

Appel à projets Thématiques Spécifiques en Intelligence Artificielle (TSIA) – Edition 2024

Acronym	FATES-MLOps		
Project Title	Incorporating FATES Principles in Continuous Development of ML-Integrated Systems: A MLOps Perspective		
Total requested budget	603.8K	Duration	48 mois
Keywords	Fairness Accountability Transparency Ethics Security Continuous Integration		

Project Coordinator (main contact for the proposal)

Name	Brueel Jean-Michel
Institution/Department	Institut de Recherche en Informatique de Toulouse (IRIT)
Address	118 rte de Narbonne 31062 Toulouse
Country	France
Phone	+33 686 00 29 02
E-mail	Jean-Michel.Brueel@irit.fr

Table of persons involved in the project

Partner (Institution / Department)	Last Name	First name	Current position	Role & responsibilities in the project (4 lines max)
1. Lab. IRIT	Brueel	Jean-Michel	Professeur	Project Coordinator
	Teste	Olivier	Professeur	WP1 & 3 contributor
	Pantel	Marc	Maitre de Conférences	WP1 Leader
2. Lab. I3S	Blay-Fornarino	Mireille	Professeur	WP2 Leader
	Collet	Philippe	Professeur	WP2 contributor
3. INRIA Sophia	Precioso	Frédéric	Professeur	WP3 Leader
	Riveill	Michel	Professeur	WP3 contributor
4. McSCert	Mosser	Sébastien	Associate Prof.	WP1 contributor
	Paige	Richard	Joseph Ip Distinguished Engineering Professor	WP1 contributor

Summary of the project in French (publishable non-confidential abstract, max. 1/2 page):

Le mouvement **MLOps** reprend les objectifs DevOps de réduction des écarts entre les équipes de développement et d'opérations en intégrant la collaboration avec les équipes de data scientists et les phases liées à la construction et le déploiement des modèles de Machine Learning (ML).

Notre projet a pour ambition d'étudier les propriétés extra-fonctionnelles telles que l'équité, la responsabilité, la transparence et la sécurité, regroupées en anglais sous l'acronyme **FATES**. En nous appuyant et en affinant les concepts et outils éprouvés du génie logiciel, nous souhaitons proposer une approche systématique et outillée pour la prise en compte de ces propriétés fondamentales dans le cycle de vie d'un logiciel développé en suivant une approche MLOps.

Les verrous technologiques portent sur la formalisation et la mesure de ces propriétés en fonction des contextes et leur prise en charge systématique dans le processus MLOps par des mécanismes et algorithmes adaptés. Cela implique l'analyse et la conception des workflows de construction des modèles, les processus d'intégration et de déploiement, ainsi que la justification du respect de ces propriétés.

Appel à projets Thématiques Spécifiques en Intelligence Artificielle (TSIA) – Edition 2024

Summary of the project in English (publishable non-confidential abstract, max. 1/2 page):

The **MLOps** movement adopts the DevOps objective of reducing the gaps between development and operations teams by integrating data scientist teams and Machine Learning (ML) models. In this project, we wish to apply and adapt good software engineering practices to strengthen both the overall quality of the ML model construction processes and the quality of the software systems produced, particularly in terms of extra-functional properties that will become crucial issues: Fairness, Accountability, Transparency, Ethics, and Security (**FATES**). The key concerns will tackle the study, formalization, measurement, and management of these properties throughout the continuous MLOps process. Indeed, more than traditional Key Performance Indicators (KPIs), such as precision and recall, are required to evaluate models' robustness in practical applications. Our project aims to study the FATES properties and, by refining proven software engineering concepts and tools, propose a systematic and tailored approach for considering those properties, particularly from the lens of ML Scientists or ML Engineers, throughout the lifecycle of the software developed following an MLOps approach.

1. Excellence scientifique et technique, innovation et avancées en R&D

1.1. Motivations et Pertinence pour l'Appel à Projet

Le mouvement **MLOps**¹ reprend les objectifs du mouvement DevOps [DevOps2021], qui est né de la nécessité de réduire les écarts entre les équipes de développement et d'opérations, en y intégrant la collaboration avec les équipes de data scientists et les phases liées à la construction des modèles de Machine Learning (ML) [Testi2022]. Ainsi mener un projet qui intègre du ML en suivant une démarche MLOps implique l'automatisation, l'intégration et la surveillance à toutes les étapes de la construction d'un système ML, y compris l'entraînement, l'intégration, les tests, la publication, le déploiement et la gestion de l'infrastructure. Cette systématisation des processus de construction des modèles de ML s'accompagne d'une exigence sur la qualité des systèmes logiciels produits. Cependant, cette qualité reste à définir, étudier, formaliser, mesurer, notamment dans le contexte des systèmes intégrant du ML. La surveillance continue des modèles ML est cruciale pour garantir leur performance et leur qualité dans des contextes réels, notamment leur adaptation quand les données évoluent.

Le mouvement international pour passer du "Model-centric AI" au "Data-Centric AI" met en exergue la nécessité de vérifier et justifier que les systèmes respectent notamment les propriétés FATES tout au long du processus de construction de systèmes intégrant du ML. Nées en 2014 de l'initiative [FAT/ML](#), les propriétés d'équité (*Fairness*), responsabilité (*Accountability*), et de transparence (*Transparency*), ont été complétées par l'éthique (*Ethics*) pour donner le groupe de recherche Microsoft [FATE](#), puis plus récemment par la sécurité et la sûreté (*Security/Safety*) pour donner les propriétés **FATES**, et le mouvement [Data for Good](#) de l'Université de Columbia. Pour répondre à ces exigences, plusieurs algorithmes ont été développés pour aborder les propriétés à différents degrés (F, T, et S), tandis que d'autres propriétés reposent plus sur l'engagement (A et E). Les réglementations internationales et la société en général exigent, de manière croissante, de la transparence et des responsabilités de la part des « développeurs » de systèmes utilisant ces modèles. Les Etats s'attachent aujourd'hui à proposer des cadres pour aider les organisations et les individus « à favoriser la conception, le développement, le déploiement et l'utilisation responsables des systèmes

¹ <https://ml-ops.org/>

Appel à projets Thématiques Spécifiques en Intelligence Artificielle (TSIA) – Edition 2024

d'IA au fil du temps »^{2,3}. Actuellement, il n'existe pas, à notre connaissance, d'étude systématique ni de support pour guider les scientifiques et/ou les ingénieurs en ML sur des indicateurs permettant le suivi des propriétés FATES. Il impacte l'ensemble du cycle de vie du logiciel de manière variable, en fonction des problèmes et des avancées dans le domaine. En s'appuyant et en affinant les concepts et outils du génie logiciel, notre projet a pour ambition d'étudier les propriétés FATES, de proposer une démarche outillée systématique pour la prise en compte de ces propriétés fondamentales dans le cycle de vie d'un logiciel développé en suivant une approche MLOps. Le verrou auquel ce projet s'attaque donc est le suivant :

Peut-on inclure de manière systématique et évolutive la justification d'une construction et d'une exploitation FATES d'un système logiciel intégrant du ML ?

1.2. État de l'art

Nous abordons l'état de l'art selon deux axes. D'une part les propriétés FATES en mettant l'accent sur les points à vérifier et d'autre part les outils qui doivent être adaptés pour aider à prendre en charge ces propriétés dans un processus MLOps.

1.2.a. Les propriétés FATES

Les propriétés FATES se recoupent. Pour garantir l'équité, il est essentiel de pouvoir expliquer le modèle, d'assurer la fiabilité des données utilisées et de surveiller les dérives potentielles lors de l'exploitation du modèle. Sans transparence, la responsabilité devient plus complexe à définir. Dans cette section, nous présentons ces propriétés séquentiellement dans l'ordre de l'acronyme FATES, en mettant en évidence les mécanismes et algorithmes, lorsqu'ils existent, à intégrer dans un processus MLOps pour garantir ces propriétés.

Fairness/Équité – La recherche sur l'équité dans l'apprentissage automatique vise à garantir l'impartialité des décisions ou prédictions des modèles construits [FAPFID23]. Définir formellement l'équité est un domaine de recherche actif, impliquant des spécialistes en mathématiques, en informatique, en sciences sociales, et des juristes. Les biais peuvent apparaître à diverses étapes d'un processus ML [Suresh2021]. Des algorithmes sont proposés pour atténuer les biais, notamment le débiaisement des données lors de la collecte et l'analyse des modèles de ML [feldman2015]. Pour détecter les biais, de nombreuses métriques sont proposées, récemment Wachter et al. [Wachter2021] proposent la disparité démographique conditionnelle (CDD) comme référence statistique pour évaluer la discrimination potentielle dans les systèmes automatisés. Dans [Breck2017], les auteurs identifient différentes formes de tests pour détecter ces dérives. Plusieurs travaux sur les LLMs mettent en avant une combinaison de ces approches [Brown2020, Ferrara2023].

Accountability/Responsabilités – Aujourd'hui, la nécessité pour les producteurs de systèmes logiciels d'assumer la responsabilité des choix effectués est largement discutée tant les parties prenantes sont nombreuses et impactantes à des niveaux différents⁴. La responsabilité signifie que la manière dont un résultat d'un modèle a été obtenu grâce à un système de bout en bout, est compréhensible/explicable, est vérifiable et est reproductible⁵. Le versionnement des modèles comprenant les informations sur les données d'apprentissage, les résultats des tests, ainsi que des environnements de calcul, est un moyen courant pour garantir la traçabilité. Des frameworks comme

² AI, NIST. (2023). Artificial Intelligence Risk Management Framework (AI RMF1.0) <https://doi.org/10.6028/NIST.AI.100-1>

³ European AI Act: Regulation of The European Parliament and of The Council Laying Down Harmonised Rules On Artificial Intelligence and Amending Certain Union Legislative Acts, Document 52021PC0206, COM/2021/206 final

⁴ Air Canada, <https://intelligence-artificielle.com/chatbot-air-canada-hallucine/>

⁵ La CNIL en donne une définition différente, dans ce projet, nous nous limiterons à la définition donnée ici : <https://www.cnil.fr/fr/developpement-des-systemes-dia-les-recommandations-de-la-cnil-pour-respecter-le-rgpd>

Appel à projets Thématiques Spécifiques en Intelligence Artificielle (TSIA) – Edition 2024

MLFlows et les approches par conteneurs sont utilisés dans le contexte MLOps pour améliorer cette traçabilité qui reste cependant à renforcer [Chen2020]. La réutilisation des modèles pré-entraînés est devenue indispensable dans la construction de nouveaux modèles, en particulier pour les approches basées sur les LLMs. En explicitant les dépendances à la version de ces modèles, la traçabilité est renforcée, mais la question de l'introspection des modèles, notamment en ce qui concerne les données utilisées pendant la phase de pré-entraînement, devient plus forte [Liu2024].

Transparency/Transparence – L'IA explicable (XAI) est un champ de recherches très intenses visant à rendre les décisions des modèles d'IA compréhensibles par les humains, même si la pleine explication des modèles reste un défi [CIKM22]. Les algorithmes d'explication peuvent être classés en méthodes *ante-hoc*, nécessitant l'accès aux mécanismes internes du modèle, et en méthodes *post-hoc* n'accédant qu'aux prédictions du modèle [Lopardo2023, Lopardo2024]. En production, l'utilisation d'algorithmes *post-hoc* est privilégiée. L'utilisation d'architectures à base d'événements est une solution pour atteindre le double objectif d'indépendance, permettant des traitements de surveillance adaptés, et une montée en charge [Klaise2020]. Des défis persistent dans la surveillance et l'explication des modèles déployés, des solutions sont déjà disponibles pour en relever certains [Wang2024], mais déterminer les solutions techniques à partir des spécifications d'un problème reste une difficulté majeure qui entrave la production de systèmes d'IA transparents [Mill2024].

Ethics/Éthique – La question de l'éthique est intrinsèquement liée à la philosophie, et déterminer si un système est éthique ou acceptable dépend souvent du point de vue adopté, qui peut varier d'un individu à un autre, voire d'un contexte à un autre. En conséquence, évaluer l'éthique d'un système peut se situer en amont du projet. Dans le cadre de ce projet, nous nous inscrivons dans la logique des Responsible AI Licences (RAIL) [Contractor2022].

Safety & Security/Sécurité – La prise en compte de la sécurité est une préoccupation largement documentée, y compris récemment dans le cadre du DevOps, par un focus appelé DevSecOps [VeriDevOps23, RQCODE22]. Les spécificités de la sécurité dans MLOps vont surtout concerner la *privacy*. Les utilisateurs de solutions basées ML sont légitimement en interrogation du devenir des données (où sont stockées les données, qui y a accès, etc.). Nous utiliserons plus particulièrement l'exemple prégnant de l'anonymisation des données. L'autre facette de la sécurité en français (au sens de la *safety* en anglais), par exemple qui est responsable en cas de problème de sécurité, concerne plus les propriétés de responsabilité (*Accountability*) et sera donc abordée dans cette propriété. Enfin, la sécurité doit également garantir que les modèles de ML sont robustes face aux attaques et ne peuvent pas être utilisés à des fins malveillantes. Ce dernier point, bien que très actuel avec l'injection de codes malveillants par les prompts dans les LLMs, ne sera pas abordé parce qu'il pourrait représenter un projet à lui seul.

Les propriétés FATES sont contextuelles au système logiciel. On ne recherche pas à garantir les mêmes propriétés, ou du moins pas à un même degré selon l'usage et le domaine (critique ou non, impliquant l'humain ou non, spécifique ou général, etc.). **Ces propriétés sont invasives** dans l'ensemble du cycle de vie d'un logiciel, par exemples, dans l'analyse du problème (est-il éthique d'aborder cette question?), dans la collecte des données (est-ce que les données collectées sont équitables?), dans les choix des modèles (que sait-on des décisions prises par les modèles produits?), dans les traitements opérés sur les données pour apprendre (est-ce que les choix pour améliorer l'efficacité de l'entraînement sont pris en responsabilité?), dans les traitements réalisés en exploitation (est-ce que les données utilisées pour renforcer l'apprentissage ne brise pas l'équité du modèle?), etc. **La surveillance de ces propriétés évolue en fonction des connaissances** que nous avons des systèmes de ML et des cas réels observés. Par exemple, si certains types de biais sont connus et des algorithmes ont été définis pour pallier ces biais, d'autres tels que la production

Appel à projets Thématiques Spécifiques en Intelligence Artificielle (TSIA) – Edition 2024

d'exemples adversaires sont proposés chaque jour. Cette évolution est si forte que le document de référence produit par le NIST⁶ en matière des risques liés à l'IA, est conçu comme un document vivant⁷. Dans [Sculley2015, Breck2017], les auteurs mettent en évidence différents facteurs de risque spécifiques au ML à prendre en compte dans la conception du système. Bien que nous ayons choisi un angle d'attaque différent, l'étude des propriétés FATES adresse différents éléments de dettes, dont ceux dits liés aux changements dans le monde extérieur tels que le monitoring et le test, le choix de métriques, mais aussi la gestion du processus mise à jour et de reconstruction des modèles.

1.2.b. Les outils en support au FATES MLOps

Il existe, comme nous l'avons vu, de nombreux algorithmes et outils qui visent à mesurer et garantir les propriétés FATES et ceux-ci sont en plein développements rapides et concurrents. Nous abordons ici la question davantage d'un point de vue GL et intégrateur.

Le test et la surveillance

Dans [Breck2017], les auteurs mettent en exergue la difficulté de formuler des tests spécifiques, puisque le comportement réel d'un modèle de prédiction donné est difficile à spécifier a priori. En comparant l'entraînement d'un modèle à de la compilation, ils proposent différentes approches du test complémentaires où la source est à la fois le code et les données d'entraînement. Même si ces tests ne sont pas liés aux propriétés FATES, il est intéressant de reprendre certaines d'entre elles comme les exigences de méta-niveaux pour réduire les biais ou les contrôles de privacy.

Les environnements

L'un des grands défis du MLOps dans le contexte du suivi des propriétés FATES est de concevoir des systèmes qui intègrent à la fois les bons composants pour adapter les données, de la surveillance adaptée des déploiements, déclenchent des alertes, assurent un versionnement et une traçabilité des modèles. Actuellement, plusieurs outils d'automatisation du machine learning sont disponibles, tels que MLFlow [Zaharia2018], SageMaker et Kubeflow. Cependant, à notre connaissance, aucun de ces outils n'aborde explicitement la question du support à la vérification des propriétés FATES, que ce soit lors de la production des modèles ou de la vérification des composants de surveillance de ces propriétés. En intégrant des solutions de surveillance des propriétés FATES dans des piles logicielles telles qu'Hugging Face et Langchain, nous visons à répondre aux besoins croissants de la communauté en matière de contrôle qualité et de fiabilité des systèmes intégrant du ML.

1.2.c. Des exigences aux Justifications

Modélisation et variabilité

Construire des systèmes efficaces de science des données est d'autant plus difficile que les solutions ML disponibles ne cessent de croître [Zaharia2018]. Pour aider les DS à sélectionner des pipelines cohérents en fonction de leur problème, nous avons appréhendé cette diversité sous la forme d'une ligne de produit [Amraoui2022], que nous exploitons pour identifier des solutions à réutiliser [Brault2023] et avons modélisé un méta-modèle de pipeline de ML pour les prendre en charge dans un contexte DevOps [Benni2019]. Sur la base de ces travaux préliminaires, nous visons à étendre notre approche pour intégrer spécifiquement les propriétés FATES dans les workflows de ML [Vasudevan2020], en capturant la variabilité des algorithmes avec la logique des feature models.

Justifications

Dans les contextes critiques, la documentation joue un rôle essentiel dans l'accréditation des produits en établissant la confiance dans leur processus de développement et de conception. L'objectif est d'élaborer des justifications pour rassurer sur la gestion appropriée du processus de développement

⁶ National Institute of Standards and Technology, U.S. Department of Commerce

⁷ AI, NIST. (2023). Artificial Intelligence Risk Management Framework (AI RMF1.0) <https://doi.org/10.6028/NIST.AI.100-1>

Appel à projets Thématiques Spécifiques en Intelligence Artificielle (TSIA) – Edition 2024

et le respect des normes. Justifier qu'un système logiciel respecte les propriétés FATES⁸, rejoint le même objectif. Dans [Polacsek2018], Polacsek et al. ont introduit les diagrammes de justification (JD), en conformité avec l'IEC 62304 pour organiser les éléments contribuant à la justification d'un résultat. Dans [Duffau2018], nous avons étendu et appliqué ces diagrammes à l'élaboration de dispositifs médicaux critiques, puis plus récemment, à la justification de pipelines DevOps à grande échelle [Mosser2023]. Nous proposons de poursuivre ces travaux dans le cadre du MLOps.

1.3. Approche/Solution et plus-value scientifique

Dans le domaine dynamique et en constante évolution du ML, et plus spécifiquement dans le cadre d'une approche responsable des systèmes intégrant des LLMs, les recherches menées par les juristes, philosophes et sociologues revêtent une importance capitale. Cependant, pour que ces avancées puissent profiter à un large éventail d'acteurs, y compris les entreprises de taille plus modeste, il est essentiel de rendre les concepts et les pratiques accessibles et intégrables dans les processus de développement logiciel. C'est précisément l'objectif de notre projet, qui s'attaque de manière pragmatique à un problème complexe, embrassant de multiples dimensions. En alignant les développements algorithmiques sur les besoins concrets de la production logicielle, notre ambition est de fournir à la communauté scientifique des principes, des théories et des outils favorisant une approche systématique de l'évaluation et de la traçabilité des propriétés FATES, de la phase d'analyse jusqu'à la mise en exploitation. Dans cette démarche, nous nous engageons également à mettre en évidence les limites ainsi que les progrès réalisés dans ce domaine.

1.4. Sorties attendues du projet et mesure objective de qualité

Les sorties attendues du projet sont directement liées aux cas d'usage du WP3, avec d'une part un ChatBot en Wolof, dont une évaluation tout au long du développement sera réalisée, et d'autre part des composants intégrés à StarCoder et évalués sur la génération des codes. Nous évaluerons par exemple si les codes générés par StarCoder avec nos propriétés FATES ont récupéré, eux aussi, des propriétés FATES.

La caractérisation des propriétés FATES dans leur globalité servira de guidelines pour les implémentations FATES futures. Nous produirons des exemples de prise en compte des propriétés FATES et de leur justification dans un processus d'intégration et de déploiement continu.

Nous produirons des éléments de mesures sur les exigences et la qualité des propriétés en utilisant au moins des métriques de l'état de l'art. Nous distribuerons en open source les artefacts logiciels de définition, de mesure, d'intégration d'algorithmes, de trace qui seront évalués à la fois dans la diversité des mécanismes pris en compte et des applications considérées.

2. Faisabilité

2.1. Démarche scientifique

2.1.a. Complémentarité et principes généraux

Nous présentons à présent les bases de notre démarche, qui repose sur une approche transversale des propriétés FATES : de leur analyse en tant qu'exigences à leur surveillance dans les applications intégrant des composants de ML. Cette approche nécessite une collaboration étroite entre les chercheurs en génie logiciel et en sciences des données, collaboration déjà existante et fructueuse [Benni2019, Brault2023]. Pour atténuer les risques inhérents à un domaine aussi dynamique et aux multiples applications, nous adopterons une approche itérative, cohérente avec MLOps, en

⁸ Cf. le début de formalisation Microsoft Responsible AI Standard, v2 GENERAL REQUIREMENTS <https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE5cmFI?culture=fr-fr&country=fr>

Appel à projets Thématiques Spécifiques en Intelligence Artificielle (TSIA) – Edition 2024

enrichissant progressivement les processus d'analyse avec de nouvelles propriétés et mécanismes. L'opérationnalisation de ces propriétés sera ainsi au cœur de notre démarche.

En intégrant les propriétés FATES dès la phase d'analyse d'un problème et en les suivant tout au long du cycle de vie du logiciel, notre projet vise à améliorer les développements en IA, ainsi que les systèmes qui incorporent ces technologies.

Notre approche se distingue en ce sens que nous ne cherchons pas à développer un nouveau framework, mais plutôt à mener une étude approfondie pour comprendre les interdépendances entre les propriétés, les outils et les objectifs. Nous proposerons des versions outillées des points abordés, tout en reconnaissant que nous ne pourrions pas couvrir tout l'espace des propriétés FATES, qui est très vaste.

2.1.b. Qualification/modélisation/formalisation des propriétés FATES

Nous visons à formaliser les propriétés FATES pour guider leur analyse et faciliter leur intégration dans le processus de développement en tenant compte des exigences qui portent sur un système donné. Nous établirons des relations logiques entre les exigences et les algorithmes/recommandations existants, afin de guider le choix des composants algorithmiques à intégrer dans le développement, qu'il s'agisse des codes d'entraînements, des chaînes d'intégration continue, déploiement ou des workflows tels que définis par Langchain. En nous basant d'une part sur la formalisation des propriétés et d'autres parts sur les algorithmes existants ou recommandations, nous visons à guider l'analyse des propriétés FATES et à faciliter leur intégration dans le processus de développement. Nous ne développerons pas d'algorithmes, ni ne proposerons de nouvelles recommandations. Nous nous plaçons en aval de ces recherches ; nous nous focaliserons sur leur exploitation systématique dans le développement des applications.

2.1.c. Intégration dans le Processus MLOps

Dans la mesure des artefacts logiciels dont nous disposons, nous analyserons les différentes étapes du processus de développement pour intégrer et vérifier la présence de composants utiles au suivi et au respect des propriétés FATES. Nous nous concentrerons sur les workflows d'entraînement des modèles de ML (e.g., pour suréchantillonner les groupes sous-représentés), les compositions de pipelines dans LangChain (e.g., pour introduire une étape de débiaisage sur les données de renforcement) et des workflows de CI/CD (e.g., pour déployer un modèle d'explication parallèle au système ou un système de journalisation des événements). Nous développerons des justifications automatiques pour suivre et documenter les compromis et les vérifications effectués. Cette étape intégrera le développement de COTS réutilisables, indépendants et composables, permettant à d'autres travaux de minimiser leur effort d'intégration de ces bonnes propriétés. Nous mettrons en œuvre notre approche dans des applications MLOps, qui serviront également de démonstrateur.

En résumé, notre approche vise à formaliser les propriétés FATES, à les intégrer de manière systématique dans le processus de développement, et à les appliquer dans des applications MLOps réelles pour garantir des systèmes plus responsables et fiables.

2.2. Méthodologie et gestion des risques

La pertinence et l'originalité de notre projet résident dans notre approche intégrative des propriétés FATES dans le processus de développement. Ce qui distingue notre consortium, c'est :

1. Une approche interdisciplinaire qui évalue et intègre ces propriétés du point de vue des *Data Scientists* et des développeurs (*ML Engineer*);
2. L'utilisation de paradigmes et d'outils complémentaires pour vérifier, mesurer et visualiser la prise en compte des propriétés FATES sous différentes perspectives telles que l'architecture, les métriques et les algorithmes palliatifs ;

Appel à projets Thématiques Spécifiques en Intelligence Artificielle (TSIA) – Edition 2024

3. Une démarche proactive de diffusion à travers l'open-source et diverses communautés, notamment via Hugging Face pour les grands modèles, l'implication du McSCert à l'international et l'interaction avec les communautés de développeurs en rendant nos codes disponibles en open-source et en s'intéressant à LangChain.

Nous adoptons *une approche scientifique centrée sur les modèles* au sens de représentations multiples d'un même objet d'étude combinant plusieurs formalismes et techniques. Cette approche nous offre les outils nécessaires pour adapter les représentations aux divers points de vue au fil du temps, tout en permettant de tracer les exigences depuis leur formalisation initiale jusqu'à leur intégration dans les processus de développement et les architectures logicielles, et enfin jusqu'à la production des justifications nécessaires. En suivant une *démarche itérative et incrémentale, fondée sur la littérature et les applications*, nous cherchons à atténuer les risques liés à la complémentarité entre génie logiciel et sciences de la donnée ainsi qu'à la complexité du domaine, tout en assurant une production continue et rapide. Notre approche repose sur l'utilisation combinée de formalismes et de techniques du Génie Logiciel pour formaliser les exigences, modéliser les workflows, capturer la variabilité des algorithmes et métriques associés aux propriétés FATES, et justifier de la prise en compte de ces propriétés dans le cycle de vie du logiciel. Nous utilisons notamment l'expression des exigences, les modèles de variabilité tels que les feature models et les systèmes de contraintes, les diagrammes de justification, ainsi que les principes de configurations dans les lignes de produits logiciels. En outre, nous faisons appel à l'ingénierie dirigée par les modèles pour établir des ponts entre ces formalismes et assurer une intégration cohérente des propriétés FATES dans le cycle de vie du logiciel.

Atténuation des risques

Pour atténuer les risques liés à l'intégration de différentes formalisations et construire rapidement des ponts entre les chercheurs en Data Science et en GL, ainsi que pour assurer une modélisation précise des interactions entre les différentes approches (formalisation des exigences, modélisation des workflows, exploitation de la variabilité pour représenter les techniques de prise en compte des propriétés FATES, construction des diagrammes de justifications), nous consacrerons les six premiers mois du projet au développement d'un premier Produit Minimum Viable (MVP). Les équipes ayant déjà travaillé ensemble et connaissant bien leurs outils respectifs, cette étape permettra d'approfondir les automatismes nécessaires à la mise en œuvre de la chaîne logicielle. La tâche T0.2 a pour but de minimiser les risques sur le reste du projet en développant une version simplifiée, mais complète de la chaîne de raisonnement attendue. L'expérience de l'équipe UniCA dans la collaboration avec les autres équipes est un avantage majeur pour appréhender l'interdisciplinarité du projet. Pour appréhender la diversité des systèmes exigeant la prise en compte des propriétés FATES, nous réduisons l'espace aux systèmes basés sur des LLMs et disponibles en open-source. Nous suivrons une approche itérative par intégration progressive des applications et des propriétés traitées. Les études sur les architectures de StarCoder et sur des applications LangChain ont été commencées dans des contextes d'enseignements et de partages de connaissances. Nous sommes confiants dans l'applicabilité des principes énoncés dans le projet. Pour s'assurer de l'adéquation des modélisations, notamment des exigences en amont et des justifications en aval, avec les besoins des différentes parties prenantes, nous viserons des productions régulières des sorties du projet avec les différentes communautés notamment avec des échanges privilégiés avec Hugging Face.

2.3. Organisation en WorkPackage

Notre projet compte trois *Work Packages* WPs, présentés ci-après. La complémentarité des axes de recherche des partenaires les amène à participer à tous les WPs. Ainsi, pour chaque WP, sont précisés

Appel à projets Thématiques Spécifiques en Intelligence Artificielle (TSIA) – Edition 2024

la période sur laquelle il sera traité et sa coordination. Pour chaque tâche sont précisés les partenaires impliqués avec, en gras, le responsable de la tâche et les livrables attendus.

WP0 [0-48] : Veille, Coordination et Diffusion (coord. IRIT)

En plus de la coordination du projet et de sa dissémination (e.g., diffusion des composants à intégrer dans les processus d'intégration en open-source; analyses d'applications), ce WP comprend une tâche initiale d'intégration (équipes et outils) ainsi que de positionnement par rapport à l'état de l'art.

Tâche 0.1 - Coordination (IRIT)

Cette tâche se concentrera d'abord sur toutes les activités nécessaires au bon démarrage du projet (préparation de l'accord de consortium, mise en place des outils de gestion de projet et de communication nécessaires, etc.). Le processus de gestion du projet assurera : (a) la coordination et le fonctionnement efficace de toutes les activités scientifiques, techniques et de mise en réseau du projet (y compris l'organisation de réunions virtuelles et physiques régulières); (b) le suivi continu des progrès et l'établissement de rapports réguliers, avec au moins un livrable tous les 6 mois, à la fois à des fins internes et pour les rapports externes à l'ANR; (c) la définition du Plan de données assurera au projet de respecter toutes les exigences légales (workflows qui ne seront pas tous créés par le consortium, ceux collectés sur des dépôts de codes), et produira des données telles que des artefacts liés à l'analyse de workflows ou des formalisations d'exigences.

Acteurs : tous les partenaires

Livrables : L0.1 (M6) : Site Web (onboarding, partage des exigences, des codes et autres livrables, ...).

Tâche 0.2 - Intégration par chaîne outillée (UniCA)

Cette tâche a pour but de produire un produit minimum viable. Ce produit prendra en charge (i) une formalisation simple de chacune des propriétés (e.g., les données sont équilibrées (F), les métriques relatives aux données sont accessibles (A), l'algorithme est public (T), le code est sous licence RAIL (E), les données sont anonymisées (S)), (ii) une chaîne d'intégration et un pipeline qui intègrent les composants nécessaires au suivi des propriétés FATES conformément à ces exigences, (iii) des diagrammes de justification seront produits pour expliciter le respect de ces premières exigences. Pour cette tâche, une partie de la chaîne restera manuelle, et l'application servira d'exemples fil rouge sur lequel nous souhaitons pouvoir facilement communiquer..

Acteurs : tous les partenaires

Livrables : L0.2 (M6) Chaîne minimale outillée dans un processus MLOps minimal.

Tâche 0.3 - Dissémination et Exploitation (Inria Sophia)

Cette tâche a pour but de diffuser les résultats du projet auprès des communautés académiques en GL et en sciences des données ainsi que des industriels qui souhaitent vérifier et afficher des propriétés FATES. Nous ne produirons des résultats que sur des données qui se trouvent dans le domaine public ou pour lesquelles nous avons obtenu le consentement explicite de leurs détenteurs de droits pour l'objectif visé. Un contact étroit sera maintenu avec le service de la protection des données d'Inria afin de garantir une gestion sûre et responsable des données. Toutes les données produites dans le cadre du projet seront mises à disposition conformément aux directives de l'Open Science et en accord avec le Plan de données établi en Tâche 0.1.

Acteurs : tous les partenaires

Livrables : L0.3(M48) La chaîne outillée dans sa globalité : elle intègre le siteweb, les composants, les exemples, les liens vers les applications...

Appel à projets Thématiques Spécifiques en Intelligence Artificielle (TSIA) – Edition 2024

WP1 [0-42]: Modélisation des propriétés FATES (IRIT)

Ce WP est dédié à l'étude, la formalisation, la mesure des propriétés FATES afin de les intégrer dans le cycle MLOPs (T1.1). L'objectif, outre de caractériser les exigences FATES pour anticiper leurs évolutions, est de faire le lien entre les exigences et leur gestion par les algorithmes, dans un souci permanent d'être en mesure de proposer des justifications (T1.2). Cela implique une gestion et une traçabilité de ces exigences sur l'ensemble de la chaîne MLOPs avec des décisions tracées et intégrées dans le système de justification (T3.3).

Tâche 1.1 - Des exigences FATES au respect des propriétés (IRIT)

Cette tâche réalisera l'étude et la caractérisation des propriétés FATES par la définition de leurs exigences (qui seront amenées à évoluer avec le temps et leur prise en compte de plus en plus systématique dans le futur). Nous établirons de plus des liens entre ces exigences de haut niveau définies sur les propriétés FATES, les composants algorithmiques et les métriques ou artefacts. Ces liens nécessiteront parfois la réalisation de composants dédiés chargés de vérifier, voire, de mettre en œuvre certaines caractéristiques des propriétés (e.g. anonymisation). Cette tâche nécessite une étroite collaboration avec la T2.2 ci-après (le livrable correspondant étant commun L1.1c/ L2.2a).

Acteurs : IRIT, UniCA, McSCert

Livrables : L1.1a(M6) Artefacts représentant les exigences FATES, L1.1b(M12) KPIs de mesures, L1.1c/L2.2a(M18) Composants dédiés.

Tâche 1.2 - Modélisation des justifications pour propriétés FATES (IRIT)

Les propriétés FATES ne sont pas indépendantes (par exemple la Transparency participe à renforcer la Fairness) et leur prise en compte n'est pas binaire. Il convient d'être capable de fournir la justification du niveau de prise en charge (qui peut bien sûr atteindre les 100%) de ces propriétés, voire de définir, pour un projet ou une application donnée, les priorités données par l'organisme responsable du développement basé ML. Les systèmes de justifications permettent de répondre à cette problématique. Cette tâche consistera à définir le modèle générique associé aux propriétés FATES.

Acteurs : IRIT, Inria Sophia, McSCert

Livrables : L1.2(M18) Le modèle de justification intégrée des propriétés FATES.

Tâche 1.3 - Opérationnalisation des justifications pour propriétés FATES (IRIT)

En s'appuyant sur les artefacts et composants dédiés (issus de T1.1) et sur les modèles de justifications (issus de T1.2), cette tâche réalisera la mise en application concrète en intégrant la génération des modèles de justification dans les workflow intégrant du ML issus de la tâche T2.1 et en prenant en compte les mesures fournies grâce aux livrables de la T2.2.

Acteurs : UniCA, McSCert

Livrables : L1.3a(M24) Modèle de variabilité de propriétés FATES, L1.3b(M24) Génération automatique de modèles de Justification.

WP2 [0-42]: FATES Integration (UniCA)

L'objectif de ce WP est de modéliser les process et outils utilisés dans les processus MLOPs. La modélisation des workflows de construction et d'exploitation des modèles de machine learning vise à caractériser les points d'accroche des mitigations, vérifications et mesures des propriétés FATES (T2.1). Des composants logiciels dédiés à la production des éléments de preuve nécessaire aux justifications automatiques sur l'évaluation des propriétés FATES seront développés (T2.2).

Tâche 2.1 - Modélisation des workflows intégrant du ML (UniCA)

Cette tâche vise à élaborer un méta-modèle de workflow pour caractériser les étapes impliquées dans la vérification et la justification des propriétés FATES. Ce méta-modèle permettra de raisonner

Appel à projets Thématiques Spécifiques en Intelligence Artificielle (TSIA) – Edition 2024

sur ces processus et de visualiser les workflows résultants avec leurs propriétés. Les résultats des tâches T1.1 et T1.2, seront utilisés pour annoter les modèles en relation avec les exigences FATES et les éléments de preuve produits. Les opérationnalisations développées dans T1.3 et les applications du WP3 devront être capturées par ces modèles.

Acteurs : UniCA et IRIT pour la modélisation, Inria Sophia pour la validation

Livrables : L2.1a (M12) : métamodèle de WF et bases d'exemples; L2.1b(M30) : Validation de WFs

Tâche 2.2 - Outillage pour produire les éléments de preuve (UniCA)

Les algorithmes, les métriques, mais également des mécanismes tels que les Logs sont des moyens d'évaluer, mesurer, vérifier les propriétés FATES. Il s'agit dans le cadre de cette tâche de proposer des composants d'intégration, en particulier dans les scripts d'intégration. Cette tâche nécessite une étroite collaboration avec la T1.1 du WP1 (le livrable correspondant étant commun L2.2a/L1.1c).

Par exemple, nous prévoyons de sauvegarder, automatiquement, sous forme d'artefacts exploitables, les métriques liées à la Fairness, ainsi que les résultats de validation comme ceux publiés par StarCoder [Lozhkov2024]. De plus, nous intégrerons les retours des composants pour évaluer la dérive des données, comme décrit dans [Mayaki2022], dans les pipelines de production pour générer dynamiquement des éléments de justification ou des alertes. Ces preuves et leur composition sous la forme de justifications (T1.3) en étant produites de manière systématique sont plus faciles à maintenir et servent à communiquer, mesurer, comprendre.

Acteurs : UniCA pour le développement et Inria Sophia pour le choix et la préparation des algorithmes et métriques.

Livrables : L2.2a/L1.1c(M18) composants dédiés CI/CD; L2.2b(M42) Bibliothèque de composants dédiés.

WP3 [0 - 48]: FATES Applications et Validation (Inria Sophia)

Ce workpackage vise à appliquer et valider nos travaux dans différents contextes impliquant des LLMs. Une analyse de l'existant nous permettra de valider nos modélisations et raisonnements sur des applications concrètes sans nécessairement exiger un contrôle à l'exécution (T3.1). Une opérationnalisation des modélisations dans le processus d'intégration continue sera réalisée sur des applications en cours de construction (T3.2).

Tâche 3.1 - Validation par Analyse des propriétés FATES (Inria Sophia)

Les biais peuvent se glisser dans les codes générés (et bien sûr les codes utilisés pour entraîner le modèle) [Liu2024]. Nous étudierons en particulier l'application StarCoder [Lozhkov2024] pour lesquelles une grande partie des informations nécessaires à son évaluation sous l'axe des propriétés FATES sont disponibles, avec en plus la possibilité de bénéficier d'une collaboration étroite avec Hugging Face, comme mentionné dans leur lettre de soutien⁹.

Les modélisations des workflows, l'identification des composants pour mesurer, évaluer, corriger les propriétés FATES de même que les justifications associées valideront nos travaux. Ces travaux permettront non seulement de valider notre approche, mais également de diffuser la prise en compte des propriétés FATES dans le développement des LLMs et la possibilité de justifier de cette prise en compte.

Acteurs : tous les partenaires

Livrables : L3.1 (M18) les analyses (modèles de WF, Justifications).

⁹ Accessible ici : <https://capture.dropbox.com/YJaDG9w9diYGGEJm>

Appel à projets Thématiques Spécifiques en Intelligence Artificielle (TSIA) – Edition 2024

Tâche 3.2 - Validation par Opérationnalisation des propriétés FATES (UniCA)

L'opérationnalisation pour évaluer dynamiquement et construire les justifications dans les chaînes d'intégrations reposera sur deux approches.

D'une part, une analyse dès la conception du système que nous appliquerons à la construction d'un agent conversationnel pour le Wolof (projet de l'équipe Inria Sophia en partenariat avec deux universités au Sénégal). Cet agent permettra à terme l'accès à différents services publics ou de première nécessité. Il s'agit d'aider à la mise en place des composants permettant de prendre les propriétés FATES dès les phases de conception et tout au long de l'élaboration d'un tel système. La collecte des données respecte les récentes recommandations FATES de la littérature [Gebreu2021].

De plus, nous nous intéresserons à l'étude des architectures impliquant des LLMs dans les tutoriaux ou les systèmes en production que l'on peut trouver dans l'éco-système de Langchain par exemple sur la construction des chatBots [Topsakal2023]. Il s'agira d'étudier les propriétés dans un système logiciel composite. Ces applications et le framework associé sont davantage associés à la communauté des développeurs qu'au développement des LLMs, et mettent en jeu d'autres types d'algorithmes et métriques. Nous renforcerons ainsi nos validations notamment en termes de workflows, ces architectures impliquant parfois plusieurs LLMs (e.g., NLP, Génération de codes, analyse d'images). D'autres parts, nous produirons des preuves de concepts (POC) pour évaluer les différents composants et démontrer la faisabilité et l'intérêt de l'approche sur d'autres cas d'études comme celles menées au sein de McSCert, qui dispose d'une expertise de plusieurs années en l'observation de *ML Engineering* (en lien avec le CIRST, Centre Interuniversitaire de Recherche sur la Science et la Technologie, Montréal, Québec) développant des modèles de support au diagnostic de troubles de la santé mentale [CLEF20, CLEF22, CLEF23].

Acteurs : tous les partenaires

Livrables : L3.2.a(M30) Codes pour validation/intégration FATES dans LangChain; L3.2.b(M46) Codes pour validation/intégration agent conversationnel en Wolof.

3. Partenariat (consortium)

Le coordinateur scientifique du projet, J.-M. Bruel du laboratoire IRIT possède une solide expérience dans le développement de systèmes cyber-physiques à forte composante logicielle, et l'intégration méthodes/modèles/langages, avec un focus sur l'ingénierie des systèmes basée sur les exigences et les modèles. Il est co-responsable du Pôle "Développement de Logiciel" au sein du GDR GPL (Génie de la Programmation et du Logiciel) du CNRS. Il a été en charge de plusieurs projets avec des partenaires industriels, notamment la chaire d'ingénierie des systèmes pilotés par les modèles entre AIRBUS et l'Université Toulouse 2 Jean Jaurès, dont il est titulaire depuis 2022. Il a rejoint le Laboratoire international de recherche en IA (IPAL) en 2023. À l'IRIT, participeront également **M. Pantel**, qui travaille depuis plus de 20 ans sur l'utilisation des langages formels pour la spécification des systèmes complexes en contexte industriel, et **O. Teste**, spécialiste en Data Science qui possède également de nombreux contrats industriels notamment avec Airbus et sur l'analyse des données en contexte ML. L'équipe sera leader sur les WP0 et WP1.

M. Blay-Fornarino (resp. de site) et **P. Collet** du laboratoire I3S (UniCA) sont experts en lignes de produits logiciels (SPL) et plus spécifiquement sur la capture des connaissances dans les systèmes ML. M. Blay-Fornarino est co-directrice du GDR Génie de la Programmation et du logiciel, tandis que P. Collet est un membre éminent de la communauté SPL. L'équipe sera responsable du WP2.

F. Precioso (resp. de site) et **M. Riveill** avec Inria/MAASAI (Inria Sophia) sont experts en science des données. Ils conçoivent des systèmes intégrant du ML pour de nombreux domaines d'application. En particulier, l'équipe a commencé depuis quelques mois un partenariat avec les Universités Gaston

Appel à projets Thématiques Spécifiques en Intelligence Artificielle (TSIA) – Edition 2024

Berger et Cheikh Anta Diop pour la conception d'un Agent Conversationnel afin d'aider les personnes ne parlant que Wolof à accéder à des services de santé, bancaires, d'emploi au Sénégal. L'équipe sera responsable du WP3.

S. Mosser et **R. Paige** sont membres du McMaster Centre for Software Certification (McSCert) à l'Université McMaster, au Canada. S. Mosser est un expert en modélisation et possède une expertise de cinq années à collaborer avec des ML Engineers dans un contexte de santé mentale. R. Paige est un expert en modélisation appliquée à la sécurité et sûreté des systèmes critiques. McSCert, qui vient d'être récompensé en 2024 pendant la conférence ICSE par le prestigieux IEEE TCSE Synergy Awards, est le seul centre de recherche universitaire canadien dédié à la certification du logiciel.

L'ensemble des membres du consortium a l'habitude de collaborer sur des projets de recherche. Les compétences couvertes par le consortium sont une garantie de contributions significatives.

4. Impact, dissemination et exploitation

Les applications de l'IA ont un impact important dans la société. Dans son ambition de souveraineté et de compétitivité, la France lance de nombreux investissements et chantiers autour de l'IA. Nous sommes persuadés qu'elle se doit d'être exemplaire dans ces efforts en investissant également dans la maîtrise des propriétés FATES afin de minimiser les biais et les risques en matière de déploiement de du Machine Learning. En effet, de l'introduction de biais de recrutement chez Amazon lors de l'automatisation de la lecture des CVs, à la non-reconnaissance de personnes racisées par les algorithmes de détection de piétons de Tesla, la non-prise en compte des propriétés FATES lors du développement de produits basés sur de l'IA conduit inévitablement à des situations dramatiques.

Par ses applications pratiques (WP3), ce projet a pour ambition de **démontrer concrètement le ratio coût-bénéfice** de la prise en compte systématique des propriétés FATES lors de la production de logiciels: coût de la documentation, possibilité d'automatisation, impact sur les processus de développement. Les applications visées couvrent différents domaines (Génération de code, agent conversationnel pour langues sous-représentées, Santé mentale). Par ses contributions fondamentales (WP1, WP2), le projet proposera un **cadre conceptuel et outillé** permettant de supporter les ingénieurs logiciels lors de la mise en place de chaînes de production de nouveaux produits logiciels. L'ambition est ici de fournir des **modèles réutilisables et open-source** à la communauté, en se reposant sur l'expertise pré-existante au sein du consortium, sur la publication et maintenance de logiciels "open-source" et de jeux de données "open-data". La compagnie Hugging Face, dans sa lettre de soutien, indique un intérêt tout particulier sur le **transfert technologique** de ces résultats fondamentaux et appliqués à leur propre chaîne de production et d'entraînement de LLMs. Si une exploitation industrielle est hors du périmètre de ce projet (100% académique), l'intérêt d'un acteur majeur du domaine pour valider les résultats obtenus est un atout supplémentaire à la validité, la pérennité des résultats en dehors du projet lui-même, et surtout à sa visibilité qui bénéficiera du rôle central que joue Hugging Face au sein de la communauté.

Conscients de l'ampleur de la tâche qui relève de la prise en compte des propriétés FATES et du besoin profond et impérieux de cette prise en compte, que ce soit dans l'industrie, mais aussi dans le monde académique, nous souhaitons que ce projet soit un démonstrateur et une illustration concrète que non seulement c'est possible, mais extrêmement bénéfique. Nous anticipons des impacts à court, moyen et long terme, chaque étape étant associée à des livrables spécifiques et présentant des risques et des opportunités croissants. À court terme, nous apportons (i) une compréhension améliorée des risques et des solutions FATES (par exemple l'amélioration des futurs générateurs de code comme Starcoder contre les biais), et (ii) une démocratisation du FATES-MLOps (accessibilité des codes et artefacts en open-source). À moyen terme, la réalisation concrète d'un chatbot "FATES" en

Appel à projets Thématiques Spécifiques en Intelligence Artificielle (TSIA) – Edition 2024

Wolof constituera une vitrine pour une diffusion plus large des principes FATES. Enfin à long terme, via la formalisation des propriétés FATES, leur alignement sur les normes existantes, la prise en compte de leur évolution et leur opérationnalisation pour guider leur intégration dans les processus MLOps, nous fournirons des outils précieux pour les Data Scientists / ML Engineers de demain.

5. Tableau Financier : Résumé des coûts et des efforts

Partenaire / Budget Total	Nom/Fonction	Budget demandé	Rôle dans le projet *	Effort
IRIT 274 800 €	Jean-Michel Bruel / Pr	Missions : 20 k€ Equip.: 3 k€	Coordinateur Scientifique, Contributeur WP1, WP2, et WP3	16.8 p.mois
	Olivier Teste / Pr		Contributeur WP1 et WP3	5 p.mois
	Marc Pantel / MCF		Resp. Scientifique IRIT, Resp. WP1, Contributeur WP2 et WP3	15 p.mois
	Doc (à recruter)	130 k€	T1.1, T1.2, T1.3, T2.1, T0.3	36 p.mois
	Post-doc (à recruter)	112 k€	T1.2, T1.3, T3.1, T2.2, T0.3	18 p.mois
I3S 186 099 €	Mireille Blay-Fornarino / Pr	Missions : 15 k€ Stages : 5.5 k€ Equip.: 3 k€	Resp. Scientifique I3S, Resp. WP2, Contributrice WP1 et WP3	14.4 p.mois
	Philippe Collet / Pr		Contributeur WP2 et WP3	7.2 p.mois
	Doc (à recruter)	138 k€	WP2, T1.2, T1.3, T3.1, T0.3	36 p.mois
INRIA Sophia 142 896 €	Frédéric Precioso / Pr	Missions : 10 k€ Stages : 11 k€ Equip.: 4 k€	Resp. Scientifique INRIA, Resp. WP3, Contributeur WP1 et WP2	15 p.mois
	Michel Riveill / Pr		Contributeur WP2 et WP3	15 p.mois
	Post-doc (à recruter)	100 k€	WP3, T0.3	24 p.mois
McSCert 0 € (eq. 300 k€)	Sébastien Mosser /As. Pr	Missions : 0 (eq. 10 k€) Equip.: 0 (eq. 8 k€)	Contributeur WP1, WP3	17 p.mois
	Richard Paige / Pr		Contributeur WP1, WP2	7 p.mois
	Post-doc (à recruter)	0 (eq. 102 k€)	T0.2, T1.2, T1.3, T2.1, T0.3	24 p.mois
	2 Doc (à recruter)	0 (eq. 170 k€)	T1.2, T1.3, T2.1, T3.2, T0.3	96 p.mois
Total		603 795 €		346.4 p.mois

*Tous les partenaires participent au WP0, notamment la tâche T0.3 de dissémination et exploitation.

6. Bibliographie

[Amraoui2022] Amraoui, Y. el, **Blay-Fornarino, M., Collet, P., Precioso, F., & Muller, J.** (2022). Evolvable SPL management with partial knowledge: an application to anomaly detection in time series. In ACM SPLC, Vol. A, 222–233.

[Benni2019] Benni, B., **Blay Fornarino, M., Mosser, S., Precioso, F., & Jungbluth, G.** (2019). When DevOps meets meta-learning: A portfolio to rule them all. In ACM/IEEE MODELS, 605–612. <https://doi.org/10.1109/MODELS-C.2019.00092>

[Brault2023] Brault, Y., El Amraoui, Y., **Blay-Fornarino, M., Collet, P., Jaillet, F., & Precioso, F.** (2023, August). Taming the Diversity of Computational Notebooks. In ACM SPLC, Vol. A, pp. 27-33.

[Breck2017] Breck, E., et al. (2017). The ML test score: A rubric for ML production readiness and technical debt reduction. IEEE International Conference on Big Data (Big Data), 1123–1132.

Appel à projets Thématiques Spécifiques en Intelligence Artificielle (TSIA) – Edition 2024

- [Brown2020] Brown, et al. (2020). Language models are few-shot learners. *NeurIPS*, 33, 1877-1901.
- [Chen2020] Chen, A., et al. 2020. Developments in MLflow: A System to Accelerate the Machine Learning Lifecycle. In *International Workshop on Data Management for End-to-End Machine Learning (DEEM'20)*, ACM.
- [CIKM22] Cugny, R., Aligon, J., Chevalier, M., Roman-Jimenez, G., **Teste, O.**: AutoXAI: A Framework to Automatically Select the Most Adapted XAI Solution. *CIKM 2022*: 315-324
- [CLEF20] Maupomé, D., Armstrong, M.D., Belbahar, R.M., Alezot, J., Balassiano, R., Queudot, M., **Mosser, S.**, Meurs, M.-J. 2020. Early Mental Health Risk Assessment through Writing Styles, Topics and Neural Models. *CLEF (Working Notes) 2020*.
- [CLEF22] Saravani, S.H.H., Normand, L., Maupomé, D., Rancourt, F., Soulas, T., Besharati, S., Normand, A., **Mosser, S.**, Meurs, M.-J. 2022. Measuring the Severity of the Signs of Eating Disorders Using Similarity-Based Models. *CLEF (Working Notes)*.
- [CLEF23] Maupomé, D., Soulas, T., Rancourt, F., Cantin-Savoie, G., Winterstein, G., **Mosser, S.**, Meurs, M.-J. 2023. Lightweight Methods for Early Risk Detection. *CLEF (Working Notes) 2023*: 718-726.
- [Contractor2022] D. Contractor, et al. (2022). Behavioral Use Licensing for Responsible AI. In *FACCT '22*.
- [DevOps2021] *The DevOps Handbook: How to Streamline IT Delivery and Management* par Gene Kim, Jeannie Kim, Kevin Behr et George Spafford. 2021.
- [Duffau2018] Duffau, C., Polacsek, T., & **Blay-Fornarino, M.** (2018). Support of justification elicitation: Two industrial reports. In *CAISE 2018*, pp. 71-86.
- [FAPFID23] Dorleon, G., Megdiche, I., Bricon-Souf, N., **Teste, O.**: FAPFID: A Fairness-Aware Approach for Protected Features and Imbalanced Data. *Trans. Large Scale Data Knowl. Centered Syst.* 53: 107-125 (2023)
- [Feldman2015] Feldman, M., et al. (2015). Certifying and Removing Disparate Impact. In *ACM SIGKDD KDD '15*.
- [Ferrara2023] Ferrara, E. (2023). Should ChatGPT be Biased? Challenges and Risks of Bias in Large Language Models. *ArXiv Preprint ArXiv:2304.03738*, 28(11). <https://doi.org/10.5210/fm.v28i11.13346>
- [Gebru2021] Gebru, T., et al. 2021. Datasheets for datasets. In *Comm. of the ACM* 64, 12 (December 2021), 86–92.
- [Klaise2020] Klaise, J., et al. (2020). Monitoring and explainability of models in production. *arXiv preprint arXiv:2007.06299*.
- [Liu2024] Liu, Y., et al. (2024). Uncovering and quantifying social biases in code generation. *NeurIPS*, 36.
- [Lopardo2023] Lopardo, G., **Precioso, F.**, & Garreau, D. (2023, April). A Sea of Words: An In-Depth Analysis of Anchors for Text Data. In *IASTATS*, pp. 4848-4879.
- [Lopardo2024] Lopardo, G., **Precioso, F.**, & Garreau, D. (2024). Attention Meets Post-hoc Interpretability: A Mathematical Perspective. *arXiv preprint arXiv:2402.03485*.
- [Lozhkov2024] Lozhkov, A., et al. (2024). StarCoder 2 and The Stack v2: The Next Generation. <http://arxiv.org/abs/2402.19173>
- [Mayaki2022] Mayaki, M. Z. A., & **Riveill, M.** (2022). Autoregressive based drift detection method. In *IEEE IJCNN*, pp. 1-8.
- [Mill2024] Mill, E., et al. (2024). The SAGE Framework for Explaining Context in XAI. *Applied Artificial Intelligence*, 38(1).
- [Mosser2023] **Mosser, S.**, Pulgar, C., **Blay-Fornarino, M.**, Patel, D., Loh, A., & **Bruel, J.-M.** (2023, October). Yes, Configuring is Good, But Have You Ever Tried Justifying? *CONFLANG Workshop (co-located with SPLASH)*.
- [Polacsek2018] Polacsek, T., et al. (2018). The need of diagrams based on toulmin schema application: an aeronautical case study. *EURO Journal on Decision Processes*, 6(3-4), 257-282.
- [RQCODE22] Nigmatullin, I., Sadovykh, A., Messe, A., Ebersold, S., **Bruel, J.-M.** (2022). RQCODE - Towards Object-Oriented Requirements in the Software Security Domain. *ICST Workshops 2022*: 2-6.
- [Sculley2015] Sculley, D., et al. (2015). Hidden Technical Debt in Machine Learning Systems. *NeurIPS*, 28.
- [Suresh2021] Suresh, H., & Gutttag, J. (2021). A framework for understanding sources of harm throughout the machine learning life cycle. In *1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (pp. 1-9).
- [Testi2022] Testi, M., Ballabio, M., Frontoni, E., Iannello, G., Moccia, S., Soda, P., & Vessio, G. (2022). MLOps: A taxonomy and a methodology. *IEEE Access*, 10, 63606–63618.
- [Topsakal2023] Topsakal, O., & Akinci, T. C. (2023). Creating large language model applications utilizing langchain: A primer on developing LLM apps fast. In *Int. Conf. on Applied Engineering and Natural Sciences (Vol. 1, No. 1, pp. 1050-1056)*.
- [Vasudevan2020] Vasudevan, S., et al. (2020). Lift: A scalable framework for measuring fairness in ML applications. In *ACM CIKM*, pp. 2773-2780.
- [VeriDevOps23] Enoiu, E.P., et al. (2023) VeriDevOps Software Methodology: Security Verification and Validation for DevOps Practices. *ARES 2023*: 135:1-135:9.
- [Wachter2021] Wachter, S., et al. (2021). Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. *Computer Law & Security Review, Elsevier*, 41, 105567.
- [Wang2024] Z. Wang et al. (2024). XAIport: A Service Framework for the Early Adoption of XAI in AI Model Development. In *ICSE 2024, Track New Ideas and Emerging Results*.
- [Zaharia2018] Zaharia, M., et al. (2018). Accelerating the machine learning lifecycle with MLflow. *IEEE Data Eng. Bull.*, 41(4).