

Pourquoi cette UE ?

Les langues naturelles sont omniprésentes dans notre quotidien, mais leur manipulation automatique (analyse, compréhension, génération, traduction) est un défi de recherche complexe depuis plus de 70 ans, à l’intersection de la linguistique, de l’informatique et des sciences cognitives : c’est le Traitement Automatique des Langues Naturelles(TALN). Avec l’essor de l’apprentissage profond et d’applications comme les assistants vocaux, chatbots ou résumés automatiques, le TALN suscite un intérêt croissant en recherche et en industrie. Ce module offre aux étudiants de l’option IASD, spécialité « Connaissance et Texte », des bases solides pour appliquer ces modèles théoriques à divers cas réels.

Eléments constitutifs de l'UE

		coefficient
2IAiasd_10_2con-1 Traitement Automatique du Langage Naturel et de la Parole: Introduction Générale, Contrôle, Projet		1
2IAiasd_10_2con-2 Apprentissage profond pour données séquentielles, audio et neuromorphiques		1
2IAiasd_10_2con-3 LLMs avancés : fine-tuning, adaptation et agents orchestrés		1
Volume d'heures d'enseignement encadré	Volume d'heures de travail personnel	Nombre d'ECTS
50	30	3

Alignement curriculaire

Parmi les compétences visées par la formation, lesquelles sont développées dans cette UE ?



**Contexte et enjeux de l'enseignement**

Les technologies en lien avec le langage naturel jouent un rôle de plus en plus prépondérant dans nos vies quotidiennes : que ce soit avec des assistants vocaux, des chatbots, des applications de traduction automatique, et au-delà. De nombreuses entreprises, grandes comme petites investissent des sommes importantes dans ces technologies du traitement automatique de la langue naturelle et cherchent de nouveaux talents ayant la capacité de naviguer parmi les subtilités d'un domaine aux frontières de l'informatique, de la linguistique et des sciences des données.

**Prise en compte des dimensions socio-environnementales**

ODD9 - Industrie, innovation et infrastructure

**Prérequis**

Introduction à l'IA, Apprentissage Automatique Avancé, Apprentissage Profond

**Modalités d'enseignement et d'évaluation**

	Nb d'heures
Cours	4
Cours intégré (cours + TD)	
TD	
TP	2
Projets	16
Travail en autonomie encadré	16
Contrôles et soutenances	2
Travail personnel	20

**Objectifs pédagogiques**

(à la fin de cet enseignement, l'étudiant sera capable de ...)

- Être capable d'invoquer des connaissances élémentaires en linguistique pour approcher des problèmes de TAL. Identifier les problèmes de TAL et être capable de proposer des stratégies de résolution afin d'aller vers l'implémentation de solutions (avec les bons outils et les bonnes ressources).
- Mettre en pratique les compétences acquises dans toutes les ECUE de l'UE Traitement Automatique de la Langue et de la Parole en semi-autonomie

**Activités**

(CM, TD, TP, projet, sortie terrain, etc. )

CM: Introduction Générale au module, Concepts de base de Linguistique et de TAL

CMTP: Modélisation du Langage et Vecteurs de Mots

Projet: Application en semi-autonomie

**Évaluations et retours faits aux élèves**

(évaluations qui comptent pour la note ou qui permettent à l'étudiant de se situer, corrigés, feedback personnalisé...)

Devoir sur table / QCM (coef 1), Projet (coef 2)

Retour sur l'évaluation fait à l'élève :  
Retours personnalisés sur le travail de TP à la demande des élèves. Corrections de l'examen et évaluation du projet.

2IAiasd_10_2con Traitement Automatique du Langage Naturel et de la Parole	FIG
2IAiasd_10_2con-1 Traitement Automatique du Langage Naturel et de la Parole: Introduction Générale, Contôle, Projet	S10

## Plan de cours

- Séance 1: CM Introduction Générale (Ressources, Tâches, Problèmes) (4h)
- Séance 2: CMTP Modélisation du langage et Vecteurs de Mots (2h)
- Projet (16h)
- Examen (2h)

## Ressources et références

Daniel Jurafsky and James H. Martin. 2024. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models, 3rd edition. Online manuscript released August 20, 2024. <https://web.stanford.edu/~jurafsky/slp3>.

2IAiasd_10_2con Traitement Automatique du Langage Naturel et de la Parole	FIG
2IAiasd_10_2con-2 Apprentissage profond pour données séquentielles, audio et neuromorphiques	S10

Contexte et enjeux de l'enseignement	Prise en compte des dimensions socio-environnementales	Modalités d'enseignement et d'évaluation																		
		<table><tr><th></th><th>Nb d'heures</th></tr><tr><td>Cours</td><td></td></tr><tr><td>Cours intégré (cours + TD)</td><td></td></tr><tr><td>TD</td><td></td></tr><tr><td>TP</td><td></td></tr><tr><td>Projets</td><td></td></tr><tr><td>Travail en autonomie encadré</td><td></td></tr><tr><td>Contrôles et soutenances</td><td></td></tr><tr><td>Travail personnel</td><td></td></tr></table>		Nb d'heures	Cours		Cours intégré (cours + TD)		TD		TP		Projets		Travail en autonomie encadré		Contrôles et soutenances		Travail personnel	
	Nb d'heures																			
Cours																				
Cours intégré (cours + TD)																				
TD																				
TP																				
Projets																				
Travail en autonomie encadré																				
Contrôles et soutenances																				
Travail personnel																				
	<div>Prérequis</div>																			

Objectifs pédagogiques	Activités	Évaluations et retours faits aux élèves
(à la fin de cet enseignement, l'étudiant sera capable de ...)	(CM, TD, TP, projet, sortie terrain, etc. )	(évaluations qui comptent pour la note ou qui permettent à l'étudiant de se situer, corrigés, feedback personnalisé...)
<p><b>Deprecated:</b> htmlspecialchars(): Passing null to parameter #1 (\$string) of type string is deprecated in <b>C:\Developpement\syllabus\public_html\views\syllabus_template.php</b> on line <b>258</b></p>	<p><b>Deprecated:</b> htmlspecialchars(): Passing null to parameter #1 (\$string) of type string is deprecated in <b>C:\Developpement\syllabus\public_html\views\syllabus_template.php</b> on line <b>261</b></p>	<p><b>Deprecated:</b> htmlspecialchars(): Passing null to parameter #1 (\$string) of type string is deprecated in <b>C:\Developpement\syllabus\public_html\views\syllabus_template.php</b> on line <b>264</b></p>

2IAiasd_10_2con Traitement Automatique du Langage Naturel et de la Parole	FIG
2IAiasd_10_2con-2 Apprentissage profond pour données séquentielles, audio et neuromorphiques	S10

### Plan de cours

**Deprecated:** htmlspecialchars(): Passing null to parameter #1 (\$string) of type string is deprecated in C:\Developpement\syllabus\public\_html\views\syllabus\_template.php on line 292

### Ressources et références

**Deprecated:** htmlspecialchars(): Passing null to parameter #1 (\$string) of type string is deprecated in C:\Developpement\syllabus\public\_html\views\syllabus\_template.php on line 297

2IAiasd_10_2con Traitement Automatique du Langage Naturel et de la Parole	FIG
2IAiasd_10_2con-3 LLMs avancés : fine-tuning, adaptation et agents orchestrés	S10

Contexte et enjeux de l'enseignement

Les LLMs ont ouvert de nouvelles possibilités en matière d'interaction homme-machine, de recherche documentaire et de raisonnement automatisé. Pour répondre aux exigences de cas d’usage métiers réels, il devient essentiel de maîtriser à la fois leur adaptation légère (prompt engineering, LoRA), leur enrichissement via des données externes (RAG), et leur orchestration dans des agents autonomes capables de planifier, d’agir et de s’adapter. Ce module offre une progression complète de l’optimisation locale (prompt) jusqu’à l’IA agentique.

Prise en compte des dimensions socio-environnementales

ODD9 - Industrie, innovation et infrastructure

**Prérequis**

Apprentissage Profond

Modalités d'enseignement et d'évaluation

	Nb d'heures
Cours	4
Cours intégré (cours + TD)	
TD	2
TP	4
Projets	
Travail en autonomie encadré	
Contrôles et soutenances	
Travail personnel	10

Objectifs pédagogiques

(à la fin de cet enseignement, l'étudiant sera capable de ...)

- À l’issue du module, l’étudiant sera capable de :
- Adapter un LLM à un cas métier via des stratégies de prompt engineering et fine-tuning léger
  - Concevoir un pipeline RAG combinant récupération de documents et génération contrôlée
  - Implémenter un agent LLM orchestré capable de réaliser des tâches complexes de manière autonome

Activités

(CM, TD, TP, projet, sortie terrain, etc. )

- Cours magistraux (CM) : introduction aux concepts et cadres d’architecture
- Travaux dirigés (TD) : études de cas, conception guidée de pipelines et agents
- Travaux pratiques (TP) : mise en œuvre de prompts, fine-tuning, RAG, agents via des frameworks actuels
- Mini-projet final : création d’un système LLM complet intégrant prompts optimisés, mémoire externe et raisonnement multi-étapes

Évaluations et retours faits aux élèves

(évaluations qui comptent pour la note ou qui permettent à l'étudiant de se situer, corrigés, feedback personnalisé...)

- Évaluation sommative :
- Mini-projet final (70 %) : création d’un système agent + RAG opérationnel, documenté et justifié
  - QCM ou quiz final (30 %) : évaluation des connaissances sur les concepts vus (fine-tuning, RAG, agents, etc.)
- Évaluation formative :
- Feedback intermédiaire sur les choix techniques du mini-projet
  - Correction collective des TD avec retours détaillés
  - Suivi individuel sur les TPs

2IAiasd_10_2con Traitement Automatique du Langage Naturel et de la Parole	FIG
2IAiasd_10_2con-3 LLMs avancés : fine-tuning, adaptation et agents orchestrés	S10

## Plan de cours

Séance 1 (CM) – Panorama des techniques d’adaptation des LLMs

- Limites des modèles généralistes
- Fine-tuning complet vs. LoRA, Adapters, Prompt tuning
- Coût, efficacité, contraintes techniques

Séance 2 (CM + TD) – Stratégies de prompt engineering avancées

- Prompt direct, instruction-based, zero/few-shot
- Chain-of-thought, prompting incrémental, prompt injection
- TD : étude comparative de prompts sur tâches variées

Séance 3 (TP) – Pratique du prompt engineering et fine-tuning léger

- Utilisation de modèles Hugging Face (PEFT, LoRA)
- Conception de prompts métier
- Évaluation qualitative/quantitative des réponses

Séance 4 (CM + TD) – RAG : LLMs augmentés par la recherche documentaire

- Introduction à la RAG : motivation, architecture (Retriever + Generator)
- Bases vectorielles, embeddings, chunking
- TD : design d’un pipeline RAG pour un assistant métier

Séance 5 (TP) – Mise en œuvre d’un pipeline RAG

- Création d’un index vectoriel à partir de documents
- Intégration avec un LLM via LangChain ou équivalent
- Tests sur des requêtes métiers réelles

Séance 6 (CM + TP) – Agents LLM : conception et orchestration

- Définition de l’IA agentique : réflexion, planification, exécution
- Outils : LangChain agents, OpenAI Functions, Tool use
- TP : création d’un agent capable d’exploiter une API ou une base documentaire pour résoudre un problème métier
- Mini-projet : présentation et feedback final

## Ressources et références

Daniel Jurafsky and James H. Martin. 2024. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models, 3rd edition. Online manuscript released August 20, 2024. <https://web.stanford.edu/~jurafsky/slp3>.