

Journée du club des partenaires du GDR TAL

13 mars 2025

Géraldine Damnati, Orange INNOV/DATA-AI

Vincent Claveau, AMIAD

Frédéric Béchet, LIS, Aix-Marseille Université

90 participants
31 organisations

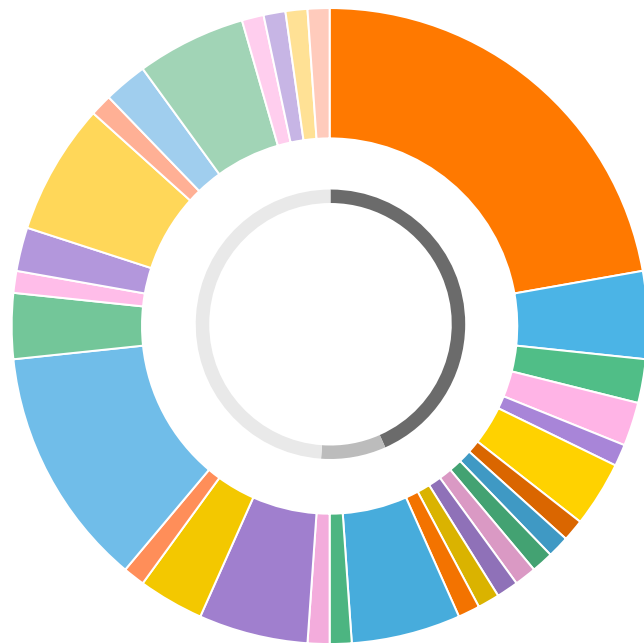
13 entreprises (39 p)

3 institutions publiques (7 p)

15 institutions académiques (44 p)

5 doctorants CIFRE

- Orange Innovation
- Ouest France (LabCom)
- NVIDIA
- HuggingFace
- LINAGORA
- Airbus
- RTE (Réseau de Transport d'Electricité)
- Lalilo by Renaissance
- TEKLIA
- SNCF
- Chapsvision (+INALCO)
- Alten
- Alphabets
- AMIAD
- CHU Nantes
- INRAE
- IRISA, Université de Rennes
- IRISA, Université Bretagne Sud
- LIDILE, Université Rennes 2
- LS2N, Nantes Université
- LIUM, Université du Mans
- FGES, Université Catholique de Lille
- Centrale Supélec
- LISN, Université Paris Saclay
- LORIA, Université de Lorraine
- LITIS, Université de Rouen Normandie
- LIS, Aix-Marseille Université
- ENS, Université Paris Sciences & Lettres
- CERES, Sorbonne Université
- LIPN, Université Sorbonne Paris Nord
- ENIT, Tunisie



Programme de la journée

- 10:00 - Introduction
- 10:30 - **Clémentine Fourier** (HuggingFace),
Panorama de l'évaluation des LLM
- *11:15 - pause café*
- 11:45 - **Christophe Cerisara** (LORIA) et **Jean-Pierre Lorre** (Linagora),
Présentation du modèle Lucie 7B et panorama des travaux du projet LLM4ALL
- *12:30 - repas*
- 13:45 - **Laurent Amsaleg** (IRISA) et **Michel Le Nouy** (Ouest France)
Présentation du labcom Synapses
- 14:30 - **Meriem Bendris** (NVIDIA),
Tour d'horizon des solutions Generative AI de NVIDIA
- 15:15 – Posters / démos
- 16:30 – **Delphine Charlet** (Orange),
Outils d'évaluation des LLMs pour les cas d'usage métier
- *17:15 - le mot de la fin*

Projets collaboratifs ANR - TSIA

Projet	Partenaires	Responsable
FINLAM Modèles de fondation intégrés pour les Bibliothèque, Archives et Musées* TSIA 2023	• TEKLIA TEKLIA • BnF Bibliothèque nationale de France • LITIS Laboratoire d'Informatique Traitement de l'Information et des Systèmes	Thierry Paquet
InExtenso Evaluation intrinsèque et extrinsèque des biais dans les grands modèles de langue TSIA 2023	• LORIA Laboratoire lorrain de recherche en informatique et ses applications • CHU Rouen CENTRE HOSPITALIER UNIVERSITAIRE DE ROUEN • LISN Laboratoire Interdisciplinaire des Sciences du Numérique	Karën Fort
LLM4All LLM à jour pour tous TSIA 2023	• LINAGORA LINAGORA GSO • PRIME APHP.Centre (Prévention-Recherche-Innovation-Médicaments et Ethique) • LORIA Laboratoire lorrain de recherche en informatique et ses applications • LIX Laboratoire d'informatique de l'École polytechnique	Christophe CERISARA
MALADES Grands modèles de langue adaptables et souverains pour le domaine médical français TSIA 2023	• CHUN CHU de Nantes • LIS Laboratoire d'Informatique et Systèmes • LIA Laboratoire Informatique d'Avignon • LS2N Laboratoire des Sciences du Numérique de Nantes	Richard Dufour
Pantagruel Construction et d'évaluation de grands modèles de langue multimodaux et inclusifs (écrit, oral, pictogrammes) pour le français général et clinique TSIA 2023	• LLF laboratoire de linguistique formelle • CREST Centre de Recherche en Economie et Statistique – CREST • INA Institut national de l'audiovisuel • LIG Laboratoire d'Informatique de Grenoble • LIA Laboratoire d'Informatique d'Avignon	Didier Schwab
TRaLaLaM Traduire avec les Large Language Models TSIA 2023	• ISIR Institut des Systèmes Intelligents et de Robotique • SYSTRAN • INRIA Paris Centre de Recherche Inria de Paris	Josep CREGO

Projets collaboratifs ANR - AAPG

Projet	Partenaires	Responsable
CHICA-AI Entraîner les compétences de compréhension en lecture des enfants avec une intelligence artificielle adaptée AAPG2023	• EMC LABORATOIRE D'ETUDE DES MECANISMES COGNITIFS • LALILO • LIUM LABORATOIRE D'INFORMATIQUE DE L'UNIVERSITE DU MANS (LIUM) • IRIT Institut de Recherche en Informatique de Toulouse	Lucile GELIN
RENAISSANCE Synthèse Texte-Vidéo pour Applications Créatives AAPG2023	• Obvious • ISIR Institut des Systèmes Intelligents et de Robotique	Hugo CASELLES-DUPRÉ
EVA Expliciter les Attributs de la Voix AAPG2023	• LIA Laboratoire d'Informatique d'Avignon • ORANGE • INSTITUT DE RECHERCHE ET COORDINATION ACOUSTIQUE MUSICALE (IRCAM) • IRISA Institut de Recherche en Informatique et Systèmes Aléatoires • LPP Laboratoire de Phonétique et Phonologie	Olivier LE BLOUCH
ECLADATTA Extraction de Connaissances Latentes dans les Documents par une Analyse conjointe des Textes et des Tables AAPG2022	• EURECOM EURECOM • Orange ORANGE SA • IRIT Université Toulouse 3 - Paul Sabatier	Yoan CHABOT
MINERAL Représentation multi dimensionnelle pour la modélisation des conversations AAPG2024	• LISN Laboratoire Interdisciplinaire des Sciences du Numérique • LIS Laboratoire d'Informatique et Systèmes • Orange ORANGE SA	Camille GUINAUDEAU

Posters / démos

AdminSet and AdminBERT: a Dataset and a Pre-trained Language Model to Explore the Unstructured Maze of French Administrative Documents,

Sebbag Thomas, Quiniou Solen, Stucky Nicolas, Morin Emmanuel, 1 - TALN-LS2N (Nantes Université)

Cadre de rédaction de tests logiciels en langage naturel

Ronzon Mathis, 1 - GESTION DES DONNÉES ET DE LA CONNAISSANCE (France), 2 - Alten (France)

Factual Knowledge Assessment of Language Models Using Distractors

Ammar Khodja Hichem, 1 - Orange Innovation [Lannion], 2- Laboratoire d'Informatique et des Systèmes (LIS) (Aix-Marseille Université)

Génération automatique de datasets de questions diversifiées pour l'évaluation des systèmes RAG à partir de corpus variés

Jourdain Louis, Hellal Skander, 1 - Chapsvision (France)

Part-Of-Speech Sensitivity of Routers in Mixture of Experts Models

Antoine Elie, 1 - Laboratoire d'Informatique et des Systèmes (LIS) (Aix-Marseille Université)

Projet POPCORN: POPulation de bases de CONnaissances par Réseaux Neuronaux

Oliveri Ulysse, 1 - Airbus Defence & Space (France), IRISA, Lannion

Statistical Deficiency for Task Inclusion Estimation

Fosse Loïc, 1 - Orange Innovation [Lannion], 2- Laboratoire d'Informatique et des Systèmes (LIS) (Aix-Marseille Université)

The Lucie 7B LLM and the Lucie training dataset: open resources for multilingual language generation

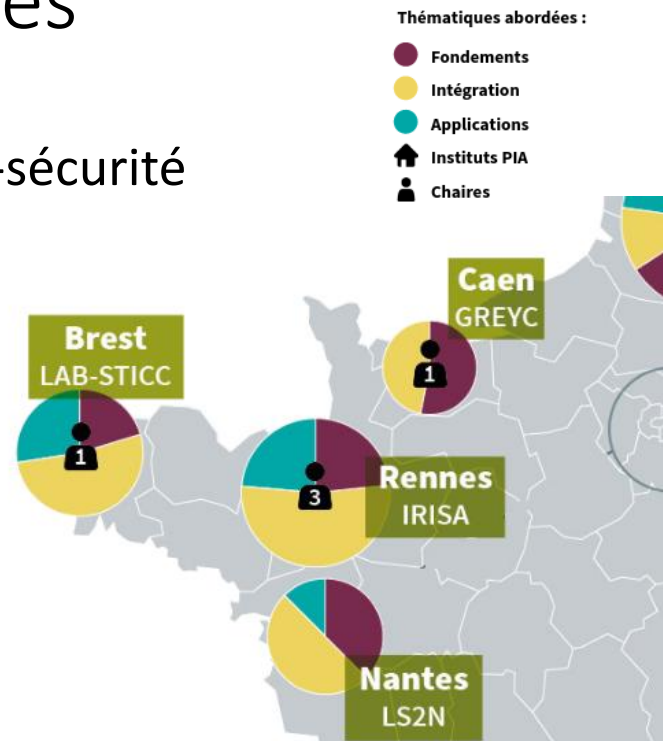
Lorré Jean-Pierre, 1 - LINAGORA (France)

Tired Transformers for Hierarchical Structure Understanding

Mocaer William , 1 - LITIS EA4108, University of Rouen Normandie, France (France)

TAL et IA en Bretagne et à Rennes

- Fort historique en télécommunication et cyber-sécurité
 - Entreprises innovantes
 - Pole image et réseaux
 - Pôle Excellence cyber
- Accélération sur l'IA:
 - IA-Cluster SequoIA (IRISA, Orange...)
 - TIARe (Univ Rennes, ENS, INSA, ENSAI, Agro, U. Re
 - Le Pool French Tech



+ Vannes Lannion : parole expressive

+ Angers : génération de texte

Source : CNRS

TAL et IA au Ministère des Armées



AMIAD

- accélérer le développement, l'intégration et le déploiement à grande échelle des technologies de l'IA
- Pôle technique à Bruz, pôle recherche à Saclay
- TAL, parole, multimédia, parole ; domaines de spécialité, langues rares, données bruitées ; de l'embarqué au clusters GPU

+ 150 agents à
embaucher
d'ici fin 2026 !

Financements pour la recherche

- Pôle recherche de l'AMIAD
- Thèse, projets RAPID ou ASTRID de l'AID <https://www.defense.gouv.fr/aid/appels-a-projets-0>

Réserve opérationnelle numérique

- vos compétences en IA quelques jours par an

Actualités

Atelier evalLLM – évaluation des LLM

- 30 juin à Marseille, conjoint à CORIA-TALN
- Appel à communication en cours
- 2 challenges
 - extraction d'information (entités, relations) en few-shot
 - fine-tuning de LLM sur un domaine spécialisé
 - places CAID à gagner
- <https://evalllm2025.sciencesconf.org/>