	Nom de l'entreprise	Location	Contact E/se	Superviseur	Affectation	Sujet du Projet	Description	Missions	Outils et Technologies Utilisés (Si Applicable)
Sujet N°1	**AITELA BRANDING**	Marrakech	IMAD AIT EL ARABI		**2nd**	Xenith Pulse: Development of an Autonomous Deep Research Agent for B2B Intelligence and Lead Enrichment	Build an AI-driven research engine capable of conducting autonomous web searches to generate high-fidelity, structured reports on companies, prospects, or market trends. The goal is to provide sales teams with "Golden Profiles"—leads enriched with real-time news, financial data, and pain points—moving beyond static contact info to actionable business intelligence.	The Deep Research Agent will function as a three-stage pipeline: Autonomous Query Expansion: Using an LLM (e.g., GPT-4o) to break a single user prompt (e.g., "Find the latest expansion plans for logistics companies in Germany") into multiple specialized search queries. Resilient Web Harvesting: Utilizing Web Search APIs (such as Tavily, Brave Search, or Serper) combined with a Rotating Proxy Layer to bypass anti-bot protections and access localized data. Synthesis & Formatting: Aggregating raw data into a structured Markdown report for human reading or a JSON object for seamless injection into the Xenith AI CRM and Outreach modules.	Technical Architecture: Orchestration Layer: A Fast API microservice that manages the "Agentic Loop"—deciding when it has enough information to stop searching. Search & Retrieval: * Reverse Engineering Tavily or Exa (optimized for AI agents) for search capabilties. Proxies: Implementation of residential proxies (e.g., Bright Data or Oxylabs) to ensure high success rates across diverse geographic regions.
Sujet N°2	**ATOS**	Casablanca	Kawthar EL GAMAH		**2nd**	Conception d'un système Agentic AI basé sur le Model Context Protocol (MCP) pour l'orchestration intelligente de services IA.	"Comprendre et appliquer les concepts de l'IA agentique et du Model Context Protocol (MCP). Concevoir une architecture d'orchestration intelligente adaptée aux enjeux du billing Telco . Mettre en œuvre des agents IA Intégrer les contraintes clés d'un environnement opérateur : sécurité, traçabilité, gouvernance des données et performance."	"Tout au long du projet, les étudiants seront amenés à : Réaliser une analyse de l'existant et des technologies associées (MCP, agents IA, orchestration). Définir l'architecture fonctionnelle et technique de la solution cible. Implémenter un prototype (preuve de concept) incluant : Des agents IA spécialisés (analyse, décision, orchestration), Des connecteurs vers des services IA et des sources de données, Une logique d'orchestration basée sur le contexte métier. Tester, évaluer et documenter la solution. Présenter les résultats et les perspectives d'industrialisation pour un déploiement local."	-
Sujet N°3	**Duferco Group**	Italie	Riccardo Necrisi		**2nd**	Building an Intelligent BESS Digital Twin (Design & monitoring).	HVGM: realization of a green H2 production plant coupled with PV plant and BESS to ensure cost effective H2 for automotive and transport		Python, Autocad, SimFlow H2 Production System (AEM), BESS, PV (on roof and on ground with tracker)
Sujet N°4	**Forschungszentrum Jülich**	Germany	Maik Boltes	Mohcine Chraibi	**2nd**	Trajectory optimization with the use of an existing foreground/background mask	Based on an existing foreground/background mask, the head trajectories of people in crowds are to be optimized. Trajectories of individual people extracted from overhead video recordings contain errors, causing the trajectories in the image to shift from the head to the background. This should be detected and corrected automatically if possible. Sample data from an experiment: https://doi.org/10.34735/ped.2018.1	- Development of optimization strategies (e.g., extrapolation of a tracked point that lies in the background) - Implementation of the strategy in the C++ software PeTrack - Documentation and presentation of the strategy and its implementation	C++, Git, OpenCV, Qt
Sujet N°5	**Forschungszentrum Jülich**	Germany	Khalfaoui, Ismail	Mohcine Chraibi	**2nd**	Efficient Implementation and Scaling of Learnable Activation Functions via Optimized GPU Kernels	Recent advances in learnable activation functions have demonstrated their potential to outperform tradi tional fixed activations such as GELU, SiLU, and SwigGLU. In particular, the paper Polynomial, Trigono metric, and Tropical Activations [1] presents a rigorous framework for variance-preserving initialization and empirical validation of Hermite, Fourier, and Tropical activations across deep learning models, including ConvNeXt and GPT-2. However, as identified in this work, the primary limitation preventing widespread adoption of learnable activations is not the theoretical floating-point operation (FLOPs) count, but rather the efficiency with which these operations are parallelized and implemented on modern hardware.	The primary objective of this project is to investigate whether the optimization techniques developed for rational activations in FlashKAT can be generalized and applied to alternative learnable activation families, specifically polynomial-based activations such as Hermite polynomials. The project aims to: • Adapt FlashKAT-style gradient accumulation and memory-efficient backward passes to polynomial activation functions. • Develop high-performance GPU kernels for Hermite, Fourier, and Tropical activations. • Evaluate training speed, numerical stability, and scalability relative to existing implementations. • Assess whether polynomial activations can	The project stems from the observation that rational activations used in KAT reduce to polynomial activa tions when the denominator degree is zero. This implies that the optimization strategies used for rational functions are directly applicable to the polynomial case. We have already initiated the development of CUDA kernels for Hermite activations. The current implementation is written in Numba and exploits a recursive formulation of Hermite polynomials. This recursive evaluation strategy can be viewed as an analogue to Horner's method used in KAT Triton kernels. As part of this project, the Numba-written CUDA kernels will be reworked into Triton, incorporating optimized gradient accumulation strategies inspired by FlashKAT. Particular emphasis will be placed on minimizing memory traffic and improving parallelism in the backward pass. Visualization tools such as Nsight will be leveraged to benchmark and diagnose compute and memory bottlenecks within the designed GPU kernels.
Sujet N°6	**Forschungszentrum Jülich**	Germany	Khalfaoui, Ismail	Mohcine Chraibi	**2nd**	Scalable Data Curation for Low-Resource Languages using NeMo Curator	The performance, robustness, and safety of large language models (LLMs) depend critically on the quality of their training data. While model architectures and optimization techniques have advanced rapidly, data curation remains a major bottleneck due to the scale, heterogeneity, and noise present in modern text corpora. Raw data collected from cloud storage, internet sources, and open repositories often contains duplication, low-quality content, personally identifiable information (PII), and domain irrelevant samples. In low-resource languages, this challenge is amplified due to limited data availability, noisier content, and fewer pre-existing filtering tools. Addressing these issues is key to improving model generalization and enabling equitable NLP research across languages. NVIDIA NeMo Curator provides a modular framework for data curation pipelines, supporting inges tion, cleaning, heuristic and model-based filtering, deduplication, and quality control. NeMo Curator is flexible, enabling both CPU and GPU-accelerated workflows.	The project aims to design a reproducible, scalable data curation pipeline for a selected low-resource language, with goals including: • Implementing end-to-end data curation workflows in NeMo Curator. • Performing heuristic and model-based quality filtering. • Applying deduplication (exact, fuzzy, semantic) to limited datasets. • Adapting distributed processing from Dask to Ray for CPU-based execution. • Evaluating curated datasets for quality, diversity, and downstream usability.	1. Language Selection: Choose a low-resource language (Faroese, Moroccan Arabic, ...) to target. 2. Data Ingestion: Collect raw text from public corpora or web sources. 3. Cleaning and Preprocessing: Normalize text, remove malformed or irrelevant samples, and apply language-specific filters. 4. Heuristic Filtering: Remove low-information content using rule-based approaches. 5. Deduplication: Apply exact, fuzzy, and semantic deduplication. 6. Model-Based Filtering: Use lightweight classifiers for PII and toxicity removal. 7. Dataset Preparation: Shuffle and output a clean dataset for model training or fine-tunin
Sujet N°7	**INTIS & EPSIMO AI**	Marrakech / France	Jérôme BOULOTON		**2nd**	Automatisation de la prospection avec un agent IA	Dans le cadre de notre partenariat avec la société EPSIMO AI, éditeur d'une plateforme d'agents AI, nous souhaitons enrichir les capacités de notre agent de prospection	Enrichir les capacités d'automatisation de la prospection (contacts, signaux faibles, enrichissement, scoring, copyright, nurturing, etc.) de notre agent EPSIMO.AI	RAS
Sujet N°8	**Leyton CognitX**	Casablanca	Youssef Bennani	Hicham Azeroual	**2nd**	Automated Evaluation Pipeline for AI-Generated R&D Content	As the use of generative AI expands, many organizations rely on language models to produce complex documents such as R&D reports, technical summaries, and structured project descriptions. While these systems can generate high-quality content at scale, their outputs vary depending on the model, prompt design, and input data. Without a standardized evaluation framework, it becomes difficult to assess whether the generated content is accurate, complete, aligned with source material, or written in a convincingly human style. Inconsistencies, hallucinations, and overly "AI-sounding" text can reduce trust and create additional review work. A dedicated evaluation pipeline is therefore essential to measure the quality of generative AI outputs, compare model performance, and ensure that the generated documents meet scientific and professional expectations.	The scope may evolve depending on the project duration, available time, and student workload. The foundations described below represent the target objectives, which can be adjusted based on feasibility and progress during the project period.Develop an automated evaluation pipeline capable of scoring the quality of AI-generated R&D content according to predefined criteria. Define clear, domain-specific evaluation metrics, including completeness, scientific accuracy, clarity of innovation, coherence, and consistency with the source notes or context. Benchmark multiple LLMs to identify which models produce the most reliable technical and scientific content. Detect AI-generated writing patterns, such as repetitive, generic, or unnatural phrasing, to flag content that may require humanization. Design methods to humanize the generated text when necessary, ensuring the final output reads naturally and aligns with professional writing expectations. Introduce rule-based and semantic validation checks to detect hallucinations, contradictions, missing elements, or terminology issues.	-
Sujet N°9	**Leyton CognitX**	Casablanca	Youssef Bennani	Kawtar Aberkchane	**2nd**	Greenwashing Detection in Corporate Documents	As sustainability becomes a central pillar of corporate strategy, many companies publish reports highlighting their environmental commitment. However, some engage in greenwashing, communicating exaggerated or misleading environmental claims to appear more eco-friendly than they actually are. Early detection of greenwashing is essential for: Investors evaluating sustainability claims Governments monitoring ESG compliance Consulting firms verifying the credibility of corporate disclosures Citizens seeking transparency This project aims to leverage Natural Language Processing (NLP) and Machine Learning (ML) to detect and flag marketing-oriented or deceptive sustainability statements.	The project aims to develop a robust ML/NLP system capable of detecting greenwashing signals in corporate documents and ensure the following points Build a labeled dataset of sustainability-related paragraphs Develop a paragraph-level classifier Identify linguistic patterns of greenwashing	-
Sujet N°10	**OCP Mining**	Benguerir	Zakaria Mourchid	Pr. Lamiae Azizi	**2nd**	AI empowered digital twin (4ème année)	Digital twins are already powerful technologies. But combined with AI and machine learning, they evolve from passive replicas to intelligent systems that learn, predict, and optimize autonomously. The aim of this project is to investigate and develop a prototype AI empowered digital twin for OCP mining operations.	-	-
Sujet N°11	**Syensqo**	Belgique	Nassim Bensaoud	Pr. Lamiae Azizi	**2nd**	AI for industrial predictive maintenance	This project is about developing AI algorithms to analyze data from sensors embedded in machinery to monitor operational conditions and detect anomalies. These sensors collect data on various parameters like temperature, vibration, pressure, and fluid levels. AI tools can compare this data against baseline performance metrics to identify even the smallest dips in efficiency. This real-time analysis allows our industries to predict failures more accurately and open maintenance tickets.	-	-
Sujet N°12	**Stellantis**	Casablanca	Mohamed OUMRI	Mohamed OUMRI	**2nd**	Mise en œuvre d'un Hybrid Twin pour la simulation d'un crash latéral avec obstacle fixe	Mise en œuvre d'un cas d'étude Hybrid Twin sur crash latéral : génération d'une simulation haute-fidélité de référence, création de données pseudo-expérimentales bruitées, puis développement d'un modèle hybride couplant jumeau numérique et assimilation de données.	Réduction de l'écart simulation-mesure Amélioration de la prédiction en dynamique rapide Meilleure robustesse face au bruit Valorisation conjointe modèle + données..	Simulation EF / crash Algorithmes ML/DL Hybrid Twin & assimilation de données Modélisation statistique & bruit Analyse d'incertitudes

	Nom de l'entreprise	Location	Contact E/se	Superviseur	Affectation	Sujet du Projet	Description	Missions	Outils et Technologies Utilisés (Si Applicable)
Sujet N°13	**Artefact**	Casablanca	Aymane GOURIMATE	Salah Eddine LABIAD.	**2nd**	Real-Time Voice Agent : La Course à la Latence	1. Implémenter un pipeline vocal temps réel avec latence < 500ms 2. Comparer les approches STT+LLM+TTS vs Speech-to-Speech natif 3. Concevoir un agent vocal plug-and-play, avec composants interchangeables 4. Intégrer des outils métier (RAG, base de données, API) via un mécanisme de tool calling 5. Supporter le multilingue (FR/EN/AR) avec switching automatique 6. Gérer interruptions, corrections utilisateur et validation des informations collectées	Phase 1 : Architecture & Setup (Semaines 1-3) → Analyse des composants SOTA : Whisper, Faster-Whisper, Groq Whisper, Deepgram → Evaluation des moteurs TTS : ElevenLabs Turbo, OpenAI TTS, XTTS, Coqui → Setup de l'infrastructure streaming audio (WebRTC, WebSocket) → Benchmark initial des latences de chaque composant Phase 2 : Pipeline STT + LLM + TTS (Semaines 4-7) → Implémentation du streaming STT avec Faster-Whisper (optimisation CUDA) → Intégration d'un LLM rapide : Groq (Llama 3/Mixtral) ou Cerebras → Connexion TTS avec streaming token-by-token (ElevenLabs Turbo) → Optimisation asyncio pour parallélisation maximale → Mesure et optimisation de chaque milliseconde de la chaîne Phase 3 : Fonctionnalités Avancées (Semaines 8-11) → Intégration (RAG, Accès base de données, Appels API) comme tool de l'agent → Gestion des interruptions et corrections utilisateur → Detection de langue automatique et switching multilingue → Implémentation Speech-to-Speech avec OpenAI Realtime API (comparaison) → Gestion du contexte conversationnel et mémoire de session Phase 4 : Démo & Production (Semaines 12-14) → Dashboard de monitoring temps réel (latences, qualité audio) → Containerisation et déploiement micro-services → Tests utilisateurs et collecte de feedback → Documentation et présentation finale avec démo live	● Le pipeline classique STT + LLM + TTS, offrant flexibilité, contrôle métier et auditabilité ● Les nouveaux modèles Speech-to-Speech (STS) en temps réel, offrant une latence minimale mais moins de contrôle et de personnalisation
Sujet N°14	**Think Forward AI**	USA	Youness Mouqadem	Youness Mouqadem/ El Houcine Bergou	**2nd**	Voice Food Ordering System	Build and evaluate an AI-powered voice ordering system for food service environments, comparing different deployment approaches and AI models.	Model Evaluation: Test and compare available AI voice models (open-source vs proprietary) Deployment Analysis: Compare edge vs cloud solutions for latency, cost, and reliability Domain Optimization: Test fine-tuning effectiveness for specific contexts (coffee shops, burger joints) to reduce ordering errors Feasibility Assessment: Determine technical and economic viability of edge deployment	-
Sujet N°15	**Think Forward AI**	USA	Youness Mouqadem	Youness Mouqadem/ El Houcine Bergou	**2nd**	EzDeal - AI-Powered Investment Analysis Platform	EzDeal is a tool designed to help angel investor networks evaluate startup investment opportunities. It uses AI to analyze key startup materials including pitch decks, financial statements, patents, and interview transcripts with founders and executives. The tool converts this complex information into clear, digestible insights that enable investor groups to make informed investment decisions. The platform is in active development and technical decisions are still open. This project is well suited for students interested in applied AI, backend systems, and product development who want hands-on experience building a real product, contributing to core features, and helping shape the direction of an evolving platform. What You'll Work On: AI document processing and analysis systems Backend architecture for handling complex financial data User interface design for investor decision-making workflows Integration with existing angel network processes Ideal For Students With Interest In: Applied AI and machine learning Backend development and system architecture Product development and user experience Fintech and investment technology This is an opportunity to work on a real product that will be used by actual angel investors to make investment decisions, giving you direct exposure to both the technical and business sides of startup evaluation.	-	-
Sujet N°16	**UM6P bioinformatics laboratory**	Benguerir	Achraf El Allali		**2nd**	Agentic System for MEGA (UM6P's Multi-service Environment for Genomic Applications)	This project focuses on the co-development of an agentic system integrated into MEGA, UM6P's multi-service environment for genomic applications. The system autonomously designs, configures, and executes bioinformatics pipelines based on the user's input data and analysis goals. By leveraging intelligent agents, it adapts workflows dynamically, selects appropriate tools, manages dependencies, and monitors execution, enabling students to gain hands-on experience in automated, scalable, and user-driven genomic data analysis.	The students will co-develop an agentic system that creates and runs bioinformatics pipelines automatically based on the user's data.	Python: The core language for agent logic, pipeline orchestration, and bioinformatics scripting Agentic Frameworks (e.g., LangGraph / AutoGen): To be used to design and coordinate intelligent, autonomous agents Workflow Management Systems (e.g., Nextflow or Snakemake): They are used for building, executing, and scaling bioinformatics pipelines Containerization (Docker / Singularity) in order to ensure reproducibility and manage bioinformatics tool dependencies Bioinformatics Tools & Libraries which will be limited to well known tools such as BLAST, BWA, GATK, and Biopython, depending on the analysis tasks Cloud or HPC Infrastructure deployed for scalable computation and pipeline execution within MEGA
Sujet N°17	**ALGODOERS**	France	Hatem LTAIEF, CTO	Hatem LTAIEF & Pr. Imad KISSAMI	**2nd**	Accelerating Dense Linear Algebra Using Mixed-Precision Algorithms with Ozaki Scheme	High-Performance Computing (HPC) and Artificial Intelligence (AI) applications rely heavily on large-scale linear algebra operations, including dense and sparse matrix computations with direct/iterative solvers. These workloads are often limited by either compute throughput or memory bandwidth on modern architectures. Mixed-precision computing offers a powerful solution by combining low-precision arithmetic for performance with high-precision corrections for numerical accuracy. The Ozaki scheme provides a robust framework for compensated computations by engaging low-precision hardware units	Objectives - Analyze numerical behavior of mixed-precision algorithms - Deploy Ozaki-based mixed-precision kernels into matrix computations - Evaluate performance, accuracy, and energy efficiency Scope of Work - Dense matrix multiplication (GEMM) - Direct (Cholesky, QR, LU, Eigenvalues, SVD) solvers Timeline 14 weeks covering review, implementation, benchmarking, and reporting. Impact Relevant to HPC/AI acceleration for computational science and engineering	Methodology - Literature review - Implementation using C/C++ or Python - Benchmarking and profiling on Nvidia GPUs - Validation against native precision Expected Outcomes - Working prototype - Performance vs accuracy analysis - Final report and presentation Required Skills - Linear algebra and numerical methods - Programming in C/C++ or Python - Basic parallel computing knowledge
Sujet N°18	**ALGODOERS**	France	Hatem LTAIEF, CTO	Hatem LTAIEF & Pr. Imad KISSAMI	**2nd**	Accelerating Sparse Linear Algebra Using Mixed-Precision Adaptive Storage with Fixed-Point Formats	High-Performance Computing (HPC) and Artificial Intelligence (AI) applications rely heavily on large-scale linear algebra operations, including dense and sparse matrix computations with direct/iterative solvers. These workloads are often limited by either compute throughput or memory bandwidth on modern architectures. This project will focus on deploying adaptive mixed precisions for storage in the context of sparse iterative solvers. By relying on fixed-point storage, we can reduce memory footprint and bandwidth usage and significantly increase the throughput of the applications, while maintaining the numerical integrity of the application.	Objectives - Analyze numerical behavior of mixed-precision algorithms - Design fixed-point storage tools to assess the data representation format - Evaluate performance, accuracy, and energy efficiency Scope of Work - Dense matrix multiplication (GEMM) - Sparse matrix-vector multiplication (SpMV) - Sparse Iterative (CG / GMRES) solvers Timeline 14 weeks covering review, implementation, benchmarking, and reporting. Impact Relevant to HPC/AI acceleration for computational science and engineering	Methodology - Literature review - Implementation using C/C++ or Python - Benchmarking and profiling on Nvidia GPUs - Validation against native precision Expected Outcomes - Working prototype - Performance vs accuracy analysis - Final report and presentation Required Skills - Linear algebra and numerical methods - Programming in C/C++ or Python - Basic parallel computing knowledge
Sujet N°19	**ByteDoc**	Germany		Amine abou omar	**2nd**	AI powered to detect full body cancer	"Detect the full body cancer using AI that trained on million CT images to detect this disease and tumour. ByteDoc is an AI-assisted medical imaging platform that analyzes CT scans to help clinicians detect suspicious cancer-related findings across the body. It is designed as a decision-support tool: it highlights abnormal regions, estimates risk, and generates structured results to support radiologists and doctors—not replace them. Because "full-body cancer detection" is a very big problem, ByteDoc is built as a multi-stage system: starts with one or two high-impact cancers (e.g., lung nodules / liver tumors), then expands to multi-organ detection, and finally moves toward whole-body triage (flagging "suspicious" cases for faster review). Main outputs: heatmaps / boxes / segmentation masks showing suspicious regions, risk scores per organ or per lesion,	"mission to build this modal: 0) Decide what the model must do (the "task") "Full-body cancer detection" is too big as a first model, so you choose one V1 task: Pick ONE: Classification (scan-level): "Is there any suspicious tumor in this CT? yes/no" Detection (lesion-level): "Where is the tumor?" (boxes/points) Segmentation (pixel-level): "Exact tumor shape" (mask) Best V1 to start fast: Classification or Segmentation (if you have masks). Deliverable: 1 paragraph: input, output, success metric. 1) Get data + labels (you can't train without labels) A) Start with public datasets (recommended) This lets you build everything legally and quickly. B) Later: local/hospital data (only with permissions + anonymization) This is for real-world Moroccan performance later. You must have labels: classification labels (0/1) or bounding boxes or segmentation masks Deliverable: dataset folder + label format. 2) Build the CT preprocessing pipeline (most important part) CT scans are not like normal images. You need to standardize them. Typical pipeline: Load CT (DICOM → volume)	AI modal, HPC..

	Nom de l'entreprise	Location	Contact E/se	Superviseur	Affectation	Sujet du Projet	Description	Missions	Outils et Technologies Utilisés (Si Applicable)
Sujet N°20	**AISIN Europe**	Belgique	Najwa Madaoui		**2nd**	Detect Illegal dumping & street cleanliness using AI on dashcam images	Detect the following situations on images collected by vehicle's dashcam: Illegal dumping spots (construction debris, bulky waste, rubble piles, inert waste in peri-urban zones). Household waste bags left outside designated times/places. Open burning of waste (smoke plumes, visible fires). Overflowing or broken municipal bins and underground containers. Litter hotspots (plastic bags, bottles, packaging on verges and vacant plots). If the objective is reached, other types of situations will be added to the detection model (traffic signs, road marking, equipment, lighting poles, etc).	The goal is to automate the detection of waste related issues in the streets. The image dataset will be provided by the end of January, but preliminary images could be collected by the students for early tests. The mission is to reach a high level of confidence in the detection (>80%).	VLM, SAM3 and other recent AI techniques, and standard vision computing algorithms. The models can first run on local machines, but will eventually be automated pipelines on cloud.
Sujet N°21	**GEP**	Benguerir	Amine ait oufkir		**2nd**	Développement d'un module d'interaction en langage naturel basé sur un LLM pour la plateforme Digital Twin	"L'objectif du stage est d'ajouter à la plateforme existante un module permettant aux utilisateurs d'interagir avec le jumeau numérique en langage naturel, à l'aide d'un modèle de langage (LLM) comme GPT ou LLaMA. L'étudiant développera un microservice ou une API capable de : · recevoir une question utilisateur (ex. « Quel est l'état de l'équipement X ? ») · reformuler/structurer la question pour la plateforme · extraire les données pertinentes · générer une réponse claire et contextualisée Ce module rendra la plateforme plus intuitive, plus accessible et plus intelligente."	"· Analyse de l'architecture de la plateforme existante. · Intégration d'un LLM (API ou modèle local). · Création d'un pipeline d'interprétation : utilisateur → LLM → requête interne → réponse. · Développement d'une interface (chat Web ou widget intégré). · Tests et validation sur plusieurs scénarios. · Documentation technique du module."	"· Développement Web/API · Intégration logicielle · Appels API LLM · Notions de Digital Twin"
Sujet N°22	**GEP**	Benguerir	Amine ait oufkir		**2nd**	Conception d'un module d'analyse intelligente des séries temporelles basé sur la semanticisation et les LLM	"Le but est de développer un module IA capable d'enrichir la plateforme Digital Twin avec une analyse intelligente des données capteurs. Ce module appliquera les techniques décrites dans l'article scientifique ARTICLE LLM · Décomposition temporelle (tendance, saisonnalité, résidu) · Extraction automatique de caractéristiques · Génération d'une description textuelle des comportements (semanticization) · Détection et explication d'anomalies via un LLM Le résultat : un système capable d'interpréter les données du bâtiment comme le ferait un expert humain."	"· Collecte ou sélection des séries temporelles existantes dans la plateforme. · Implémentation de la décomposition Prophet. · Extraction de caractéristiques (moyenne, variance, pics, anomalies…). · Génération automatique de résumés sémantiques des séries. · Utilisation d'un LLM pour : o analyser ces descriptions o expliquer les anomalies ou variations o proposer des causes probables · Validation du module sur différents jeux de données."	"· Data Science · Analyse de séries temporelles · Modèles de langage (LLM) · Python / ML"