

Rapport d'activité 2025

Équipe PASTIS

1 Composition et domaines de recherche de l'équipe

Le groupe de recherche PASTIS est composé de 12 titulaires, 3 doctorants :

- **Titulaires :**
 - Adrien Revault d'Allonnes (MCF)
 - Alexandros Singh (MCF)
 - Alice Millour (MCF)
 - Anna Pappa (PU - Responsable de l'équipe)
 - Benjamin Dupont (PRAG)
 - Farès Belhadj (MCF)
 - Françoise Balmas (MCF-HDR)
 - *Jean-Jacques Bourdin* professeur émérite (retraite depuis oct 2025 - PU poste vacant)
 - Louis Falissard (MCF)
 - Nicolas Jouandeau (PU - Directeur LIASD)
 - Pablo Rauzy (MCF)
 - Revekka Kyriakoglou (MCF)
- **Doctorant·es :**
 - 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)
- **Stagiaires :**
 - Gaëlle Klos, L2 - Université Paris 8 Saint-Denis, Licence informatique et vidéoludisme, 1 mois.
 - César Pichon, M1 - Traitement automatique des langues : recherche et développement, Université Sorbonne Nouvelle, 6 mois.
 - Ziyang Xu, M2 - Université de Lorraine, Institut des Sciences du Digital, Management et Cognition, 6 mois.
 - Wassim Zemouri, M2 - École Supérieure en Informatique 8 Mai 1945 de Sidi Bel-Abbès, Algérie, 6 mois.

Le groupe possède des compétences en 1) Synthèse d'image expressive, 2) Enseignement de l'informatique graphique, 3) Combinatoire et jeux, 4) Création de corpus, analyse de textes, ressources et modèles de TALN, 5) Modélisation de la confiance en 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 ».

*version 250122-1609

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 ».

2 Bilan en chiffres

Pour l'année 2025, le groupe PASTIS compte :

- publications en conférence et revues : 12
- interventions académiques sans publication : 6
- publications grand public/médiation/valorisation : 1
- interventions grand public/médiation/valorisation : 17
- soutenances de thèse : 0
- encadrements de stage : 4
- organisations d'événements scientifiques : 3
- participations à des comités de programme/review : 13
- projets en cours : 4 et 1 délégation
- nouveaux projets déposés : 5
- ressources scientifiques ouvertes (corpus, données, outils) : 1
- implications dans les sociétés savantes, GdR et centres de recherche : 4

Le groupe est impliqué dans le Conseil d'administration de l'AFIG¹, dirige le chapitre de IEEE CIS France², est impliqué dans le centre de recherche Geopolitique de la *datasphere* (GEODE)³, est impliqué dans le Conseil enseignement de EUROGRAPHICS⁴.

3 Soutenances de thèse

Pas de soutenance en 2025

4 Stages

- Gaëlle Klos, L2 - Université Paris 8 Saint-Denis, Licence informatique et vidéoludisme : "Pensée informatique à bord" : s'initier à la pensée informatique à travers le jeu dans les transports en commun. Encadrement **Revekka Kyriakoglou**. 1 mois.
- César Pichon, M1 - Traitement automatique des langues : recherche et développement, Université Sorbonne Nouvelle, Dpt Linguistique et Phonétique Gen. et Appl. : Création corpus 'presse', à partir de sources numérisées pour le projet EUR ARTEC "Les femmes dans les métiers du cinéma", encadrement **Anna Pappa**. 6 mois.

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

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

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

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

- Ziyang Xu, M2 - Université de Lorraine, Institut des Sciences du Digital, Management et Cognition : Évaluation de l’adaptabilité des grands modèles de langue à la variation linguistique, encadrement **Alice Millour**, Carlos Gonzales Gallardo (Université de Tours) et Jean-Yves Antoine (Université de Blois). 6 mois.
- Wassim Zemouri, M2 - École Supérieure en Informatique 8 Mai 1945 de Sidi Bel-Abbès, Algérie : Développement d’un système de *speech-to-text* pour la langue bretonne, encadrement **Alice Millour** et Loïc Grobol (Université Paris Nanterre). 6 mois.

5 Organisations d’événements scientifiques

1. **2ème** édition de la conférence internationale **LT4All** - Language Technology for All. **S. Sakti**, **G. Adda**, **A. Millour**, **J. Mariani**, **P. Paroubek**, **M. Melero**, **C. Soria**, Paris, UNESCO, du 24 au 26 février 2025
2. **9ème** édition du concours national **API8** (api8.fr). **F. Belhadj**, **B. Dupont** et **A. Revault d’Allonnes** journée de présentation des travaux, Université Paris 8, 17 mai 2025.
3. **1er** atelier **ParCoL** : Science Participative pour les Données et Corpus Linguistiques / *Crowd-sourced Linguistic Data Collection and Augmentation*, en marge de la conférence TALN 2025. **A. Millour** et **M. Hutin**, campus Saint-Charles de l’Université Aix-Marseille, 30 juin 2025.

6 Participations à des comités de programme/review

- Expert Systems with Applications journal (Elsevier-ESWA)
- International Conference on Agents and Artificial Intelligence (ICAART)
- International Federation for Information Processing – International Conference on Entertainment Computing (IFIP-ICEC)
- Plate-Forme Intelligence Artificielle - Conférence Nationale en Intelligence Artificielle - (PFIA-CNIA)
- IADIS International Conference Game and Entertainment Technologies (IADIS-GET)
- Advances in Computer Games conference (ICGA-ACG)
- International Computer Games Association Journal (IOSPress-ICGA)
- Undone Computer Science 2026
- International Conference on Language Resources and Evaluation (LREC)
- Association for Computational Linguistics (ACL)
- International Conference on Computational Linguistics (COLING)
- Eurographics Education program
- ACM SIGGRAPH reviews

7 Projets

7.1 En cours

7.1.1 Projet délégation CNRS

Chercheuse en délégation : Anna Pappa

Collaboratrice CNRS : Florence Dupin de Saint Cyr–Bannay

Lieu : IRIT Toulouse

Équipes partenaires : ADRIA et PASTIS

Date de début : 01/09/2025

Date de fin : 31/08/2026

Titre du projet : Modération de l’intégration de données textuelles à l’aide de graphes de connaissances dynamiques et de raisonnement non monotone

Description : Ce projet propose un système d’analyse et de signalement d’inexactitudes dans les informations textuelles, avec une application particulière à la détection des *fake news*. Il repose sur une combinaison de méthodes issues du TALN et de la logique pour garantir la cohérence d’un texte et sa relation avec des connaissances de sens commun extraites de corpus multilingues. L’originalité de ce projet réside dans l’utilisation du raisonnement non monotone, pour traiter des informations évolutives et contradictoires, ainsi que dans la mise à jour dynamique de graphes de connaissances.

7.1.2 Projet ArTeC : Les femmes dans les métiers du cinéma français

Porteuse : Tifenn Martinot-Lagarde BNF

Partenaires LIASD : Anna Pappa (**Resp scientifique**), Revekka Kyriakoglou et Alice Millour.

Partenaires : ESTCA Université Paris 8; 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 : 2025

Date de fin : 2026

URL : <https://eur-artec.fr/projets/les-femmes-dans-les-metiers-du-cinema-francais/>

Description projet : Ce programme de recherche vise à rendre visibles les femmes professionnelles du cinéma, longtemps marginalisées dans les récits historiographiques, en permettant de les identifier, de les quantifier et de documenter leurs trajectoires. Il repose sur l’exploration systématique d’archives issues de plusieurs institutions patrimoniales de référence, afin de produire une cartographie inédite des métiers du cinéma occupés par des femmes depuis le début du cinéma.

Le rôle de LIASD : il traite par des méthodes d’intelligence artificielle, combinant apprentissage automatique et approches linguistiques la presse de cinéma conservés à la Bibliothèque nationale de France (BnF) et à la Cinémathèque française, pour en extraire les noms de femmes liés à des métiers du cinéma et, les compter par catégories de métiers. Ces sources seront croisées avec les cartes professionnelles (archives du CNC) et les dossiers de scolarité d’écoles de cinéma conservés aux Archives nationales.

Travaux réalisés en 2025 : les travaux menés en 2025 ont porté sur la mise en place de la chaîne de traitement des données et sur la constitution d’un corpus de référence associant

des femmes et des métiers du cinéma. À partir de textes issus de la presse cinématographique numérisée (1920–1965), une chaîne de traitement séquentielle a été développée, comprenant l'extraction du texte OCRisé en ligne, la détection automatique des entités nommées, le filtrage des entités correspondant à des personnes féminines, l'annotation des métiers et la structuration des résultats dans un corpus exploitable. La détection des entités nommées repose sur des outils de traitement automatique des langues intégrant des modèles de langage pré-entraînés, tandis que l'identification des femmes s'appuie sur des lexiques de prénoms et de titres. L'annotation des métiers est réalisée à partir d'un lexique métier de base, enrichi de manière itérative au niveau des formes, et de patrons linguistiques permettant de repérer les formulations professionnelles dans les textes. Le couplage entre femmes et métiers repose sur une approche empirique fondée sur la proximité textuelle, via le calcul de distances linéaires entre entités dans des fenêtres contextuelles paramétrables. Les associations obtenues sont intégrées dans un corpus de référence structuré sous forme tabulaire (CSV), comprenant notamment les noms, métiers, sources et contextes d'apparition. En complément de cette approche par règles, une vérification ponctuelle de certains appariements a été menée en aval du pipeline à l'aide de modèles de langage pré-entraînés de type Transformer, en particulier DistilBERT et CamemBERT, utilisés en inférence comme classifieurs contextuels. Ces modèles sont exploités dans un cadre contraint et non génératif, via des requêtes fermées produisant une décision discrète (association plausible / non plausible), sans apprentissage supplémentaire. Enfin, un premier essai de typologie des contextes d'association femme–métier a été engagé, afin de caractériser les différents schémas linguistiques observés. L'ensemble de ces travaux constitue une base méthodologique destinée à être consolidée et étendue lors des phases suivantes du projet.

7.1.3 Projet GENEPI

Porteuses : Alice Millour et Revekka Kyriakoglou

Partenaires :

- Anis Boubaker, PU à l'École de technologie supérieure (ÉTS), Montréal, Canada.
- Sébastien Lallé, MCF à Sorbonne Université, LIP6, France.
- Guilherme M. Machado, MCF à ECE Paris, France.
- Nour El Mawas, PU à l'Université de Lorraine, CREM, France.
- Sonia Proust-Andrewkha, MCF à l'Université de Sherbrooke, Canada.
- Valéry Psyché PU à TÉLUQ, Canada.

Date de début : 2024

Date de fin : 2025

Description : GENEPI est un projet interdisciplinaire et inter-universitaire associant des laboratoires en France et au Canada. Il vise à étudier l'intégration de l'intelligence artificielle générative (IAGén) dans l'enseignement supérieur, avec un focus particulier sur l'enseignement de l'informatique. Si l'usage des outils d'IAGén par les étudiant-es est désormais largement répandu, leurs effets sur les apprentissages, les compétences de résolution de problèmes et les pratiques pédagogiques en informatique restent encore peu documentés.

Le projet GENEPI a pour objectif principal d'analyser l'impact de l'utilisation de l'IAGén sur les performances des étudiant-es, leur capacité à raisonner de manière critique, ainsi que les enjeux éthiques liés à une dépendance accrue à ces outils, notamment dans les cours de programmation. EPIAGEN se distingue par une approche comparative originale, étudiant les usages de l'IAG dans différents contextes pédagogiques (présentiel vs distanciel) et dans deux systèmes éducatifs (France et Québec), un angle encore peu exploré dans la littérature. En 2025, le projet a contribué à l'état de l'art par la publication d'une *scoping review* [p1] présentée

à l'International Conference on Artificial Intelligence in Education, portant sur l'intégration de l'IA générative dans l'enseignement de l'informatique au-delà de l'assistance à la programmation. En plus, Kyriakoglou et Machado ont été invitée comme intervenantes lors du workshop AIED 2025 « *Generative AI in Education : Exploring Human-AI Collaboration and Culture to Promote Equity and Inclusion* » [i1], afin de présenter les résultats de leurs travaux sur les enjeux éthiques de l'IAGén en enseignement de l'informatique.

7.1.4 Projet Yezh Ar Vro - La Langue du Pays

Porteuse : Mélanie Joutteau

Responsable scientifique pour le LIASD : Alice Millour.

Type de projet : Projet ANR Science avec et pour la société – Recherches participatives 2.

Partenaires :

- Mélanie Joutteau, CNRS, IKER
- Alice Millour, Université Paris 8 Vincennes Saint-Denis, LIASD
- Morgan Grobol, Université Paris Nanterre, Modyco
- Jean-Yves Antoine, Université de Tours

Date de début : 30/06/2025

Date de fin : 30/01/2027

Description : La recherche pilote que constitue le projet YAR est une collaboration entre acteurs de la société civile et chercheur-euses en linguistique et informatique (traitement automatique des langues, ergonomie) dans le cadre du développement de technologies œuvrant à la préservation linguistique. Elle s'inscrit en Bretagne, et se construit avec les locuteurs, locutrices et acteurs de la préservation d'une langue celtique en très grave danger d'extinction d'après l'UNESCO : le breton. L'association Dastum (collecte sauvegarde et diffusion du patrimoine oral breton depuis 1972) et le pôle scientifique (CNRS et trois universités) mobilisent trois associations d'enseignant-es et se coordonnent avec le soutien du fond de dotation Bretagne Numérique. YAR comprend la mise à disposition (i) d'une application smartphone de collecte de la parole géolocalisée pour le grand public, (ii) d'une plateforme web de transcription participative intégrant le son et le texte, utilisable en contexte pédagogique. Le matériel sonore collecté alimentera la plateforme de transcription participative conçue avec et pour le pôle pédagogique. Le corpus oral transcrit soutiendra le développement et le perfectionnement de la reconnaissance vocale automatisée (ASR) pour le breton.

7.1.5 Projet ComTE

Porteurs : Alexandros Singh, Revekka Kyriakoglou

Partenaire : Université de l'Égée

Date de début : 01/01/2024

Date de fin : 31/12/2025

Description : Le projet a donné lieu à une prépublication sur arXiv et à plusieurs visites d'une semaine (été 2024, hiver 2024, hiver 2025). <https://arxiv.org/abs/2510.10192>

7.2 Nouveaux projets déposés

Type de projet : ANR – PRCI (Projet de recherche collaborative internationale, Phase 1 déposée – AAPG 2026)

Intitulé du projet : EPIAGEN (Teaching and Learning Computer Science in the era of Generative AI)

Porteuse : Anna Pappa

Partenaires : Université Paris 8 (LIASD); Sorbonne Université (LIP6); Université de Lorraine (Centre de Recherche sur les Médiations); Université TÉLUQ (Canada); École de Technologie Supérieure (Canada); Université de Sherbrooke (Canada)

Durée : 36 mois

Description : Le projet EPIAGEN porte sur l’analyse des usages et des effets pédagogiques de l’intelligence artificielle générative dans l’enseignement supérieur, en particulier dans l’enseignement de l’informatique et de la programmation. Il vise à étudier les pratiques réelles d’utilisation de l’IA générative par les étudiant·e·s et les enseignant·e·s, leurs impacts sur les apprentissages et les compétences, ainsi que les transformations des pratiques pédagogiques induites par ces outils.

Le projet adopte une approche interdisciplinaire combinant informatique, traitement automatique du langage naturel et sciences de l’éducation, et s’appuie sur une méthodologie de type *design-based research*. Il prévoit des observations en contexte d’enseignement, des entretiens et des enquêtes longitudinales menées en France et au Québec. Les travaux visent également à concevoir et évaluer des scénarios et outils pédagogiques favorisant des usages responsables et éthiques de l’IA générative, en tenant compte des inégalités d’accès, des enjeux sociaux et des contraintes environnementales.

Type de projet : ANR – JCJC (Programme Jeunes Chercheuses et Jeunes Chercheurs, Phase 1 déposée – AAPG 2026)

Intitulé du projet : MEC-PLLM : Motivation, Engagement, and Collaboration in Programming with Large Language Models

Porteuse : Revekka Kyriakoglou

Partenaires : –

Durée : 36 mois

Description : Le projet *Motivation, Engagement et Collaboration en Programmation avec les Grands Modèles de Langage* (LLMs) étudie la manière dont les LLMs transforment les dynamiques d’apprentissage et d’interaction dans l’enseignement supérieur en informatique. Bien que ces modèles soient de plus en plus utilisés pour accompagner les étudiants dans leurs activités de programmation, on connaît encore peu leur impact réel sur la motivation, l’engagement et la collaboration (MEC) des apprenants, ni comment les caractéristiques propres à chaque modèle influencent ces processus. Le projet poursuit trois objectifs principaux : (i) analyser les interactions étudiant–LLM afin d’identifier des indicateurs de MEC ; (ii) étudier les « traits de personnalité » des LLMs et leurs effets sur la motivation, l’engagement et la collaboration ; (iii) concevoir et valider un cadre d’adaptation de ces traits de personnalité aux besoins pédagogiques. Ancré dans une approche de recherche orientée par la conception (Design-Based Research), le projet combine des analyses TALN (traitement automatique du langage naturel) des traces d’interaction, du code, des observations et des questionnaires recueillis dans des contextes d’apprentissage authentiques : hackathons collaboratifs et projets de programmation menés en France et au Canada. Les résultats attendus incluent : (1) une typologie des dynamiques d’interaction et de collaboration étudiant–LLM ; (2) des métriques TALN permettant de prédire des indicateurs comportementaux de motivation, d’engagement et de collaboration ; et (3) une couche de personnalisation des LLMs adaptée aux besoins pédagogiques.

Type de projet : ANR – PRC (Projet de Recherche Collaborative, Phase 1 déposée – AAPG 2026)

Intitulé du projet : HEMERA (De l’ombre à la lumière : l’IA pour exhumer les femmes professionnelles du cinéma et de l’audiovisuel français)

Coordination scientifique : Tifenn Martinot-Lagarde (Bibliothèque nationale de France)

Partenaires principaux : Bibliothèque Nationale de France ; **Université Paris 8 (LIASD, ESTCA)** ; EPITA (LRE) ; Université de La Rochelle (L3I) ; École nationale des chartes ; Université Sorbonne Nouvelle (IRCAV) ; Archives nationales ; Centre national du cinéma et de l’image animée (CNC) ; Institut national de l’audiovisuel (INA), ainsi que plusieurs partenaires académiques français et internationaux

Responsable scientifique LIASD : Anna Pappa

Durée prévisionnelle : 48 mois

Description : Le projet HEMERA vise à rendre visibles les femmes professionnelles du cinéma et de l’audiovisuel français sur la période 1917–1974, en développant des outils d’intelligence artificielle pour l’exploration et l’analyse de grands volumes d’archives patrimoniales hétérogènes. Il s’inscrit dans le prolongement des travaux engagés dans le cadre du programme ArTeC « Les femmes dans les métiers du cinéma français » et ambitionne d’élargir l’analyse à de nouveaux corpus (annuaires professionnels, presse spécialisée, génériques de films, contenus audiovisuels et iconographiques).

La contribution du LIASD porte sur le traitement automatique des corpus textuels patrimoniaux, l’extraction et la structuration d’entités nommées, les problématiques de désambiguïsation et de liage de données, ainsi que l’analyse diachronique des associations femme–métier. Ces travaux s’inscrivent dans une approche interdisciplinaire associant sciences du numérique, humanités numériques et sciences humaines et sociales, et visent la production de données structurées interoperables destinées à la recherche et à la valorisation patrimoniale.

Type de projet : ANR – PRC (Projet de recherche collaborative, Phase 1 déposée – AAPG 2026)

Intitulé du projet : DIVÆ - Diversité linguistique, VARIation et Évaluation des grands modèles de langues

Coordination scientifique : Marianne Vergez-Couret

Partenaires principaux : Université de Montpellier Paul Valéry (EPE) - DIPRALANG - Laboratoire de sociolinguistique, d’anthropologie des pratiques langagières et de didactique des langues-cultures ; Université de Strasbourg - LiLPa - Linguistique, Langues et Parole (EA 1339 - UR 1339 depuis 01.01.2020) ; Université Paris 8 Vincennes Saint-Denis - LIASD.

Responsable scientifique LIASD : Alice Millour

Durée prévisionnelle : 48 mois

Description : Ces dernières années, les efforts se multiplient pour promouvoir une meilleure diversité linguistique dans le monde numérique. Des ressources ont vu le jour pour une grande diversité de langues, en témoignent les corpus de textes alignés multilingues (comme la plateforme Parcolab) ou les corpus annotés en dépendances syntaxiques dans la communauté Universal Dependencies. Cependant, l’écart demeure réel et l’avènement des grands modèles de langue (GML) via l’essor de l’Intelligence Artificielle apporte autant de promesses que de nouvelles problématiques. Dans des contextes linguistiques où la visibilité et l’apprentissage sont des enjeux vitaux, les opportunités sont nombreuses. Cette technologie porte néanmoins également un risque d’uniformisation voire de déformation de langues d’autant plus vulnérables qu’elles sont variées et peu dotées en ressources disponibles pour l’entraînement des modèles. A travers ce projet, nous proposons de nous doter des moyens de mesurer d’un point de vue linguistique et pédagogique

l'impact de cette nouvelle technologie sur l'inclusivité des langues peu dotées dans l'espace numérique.

Les langues visées sont des langues peu dotées de France (comme l'alsacien et l'occitan) et d'Europe (comme le macédonien et le serbe). Elles appartiennent à des familles de langues différentes (romanes, germaniques, slaves) et ont des statuts sociolinguistiques différents.

L'objectif principal du projet est d'évaluer la compétence des GML dans le contexte de ces langues, notamment en lien avec leur utilisation par des apprenants, en considérant deux axes : (i) le caractère linguistiquement correct de ce qui est produit et (ii) la représentativité dialectale des contenus produits. Nous formulons ainsi les hypothèses suivantes (i) les productions des GML ne sont pas représentatives des pratiques et (ii) les ressources linguistiques de qualité, avec métadonnées et annotations fines, peuvent nous renseigner sur la nature de cette divergence.

Type de projet : LINGUA : An Open Call from Microsoft's AI for Good Lab

Intitulé du projet : ASR for Minoritized Languages in Contact with French : Lessons from the Breton Case **Coordination scientifique :** Alice Millour

Partenaires principaux : CNRS - IKER, Université Paris Nanterre - Modyco, Université Paris 8 Vincennes Saint-Denis - LIASD.

Responsable scientifique LIASD : Alice Millour

Durée prévisionnelle : 24 mois

Ce projet vise à favoriser l'inclusion numérique des langues minoritaires parlées dans les régions francophones, notamment celles où les locuteurs sont bilingues et en situation de diglossie, grâce au développement de technologies de reconnaissance vocale automatique incluant l'alternance codique (alternance entre français et une langue minorisée).

Notre objectif est de remédier à ce problème en développant un outil ciblant les passages en français présents dans les enregistrements vocaux afin d'accélérer la transcription assistée par ordinateur et, par conséquent, d'alléger la charge de travail des linguistes et transcripteur·ices dans des centaines de langues minoritaires.

L'efficacité sociotechnique de notre méthode sera évaluée dans le cadre de la collecte de corpus de parole naturelle en breton et de la production de leur transcription.

7.3 Nouveaux projets industriels

Pas de nouveaux projet industriels pour cette année.

8 Activité du groupe par domaine

En enseignement de l'Informatique Graphique (IG), nos contributions ont commencé par l'écriture d'un article [p2] pour la conférence EUROGRAPHICS dans la session Enseignement. Ce papier a été accepté et présenté. Il montre l'intérêt d'ajouter à nos conférences nationales des sessions (voire ateliers) "enseignement" et plusieurs collègues de pays différents ont commencé à discuter de cette idée.

Toujours dans la même conférence, au cours de la réunion du *board education* M. Bourdin a participé à la mise en route d'un prix pour le meilleur article de cette catégorie. Le prix sera nommé Steve Cunningham du nom d'un acteur important de notre communauté, maintenant décédé.

Nous avons également participé à la préparation de la table-ronde enseignement des journées JFIG de Calais (novembre 2025) où la thématique a porté sur la place de l'art dans l'enseignement de l'informatique graphique, et comment sont équilibrés les enseignements entre aspects

artistiques et connaissances techniques. La formation de Paris 8, ATI, était représentée par Manu Cherem, doctorant de Cédric Plessiet.

En résolution des jeux et apprentissage automatique, des travaux [p3] ont porté sur l'utilisation d'algorithmes d'apprentissage par renforcement sur des jeux, des problèmes de prédiction ou d'optimisation. Les modèles LSTM+Transformer, Transformer et LSTM bidirectionnel ont été étudiés pour des prédictions d'indices financiers (étalon or, paire de devises, ...), en utilisant l'environnement Gymnasium. La combinaison LSTM+Transformer donne les meilleurs résultats. Des travaux [p4] ont débuté sur la question de la généralisation de l'apprentissage pour obtenir des solutions s'adaptant plus aisément aux environnements non inclus dans les ensembles d'entraînement, en utilisant la randomisation de domaine pendant la phase d'entraînement de solution utilisant les algorithmes Deep Q-learning (DQN) et Proximal Policy Optimization (PPO). Les expérimentations ont été réalisées sur la prédiction d'indices financiers (BeshaGym) et l'optimisation de la culture du blé en agriculture (CropGym). Les gains sur les indicateurs utilisés sont de l'ordre de 30%. Des travaux de résolution par apprentissage par renforcement sur les puzzles Sokoban, 2048 et Morpion Solitaire ont été menés sans aboutir à des publications. Des travaux par auto-apprentissage (AlphaZero selfplay learning) ont été réalisés sur le jeu à information complète Breakthrough sans arriver à obtenir de meilleurs résultats qu'une recherche arborescente parallèle MCTS sans apprentissage. Des travaux de recherche arborescente sans apprentissage ont été menés sur Breakthrough et sur le jeu à information imparfaite Chinese Dark Chess. La participation aux computer olympiad 2025 a été réalisée dans la continuité de la collaboration établie avec Jean-Noël Vittaut de l'équipe LFI du LIP6. L'équipe a réalisé les premiers bêta tests de mise en place de la nouvelle plateforme de communication pour la compétition, appelée Discord Computer Olympiad Interface (DCOI). L'ancienne plateforme Ludii a été utilisée pour quelques matchs.

Des travaux effectués sur le développement des notions de connexité de mode intrinsèques et leurs applications au finetuning de modèles fondations ont donné lieu à des résultats positifs et à la rédaction d'un manuscrit actuellement en cours de soumission. La mise en oeuvre d'études sur les mécanismes d'apprentissages des modèles à flot génératifs continuent leur cours, et sont encourageant.

Une intervention grand publique en vulgarisation des méthodes d'IA générative a été réalisée sous forme d'interview dans une série de reportages vidéos sur le sujet au sein du journal Politis [m1].

En traitement automatique du langage naturel et apprentissage automatique, les travaux ont porté sur la conception, la structuration et l'analyse de corpus textuels spécialisés, avec un accent particulier sur les enjeux méthodologiques liés à la collecte, au nettoyage, à l'annotation et à la diffusion de données textuelles issues du web et de sources institutionnelles. Ces travaux ont donné lieu à une publication abordant les défis scientifiques, techniques et éthiques associés à la constitution de ressources textuelles exploitables [p5] pour l'apprentissage automatique. Par ailleurs, les travaux menés dans le cadre du projet MALANTIN ont abouti à la mise en ligne⁵ d'un corpus textuel structuré et documenté [p6], [m2], destiné à l'innovation non technologique à partir de rapports d'entreprises, et accompagné de ressources dérivées exploitables pour des tâches de fouille de textes et de modélisation. À la suite de ces travaux, une action de diffusion scientifique a également été menée à destination des communautés patrimoniales et documentaires, avec une intervention dans un podcast européen [m3], organisé par AI4LAM et une présentation du projet Little_Bot lors de la conférence LIBER, portant sur les usages de l'IA dans les institutions GLAM (Galleries, Libraries, Archives and Museums), en lien avec les archives et la mémoire collective.

5. <https://www.ortolang.fr/workspaces/malantin>, <https://entrepot.recherche.data.gouv.fr/dataset.xhtml?persistentId=doi:10.57745/M9UHMx>

En plus, un travail a été publié par Kyriakoglou et C. N. Zacharopoulos, intitulé *Decoding Emergent Big Five Traits in Large Language Models : Temperature-Dependent Expression and Architectural Clustering* [p7], dans le cadre de conférence IJNLP-AAACL 2025. Ce travail étudie l'émergence de comportements de type « personnalité » dans les grands modèles de langage (LLM), un enjeu important pour le développement et le déploiement responsables des systèmes d'intelligence artificielle. L'étude analyse six LLM à l'aide du cadre psychométrique Big Five Inventory-2 (BFI-2), en évaluant l'expression des traits de personnalité selon différents réglages de température d'échantillonnage. Les résultats mettent en évidence des variations significatives sur quatre des cinq dimensions du modèle des Big Five, en particulier pour le Névrosisme et l'Extraversion, qui se révèlent fortement sensibles aux paramètres de décodage. Une analyse de classification hiérarchique révèle également l'existence de groupes distincts de modèles présentant des profils de traits similaires, suggérant que certaines caractéristiques architecturales influencent la stabilité de ces profils. Ce travail contribue ainsi à une meilleure compréhension des comportements émergents des LLM et de leurs implications pour le réglage des modèles, leur sélection et la gouvernance éthique de l'IA.

Un second travail de Zacharopoulos et Kyriakoglou, *In Machina N400 : Pinpointing Where a Causal Language Model Detects Semantic Violations* [p8] (AICS 2025), s'intéresse aux mécanismes par lesquels un modèle de langage causal détecte des violations sémantiques au cours du traitement d'une phrase. À partir d'un corpus contrôlé de phrases se terminant de manière plausible ou implausible, l'étude analyse les états cachés du modèle phi-2 couche par couche, à l'aide de deux approches complémentaires : un classifieur linéaire par couche et une analyse de la dimension effective des représentations. Les résultats montrent que la détection des anomalies sémantiques émerge principalement dans les couches intermédiaires du modèle, suivie d'une consolidation rapide dans les couches supérieures. Ces observations suggèrent une dynamique de traitement comparable à certains résultats classiques en psycholinguistique humaine, notamment l'idée que la détection des anomalies sémantiques intervient après la résolution syntaxique.

Ensemble, ces deux travaux contribuent à une meilleure compréhension des comportements émergents des LLM, tant du point de vue de leurs propriétés internes que de leurs implications pour l'interprétabilité et les liens entre modèles de langage et cognition humaine.

Les travaux sur les langues minorisées et l'apport des sciences participatives [p9] se poursuivent, en particulier dans le cadre projet participatif Yezh Ar Vro [i2], [p10], dont A. Millour est co-porteuse. Au cours des 6 premiers mois du projet, un stage de 6 mois autour de la reconnaissance vocale du breton a été réalisé (publication soumise à LREC 2026), une développeuse web a été recrutée pour initier l'outillage de la collecte (collecte de données sonores *via* une application mobile et transcription *via* une interface Web participative). Plusieurs focus groups ont été réalisés avec la communauté des utilisateur·ices en amont des développements. Le projet Divæ qui a été déposé à l'ANR entend questionner l'influence des technologies issues de l'IA générative sur la diversité linguistique.

Enfin, les recherches sur l'évaluation des outils de reconnaissance des entités nommées en français en contexte linguistique varié (variation en genres textuels) se sont poursuivies, avec notamment l'encadrement d'un stage de 6 mois [p11] qui a donné lieu à une soumission à LREC 2026 autour de la robustesse des LLMs face à la variation textuelle. M. Seghier a poursuivi ses travaux d'annotation et d'évaluation. Une collaboration avec des chercheurs de Carnegie Mellon University a donné lieu à l'adaptation au français de la librairie `pseudobibeR` pour le français (disponible ici : <https://github.com/browndw/pseudobibeR.fr/tree/main>). Cette librairie permet la vectorisation de corpus en français selon les traits linguistiques définis par Douglas Biber. Le corpus NEM.fr a été mis à disposition. Une publication associée a été soumise à LREC 2026.

En combinatoire, Dupont, Kyriakoglou, Metaftsis, Prassidis et Singh ont travaillé sur le

préprint intitulé *Pairs of tree dessins, their Shabat polynomials, and monodromy groups* [p12]. Cet article étudie une classe particulière de revêtements ramifiés de la sphère de Riemann, donnée par les polynômes de Shabat, et leur correspondance avec les dessins d'enfants, qui sont ici des graphes planaires connexes, bicoloriés, dont la structure est celle d'un arbre. Les auteur-es analysent les passeports produisant des paires de dessins non isomorphes partageant les mêmes données combinatoires, déterminent les polynômes de Shabat associés et étudient les groupes de monodromie correspondants. Ce travail contribue à une meilleure compréhension des relations entre invariants combinatoires, symétries algébriques et structures de revêtements, dans le cadre de la théorie des dessins d'enfants et de la géométrie algébrique.

Enfin, il y a eu la publication de « Asymptotic distribution of parameters in trivalent maps and linear lambda terms » dans le journal *Combinatorial Theory*.

En sécurité et privacy, les travaux de médiations technocritiques des années précédentes se sont poursuivis. Concernant les limites des blockchains, plusieurs interventions universitaires ont été réalisées, aussi bien auprès de collègues en séminaire de recherche [i3] qu'auprès d'étudiant-es de différentes formations de master et ingénieur à Paris et Rennes [i4, i5]. Plusieurs interventions auprès du grand public ont également eu lieu, notamment une invitation dans un podcast [m4] ainsi qu'une intervention plus institutionnelle auprès de la commission des finances du parti socialiste au sujet des enjeux de la fiscalité des cryptomonnaies [m5]. À cela s'ajoute un nouvel article de médiation grand public [a1] ainsi que le montage et la publication des vidéos et podcasts de 7 des 8 interventions du *Unblocked & Unchained Symposium*⁶ organisé fin 2024 [m6, m7, m8, m9, m10, m11, m12]. En dehors de la technocritique des blockchains, il y a eu une intervention en séminaire de recherche sur des questions de privacy et de modélisation du contrôle sur les données personnelles [i6], une intervention en table ronde auprès d'étudiant-es ingénieur sur des questions autour des logiciels libres [m13], et une intervention sur un live de *La Quadrature du Net* au sujet de la politique des technologies [m14]. Enfin, un atelier de médiation sous forme de jeu de rôle grandeur nature sur le fonctionnement d'Internet, avec des notions de cybersécurité, a été testé avec une classe de lycéen-nes dans le cadre d'une collaboration avec l'équipe de médiation de La Cité des Sciences et de l'Industrie [m15].

En enseignement de l'informatique, les travaux du projet GENEPI et EPIAGEN ont donné lieu à la publication d'une *scoping review* [p1] présentée à l'International Conference on Artificial Intelligence in Education (AIED 2025), portant sur l'intégration de l'IA générative dans l'enseignement de l'informatique. En plus, Kyriakoglou et Machado ont également été invitées comme intervenantes lors du workshop AIED « *Generative AI in Education : Exploring Human-AI Collaboration and Culture to Promote Equity and Inclusion* » [i1], afin de parler des enjeux éthiques de l'IA Gén en éducation de l'informatique.

Concernant l'informatique débranchée, A. Millour et J. Lavauzelle ont pour la seconde année organisé et animé un atelier "débranché" autour de l'apprentissage automatique et du jeu de Nim dans le cadre de la Fête de la Science 2025.

Dans le cadre de la semaine de la Fête de la science 2025, une conférence grand public intitulée « IA générative : super outil ou super illusion ? » [m16] a été donnée par A. Pappa, le lundi 6 octobre 2025 à l'Université Paris 8. Cette intervention visait à présenter de manière accessible les principes, les usages et les limites de l'IA générative, en s'appuyant sur des exemples concrets et des enjeux sociétaux actuels. Dans le même cadre, un atelier participatif intitulé « Jouer à débattre » a été animé par A. Pappa, le 7 octobre 2025. Cet atelier proposait une approche ludique et interactive destinée à encourager la discussion et la réflexion critique autour des usages de l'IA et de leurs impacts [m17].

6. <https://unblockchained.sciencesconf.org/>

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

Dans la continuité des années précédentes, le Séminaire d'Ouverture aux Informatiques et de Formation (**SOIF**) a été maintenu avec des présentations portant sur les travaux en cours des membres de l'équipe et de leurs invités. La fréquence des séminaires est restée d'une fois par semaine pour favoriser les échanges autour des différentes questions liées à la formation et à la recherche. Seuls les séminaires avec la présentation de travaux spécifiques sont listés ci-dessous.

Daniel RUIRU (Doctorant dans l'équipe PASTIS du LIASD, Université Paris 8)

Date : 27 novembre 2025

Titre : The Struggle of Generalizing Reinforcement Learning Agents – Is there Hope?

Résumé : One of the biggest challenges facing reinforcement learning is developing agents that generalise in novel scenarios. Often, this problem is driven by the training processes that focus on specific environments, making it a challenge to transfer to new ones. At a high level, this training problem can be solved by exposing agents to a set of experiences. But this advancement requires a reward function that captures real-world domains, which is another tough problem. Moreover, generalisation is largely specific to complex issues (like those of resource allocation in real-world problems). These worldly problems tend to require complex models, which complicates the training and ultimately the generalisation needed. So, is there a solution?

Axel CHEMLA–ROMEU–SANTOS (Post-doctorant en musicologie à l'Université Paris 8)

Date : 20 novembre 2025

Titre : Approches expérimentales et divergentes de la synthèse audio neurale pour la création

Résumé : Alors que l'apprentissage automatique génératif tend à s'intégrer plus largement dans nos écosystèmes numériques, la plupart de ses applications actuelles sont conçues pour des industries axées sur les services, l'automatisation, la normativité et l'imitation efficace. Cette orientation, fondée sur l'extraction et la reproduction à l'échelle mondiale de données produites par l'homme, tend à diviser son accueil entre le réductionnisme techno-positiviste et l'idéalisme anthropocentrique de la création ; dans le même temps, elle entrave l'exploration et le développement d'utilisations locales et créatives de ces technologies. En revenant à une « expérimentation » radicale, nous proposerons une manière qui pourrait aider à réaffilier ces nouvelles techniques à la création artistique. Cette approche, basée sur la déconstruction et la réutilisation des techniques typiques d'apprentissage automatique axées sur les objectifs, vise à encourager la co-création homme-machine et la surprise mutuelle grâce à un échange dynamique entre l'exploration et l'attente dans le processus d'apprentissage. Plutôt que d'imposer l'uniformité et la conformité, la complexité de ces outils peut devenir une opportunité d'élargir notre utilisation des technologies génératives et d'embrasser leur potentiel co-créatif en tant que nouvelles formes de matérialité cybernétique. En combinant des approches techniques et artistiques à la pointe de la technologie, nous visons à souligner les implications artistiques, esthétiques et scientifiques de ces utilisations. À partir d'une description des outils que nous avons développés à l'IRCAM au sein du collectif ACIDS et de leur intégration dans des projets artistiques personnels et collaboratifs, nous souhaitons développer une perspective critique sur la manière dont ce changement d'approche pourrait favoriser l'intégration significative des technologies génératives dans nos écologies créatives.

Gaël LEJEUNE (Maître de conférences en informatique, STIH, Sorbonne Université)

Date : 26 juin 2025

Titre : Pour une approche descendante en TAL : des modèles de documents qui guident les modèles de langue

Résumé : Les grands modèles de langue sont devenus ubiquitaires dans la recherche en TAL mais ils restent généralement exploités pour exploiter des relations locales entre les mots. Les dimensions du corpus traité, du genre textuel auquel il appartient et aussi de la structure des documents sont globalement ignorés. Pourtant, de nombreux indices utiles peuvent être dérivés du grain corpus et du grain document que ce soit pour des tâches de classification, de recherche ou d'extraction d'information. Je présenterai des exemples de recherches menées en ce sens sur des corpus d'articles scientifiques et de presse d'une part et sur des corpus de documents anciens d'autre part.

Cédric PLESSIET (Professeur en art numérique, AIAC, Université Paris 8)

Date : 5 juin 2025

Titre : Projet Kellynoïde : de la création d'un double virtuel à son utilisation dans les arts numériques

Résumé : Depuis plus de dix ans, Cédric Plessiet travaille sur les personnages virtuels photo-réalistes interactifs. Dans le cadre du projet Kellynoïde, il collabore avec l'artiste et performeuse Kelly Mezino pour créer un double virtuel semi-autonome de cette dernière, en utilisant des technologies issues des effets spéciaux et des jeux vidéo. Ce personnage virtuel a été utilisé dans la création de deux œuvres d'art numérique en réalité virtuelle. La première "Line" utilise un réseau de neurones artificiels pour offrir une expérience immersive basée sur l'empathie et la confiance entre le spectateur et l'entité virtuelle. La deuxième "Amusique" fait usage d'impression 3D, de module Arduino et d'un moulage du buste de Kelly Mezino, truffé de capteurs et de contrôleurs Arduino, pour créer une expérience sensorielle interactive directement inspiré de l'œuvre "le violon d'Ingres" du photographe surréaliste Man Ray. Ces œuvres ont été présentées dans plusieurs festivals et manifestations internationales, et ont fait l'objet de publications scientifiques.

Guillaume MOINARD (Doctorant, équipe ComplexNetworks, LIP6, Sorbonne Université)

Date : 15 mai 2025

Titre : BeWater : Des rassemblements efficaces grâce aux bassins versants dans les réseaux de rues

Résumé : Lors de mouvements sociaux, tels que des protestations ou des manifestations, il arrive que des contre-forces dispersent les manifestants et confisquent ou surveillent leurs moyens de communication. Pour que leur action perdure, ces marcheurs veulent se rassembler dans la ville en groupes suffisamment importants, quand bien même ils n'ont pas de moyen de communiquer. Dans ce scénario, ils ont besoin d'un protocole distribué et aussi simple que possible, qui utilise exclusivement les éléments observables dans leur environnement direct. Nous présentons un nouvel algorithme de marche, BeWater, pour les rassemblements sur des réseaux urbains pondérés. Nous montrons que son efficacité dépend de l'observable utilisé par les marcheurs pour pondérer les liens (sa longueur, sa largeur, etc.). Pour une ville donnée, nous sommes en mesure d'expliquer l'efficacité d'une observable en étudiant les rassemblements de marcheurs comme des gouttes d'eau suivant les pentes d'un bassin versant. Nous proposons enfin des tactiques très efficaces en combinant plusieurs observables. Les groupes ainsi formés sont robustes, c'est à dire capables de se reformer rapidement en cas d'une nouvelle dispersion.

Bastien LEGAY (Doctorant, équipe ComplexNetworks, LIP6, Sorbonne Université)

Date : 15 mai 2025

Titre : Analyse de perturbations dans les graphes temporels : infrastructures de transport et mouvements sociaux

Résumé : De nombreux événements produisent des perturbations dans les infrastructures organisées en réseaux, tels que les réseaux de transport. C'est notamment vrai dans le cadre des

mouvements sociaux, qui peuvent avoir un retentissement médiatique fort (d'une manifestation intersyndicale classique au blocage plus localisé d'une route, d'un lieu). Toutefois il est en général bien plus difficile de comprendre l'impact matériel et l'étendue de ces perturbations : comment le fonctionnement normal du réseau est-il ainsi altéré ? Est-ce global ou localisé ? Quelle ampleur est-ce que cela prend ? Comment se passe le retour à l'équilibre ? Pour répondre à cela, nous proposons une méthode basée sur la différentiation et le clustering de link streams, un formalisme de graphes temporels ayant pour objectif de faire le lien entre graphes et séries temporelles. Nous souhaitons ainsi, à partir d'un formalisme intégrant des données réelles, obtenir une compréhension fine de l'étendue à la fois en temps et en espace (structure) d'une perturbation. Nous évaluons cette méthode sur différents toymodels et l'appliquons dans le cadre du trafic maritime mondial pour analyser l'impact de la grève des dockers ayant affecté les ports de la côte Ouest des États-Unis en 2002.

Alain BONARDI (Professeur en informatique et création musicale, Musidance, Université Paris 8)

Date : 10 avril 2025

Titre : Machine learning pour la synthèse et la spatialisation du son

Résumé : La grande majorité des modèles d'IA utilisés en musique (SUNO, NSynth, Jukebox, Audiocraft, MusicFX, etc.) se concentrent sur la génération de mélodies, des accords qui les accompagnent, voire de paroles, dans des styles donnés. Au-delà de la prévisibilité du contenu, leur qualité sonore est très moyenne, et leur diffusion sonore limitée à une restitution en mono ou en stéréo. Nous présenterons des travaux de recherche récents qui s'intéressent à la synthèse du son pour elle-même ou à sa spatialisation, voire à l'imbrication de ces deux dimensions, et montrerons les conceptions de la musique et du son que sous-tendent ces modèles.

Alexandros SINGH (Maître de conférences en informatique, LIASD, Université Paris 8)

Date : 6 mars 2025

Titre : Proofs and programs

Résumé : The goal of this talk is to provide a gentle introduction to the correspondence between proofs and programs. Starting with a historical overview of the debate between constructivist and classical views of mathematics, we build up to the Curry-Howard correspondence between implicational intuitionistic logic and the simply-typed lambda-calculus.

10 Productions scientifiques pour l'année 2025

10.1 Publications en revues et conférences, productions

Un astérisque marque l'auteur-ice ayant présenté les travaux lors de la conférence.

- [p1] Sonia Proust-Androwkha, Valéry Psyché, **Revekka Kyriakoglou**, Anis Boubaker, Sébastien Lallé, Lamprini Chartofylaka, Guilherme Medeiros Machado, and Nour El Mawas. Beyond programming assistance towards the integration of generative ai into computer science education : A scoping review. In *International Conference on Artificial Intelligence in Education*, pages 380–389. Springer, 2025.
- [p2] Céline Roudet, **Farès Belhadj**, Basile Sauvage, Maxime Maria, Guillaume Gilet, and **Jean-Jacques Bourdin**. Four years of sharing teaching practices within the French Computer Graphics community. In Rafael Kuffner dos Anjos and Karina Rodriguez Echavarria, editors, *Eurographics 2025 - Education Papers*. The Eurographics Association, 2025.
- [p3] **Daniel Ruiru***, **Nicolas Jouandeau**, and Dixon Owuor. Recurrent Neural Networks with Transformers to Trade Financial Instruments. In *Science Nature Computer Science (SNCS)*, pages 1–23, 2025.
- [p4] **Daniel Ruiru***, **Nicolas Jouandeau**, and Dixon Owuor. Towards Generalizing Deep Reinforcement Learning Algorithms For Real World Applications. In *International Joint Conference on Computational Intelligence SCITEPRESS (IJCCI)*, pages 1–18, 2025.
- [p5] **Anna Pappa**. Corpus textuels : enjeux et défis. *1024*, 25 :109–121, 06 2025.
- [p6] **Anna Pappa** and **Revekka Kyriakoglou**. Données de répliation pour : MALANTIN, 2025.
- [p7] Christos-Nikolaos Zacharopoulos and **Revekka Kyriakoglou**. Decoding emergent big five traits in large language models : Temperature-dependent expression and architectural clustering. *International Joint Conference on Natural Language Processing (IJCNLP)*, 2025.
- [p8] Christos-Nikolaos Zacharopoulos and **Revekka Kyriakoglou**. In machina n400 : Pinpointing where a causal language model detects semantic violations. *33rd International Conference on Artificial Intelligence and Cognitive Science*, 2025.
- [p9] Laure Turcati, **Alice Millour**, Renaud Debailly, Karën Fort, Asma Steinhausser, Corentin Biets, and Anne Dozières. Citizen Science in Practice : How (not) to Fail? In *Citizen Science : Theory and Practice*. 2025. 11p. SJR "Q1".
- [p10] **Alice Millour***, Mélanie Jouitteau, and Loïc Grobol. Yezh Ar Vro-La langue du pays : construire l'appropriation des applications de collecte. In *Language Technologies for All, LT4All 2025*, UNESCO, Paris, France, 2025.
- [p11] Ziyang Xu, **Alice Millour**, Carlos-Emiliano Gonzalez-Gallardo*, and Jean-Yves Antoine. Adaptabilité des Grands modèles de Langage Aux Genres Linguistiques Attestés (AGLAGLA). In *JIRC 2025 : Journées Informatique en Région Centre-Val de Loire*, Blois, France, 2025.
- [p12] **Benjamin Dupont**, **Revekka Kyriakoglou**, Vassilis Metaftsis, Efstratios Prassidis, and **Alexandros Singh**. Pairs of tree dessins, their shabat polynomials, and monodromy groups. *arXiv preprint arXiv :2510.10192*, 2025.

10.2 Interventions académiques sans publication

- [i1] **Revekka Kyriakoglou** and Guilherme Medeiros Machado. Ethical Topics in AI in Education : Insights from a Scoping Review. Workshop Generative AI in Education : Exploring

Human-AI Collaboration and Culture to Promote Equity and Inclusion, AIED. 22 juillet 2025.

- [i2] **Alice Millour**. Variation(s) : Produire des Ressources au Service de l'Évaluation en TAL. In *Séminaire du GETALP : Groupe d'Étude en Traduction Automatique/Traitement Automatisé des Langues et de la Parole*, Grenoble, France, 2025.
- [i3] **Pablo Rauzy**. La vérité sur la blockchain. Séminaire de l'équipe ComplexNetworks du LIP6. 21 janvier 2025.
- [i4] **Pablo Rauzy**. Introduction technocritique aux blockchains. Séminaire dans le cadre du cours "Blockchain & Consensus : Coopération dans les systèmes distribués" à l'IMT Atlantique. 4 mars 2025.
- [i5] **Pablo Rauzy**. Introduction technocritique aux blockchains. Séminaire auprès des étudiant-es du parcours "Sécurité de Systèmes Informatiques" du Master informatique de Sorbonne Université. 11 avril 2025.
- [i6] **Pablo Rauzy**. Modéliser le contrôle des données personnelles avec Capacity. Séminaire de l'équipe DALGO du LIS à Marseille. 6 mai 2025.

10.3 Publications grand public / médiation / valorisation

- [a1] **Pablo Rauzy**. Les calculs d'une preuve de travail doivent être inutiles. article de blog. 14 janvier 2025.

10.4 Interventions grand public / médiation / valorisation

- [m1] **Louis Falissard**. IA quoi ? Les vrais dangers de l'intelligence artificielle. Interview/vidéo par Politis. 7 février 2025.
- [m2] **Anna Pappa** and **Revekka Kyriakoglou**. Malantin, 2025. ORTOLANG (Open Resources and TOols for LANGuage) www.ortolang.fr.
- [m3] **Anna Pappa**. The Use of AI in GLAM. Podcast, Intervention dans la série *Humans of AI*, AI4LAM. 7 octobre 2025.
- [m4] **Pablo Rauzy**. Se libérer de la blockchain. Invité d'un épisode du podcast *La Cantine de Brest*. 28 octobre 2025.
- [m5] **Pablo Rauzy**. Introduction technocritique aux blockchains. Intervention auprès de la commission finance du PS au sujet des enjeux de la fiscalité des cryptos. 28 octobre 2025.
- [m6] **Pablo Rauzy**. Introduction technocritique aux blockchains (UnU 1/8). Montage vidéo et podcast des interventions du Unblocked & Unchained Symposium (13/12/2024) pour son blog par Pablo Rauzy. 6 août 2025.
- [m7] Levy, Arnaud. Alternative low-tech aux blockchains : l'exemple des diplômes (UnU 2/8). Montage vidéo et podcast des interventions du Unblocked & Unchained Symposium (13/12/2024) pour son blog par Pablo Rauzy. 20 août 2025.
- [m8] Salamatian, Kavé. La chaîne de blocs : imaginaire géopolitique, et/ou mirage technologique ? (UnU 3/8). Montage vidéo et podcast des interventions du Unblocked & Unchained Symposium (13/12/2024) pour son blog par Pablo Rauzy. 3 septembre 2025.
- [m9] Hadjadji, Nastasia. Comment Bitcoin et les cryptos ont envoûté la planète ? (UnU 4/8). Montage vidéo et podcast des interventions du Unblocked & Unchained Symposium (13/12/2024) pour son blog par Pablo Rauzy. 17 septembre 2025.

- [m10] Cayla, David. Peut-on dépolitiser la monnaie? (UnU 6/8). Montage vidéo et podcast des interventions du Unblocked & Unchained Symposium (13/12/2024) pour son blog par Pablo Rauzy. 1 octobre 2025.
- [m11] Lakowski-Laguerre, Odile. Analyse de Bitcoin au filtre d'une théorie institutionnaliste de la monnaie (UnU 7/8). Montage vidéo et podcast des interventions du Unblocked & Unchained Symposium (13/12/2024) pour son blog par Pablo Rauzy. 15 octobre 2025.
- [m12] Mitteau, Gilles. Les limites d'une économie disposant d'une quantité finie de monnaie (UnU 8/8). Montage vidéo et podcast des interventions du Unblocked & Unchained Symposium (13/12/2024) pour son blog par Pablo Rauzy. 29 octobre 2025.
- [m13] **Pablo Rauzy** and Nicolas Vivant. Logiciels libres : quels enjeux? Participation à une table ronde sur les logiciels libres auprès des étudiant·es de l'ENSIMAG à Grenoble. 2 décembre 2025.
- [m14] **Pablo Rauzy** and Mathilde Saliou. La fasciation est-elle inscrite dans le code? Invité dans un live de La Quadrature du Net sur la question de la politique de la technique. 3 décembre 2025.
- [m15] **Pablo Rauzy** and Laetitia Ferreri. OSI : on simule internet. Création et encadrement d'un atelier de médiation à La Cité des Sciences auprès d'une classe de lycéen·nes. 15 mai 2025.
- [m16] **Anna Pappa**. IA générative : super outil ou super illusion? Conférence grand public, Fête de la science 2025, Université Paris 8. 6 octobre 2025.
- [m17] **Anna Pappa**. Jouer à débattre. Animation atelier participatif pour les usages de l'IA et leurs impacts, Fête de la science 2025, Université Paris 8. 13 octobre 2025.