

Programme JSALT – Version Non détaillée

Pour l'anniversaire de ses 30 ans, le workshop JSALT s'invite pour la seconde fois en Europe, à l'Université du Mans. Après deux semaines d'école d'été, ce sont plus de 80 chercheur.e.s provenant de 30 institutions dans 17 pays qui travailleront de concert pendant 6 semaines pour construire le futur de l'intelligence artificielle appliquée au langage écrit et oral (<https://jsalt2023.univ-lemans.fr/>). À cette occasion, l'Institut Informatique Claude Chappe de l'Université du Mans, porteur de l'évènement organise une journée de sensibilisation à l'IA pour le grand public le 26 juin au cinéma Pathé Quinconces. Avec les modèles génératifs comme MidJourney, Dall-e et ChatGPT, l'Intelligence artificielle a récemment attiré beaucoup d'attention et les changements que laissent présager ces avancés technologiques concernent l'ensemble de la société. Le 26 juin, nous vous proposons, nous, citoyen.ne.s, chercheur.e.s, expert.e.s, décideur.e.s, de nous réapproprier cette actualité sur les avancées de l'IA. Après avoir différencié les fantasmes de la réalité, nous envisagerons la valeur ajoutée des récentes avancées pour les entreprises et considérerons les challenges éthiques, juridiques et environnementaux à relever pour leur acceptation et leur intégration vertueuse dans la société.

Session 1 - Matin : 9h-12h30

IA : De nouveaux usages permis par les progrès récents ?

- Faut-il avoir peur du grand méchant GPT ? Démystifions les modèles de langues.
Djamé Seddah, INRIA Paris
- Réflexions sur une recherche privée en TAL à l'ère de GPT4.
Laurent Besacier, Chef de l'équipe TAL de Naver.Labs
- Quelques challenges pour évaluer quantitativement les sujets abordés par nos clients lors des enquêtes de satisfaction
Maxence Jeunesse, Directeur de la recherche en IA à Covea
- Le LIUM et JSALT : le futur de l'IA pour le texte et la parole
Anthony Larcher, LIUM, Professeur à Le Mans Université
- L'IA au service de la recherche - exemples à Le Mans Université
 - Explorer et annoter les corpus textuels avec l'IA : le projet READ-IT
François Vignale, 3.LAM, Conservateur en chef des bibliothèques
 - L'IA embarquée au service des données hétérogènes : le projet TIFAIFAI
Valérie Renault et Florent Carlier, CREM, Maîtres de Conférences à Le Mans Université

- Temps d'échanges animé par Josep Crego, Directeur de la recherche de Systran et Romain Sambarino, CEO d'Allomédia
Avec Laurent Besacier, Hatim Bourfoune (Ingénieur de recherche à l'IDRIS), Maxence Jeunesse, Djamé Seddah.

Session 2 - Après-midi : 14h-17h30

IA : Quels challenges pour l'acceptabilité ?

- La couverture de l'IA par les médias, une économie de la promesse et de la critique
Jean-Philippe Cointet, Enseignant chercheur Sciences po Paris, MediaLab
- L'IA responsable. Rêve ou réalité ?
Vincent Courboulay, Maître de Conférences HDR à La Rochelle Université et Directeur scientifique de l'Institut du Numérique Responsable
- Les enjeux éthiques de l'IA vus par le prisme du traitement automatique des langues
Karèn Fort, Maîtresse de Conférences HDR à Sorbonne Université, Loria, Co-chair du comité d'éthique ACL
- Enjeux juridiques et éthiques des systèmes d'IA : personnalité, vie privée, projets de législation
Céline Béguin, Maître de Conférences au Themis, Le Mans Université
Magali Bouteille-Brigant, Maître de Conférences en droit privé – HDR, Co-directrice du Themis, Le Mans Université
- Régulation des systèmes d'IA : l'action de la CNIL
Félicien Vallet, Responsable IA à la CNIL
- Temps d'échanges animé par Céline Béguin
Avec Magali Bouteille-Brigant, Jean-Philippe Cointet, Vincent Courboulay, Karèn Fort, Félicien Vallet

Programme JSALT – Version détaillée

Pour l'anniversaire de ses 30 ans, le workshop JSALT s'invite pour la seconde fois en Europe, à l'Université du Mans. Après deux semaines d'école d'été, ce sont plus de 80 chercheur·e·s provenant de 30 institutions dans 17 pays qui travailleront de concert pendant 6 semaines pour construire le futur de l'intelligence artificielle appliquée au langage écrit et oral (<https://jsalt2023.univ-lemans.fr/>). À cette occasion, l'Institut Informatique Claude Chappe de l'Université du Mans, porteur de l'évènement organise une journée de sensibilisation à l'IA pour le grand public le 26 juin au cinéma Pathé Quinconces. Avec les modèles génératifs comme MidJourney, Dall-e et ChatGPT, l'Intelligence artificielle a récemment attiré beaucoup d'attention et les changements que laissent présager ces avancées technologiques concernent l'ensemble de la société . Le 26 juin, nous vous proposons, nous, citoyen.ne.s, chercheur.e.s, expert.e.s, décideur.e.s, de nous réapproprier cette actualité sur les avancées de l'IA. Après avoir différencié les fantasmes de la réalité, nous envisagerons la valeur ajoutée des récentes avancées pour les entreprises et considérerons les challenges éthiques, juridiques et environnementaux à relever pour leur acceptation et leur intégration vertueuse dans la société.

Session 1 - Matin

IA : De nouveaux usages permis par les progrès récents ?

- **9h – Introduction de la journée**
Pascal Leroux, Président de Le Mans Université et Anthony Larcher, professeur à Le Mans Université
- **9h30 – 10h : Démystifions les modèles de langages**
Djamé Seddah, INRIA Paris

Résumé

Au moment où toute prise de positions publiques fait soit l'objet d'une mémisation instantanée ou est passée au crible d'un fact-checking tout aussi rigoureux que le plus souvent invisible, l'apparition d'outils performants de génération de texte à la simplicité d'emploi confondante interroge, questionne, voire inquiète parfois. La plupart de ces outils accessibles au grand public sont en effet des pures boîtes noires à propos desquelles on ne sait rien ou presque : leurs données d'entraînement ? rien. Leurs architectures ? si peu. Leurs performances ? le plus souvent au doigt mouillé. Est-ce qu'on peut moins détecter leur contenu et par conséquent leurs éventuelles influences ? pas vraiment, même OpenAI, créateur de ChatGPT rapporte des taux de succès de 26%. Dans cette présentation, je ferai un tour d'horizon des principaux modèles de langues, questionnerai leur pertinence dans un milieu académique et aborderai la question de leur détectabilité dans des contextes antagonistes (ou "adversariaux" en français).

Bio

Djamé Seddah est Maître de conférences à Paris Sorbonne Université et en détachement à l'INRIA Paris dans l'équipe Almanach. Ses centres d'intérêts couvrent le domaine du traitement du langage naturel, principalement l'analyse syntaxique multilingue à large couverture, l'interface syntaxe-sémantique, les modèles de langues pour les langues peu dotées, etc. Spécialiste en constitution de corpus annotés (corpus Sequoia, French Social Media Bank, French Question Bank, Narabizi Treebank, etc.), il a participé à l'élaboration des modèles de langues CamemBERT, PagnolXL, CamemBERTa et des modèles à base de caractères pour des langues dialectales et fortement bruitées.

- **10h00-10h20 -Réflexions sur une recherche privée en TAL à l'ère de GPT4.**

Laurent Besacier, Chef de l'équipe TAL de Naver.Labs

Résumé

Dans cette intervention, je partagerai mes réflexions sur la recherche (privée) en TAL à l'heure de GPT4, à partir de mon expérience et mes propres directions de recherche au sein du centre de Naver Labs Europe.

Bio

Laurent est chercheur sénior au Naver Labs Europe depuis Janvier 2021 où il dirige désormais un groupe de recherche sur les systèmes interactifs pour la robotique. Avant cela, il est devenu professeur à l'Université Grenoble Alpes (UGA) en 2009, où il a dirigé le groupe GETALP (Traitement du langage naturel et de la parole). Laurent est toujours affilié à l'UGA. Son expertise principale en recherche et ses intérêts se situent dans le domaine du traitement automatique du langage naturel, la reconnaissance automatique de la parole, la traduction automatique, les langues peu dotées et l'évaluation des systèmes de TAL.

- **10h20- 10h40 - Quelques challenges pour évaluer quantitativement les sujets abordés par nos clients lors des enquêtes de satisfaction.**

Maxence Jeunesse, Directeur de la recherche en IA à Covea

Bio

Docteur en mathématiques appliquées (CERMICS/ENPC), diplômé d'un master de finance quantitative (Paris 6), polytechnicien (X2005). Après un début de carrière dans le domaine des services financiers, en tant qu'ingénieur quantitatif, puis en tant que data-scientist développant des solutions innovantes pour la gestion et le trading, j'ai rejoint une entreprise de conception logicielle en machine learning, avant d'avoir la chance d'intégrer Covéa en septembre 2021, en tant que directeur scientifique de l'intelligence artificielle, pour contribuer au

développement des solutions d'Intelligence Artificielle au service des grands domaines d'assurance du groupe Covéa.

10h40-10h55 Pause

- **10h55-11h15 - Le LIUM et JSALT : le futur de l'IA pour le texte et la parole**

Anthony Larcher, LIUM, Professeur à Le Mans Université

Résumé

Nous présenterons dans un premier temps les activités de recherche de l'équipe LST du Laboratoire d'Informatique de l'Université du Mans (LIUM). L'équipe LST exploite les techniques d'apprentissage profond dans de nombreux domaines du traitement du langage incluant la reconnaissance du locuteur et des langues, la transcription, la traduction, la compréhension, la caractérisation et la synthèse de la parole expressive ou le traitement du langage naturel. À travers ses activités nous discuterons des enjeux présent et à venir dans ces domaines. Nous présenterons ensuite le workshop JSALT dont la 30e édition est hébergée par le LIUM du 26 juin au 4 août. Pendant cette période, plus de 70 chercheurs de 16 pays et plus de 25 institutions collaboreront pour faire progresser les connaissances autour du langage.

Bio

Après avoir obtenu ma thèse en informatique à l'Université d'Avignon et travaillé 4 ans pour l'Institute for Infocomm Research de Singapour (A*STAR), j'ai rejoint Le Mans Université où je suis professeur en informatique et directeur de l'Institut Informatique Claude Chappe. Mes recherches portent sur le traitement automatique de la parole, la reconnaissance du locuteur et des langues, l'apprentissage continu et l'interprétabilité. Je coordonne le projet Européen H2020 Esperanto et organise JSALT 2023 pour faire progresser la recherche en parole en collaboration avec de nombreux partenaires internationaux.

- **11h15-11h45 - L'IA au service de la recherche - exemples à LMU**

- Explorer et annoter les corpus textuels avec l'IA : **le projet READ-IT**

François Vignale, 3.LAM, Conservateur en chef des bibliothèques

Résumé

Le projet européen READ-IT (2017-2022) coordonné par Le Mans Université a eu pour objectif de mettre au point une suite d'outils permettant aux chercheurs de SHS d'explorer automatiquement des textes numérisés en 6 langues du 18e siècle à aujourd'hui à la recherche de témoignages de lecteurs ordinaires afin de répondre aux questions encore non résolues de savoir pourquoi et comment on lit en Europe. Il aurait été impossible d'y parvenir sans mobiliser l'IA et notamment le TAL dans une approche

interdisciplinaire associant sciences humaines, sciences de l'information et informatique.

Bio

François Vignale est chercheur au laboratoire 3L.AM. Ses domaines de recherches sont l'histoire des pratiques culturelles et les humanités numériques. Au sein du projet READ-IT, il a supervisé la conception du modèle de données et de l'ontologie ainsi que l'intégration des outils au sein de la plateforme d'annotation.

- L'IA embarquée au service des données hétérogènes : le **projet TIFAIFAI** Valérie Renault et Florent Carlier, CREN, Maîtres de Conférences à Le Mans Université

Résumé

L'IA recouvre une grande diversité d'approches. Les travaux sur les systèmes multi-agents ont donné lieu à de nombreuses applications dans le monde industriel : cinéma, smart city, simulations covid, etc. Nos travaux portent sur l'intégration des agents au plus près des objets connectés, en particulier au travers de la conception d'un mur d'écrans interactif et résilient. Une application concrète de ces travaux sera illustrée au travers d'un exemple sur la neutralité carbone.

Bio

Florent Carlier et Valérie Renault sont Maîtres de Conférences à Le Mans Université depuis 2004 et membres du CREN. Les recherches de Florent Carlier portent sur l'intégration de l'IA, et en particulier des systèmes multi-agents, dans les systèmes embarqués. Il travaille actuellement sur les protocoles de communication entre les agents et les objets connectés hétérogènes.

Les recherches de Valérie Renault s'interrogent sur l'introduction de l'IA, et en particulier d'agents, à différents niveaux de conception des interfaces homme-machine (IHM) afin de prendre en compte les spécificités des usagers finaux et la diversité des données à traiter. Elle s'intéresse aussi à la problématique de l'éthique dans l'IA.

- **11h45-12h30 - Temps d'échange animé** par Josep Crego (Directeur de la recherche de Systran) et Romain d'Alloméda
Avec Laurent Besacier, Hatim Bourfoune (Ingénieur de recherche à l'IDRIS), Maxence Jeunesse, Djamé Seddah.

12h30-14h00 Repas

Session 2 - Après-midi

IA : Quels challenges pour l'acceptabilité ?

- **14h00-14h30 - La couverture de l'IA par les médias, une économie de la promesse et de la critique**

Jean-Philippe Cointet, Enseignant chercheur Sciences po Paris, MediaLab

Résumé

A travers une analyse empirique de la presse anglo-saxonne contemporaine, nous montrons que la critique de l'IA dans la presse s'articule autour de deux formes distinctes: d'une part les "injustices algorithmiques" dont les biais systématiques sont accusés de discriminer les minorités, et d'autre part une peur plus globale et diffuse de voir l'humanité supplantée par une technologie devenue incontrôlable.

Bio

Jean-Philippe Cointet travaille au développement de méthodes innovantes en sciences sociales computationnelles. "Associate Professor" au médialab de Sciences Po, il est également associé au laboratoire INCITE de Columbia University. Ses objets de recherche sont variés: de la cartographie des dynamiques des problèmes publics dans l'espace public numérique à l'analyse textuelle des relations internationales, en passant par l'inférence d'un espace idéologique latent depuis les traces d'interaction des réseaux sociaux.

- **14h40-15h00 - L'IA responsable. Rêve ou réalité ?**

Vincent Courboulay, Maître de Conférences HDR à La Rochelle Université et Directeur scientifique de l'Institut du Numérique Responsable

Résumé

Derrière une image immatérielle, la révolution majeure qu'est l'IA a un impact considérable tant sur l'environnement qu'au niveau social. Pourtant l'IA en soi n'est ni bonne ni mauvaise. Elle doit trouver sa place comme simple auxiliaire permettant aux citoyens de mieux vivre.

Bio

Vincent Courboulay est ingénieur et Maître de Conférences en informatique à La Rochelle Université depuis 15 ans. Depuis, il se spécialise dans le numérique responsable (NR) d'abord en créant des formations puis en orientant ses recherches sur cette thématique. Il participe en 2018 à la création de l'Institut du Numérique Responsable dont il devient directeur scientifique. Il travaille actuellement sur la notion d'intelligence artificielle responsable et est l'auteur des livres "Vers un Numérique Responsable" et "L'archipel des GAFAM" parus tous deux chez Actes Sud. Il est membre du comité de transformation écologique des Jeux Olympiques Paris 2024.

- **15h00-15h30 - Les enjeux éthiques de l'IA vus par le prisme du traitement automatique des langues**

Karën Fort, Maîtresse de Conférences HDR à Sorbonne Université, Loria,
Co-chair du comité d'éthique ACL

Résumé

En quelques années, les systèmes par apprentissage sont devenus suffisamment performants pour être mis à disposition directement aux utilisateurs finaux. C'est une révolution majeure, qui n'est pas sans poser de très nombreux problèmes, dont des problèmes éthiques. Je proposerai dans cette présentation un panorama rapide des questions éthiques qui se posent aujourd'hui en traitement automatique des langues, notamment en ce qui concerne les biais stéréotypés, mais également la consommation carbone et la place des Big Tech.

Bio

Karën Fort est Maîtresse de Conférences en informatique à Sorbonne Université et effectue ses recherches au LORIA à Nancy. Après six ans dans l'industrie du Traitement Automatique des Langues (TAL), elle a soutenu en 2012 une thèse sur l'annotation de corpus, sous la direction d'Adeline Nazarenko. Elle s'est ensuite intéressée à la myriadisation (crowdsourcing), en particulier aux jeux ayant un but. Elle a en parallèle développé un intérêt pour l'éthique du TAL et a co-organisé le premier colloque sur le sujet en 2014, suivi d'un atelier national (ETERNAL 2015 et 2020) et d'un numéro spécial de la revue TAL en 2016. Elle a été membre du comité d'éthique de la Sorbonne, co-présidente des deux premiers comités d'éthique de conférences du TAL (EMNLP 2020 et NAACL 2021) et est co-présidente du comité d'éthique de l'association internationale du TAL (ACL).

15h30-15h45 Pause

- **15h45- 16h15 - Enjeux juridiques et éthiques des systèmes d'IA : personnalité, vie privée, projets de législation**

Céline Béguin, Maître de Conférences au Themis, Le Mans Université
Magali Bouteille-Brigant, Maître de conférences en droit privé – HDR, Co-directrice du Themis, Le Mans Université

Résumé

Faut-il un droit des robots ? Quelle responsabilité des dommages créés par l'intelligence artificielle : la personnalité juridique est-elle une solution ? Faut-il un texte de loi pour encadrer la responsabilité ? Qu'en est-il des créations générées par les systèmes d'intelligence artificielle ?

Quels sont les projets de textes européens de législation réglementant les systèmes d'intelligence artificielle ? Comment protéger les données personnelles

utilisées par les systèmes d'intelligence artificielle (*chat GPT...*) ? Quelle est l'influence ou l'emprise des agents conversationnels (*chatbot...*) sur les personnes ? Savoir que l'on s'adresse à une intelligence artificielle suffit-il à protéger la personne qui l'utilise ?

Bios

- Céline Béguin est Maître de Conférences à la faculté de droit, Co-directrice du Master Droit des Assurances, responsable de la certification numérique en Master (Pix + Droit) Interventions orales lors de cinq colloques sur l'intelligence artificielle depuis 2018, organisatrice d'un colloque "IA et assurance" en janvier 2022 au Mans, auteur de trois articles sur l'approche juridique de l'intelligence artificielle

- Magali Bouteille Brigant est Maître de Conférences en droit privé, HDR, à la faculté de droit, Co-directrice du Themis-UM et directrice du Master Droit de la santé. Elle est la directrice scientifique du colloque " La personne face à l'intelligence artificielle" et de plusieurs articles consacrés à la personnalité juridique des robots, ou au transhumanisme. Elle dirige en outre une thèse sur la e-santé.

- **16h15-16h45 - Régulation des systèmes d'IA : l'action de la CNIL**
Félicien Vallet, Responsable IA à la CNIL

Résumé

Devant les récentes actualités sur l'intelligence artificielle, et en particulier des IA dites génératives telles que ChatGPT, la CNIL a publié un plan d'action pour un déploiement de systèmes d'IA respectueux de la vie privée des individus. Celui-ci vise à proposer un cadre de régulation pour ces systèmes conforme aux principes du RGPD et demain, du règlement IA en cours d'élaboration. <https://www.cnil.fr/fr/intelligence-artificielle-le-plan-daction-de-la-cnil>

Bio

Félicien Vallet est responsable IA à la Commission Nationale de l'Informatique et des Libertés (CNIL), l'autorité française de protection des données. Son rôle est de coordonner les actions liées à l'Intelligence Artificielle de manière transversale au sein de l'institution. Félicien Vallet est également un collaborateur régulier du LINC, le Laboratoire d'innovation numérique de la CNIL. Il interagit fréquemment avec la communauté de la recherche académique et ses institutions en France et en Europe. Avant de rejoindre la CNIL, il était chercheur à l'Institut national de l'audiovisuel (INA) et s'intéressait particulièrement aux questions relatives à l'analyse de contenu multimédia et au traitement du signal vocal. Il est titulaire d'un diplôme d'ingénieur et d'un doctorat en informatique de Télécom Paris (obtenus respectivement en 2007 et 2011).

- **16h45-17h20 - Temps d'échange animé par Céline Béguin**
Avec Magali Bouteille-Brigant, Jean-Philippe Cointet, Vincent Courboulay, Karèn Fort, Félicien Vallet

