Prise en compte des propriétés FATES en MLOps: perspectives et ambitions

Mireille Blay-Fornarino*, Jean-Michel Bruel**
Sébastien Mosser****
Frédéric Precioso***,*

*Université de Toulouse / CNRS-IRIT
**Laboratoire CNRS I3S
***INRIA Nice
****McMaster University
contact@fates-mlops.org

Résumé. 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). Le projet ANR FATES-MLOps 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.

1 Motivations

Le mouvement MLOps [1] reprend les objectifs du mouvement DevOps (Kim et al., 2016), 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* (DS) et les phases liées à la construction des modèles de *Machine Learning* (ML) (Testi et al., 2022). 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

1. https://ml-ops.org/

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 [2], 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 [3], puis plus récemment par la sécurité et la sûreté (*Security*/*Safety*) pour donner les propriétés FATES, et le mouvement *Data for Good* [4] 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 régulations 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 états 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 d'IA au fil du temps" (Tabassi, 2023; Garrido et al., 2023). 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 FATES-MLOps [5] 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 principal 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 ?**

Dans la section 2 nous dressons l'état de l'art couvrant le domaine du projet, à savoir les propriétés FATES, les outils existants, les aspects variabilité et justification. Dans la section 3 nous donnons les grandes lignes de nos contributions futures en la matière. Enfin, dans la section 4 nous dressons les actions à plus ou moins long termes qui nous permettrons d'atteindre nos objectifs.

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.

2. https://www.fatml.org
3. https://www.microsoft.com/en-us/research/theme/fate
4. https://datascience.columbia.edu/news/2018/data-for-good-fates-elaborated
5. Projet ANR-24-IAS2-0002-01.

2.1 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 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 (Dorleon et al., 2023). 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 (Suresh et Guttag, 2021). Des algorithmes sont proposés pour atténuer les biais, notamment le débiaisement des données lors de la collecte et l'analyse des modèles de ML (Feldman et al., 2015). Pour détecter les biais, de nombreuses métriques sont proposées, récemment Wachter et al. (2021) proposent la disparité démographique conditionnelle (CDD) comme référence statistique pour évaluer la discrimination potentielle dans les systèmes automatisés. Breck et al. (2017) identifient différentes formes de tests pour détecter ces dérives. Plusieurs travaux sur les *Large Language Models* (LLMs) mettent en avant une combinaison de ces approches (Brown et al., 2020; Ferrara, 2023).

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 différents [6]. 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 reproductible [7]. 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 MLFlows [8] et les approches par conteneurs sont utilisés dans le contexte MLOps pour améliorer cette traçabilité qui reste cependant à renforcer (Chen et al., 2020). 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 (Liu et al., 2023).

6. Air Canada, cf. `https://intelligence-artificielle.com/chatbot-air-canada-hallucine/`.
7. La CNIL en donne une définition différente, dans ce projet, nous nous limitons à la définition donnée ici : `https://www.cnil.fr/fr/developpement-des-systemes-dia-les-recommandations-de-la-cnil-pour-respecter-le-rgpd`.
8. `https://mlflow.org/`

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 (Cugny et al., 2022). 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 (Lopardo et al., 2023, 2024). 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 (Klaise et al., 2020). Des défis persistent dans la surveillance et l'explication des modèles déployés, des solutions sont déjà disponibles pour en relever certains (Wang et al., 2024), 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 (Mill et al., 2024).

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 inscrirons dans la logique des *Responsible AI Licences* (RAIL) Contractor et al. (2022).

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 (Enoiu et al., 2023; Nigmatullin et al., 2022). 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.

Contextualisation

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 d'exemples adversaires sont proposés chaque jour. Cette évolution est si forte que le document de référence produit par le NIST[9] en matière des risques liés à l'IA, est conçu comme un document vivant (Tabassi, 2023). Dans Sculley et al. (2015) et Breck et al. (2017), 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.

2.2 Les outils en support au FATES MLOps

Il existe 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, dans notre projet, la question davantage d'un point de vue Génie Logiciel et intégrateur.

Le test et la surveillance

Dans Breck et al. (2017), 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 (Zaharia et al., 2018), SageMaker[10] et Kubeflow[11]. 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

9. *National Institute of Standards and Technology, U.S. Department of Commerce.*
10. `https://aws.amazon.com/sagemaker/`
11. `https://www.kubeflow.org/`

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.

2.3 Des exigences aux justifications

Construire des systèmes efficaces de science des données est d'autant plus difficile que les solutions ML disponibles ne cesse de croître (Zaharia et al., 2018). 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 (Amraoui et al., 2022), que nous exploitons pour identifier des solutions à réutiliser (Brault et al., 2023) et avons modélisé un méta-modèle de pipeline de ML pour les prendre en charge dans un contexte DevOps (Benni et al., 2019). Sur la base de ces travaux préliminaires, nous visons à étendre notre approche pour intégrer spécifiquement les propriétés FATES dans les workflows de ML (Vasudevan et Kenthapadi, 2020), en capturant la variabilité des algorithmes avec la logique des *feature models*.

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 et le respect des normes. Justifier qu'un système logiciel respecte les propriétés FATES [12], rejoint le même objectif. Polacsek et al. (2018) 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 Duffau et al. (2018), 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 (Mosser et al., 2023). Nous proposons de poursuivre ces travaux dans le cadre du MLOps.

3 Contribution

3.1 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 (cf. Figure 1). 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.

12. Cf. le début de formalisation Microsoft Responsible AI Standard, v2 GENERAL REQUIREMENTS disponible ici : `https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE5cmFl?culture=fr-fr&country=fr`.

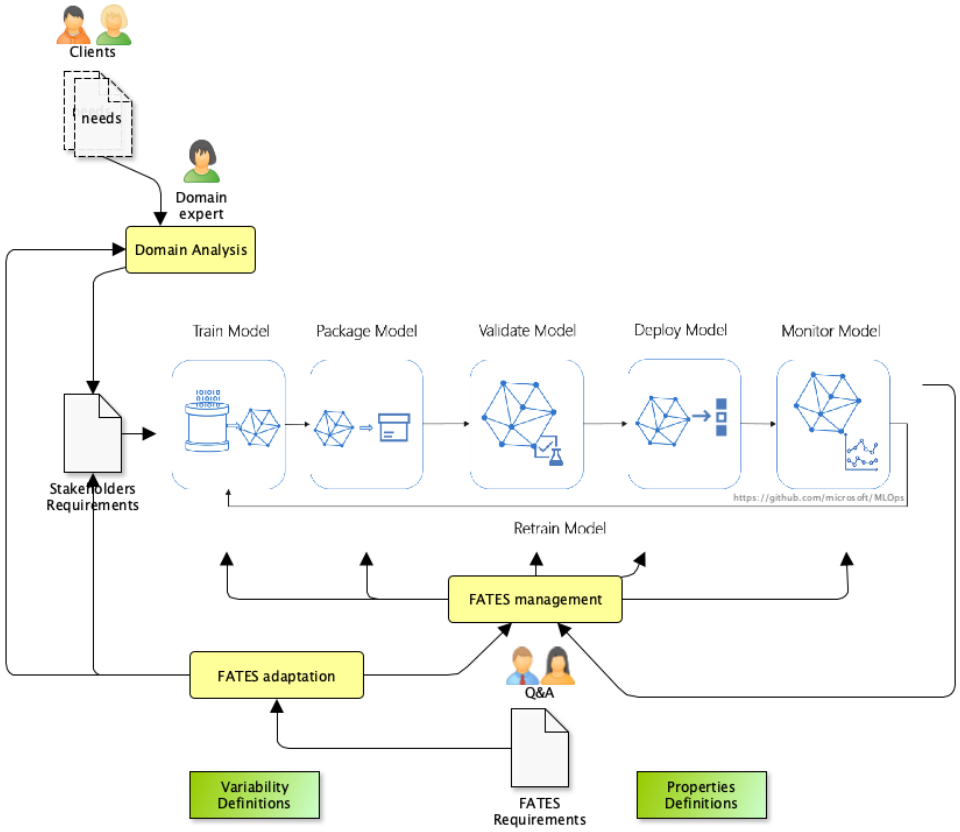

FIG. 1 – *Vue d'ensemble du projet FATES-MLOps*

3.2 Sorties attendues et mesure objective de qualité

Les sorties attendues du projet sont directement liées aux cas d'usage envisagés, 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.

3.3 Démarche scientifique

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 (Benni et al., 2019; Brault et al., 2023). Pour atténuer les risques inhérents à un domaine aussi dynamique et aux multiples applications, nous adopterons une approche itérative, cohérente avec MLOps, en 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 pourrons pas couvrir tout l'espace des propriétés FATES, qui est très vaste.

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.

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 sur-échantillonner 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 (*Components*

Off-The-Shelf) 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.

4 Conclusion

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, 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, 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 préexistante au sein du consortium, sur la publication et maintenance de logiciels "open-source" et de jeux de données "open-data". La compagnie Hugging Face, qui soutient le projet, 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 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.

Cet publication est supportée par le projet ANR-24-IAS2-0002. Les auteurs remercient les autres membres du projet pour leur participation.

Références

Amraoui, Y. E., M. Blay-Fornarino, P. Collet, F. Precioso, et J. Muller (2022). Evolvable spl management with partial knowledge : an application to anomaly detection in time series. In *Proceedings of the 26th ACM International Systems and Software Product Line Conference - Volume A*, SPLC '22, New York, NY, USA, pp. 222–233. Association for Computing Machinery.

Benni, B., M. Blay-Fornarino, S. Mosser, F. Precioso, et G. Jungbluth (2019). When DevOps Meets Meta-Learning : A Portfolio to Rule them all. In *2019 ACM/IEEE 22nd International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C)*, Munich, Germany, pp. 605–612. IEEE.

Brault, Y., Y. El Amraoui, M. Blay-Fornarino, P. Collet, F. Jaillet, et F. Precioso (2023). Taming the Diversity of Computational Notebooks. In *SPLC 2023 - 27th ACM International Systems and Software Product Line Conference*, SPLC '23 : Proceedings of the 27th ACM International Systems and Software Product Line Conference - Volume A, Tokyo, Japan, pp. 27–33. ACM.

Breck, E., S. Cai, E. Nielsen, M. Salib, et D. Sculley (2017). The ml test score : A rubric for ml production readiness and technical debt reduction. In *Proceedings of IEEE Big Data*.

Brown, T. B., B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, et D. Amodei (2020). Language models are few-shot learners.

Chen, A., A. Chow, A. Davidson, A. DCunha, A. Ghodsi, S. A. Hong, A. Konwinski, C. Mewald, S. Murching, T. Nykodym, P. Ogilvie, M. Parkhe, A. Singh, F. Xie, M. Zaharia, R. Zang, J. Zheng, et C. Zumar (2020). Developments in mlflow : A system to accelerate the machine learning lifecycle. In *Proceedings of the Fourth International Workshop on Data Management for End-to-End Machine Learning*, DEEM '20, New York, NY, USA. Association for Computing Machinery.

Contractor, D., D. McDuff, J. K. Haines, J. Lee, C. Hines, B. Hecht, N. Vincent, et H. Li (2022). Behavioral use licensing for responsible ai. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, New York, NY, USA, pp. 778–788. Association for Computing Machinery.

Cugny, R., J. Aligon, M. Chevalier, G. Roman Jimenez, et O. Teste (2022). Autoxai : A framework to automatically select the most adapted xai solution. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, CIKM '22, New York, NY, USA, pp. 315–324. Association for Computing Machinery.

Dorleon, G., I. Megdiche, N. Bricon-Souf, et O. Teste (2023). FAPFID : A Fairness-Aware Approach for Protected Features and Imbalanced Data. *Transactions on Large-Scale Data- and Knowledge-Centered Systems 13840 (TLDKS)*, 107–125. Transactions on Large-Scale Data- and Knowledge-Centered Systems (TLDKS).

Duffau, C., T. Polacsek, et M. Blay-Fornarino (2018). Support of justification elicitation : Two industrial reports. In *Advanced Information Systems Engineering : 30th International Conference, CAiSE 2018, Tallinn, Estonia, June 11-15, 2018, Proceedings*, Berlin, Heidelberg, pp. 71–86. Springer-Verlag.

Enoiu, E., D. Truscan, A. Sadovykh, et W. Mallouli (2023). Veridevops software methodology : Security verification and validation for devops practices. In *ARES '23 : Proceedings of the 18th International Conference on Availability, Reliability and Security*, pp. 1–9.

Feldman, M., S. A. Friedler, J. Moeller, C. Scheidegger, et S. Venkatasubramanian (2015). Certifying and removing disparate impact. In *proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 259–268.

Ferrara, E. (2023). Should chatgpt be biased ? challenges and risks of bias in large language models. *First Monday 28*(11).

Garrido, J. S., S. Tolan, I. H. Torres, D. F. Llorca, V. Charisi, E. G. Gutierrez, H. Junklewitz, R. Hamon, D. F. Yela, et C. Panigutti (2023). AI Watch : Artificial Intelligence Standardisation Landscape Update. (KJ-NA-31-343-EN-N (online)).

Kim, G., P. Debois, J. Willis, et J. Humble (2016). *The DevOps Handbook : How to Create World-Class Agility, Reliability, and Security in Technology Organizations*. IT Revolution Press.

Klaise, J., A. V. Looveren, C. Cox, G. Vacanti, et A. Coca (2020). Monitoring and explainability of models in production.

Liu, Y., X. Chen, Y. Gao, Z. Su, F. Zhang, D. Zan, J.-G. Lou, P.-Y. Chen, et T.-Y. Ho (2023). Uncovering and quantifying social biases in code generation. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, et S. Levine (Eds.), *Advances in Neural Information Processing Systems*, Volume 36, pp. 2368–2380. Curran Associates, Inc.

Lopardo, G., F. Precioso, et D. Garreau (2023). A sea of words : An in-depth analysis of anchors for text data.

Lopardo, G., F. Precioso, et D. Garreau (2024). Attention meets post-hoc interpretability : A mathematical perspective.

Mill, E., W. Garn, N. Ryman-Tubb, et C. Turner (2024). The sage framework for explaining context in explainable artificial intelligence. *Applied Artificial Intelligence 38*, e2318670.

Mosser, S., C. Pulgar, M. Blay-Fornarino, D. Patel, A. Loh, et J.-M. Bruel (2023). Yes, Configuring is Good, But Have You Ever Tried Justifying ? In *CONFLANG Workshop (co-located with SPLASH)*.

Nigmatullin, I., A. Sadovykh, N. Messe, S. Ebersold, et J.-M. Bruel (2022). Rqcode – towards object-oriented requirements in the software security domain. In *IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW)*, pp. 2–6.

Polacsek, T., S. Sharma, C. Cuiller, et V. Tuloup (2018). The need of diagrams based on toulmin schema application : an aeronautical case study. *EURO Journal on Decision Pro-*

cesses 6, 1–26.

Sculley, D., G. Holt, D. Golovin, E. Davydov, T. Phillips, D. Ebner, V. Chaudhary, M. Young, et D. Dennison (2015). Hidden technical debt in machine learning systems. *NIPS*, 2494–2502.

Suresh, H. et J. Guttag (2021). A framework for understanding sources of harm throughout the machine learning life cycle. In *Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, EAAMO '21, New York, NY, USA. Association for Computing Machinery.
en

Tabassi, E. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0).

Testi, M., M. Ballabio, E. Frontoni, G. Iannello, S. Moccia, P. Soda, et G. Vessio (2022). MLOps : A Taxonomy and a Methodology. *IEEE Access 10*, 63606–63618.

Vasudevan, S. et K. Kenthapadi (2020). Lift : A scalable framework for measuring fairness in ml applications. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, CIKM '20, New York, NY, USA, pp. 2773–2780. Association for Computing Machinery.

Wachter, S., B. Mittelstadt, et C. Russell (2021). Why fairness cannot be automated : Bridging the gap between eu non-discrimination law and ai. *Computer Law & Security Review 41*, 105567.

Wang, Z., Y. Liu, A. Arumugam Thiruselvi, et A. Hamou-Lhadj (2024). Xaiport : A service framework for the early adoption of xai in ai model development. In *Proceedings of the 2024 ACM/IEEE 44th International Conference on Software Engineering : New Ideas and Emerging Results*, ICSE-NIER'24, New York, NY, USA, pp. 67–71. Association for Computing Machinery.

Zaharia, M. A., A. Chen, A. Davidson, A. Ghodsi, S. A. Hong, A. Konwinski, S. Murching, T. Nykodym, P. Ogilvie, M. Parkhe, F. Xie, et C. Zumar (2018). Accelerating the machine learning lifecycle with mlflow. *IEEE Data Eng. Bull. 41*, 39–45.

Summary

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 the FATES-MLOPs ANR 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.