

Communiqué de presse

Hybridation de l'intelligence artificielle : SystemX lance « Sémantiques Métier pour l'exploitation de Données multi-sources » (SMD), 3^e projet de son programme IA2

Comment fabriquer une connaissance précise, pertinente et utile dans un contexte industriel ou commercial très ciblé à partir de volumes massifs de données hétérogènes (vidéo, image, texte, parole, graphique, etc.) provenant de sources multiples et variées ? Les technologies d'intelligence artificielle ouvrent des perspectives nouvelles sur cette problématique ancienne : dans le vocabulaire actuel, on parle d'hybrider de l'IA symbolique avec des approches de Deep learning. C'est à cet enjeu que SystemX et ses partenaires entendent répondre en croisant leurs expertises multisectorielles et en hybridant des approches récentes d'apprentissage numérique et les techniques d'IA symbolique, avec la connaissance métier.

Palaiseau, le 16 septembre 2021 - [SystemX](#), unique IRT dédié à l'ingénierie numérique des systèmes du futur, lance le projet « Sémantiques Métier pour l'exploitation de Données multi-sources » (SMD), troisième projet de son programme de recherche « [Intelligence Artificielle et Ingénierie Augmentée](#) » (IA2). Ce projet de R&D collaboratif d'une durée de 48 mois réunit 5 industriels (Airbus Defence and Space SLC, Apsys, Ecosys, EDF, RTE) et un partenaire académique (CentraleSupélec) autour d'une ambition : développer, pour les métiers de la veille ou de la gestion des connaissances, des outils hybridant l'IA symbolique et l'IA par apprentissage pour construire et exploiter des connaissances sur des données hétérogènes multi-sources, afin de favoriser l'aide à la décision en environnement statique ou dynamique.

Les entreprises disposent d'une multitude de sources de données, souvent cloisonnées et contenant des informations de nature hétérogène d'un point de vue sémantique, structurel ou syntaxique. L'hétérogénéité réside également au niveau des systèmes et technologies utilisés. L'analyse de ces données dans un but décisionnel est souvent difficile à réaliser, d'autant plus lorsqu'il s'agit de croiser ces données internes avec des dires d'expert et autres données externes (open data, web, réseaux sociaux, etc.), et d'autant plus que les traitements doivent être opérés en temps réel.

« Le contexte du projet part d'un constat réel et récurrent que toute entreprise industrielle pourrait faire face lors de son évolution numérique à l'avènement d'un volume important de données fortement hétérogènes, non-structurées et multi-sources. Avec l'accroissement de la puissance de calcul des ordinateurs, on est de plus en plus capable d'en remonter des informations. Cependant, ces attributs utilisés seuls ne permettent pas de fournir une compréhension suffisante et comparable à celle de l'être humain, d'où l'importance de coupler plusieurs approches. Le projet SMD ambitionne de lever un verrou important concernant l'hybridation des approches de représentation des connaissances et du raisonnement (IA symbolique) avec les approches récentes d'intelligence artificielle (par exemple le Deep Learning) pour l'analyse de données hétérogènes », explique Sana Tmar, cheffe de projet SMD, IRT SystemX.

Le projet SMD en quelques mots

3^{ème} projet du programme IA2

Domaines scientifiques et technologiques :
Science des données et IA

Durée : 48 mois
Effort total : 6 ETP / an

Partenaires industriels : Airbus Defence and Space Secure Land Communications, ApSys, Ecosys, EDF, RTE

Partenaire académique : CentraleSupélec (MICS)

Principal objectif du projet :
Concevoir un système hybridant l'IA symbolique et les méthodes récentes de l'apprentissage (Deep Learning) avec la connaissance métier, pour bâtir des connaissances nouvelles et intelligibles à partir de volumes très importants de données non-structurées, hétérogènes et multi-sources, présenter des recommandations pertinentes pour une meilleure prise de décision.

L'enjeu de ce projet est de prototyper des capacités à intégrer et analyser ces volumes très importants de données hétérogènes non structurées et multi-sources et de les rassembler dans un environnement commun pour réaliser des traitements sémantiques et offrir aux professionnels métier des connaissances nouvelles pertinentes, synthétiques et interprétables qui les aideront à prendre des décisions éclairées, comme pour l'identification de biais sur des ontologies de métier ou la détection d'une situation atypique par exemple. Le niveau de pertinence sera déterminé par le système en fonction des connaissances apprises à partir de données labellisées ou non, d'historiques de décisions, de feedbacks d'experts, etc. Les recommandations seront présentées avec un ensemble de commentaires ou d'explications qui leur permettront de rejeter, rectifier ou accepter les propositions remontées par l'IA. Toutes ces actions pourront être réintégrées dans la base de connaissances et réutilisées pour générer de nouvelles recommandations pour les prochaines situations identifiées.

Cinq cas d'usage ont été identifiés :

- L'évaluation d'une intelligence artificielle qui traitera en temps réel des flux de données (principalement vidéos) pour faire de la détection à la volée de situations présentant un risque. Cette IA sera destinée aux opérateurs gérant des situations très denses et parfois complexes (secours sur accidents de grande ampleur par exemple). Elle nécessitera une hybridation entre les connaissances antérieures produites par les experts du domaine et les connaissances extraites par un algorithme de type Deep Learning (Airbus Defence and Space SLC).
- La digitalisation des modèles comportementaux d'installations industrielles (analyse de dangers, vulnérabilités) et exploiter des contenus pour élaborer des diagnostic safety / security par rapport aux enjeux de conformité réglementaire. Cela passera par l'analyse de tournures de phrases, de chaînes de défaillance, de communications irrégulières en langage naturel pour détecter des situations anormales pouvant constituer un risque de sécurité (Apsys).
- L'analyse de données hétérogènes généralement définies dans un format métier et la capacité à rendre l'utilisation autonome dans la création de modèles de données (ontologies en graphe), sans besoin d'expertise particulière (Ecosys).
- La refonte et la formalisation des exigences à partir de corpus de données hétérogènes : le besoin est de rationaliser le processus de rédaction des appels d'offres et de vérification des réponses afin d'en réduire le coût et de permettre aux experts d'être plus efficaces et ce, grâce à l'hybridation des approches, pour une construction de connaissances plus efficace et pour un processus de découverte de correspondances entre ontologies (alignement d'ontologies) qui peuvent être issues de différents métiers et de différentes langues (EDF).
- L'intégration et le couplage des approches d'intelligence artificielle (IA symbolique et IA classique) avec les données métier dont sera construite une ontologie dans le but d'aider efficacement les opérateurs dans les centres d'exploitation du réseau électrique en anticipant les situations particulières pour mieux les gérer (période d'activité intense/gestion de travaux/incident (par exemple, une ligne hors tension)) dans un contexte en très forte mutation (RTE).

Au niveau académique, une thèse sera menée en collaboration avec CentraleSupélec sur le sujet de l'adaptation de domaine. Cette thèse portera sur les ontologies et les techniques de transfer learning dans l'analyse de documents multimédias et aura pour objectif de proposer des avancées en hybridation des approches de représentation des connaissances et du raisonnement (ontologies) et des approches d'apprentissage profond pour l'analyse de documents et de données multimodales hétérogènes.

A terme, les fonctions de construction, d'intégration et d'exploitation des connaissances développées au sein du projet SMD seront réunies au sein d'une plateforme technologique qui sera valorisée par SystemX.

Le programme IA2

SMD est le troisième projet du programme « Intelligence Artificielle et Ingénierie Augmentée (IA2) » lancé par l'IRT SystemX en février 2020. Ce programme est inédit à la fois par son ampleur et son format et répond aux attentes des industriels qui conçoivent, développent et valident des systèmes. A travers 6 projets de R&D collaboratifs, chapeautés par un projet amont de mise en commun des résultats scientifiques coordonné par Marc Schoenauer, chercheur renommé de Inria Saclay, l'IRT SystemX propose de développer en 5 ans des solutions hybridant trois approches de modélisation et de simulation : la modélisation physique des systèmes, l'utilisation de connaissances métier ou de modèles de comportements exprimés par des experts, et les modèles d'apprentissage machine basés sur les données réelles.

À propos de l'IRT SystemX

SystemX est un institut de recherche technologique (IRT) expert en analyse, modélisation, simulation et aide à la décision appliqués aux systèmes complexes. Seul IRT dédié à l'ingénierie numérique des systèmes, il coordonne des projets de recherche partenariale, réunissant académiques et industriels dans une perspective multi-filière. Ensemble, ils s'appliquent à lever des verrous scientifiques et technologiques majeurs de 4 secteurs applicatifs prioritaires : Mobilité et Transport autonome, Industrie du futur, Défense et Sécurité, Environnement et Développement durable. Au travers de projets orientés cas d'usage, les ingénieurs-chercheurs de SystemX répondent aux grands enjeux de notre temps, sociétaux et technologiques, et contribuent ainsi à l'accélération de la transformation numérique des industries, des services et des territoires.

Basé sur le plateau de Paris-Saclay, Lyon et Singapour, SystemX a lancé depuis sa création en 2012, 53 projets de recherche (dont 29 en cours), impliquant plus de 100 partenaires industriels et 55 laboratoires académiques, et compte actuellement 197 collaborateurs en équivalent temps plein (ETP) dont 134 ressources propres.

Pour en savoir plus : www.irt-systemx.fr | [Twitter](#) | [LinkedIn](#) | [YouTube](#)

Contacts presse

Marion Molina – Claire Flin

Tél. 06 29 11 52 08 / 06 95 41 95 90

marionmolinapro@gmail.com / claireflin@gmail.com