

Automatisation agentique des processus logiciels agiles — orchestration, prédiction et fiabilité d'agents IA sur les référentiels JIRA/SCRUM

Candidat : Anas Lahlou — Master de recherche Data Science & Big Data (2022), Ingénieur d'État en systèmes d'information · salarié à temps plein (statut doctorant salarié, autofinancé)

Structure visée : CEDoc ST2I — ENSIAS, Université Mohammed V de Rabat (équipe Alqualsadi / BPM)

1. Contexte et problématique

Les méthodes agiles (SCRUM en tête) sont devenues le mode d'organisation dominant du développement logiciel, et les outils de gestion de tickets comme JIRA en sont le système de référence : chaque ticket y laisse une trace d'exécution complète (création, estimations, transitions d'états, affectations, commentaires). Ces référentiels concentrent pourtant une charge de coordination manuelle considérable — triage et affectation des tickets, estimation de l'effort, priorisation du backlog, détection des dérives de sprint — répétitive, coûteuse en temps d'ingénierie et sujette aux biais.

L'émergence des agents fondés sur les grands modèles de langue (LLM) ouvre la voie à l'automatisation de ces tâches de connaissance : c'est l'« automatisation agentique des processus » (*Agentic Process Automation*), qui dépasse le RPA classique en confiant à l'agent la construction et l'exécution dynamique du workflow. Mais deux verrous scientifiques bloquent son adoption en entreprise :

1. Le verrou de la fiabilité. Les systèmes d'agents échouent de manière systémique et mal comprise — la taxonomie MAST recense 14 modes d'échec sur plus de 1 600 traces d'exécution multi-agents annotées, et Gartner prédit l'abandon de plus de 40 % des projets d'IA agentique d'ici 2027. Il n'existe ni cadre d'évaluation ni mécanisme de confinement des agents opérant sur des processus métier réels.

2. Le verrou du contexte processus. Les travaux existants d'IA pour le génie logiciel traitent chaque tâche isolément (estimer un ticket, router un ticket) sans exploiter la richesse du *processus* : historique de flux du ticket, comportement de l'équipe, conformité effective au rituel SCRUM — informations que le process mining sait précisément extraire des event logs.

Question centrale : comment concevoir, évaluer et fiabiliser des agents IA qui automatisent la gestion des processus logiciels agiles, en exploitant les traces d'exécution (event logs JIRA) à la fois comme **contexte** de décision et comme **garde-fou** ?

- **RQ1 (évaluation)** — Comment mesurer rigoureusement la performance et la fiabilité d'agents LLM sur des tâches de gestion de processus agiles réels ?
- **RQ2 (augmentation par le processus)** — L'enrichissement des agents par des artefacts de process mining améliore-t-il mesurablement leurs décisions par rapport aux approches ML dédiées et aux agents « naïfs » ?
- **RQ3 (fiabilité)** — Peut-on prédire et détecter les échecs d'agents en cours d'exécution, et les confiner par conformance checking et supervision humaine ciblée ?

2. État de l'art et positionnement

IA pour le génie logiciel (AI4SE / Mining Software Repositories). L'estimation d'effort et le triage de tickets sont des problèmes établis : Deep-SE (Choetkiertikul et al., IEEE TSE 2019) puis GPT2SP (Fu & Tantithamthavorn, IEEE TSE 2023) estiment les story points par apprentissage profond ; le dataset public TAWOS (Tawosi et al., MSR 2022) — ~500 000 tickets JIRA issus de 44 projets open source — sert de référence reproductible. Limite : ces approches sont mono-tâche, sans notion de processus ni d'action autonome.

Process mining prédictif. Le monitoring prédictif de processus (Teinemaa et al., ACM TKDD 2019 ; van der Aalst, 2016) prédit issue, délai et prochaine activité à partir d'event logs, avec un outillage mature (pm4py). Son application aux référentiels de tickets logiciels reste marginale, et son couplage avec des agents LLM est inexploré.

Agentic BPM. Les LLM appliqués au BPM sont une piste ouverte depuis Vidgof et al. (BPM Forum 2023) ; ProAgent (2023) a posé le concept d'*Agentic Process Automation*. Le domaine explose en 2025-2026 mais reste dominé par des démonstrations sur processus jouets — pas de validation sur des processus d'ingénierie logicielle réels avec données longitudinales.

Fiabilité des agents. MAST (Cemri et al., 2025) fournit la première taxonomie empirique des échecs multi-agents ; l'injection de prompt demeure le risque n° 1 de l'OWASP Top 10 LLM, sans solution générale. L'évaluation d'agents sur des processus métier — où une erreur a un coût organisationnel — n'a pas de benchmark de référence.

Positionnement. L'intersection des trois champs — *des agents LLM opérant sur des processus agiles réels, augmentés par le process mining et évalués/confinés systématiquement* — est inoccupée. C'est le créneau de cette thèse, avec un différenciateur rare en académie : des données industrielles réelles et longitudinales, validées en miroir sur datasets publics.

3. Contributions visées

C1 — Un cadre d'évaluation et un benchmark d'agents sur processus agiles. Suite de tâches de gestion de backlog (triage, estimation, priorisation, détection de dérive de sprint) construite sur TAWOS et sur des données industrielles anonymisées (organismes partenaires utilisant SCRUM) ; métriques à deux niveaux (qualité de la décision locale ; impact sur le processus global) ; protocole de contamination contrôlée.

Type : dataset & benchmark paper.

C2 — Des agents « conscients du processus » (process-aware agents). Architecture d'agents LLM dont le contexte est enrichi par le process mining (variantes de flux, prédictions, conformance) ; comparaison systématique aux baselines dédiées (Deep-SE, GPT2SP, modèles prédictifs classiques) pour quantifier l'apport du contexte processus.

Type : papier méthode + étude empirique.

C3 — Un cadre de fiabilité pour agents en workflow réel. Taxonomie d'échecs adaptée aux processus métier (extension de MAST), détection prédictive d'échecs en cours d'exécution, garde-fous par conformance checking et routage human-in-the-loop des cas incertains.

Type : papier fiabilité/évaluation, validation par étude de cas industrielle.

4. Méthodologie et faisabilité

Démarche : génie logiciel empirique et design science — cycles courts construire → mesurer → publier, chaque contribution étant validée à la fois sur données publiques (reproductibilité) et sur le terrain industriel (validité externe).

Données. (i) Référentiels JIRA d'organismes utilisant la méthodologie agile SCRUM (entreprises et administrations, à conventionner) : historiques réels multi-équipes, mis à disposition sous convention, anonymisation systématique et conformité à la loi 09-08 sur la protection des données personnelles ; (ii) datasets publics TAWOS et projets Apache pour la reproductibilité et la comparaison à l'état de l'art.

Moyens de calcul — sobriété assumée. Le cœur du travail (agents via API, process mining, évaluation) ne requiert aucun cluster GPU. Les expériences sur petits modèles open source tiennent sur un GPU unique ; l'accès gratuit à HPC-MARWAN (CNRST) couvre les besoins ponctuels de fine-tuning. Coût marginal : crédits API (autofinancés).

Compatibilité avec le statut de salarié. Disponibilité régulière et chiffrée (soirées et week-ends, ~15-20 h/semaine), autofinancement, et un atout structurel : une pratique professionnelle quotidienne des processus agiles (JIRA/SCRUM), qui ancre la thèse dans les problèmes réels du terrain et facilite l'identification de partenaires industriels.

Préalable en cours : identification et conventionnement d'un ou plusieurs organismes utilisant JIRA/SCRUM pour la mise à disposition de données anonymisées à des fins de recherche.

5. Calendrier prévisionnel (5 ans, régime temps aménagé)

ANNÉE	OBJECTIFS SCIENTIFIQUES	JALONS & PUBLICATIONS VISÉES
1 2026-27	Formations doctorales ; revue systématique de littérature ; conventions d'accès aux données + pipeline d'anonymisation ; constitution du benchmark (C1, v0)	Participation au MSR Mining Challenge ; communication nationale
2 2027-28	C1 complète : benchmark + évaluation de la génération d'agents actuelle	Papier conférence/atelier international (MSR, ESEM ou ICPM)
3 2028-29	C2 : agents process-aware, comparaison aux baselines	Soumission revue 1 (<i>Empirical Software Engineering</i> ou <i>Information & Software Technology</i>)

4 2029-30	C3 : taxonomie d'échecs, détection prédictive, garde-fous ; étude de cas industrielle	Soumission revue 2 ; communication internationale
5 2030-31	Consolidation, révisions des revues, rédaction du manuscrit	Soutenance — cible : 2 revues indexées (Scopus/WoS) en 1 ^{er} auteur + 2 communications

Le calendrier intègre les délais réels de review (6-18 mois par revue) en soumettant tôt (années 3 et 4).

6. Adéquation au laboratoire et au contexte national

Le sujet s'inscrit au cœur des thématiques de l'équipe Alqualsadi (ENSIAS, Pr. Karim Baïna) — Business Process Management et ingénierie des SI — en y apportant l'angle agents LLM et un terrain industriel. Il rejoint les priorités « transformation digitale » des programmes PNARDI 2025-2028 et de la feuille de route « Maroc IA 2030 », ce qui ouvre un rattachement possible à un projet financé du laboratoire.

Références principales

1. Choetkiertikul, M. et al. — *A Deep Learning Model for Estimating Story Points*, IEEE Trans. on Software Engineering, 2019 (Deep-SE).
2. Tawosi, V., Al-Subaihin, A., Moussa, R., Sarro, F. — *A Versatile Dataset of Agile Open Source Software Projects*, MSR 2022 (TAWOS).
3. Fu, M., Tantithamthavorn, C. — *GPT2SP: A Transformer-Based Agile Story Point Estimation Approach*, IEEE TSE, 2023.
4. Teinemaa, I., Dumas, M., La Rosa, M., Maggi, F.M. — *Outcome-Oriented Predictive Process Monitoring: Review and Benchmark*, ACM TKDD, 2019.
5. van der Aalst, W. — *Process Mining: Data Science in Action*, Springer, 2016.
6. Vidgof, M., Bachhofner, S., Mendling, J. — *Large Language Models for Business Process Management: Opportunities and Challenges*, BPM Forum, 2023.
7. Ye, Y. et al. — *ProAgent: From Robotic Process Automation to Agentic Process Automation*, arXiv, 2023.
8. Cemri, M. et al. — *Why Do Multi-Agent LLM Systems Fail?* (taxonomie MAST), 2025.
9. OWASP — *Top 10 for Large Language Model Applications*, 2025.

Mini-projet v1 — 14 juillet 2026 · Document de travail destiné à la prise de contact avec les encadrants (CEDoc ST2I, rentrée 2026-2027)