 à Samos, accueillie par l’Université de l’Égée). Cette école d’été d’une semaine est destinée aux étudiants en fin de licence qui s’intéressent à l’intersection des mathématiques et de l’informatique, et est envisagée comme la première d’une série d’initiatives de ce type. Outre le renforcement de nos collaborations récemment établies avec nos collègues grecs, l’école d’été constituerait une excellente source d’étudiants diplômés potentiels et pourrait servir de catalyseur pour la mobilité des étudiants entre les deux institutions de l’UP8 et de l’Université de l’Égée.

### 7.1.3 AAP Paris 8 « GENEPI »

**Porteuses** : Alice Millour et Revekka Kyriakoglou

**Partenaires** : Julien Lavauzelle

**Date de début** : 01/2024

**Date de fin** : 12/2024

**URL** : <https://glacon.up8.site/projects.html#genepi>

**Description** : L’utilisation croissante par les étudiant.es d’agents conversationnels tels que ChatGPT pour les tâches de développement informatique soulève des inquiétudes quant à leur engagement dans le processus d’apprentissage. Depuis les années 1980, des approches dites “déconnectées” ou “débranchées” ont quant à elles montré leur impact favorable sur l’implication des étudiant.es vis-à-vis de la connaissance à acquérir, mais sont peu exploitées dans l’enseignement supérieur. Ce projet de recherche vise à explorer l’intégration d’approches non conventionnelles dans l’enseignement supérieur en informatique : quels sont les risques que l’utilisation d’agents conversationnels présente pour l’implication et la motivation des étudiant.es ? Quelles sont les opportunités pédagogiques offertes par ces agents ? Que permettent les approches débranchées dans le contexte universitaire, et que peuvent-elles apporter à ce nouveau contexte pédagogique ?

Quel est l'impact de ces deux approches sur la motivation des étudiant.es, leur implication et leur responsabilisation ?

**Travaux réalisés en 2024 :**

- Un groupe de travail GT1 a été organisé à l'université Paris 8, par Revekka Kyriakoglou. Ce groupe de travail nous a permis d'initier de nouvelles collaborations en recherche. Participants : • Anis Boubaker (PR), École de technologie supérieure de Montréal (ÉTS), Canada • Lamprini Chartofylaka (ingénieur de recherche), Université Sorbonne Paris Nord, France • Revekka Kyriakoglou (mcf), LIASD, Université Paris 8 Saint-Denis, France • Sébastien Lallé (mcf), LIP6, Sorbonne Université, France • Nour El Mawas (PR), Laboratoire CREM, Université de Lorraine, France • Sonia Proust-Andrewkha (chercheuse), Université de Montreal, Canada • Valéry Psyché (PR), I2A, TÉLUQ University, Canada.
- Deux séminaires ont eu lieu le 28/11, donnés par des membres du projet GENEPI en Canada (Voir <https://informatique.up8.edu/pastis/seminaire.html>) : • Anis Boubaker, professeur-enseignant à l'ÉTS, "*Automated Generation of Challenge Questions for Student Code Evaluation Using Abstract Syntax Tree Embeddings and RAG : An Exploratory Study*" • Valéry Psyché, professeure agrégée à l'Université TÉLUQ, "*Comprendre la théorie des tuteurs intelligents et leur apport pour les chatbots pédagogiques*".
- Demande ANR PRC "EPIAGEN" (Enseignement en Pédagogie Informatique et IA Générative), porteur Revekka Kyriakoglou. EPIAGEN est un projet interdisciplinaire réunissant des universités et laboratoires en France et au Canada, dont le LIASD à l'Université Paris 8, le LIP6 à Sorbonne Université, le CREM à l'Université de Lorraine, le LIGM à l'Université Gustave Eiffel, l'I2A à l'Université TÉLUQ, et l'École de Technologie Supérieure. Il s'inscrit dans un contexte où l'intelligence artificielle générative (IAGen) s'étend rapidement à tous les niveaux académiques et dans une variété de disciplines, soulevant des questions quant à son impact sur l'apprentissage des étudiants. Il se concentre sur l'enseignement de l'informatique au niveau universitaire, notamment à la licence en France et le Baccalauréat (License) au Québec, en visant à comparer l'influence de l'IAGen sur les compétences en programmation dans différents contextes éducatifs (distanciel versus présentiel). La problématique centrale d'EPIAGEN réside dans l'impact d'une surutilisation de l'IAGen sur l'apprentissage de la programmation et des concepts algorithmiques. Le recours aux générateurs de code pourrait nuire à la pensée computationnelle, à l'engagement et à la compréhension des concepts fondamentaux de la programmation. Des préoccupations éthiques, comme le risque de plagiat, sont également pertinentes. À notre connaissance, la recherche actuelle manque d'études longitudinales comparant les effets de l'IAGen dans divers contextes éducatifs ou entre la France et le Québec. EPIAGEN vise à combler cette lacune en analysant les effets de l'IAGen sur l'enseignement de la programmation, en comparant les contextes d'enseignement entre la France et le Québec, et en promouvant une intégration éthique et responsable de l'IAGen. La méthodologie itérative, ancrée dans le Design-Based Research, inclut : une revue de la littérature ; des enquêtes auprès des étudiants/enseignants ; des expérimentations pratiques avec des tâches de codage ; et l'analyse de l'apprentissage issue des données éducatives collectées. Les principes d'IA éthique, responsable et explicable guideront chaque étape de la recherche. Les résultats attendus permettront une meilleure compréhension de l'impact de l'usage de l'IAGen sur l'apprentissage à travers.
- A. Millour a assisté à la conférence ITiCSE afin de rencontrer la communauté scientifique de l'enseignement de l'informatique à l'université. Les travaux les plus pertinents seront présentés au premier semestre 2025 lors d'un séminaire d'équipe ouvert. A. Millour et J. Lavauzelle ont expérimenté l'informatique débranchée lors d'un atelier qu'ils ont conçu pour la Fête de la Science 2024.

#### 7.1.4 AAP Paris 8 « Unblocked & Unchained Symposium »

**Porteur :** Pablo Rauzy

**Partenaires :** GÉODE

**Date du symposium :** 13/12/2024

**URL :** <https://unblockchained.sciencesconf.org/>

**Description :** Les nouvelles technologies que sont les “cryptomonnaies”, le “métavers”, les “NFT”, ou encore le “web3” sont invariablement présentées comme des innovations incontournables du monde de demain, sans que ne soit jamais vraiment expliqué ni pourquoi ni comment... sauf une chose : « c’est grâce à la “blockchain” ! ».

Mais c’est quoi, une blockchain ? À quoi ça sert ? Quelles sont les limites de cette technologie ? Qui en utilise ? Où dans le monde ? Pour quoi faire en pratique ? Quelles politiques se cachent derrière la technologie “blockchain” ?

Quid des fameuses “cryptomonnaies” ? Déjà, c’est quoi, la monnaie ? Et à quelle vision de la monnaie correspondraient les “cryptomonnaies” ? Quelles politiques se cachent derrière cette façon de voir la monnaie ?

Au travers des présentations et des échanges de cette journée d’étude, seront apportées des réponses à l’ensemble de ces questions de société. Grâce au dialogue entre universitaires de différents domaines (informatique, géopolitique, économie) et professionnel·les de la médiation ou du journalisme, ces réponses à la pointe des connaissances scientifiques resteront accessibles à un public large sans rien sacrifier de la rigueur de la recherche universitaire.

#### 7.1.5 AAP Paris 8 « Convention citoyenne étudiante de la santé mentale »

**Porteurs :** Camille Vansymaey et Louis Falissard

**Partenaires :** Marie Paule Coutot et Yannick Morvan

**Date de début :** Septembre 2024

**Date de fin :** Août 2025

**Description :** Nous souhaitons élaborer et expérimenter une méthodologie visant à constituer une convention citoyenne étudiante de la santé mentale au sein de l’Université Paris 8 qui aura pour objectif, en articulation avec le Service de Santé des Etudiants (SSE), de formuler des recommandations sur des mesures et des politiques à mettre en œuvre pour répondre aux enjeux actuels de la population étudiante de l’Université en termes de promotion et prévention de la santé mentale. La méthodologie visera à répondre aux deux freins à la représentation et à la participation étudiante : la dilution de la responsabilité d’une part, et le sentiment d’incompétence et de non-légitimité d’autre part.

## 7.2 Nouveaux projets industriels

Pas de nouveaux projet industriels pour cette année.

## 8 Activité du groupe par domaine

**En synthèse d’images expressives**, l’équipe a entrepris des travaux sur l’abstraction puis la synthèse de peintures vivantes à partir de flux vidéo de plans fixes. L’AAP 2024 de Paris 8 a permis d’obtenir un financement, via le projet Absynth, pour réaliser la captation des flux vidéos de plans fixes ; ces captations ont été effectuées dans les Jardins de Giverny. Les premiers tests et développements ont été réalisés sur la partie abstraction, ils permettent de confirmer,

temporairement de manière empirique, la validité de l’approche sur cette partie. Il reste nécessaire de généraliser l’approche afin de l’automatiser sur la problématique de préfiguration du vide avant de commencer à s’intéresser à la partie synthèse par expression fluidifiée des mouvements captés.

**En enseignement de l’Informatique Graphique (IG)**, l’équipe participe toujours aussi activement au GT-Enseignement de l’AFIG, et a aidé à l’organisation de la session enseignement des JFIG 2024 qui a eu lieu à Strasbourg. Un travail de synthèse portant sur l’organisation des sessions enseignement des JFIG sur les cinq dernières années est effectué avec plusieurs membres du GT, il aura pour objectif une publication dans la track Education d’Eurographics en 2025.

**En résolution des jeux et apprentissage automatique**, une partie importante des travaux a porté sur des applications en calculs bioinformatiques, en prévisions météorologiques et en prévisions d’indicateurs financiers. Nous avons proposé une solution pour générer des séquences d’ARN selon un modèle structurel, des contraintes d’organisation et d’énergie correspondant aux couples Adénine, Uracile, Guanine, Cytosine (AUGC) composant ces séquences. Ces travaux [p1] avec des applications possibles dans la synthèse de molécules et de médicaments, propose d’utiliser un algorithme de recherche arborescente MCTS avec une génération des séquences définie par un LLM. En parallèle, nous avons proposé dans d’autres travaux [p2] un modèle de repliement inverse de l’ARN appelé Self Play RNA (SPRNA) pour concevoir par apprentissage par renforcement des séquences plus riches en GC et donc plus stables. Pour la synthèse chimique moléculaire, nous avons proposé un modèle SMILE basé sur une décomposition en fragments [p3], combiné avec un LLM combiné avec une recherche arborescente en faisceau permet d’obtenir des propositions augmentant la validité, la nouveauté et l’unicité des molécules générées. Dans le cadre de l’utilisation de données temporelles et pour les prévisions météorologiques, un modèle combinant apprentissage actif et transfert d’apprentissage [p4] a été proposé pour obtenir des résultats équivalents aux meilleures solutions de l’état mais en utilisant moins de données. Dans le cadre de l’utilisation d’indicateurs, nous avons réalisé une étude comparative des modèles Transformer, LSTM+Transformer et LSTM bidirectionnel [p5]. Le modèle LSTM+Transformateur obtient les meilleurs résultats. Des travaux ont été réalisés sur la résolution de Breakthrough à l’aide d’analyses rétrogrades parallèles. Dans le cadre d’une collaboration avec Jean-Noël Vittaut de l’équipe LFI du LIP6, nous avons participé aux Computer Olympiads de l’ICGA 6 dans les catégories Breakthrough et Chinese Dark Chess, avec les programmes SuperPudim et SuperBaba. Il serait intéressant de réussir à intégrer de l’apprentissage par renforcement dans ces programmes pour l’an prochain.

**En représentation du raisonnement et logique non classique** une collaboration avec des chercheurs du LIFO a démarré sur l’adaptation des modèles de formalisation de la confiance en une information au cadre du Zero Trust<sup>5</sup>. Une participation à une série d’ateliers sur la manipulation de l’information a également émergé de cette collaboration. Par ailleurs, de nouveaux travaux, en collaboration avec l’équipe SMA du LIP6, portent sur la modélisation de l’informativité.

**En traitement automatique du langage naturel et apprentissage automatique**, les travaux sur la langue corse ont fait l’objet d’une publication à LREC 2024 [p6], tout comme les travaux sur l’évaluation de la tâche de reconnaissance des entités nommées en français [p7] et sur les catégories textuelles [p8]. Par ailleurs, nous avons proposé une méthodologie visant la construction d’un lexique centré sur les concepts émergents en innovation non technologique, décrite dans un article publié aux JADT 2024 [p9]. Cet article décrit une méthodologie en quatre étapes, combinant expertise humaine, statistiques et techniques d’apprentissage automatique, pour établir un modèle généralisable à plusieurs domaines. Cette procédure comprend la création d’un corpus thématique, la constitution d’un lexique de référence, l’annotation et la préparation d’un corpus d’entraînement, et enfin, l’implémentation de modèles d’apprentissage pour identifier

---

5. <https://cyber.gouv.fr/publications/le-modele-zero-trust>



de nouveaux termes. La méthodologie développée promet une applicabilité dans des domaines conceptuels. Aussi, les travaux sur BrAIcht, un agent conversationnel théâtral capable de générer des dialogues dans le style de Bertolt Brecht, réalisés par Anna Pappa et ses co-auteurs, ont fait l'objet d'une publication à la 27e conférence Generative Art [p10]. BrAIcht s'appuie sur German LeoLM, une version fine-tunée de Llama2 avec 7 milliards de paramètres, entraînée sur un corpus combinant 29 pièces de Brecht et 907 pièces allemandes stylistiquement similaires. Pour gérer les contraintes de mémoire, la méthode QLoRA a été utilisée pour un fine-tuning efficace. Les performances de BrAIcht, évaluées à l'aide de métriques telles que le score BLEU et la perplexité, montrent son aptitude à reproduire fidèlement le style dramaturgique unique de Brecht.

**En recherche d'information conversationnelle**, les travaux sur l'édition de grand modèles de langage à des fins de réductions d'hallucination ont fait l'objet d'une publication à ECIR. 2024 [p11]

**En combinatoire**, plusieurs groupes de travail ont été organisés sur l'étude des arbres, de leurs polynômes de Shabat et de leurs groupes de monodromie. L'intérêt principal de ces groupes, tel que décrit dans le AAP CoMTE, est de déterminer les polynômes de Shabat associés à des passeports spécifiques.

**En enseignement de l'informatique**, cette année, dans le cadre du projet AAP GENEPI, nous avons travaillé sur l'impact des agents conversationnels sur l'engagement des étudiants en informatique, tout en expérimentant des approches débranchées dans l'enseignement supérieur. GENEPI a permis d'initier une réflexion qui s'est élargie avec le projet EPIAGEN, visant à explorer de manière plus approfondie les enjeux pédagogiques, techniques et éthiques liés à l'intégration de l'IAGEN dans l'enseignement supérieurs. Le projet EPIAGEN, une collaboration internationale entre la France et le Canada impliquant des laboratoires tels que le LIASD (Université Paris 8), le LIP6 (Sorbonne Université), le CREM (Université de Lorraine), le LIGM (Université Gustave Eiffel), l'I2A (Université TÉLUQ) et l'École de Technologie Supérieure. Dans le cadre de ce projet, nous avons déposé une demande de financement ANR PRC, marquant une avancée majeure pour approfondir l'analyse des effets de l'IAGEN sur l'apprentissage de la programmation dans différents contextes éducatifs (présentiel versus distanciel) et entre la France et le Québec.

**En sécurité et privacy**, les activités de médiations démarrées en 2022 et 2023 concernant les limites des blockchains ont été poursuivies en 2024 jusqu'au point d'orgue que représente l'organisation de la journée d'étude interdisciplinaire "Unblocked & Unchained Symposium" le 13/12. Ainsi, un nouvel article a été publié dans la conférence JRES 2024 [p12]; quatre conférences [i1, i2, i3, i4] en séminaires de recherche et auprès d'étudiant-es de BUT et d'école d'ingénieur ont été données, ainsi qu'un entretien dans le podcast RdGP [m1]; trois conférences [m2, m3, m4] auprès du grand public, concernant les limites des blockchains mais aussi des questions plus larges sur l'informatique émancipatrice (logiciels libres, privacy); une participation à une table ronde sur l'IA et le monde du travail à la Fête de l'Humanité [m5]; deux interviews pour des médias grand public (Politis [m6] et La Tribune [m7]) sur des sujets liés à la cybersécurité; le montage d'une convention avec Universcience<sup>6</sup> a été lancée pour mettre en œuvre des activités de médiations autour du réseau et de la cybersécurité auprès de publics scolaires, et dans ce cadre Pablo Rauzy a participé à un école de médiation en informatique débranchée en novembre<sup>7</sup>; enfin, un article de blog [a1] concernant l'usage de la cryptographie par l'application de messagerie Telegram a été mis en avant sur la Une du Club Médiapart pour contribuer à l'éclairage de l'actualité dans le contexte de l'arrestation de Pavel Durov.

---

6. Universcience est l'EPIC en charge du Palais de la Découverte ainsi que de La Cité des Sciences et de l'Industrie.

7. Cf <https://emi2024.sciencesconf.org/>.

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

Dans la continuité de l'année 2023, 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. La fréquence des séminaires est restée à une fois par semaine afin d'avoir plus d'échanges dans des réunions de travail, liant enseignement et recherche. Seuls les séminaires avec la présentation de travaux spécifiques sont listés ci-dessous.

**Philippe Guillot** (ancien maître de conférences en mathématiques à Paris 8)

**Date** : 21 mars 2024

**Titre** : Le RSA est-il vraiment mort ?

**Résumé** : La fin prochaine du RSA est annoncée depuis plusieurs années. En témoigne la tenue annuelle de CATACRYPT (workshop on catastrophic events related to cryptology and their possible solutions) de 2014 à 2017, faisant suite au workshop "The End of RSA" en janvier 2013 à Menlo Park en Californie.

Les craintes étaient étayées par : 1) La prochaine factorisation du RSA 1024, rendant désormais accessible cette taille de clé aux "Nation State Resources". 2) Des progrès algorithmiques laissant espérer la découverte d'un algorithme de complexité  $L(1/4)$ , dans la foulée de MPQS  $L(1/2)$  et GNFS  $L(1/3)$  [  $L(x)$  est une fonction qui quantifie les complexités entre  $L(0)$  polynomial et  $L(1)$  exponentiel. ]. 3) L'advenue prochaine de l'ordinateur quantique et son fameux algorithme de Shor (1994) pour factoriser les entiers en temps et ressources quadratiques.

Des financements considérables sont dirigés vers le développement de cette machine révolutionnaire ainsi que vers la recherche de solutions cryptographiques qui y résisteraient. Pourtant, plus de dix ans après ce premier workshop, la factorisation du RSA 1024 se fait toujours attendre. Le record actuel reste plafonné depuis 2020 au RSA-250 (829 chiffres binaires) factorisé par GNFS, dont certains chercheurs affirment qu'il n'est guère en pratique plus efficace que MPQS. Les machines quantiques réelles peinent à dépasser les quelques dizaines de qubits intriqués, ce qui rend hautement spéculative leur menace sur le RSA. Les informations fiables et précises sont par ailleurs très rares en raison du secret industriel dont elles font l'objet, les annonces ressemblant davantage à des plaquettes commerciales creuses.

L'objectif de l'exposé est de présenter l'algorithmique de la factorisation des entiers en détaillant plus en profondeur le crible quadratique MPQS et l'algorithme de Shor de façon à ce que l'auditeur se fasse par lui-même une idée plus précise de la menace réelle.

**Xavier Pothron** (Nobody Studio)

**Date** : 11 avril 2024

**Titre** : Perspectives et considérations générales sur la 3D dans l'univers du jeu vidéo et du rendu précalculé | Présentation de Nobody Studio et de la dynamique des industries culturelles et créatives (ICC) à Montpellier

**Résumé** : Créé en 2020, Nobody Studio est une société coopérative initialement spécialisée dans les services de lighting et de compositing précalculés, offrant son soutien à d'autres studios. Avec le temps, Nobody Studio a élargi son spectre d'interventions pour inclure l'éclairage dans le secteur des jeux vidéo, la création d'environnements 3D, la prévisualisation/le layout, ainsi que la supervision intégrale de projets de leur conception à leur achèvement. Le studio aspire également à développer ses propres projets internes. À travers un résumé des diverses missions réalisées au fil des années et une présentation de Nobody Studio, basé à Montpellier, nous abordons les multiples défis auxquels nous sommes confrontés en tant que graphistes, studio et acteurs de

l'industrie. Il s'agit notamment des différentes productions sur lesquelles nous œuvrons actuellement, des domaines de recherche et développement sur lesquels nous avons ou sommes en train de travailler, ainsi que des réflexions engagées sur l'évolution de notre profession. Cela englobe les aspects relatifs à l'écosystème (ICC de Montpellier, Siggraph Occitanie, ...), au contexte industriel (grèves dans le secteur des effets visuels, licenciements dans l'industrie du jeu vidéo, écosystème éducatif, ...) ou encore aux innovations technologiques (modèles génératifs, rendu en temps réel...).

**Louis Falissard** (maître de conférence)

**Titre** : Une brève introduction aux réseaux de flots génératifs

**Résumé** : Les réseaux de flots génératifs ont récemment été introduit en apprentissage automatique, et consiste en un ensemble de méthodes permettant l'ajustement de modèles permettant l'échantillonnage d'objets composites, avec une probabilité approximativement proportionnelle à une fonction de récompense arbitraire. Ces modèles se situent conceptuellement à la croisée des chemins de domaines aussi variés que l'apprentissage par renforcement, les modèles générateurs profonds, et l'apprentissage basé sur l'énergie, et offrent notamment des possibilités de représenter des concepts liés d'analyse causale (comme par exemple l'inférence contrefactuelle) relativement naturellement.

Dans ce séminaire, nous aborderons brièvement la définition de ces objets, ainsi que les nouvelles règles d'apprentissage automatiques qu'ils proposent, et nous nous attarderont sur une potentielle piste de recherche de décomposition énergétique de ces objets inspirée des modèles contrefactuels.

**Stephen Obonyo** (doctorant)

**Date** : 5 décembre 2024

**Titre** : Bioinformatic Computation With Tree Search and Learning

**Résumé** : Bioinformatic computation uses computational methods to analyze, interpret and manage biological data such as Deoxyribonucleic Acid (DNA), Ribonucleic Acid (RNA), proteins, molecular structures, and other biological sequences. Working with biological data requires intensive computational resources in terms of time and storage due to the complexity and large size of the data. Traditionally, search algorithms such as Breadth-First Search (BFS) and Depth-First Search (DFS) have been used to solve several bioinformatic problems. However, these search algorithms are limited in their ability to solve complex bioinformatic problems due to the large search space and poor exploration. Recently, there have been significant advancements in Artificial Intelligence (AI) and Machine Learning (ML) has allowed it to develop and use new methods to solve bioinformatic problems. In particular, combining tree search with ML, where a search algorithm performs the exploration and the ML model generalizes the search solutions, has emerged as a practical approach to solving complex optimization problems in bioinformatics. This thesis aims to investigate how tree search and ML can be effectively combined to solve bioinformatic computation problems, focusing on five main problems : RNA Inverse folding, protein folding, molecular design, retrosynthetic planning, and optimization of inverse models in healthcare domain. This work is motivated by combining tree search algorithms with ML techniques ; we can overcome the limitations of traditional search methods and achieve better performance in solving these problems. Developing new and effective bioinformatic computation methods is crucial for advancing research in biology and healthcare, drug design, personalized medicine, understanding disease action mechanisms and other important application areas in biology.

## 10 Productions scientifiques pour l'année 2024

### 10.1 Publications en revues et conférences

Une astérisque marque l'auteur·ice ayant présenté les travaux lors de la conférence.

- [p1] **S. Obonyo\***, **N. Jouandeau** and D. Owuor. RNA Generative Modeling With Tree Search. In *IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pages 1–9, 2024.
- [p2] **S. Obonyo\***, **N. Jouandeau** and D. Owuor. Self-Playing RNA Inverse Folding. In *Springer Nature Computer Science (SNCS)*, pages 1–17, 2024.
- [p3] **S. Obonyo\***, **N. Jouandeau** and D. Owuor. Decoding Molecular Language Model with Beam Search. In *Workshop on ML for Chemistry and Chemical Engineering (ML4CCE) at European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD)*, pages 1–18, 2024.
- [p4] P. Gikunda and **N. Jouandeau**. Lightweight Deep Learning for Weather Prediction and Forecasting in Africa. In *Workshop on Data Science for Agriculture in Africa (DAAfrica) at African Conference on Research in Computer Science and Applied Mathematics (CARI)*, pages 1–6, 2024.
- [p5] **D. Ruiru\***, **N. Jouandeau** and D. Owuor. LSTM Versus Transformers : A Practical Comparison of Deep Learning Models for Trading Financial Instruments. In *International Joint Conference on Computational Intelligence (IJCCI)*, pages 1–7, 2024.
- [p6] **A. Millour\***, L. Kevers, L. Brasile, and A. Ghia. Agettivu, aggitivu o aghjettivu? POS Tagging Corsican Dialects. In *LREC-COLING*, pages 600–608, Turin, Italy, 2024.
- [p7] **A. Millour\***, Y. Dupont, K. Fort, and L. Duignan. Unveiling Strengths and Weaknesses of NLP Systems Based on a Rich Evaluation Corpus : the Case of NER in French. In *LREC-COLING*, pages 17217–17224, Turin, Italy, 2024.
- [p8] **M. Seghier\***, **A. Millour**, and J.-Y. Antoine. Descripteurs Linguistiques et Caractérisation Objective des Catégories Textuelles. In *5èmes journées du Groupement de Recherche CNRS “Linguistique Informatique, Formelle et de Terrain ”*, pages 106–112, Nancy, France, 2023.
- [p9] **R. Kyriakoglou\***, **A. Pappa**, J. He, A.Schoen, P. Laurens, M. Vartampetian, Ph. Laredo, and T. Kyriacopoulou. Développement automatique de lexiques pour les concepts émergents : une exploration méthodologique. In *17es Journées internationales d'Analyse statistique des Données Textuelles, JADT*, pages 1–10, 2024.
- [p10] R. Baz, K. Malyseva, **A. Pappa\***, and Tristan Cazenave. BrAIcht, a theatrical agent that speaks like Bertolt Brecht's characters. In *27th Generative Art Conference*, pages 1–5, 2024.
- [p11] P. Erbacher\*, **L. Falissard**, V. Guigue, and L. Soulier. Navigating uncertainty : Optimizing api dependency for hallucination reduction in closed-book qa. In *46th European Conference on Information Retrieval (ECIR), Glasgow, UK, March 24–28*, pages 1–7, 2024.
- [p12] **P. Rauzy**. Blockchain : une mauvaise solution à la recherche d'un problème. In *Journées réseaux de l'enseignement et de la recherche*, pages 1–10, 2024.

### 10.2 Interventions académiques sans publication

- [i1] **P. Rauzy**. La vérité sur la blockchain. Séminaire dans le cadre du cours “économie numérique et durable” en 3e année de BUT informatique à l'IUT de Villetaneuse. 12 janvier 2024.

- [i2] **P. Rauzy**. La vérité sur la blockchain. Séminaire dans le cadre du cours “Blockchain & Consensus : Coopération dans les systèmes distribués” à l’IMT Atlantique. 5 mars 2024.
- [i3] **P. Rauzy**. Blockchain : une mauvaise solution à la recherche d’un problème. JRES 2024, Rennes. 11 décembre 2024.
- [i4] **P. Rauzy**. Introduction technocritique aux blockchains. Unblocked & Unchained Symposium, Université Paris 8, Saint-Denis. 13 décembre 2024.

### 10.3 Publications grand public / médiation / valorisation

- [a1] **P. Rauzy**. Telegram n’est pas une messagerie sécurisée.

### 10.4 Interventions grand public / médiation / valorisation

- [m1] **P. Rauzy**. [Entretien] p4bl0 : À quoi sert la blockchain? Podcast RdGP, <https://rdgp.fr/@rdgp/episodes/entretien-p4bl0-a-quoi-sert-la-blockchain>, 27 novembre 2024.
- [m2] **P. Rauzy**. Paradoxe dans la pensée libriste. festival Pas Sage En Seine 2024. 31 mai 2024.
- [m3] **P. Rauzy**. Si c’est “gratuit”, c’est toi qui produis. festival Pas Sage En Seine 2024. 2 juin 2024.
- [m4] **P. Rauzy**. La vérité sur la blockchain. Conférence à la médiathèque de Saint-Denis dans le cadre de la Fête de la Science 2024. 13 octobre 2024.
- [m5] **P. Rauzy**. L’IA, le monde du travail et de la création : quelles conséquences? <https://fete.humanite.fr/blog/les-debats/?date=2024-09-14&event=66d9ed8021241f6717d0a0a1>, 14 septembre 2024.
- [m6] **P. Rauzy**. Interview pour **Politis** au sujet du dispositif MyCheckr par Thomas Lefèvre : "De l’analyse faciale dopée à l’IA dans les bureaux de tabac, « des produits illégaux »". <https://www.politis.fr/articles/2024/12/nouvelles-technologies-de-lanalyse-faciale-dopee-a-lia-dans-les-bureaux-de-tabac-des-produits>, 20 décembre 2024.
- [m7] **P. Rauzy**. Interview pour **La Tribune** au sujet de Telegram par Marine Protais : "Comment Telegram est devenu « une zone de non-droit » numérique". <https://www.latribune.fr/technos-medias/internet/comment-telegram-est-devenu-une-zone-de-non-droit-numerique-1004922.html>, 27 août 2024.