

LLMS NA GESTÃO DE EQUIPES: UMA INVESTIGAÇÃO EMPÍRICA SOBRE COMUNICAÇÃO E PREVISIBILIDADE

Mestrado

data início: agosto, 2025

data de término (planejada): junho, 2027

Eduardo André de Lima Carneiro(mestrando), Ivaldir Honório de Farias
Júnior(orientador)

PPGEC – Escola Politécnica de Pernambuco – Universidade de Pernambuco (UPE)
R. Benfica, 455 - Madalena, Recife - PE, 50720-001, Brasil

ealc@ecom.poli.br, ivaldir.farias@upe.br

***Resumo.** Este artigo investiga como os modelos de linguagem (LLMs) podem apoiar a comunicação e a previsibilidade em equipes. Apresenta-se o delineamento de um survey exploratório destinado a mapear dores, práticas e percepções de profissionais sobre o uso de IA em atividades colaborativas e rituais do time. A motivação decorre de lacunas na literatura referentes ao impacto dos LLMs em processos coletivos, especialmente aqueles que dependem de clareza de requisitos, registro consistente de decisões e estabilidade do fluxo de trabalho. O artigo também descreve um protocolo experimental complementar, planejado para etapas futuras da pesquisa. Espera-se que os resultados contribuam para orientar organizações na adoção responsável de LLMs e para fortalecer a base empírica sobre colaboração humano-IA em Engenharia de Software.*

Introdução

A adoção de métodos ágeis transformou o trabalho em equipes de software, embora os desafios persistam, especialmente nas dimensões de **comunicação** e **previsibilidade** do fluxo de trabalho. Ruídos de entendimento, desalinhamentos entre áreas e divergências na interpretação de requisitos continuam associados a retrabalho, atrasos e perda de produtividade, como destacado por Stol e Fitzgerald (2018). A previsibilidade também permanece limitada devido à variabilidade das tarefas, dependências entre equipes e ausência de mecanismos eficazes para antecipação de riscos, conforme observado por Dingsoyr, Moe e Ralph (2022). Este estudo se destina a profissionais que atuam com métodos ágeis e a pesquisadores interessados em compreender a influência da inteligência artificial sobre práticas colaborativas e sobre a gestão do fluxo de trabalho.

Simultaneamente, modelos de linguagem de grande escala (LLMs), como ChatGPT, Gemini e Claude, vêm sendo incorporados às rotinas de desenvolvimento, apoiando sínteses de reuniões, elaboração de histórias de usuário e identificação preliminar de impedimentos. Evidências iniciais sugerem que tais ferramentas podem melhorar a clareza de artefatos textuais e acelerar tarefas cognitivas (Peng et al., 2023), embora ainda exista uma lacuna empírica sobre como esses modelos influenciam comunicação e previsibilidade em ambientes ágeis. Este estudo busca preencher essa lacuna investigando percepções, práticas e impactos potenciais do uso de LLMs nessas duas dimensões cruciais, combinando um survey exploratório com a possibilidade

de um experimento controlado dedicado à análise de um ritual ágil específico. Ao esclarecer esses aspectos, espera-se gerar evidências que apoiem equipes e organizações na tomada de decisões mais informadas sobre a adoção da inteligência artificial em processos ágeis.

Fundamentação teórica

A comunicação é um elemento crítico em equipes de desenvolvimento, dada a interdependência entre membros e a necessidade de alinhamento contínuo. Ruídos, artefatos pouco claros e divergências de entendimento continuam associados a retrabalho e queda de produtividade, mesmo na presença de cerimônias estruturadas, como apontam Stol e Fitzgerald (2018). Esses desafios reforçam a importância de mecanismos que apoiem síntese, clarificação e rastreabilidade das informações. A previsibilidade, por sua vez, depende da capacidade de estimar esforço, antecipar riscos e manter estabilidade no fluxo de trabalho, mas permanece limitada por requisitos ambíguos, comunicação insuficiente e visibilidade reduzida de impedimentos, conforme discutido por Dingsøyr, Moe e Ralph (2022).

Nesse contexto, modelos de linguagem de grande escala (LLMs) passaram a apoiar atividades como síntese de reuniões, revisão de histórias e documentação, acelerando tarefas cognitivas e melhorando clareza textual (Peng et al., 2023). Ainda assim, a maior parte das evidências concentra-se em usos individuais, deixando em aberto como esses modelos influenciam comunicação, registro de decisões e previsibilidade em interações coletivas. Para investigar tais questões, a Engenharia de Software Experimental oferece métodos adequados, como surveys para mapear percepções e práticas (Kitchenham & Pfleeger, 2008) e experimentos controlados para avaliar possíveis efeitos causais de intervenções baseadas em IA.

Trabalhos relacionados

Pesquisas recentes sobre o uso de Large Language Models (LLMs) em práticas ágeis têm se concentrado na qualidade de artefatos textuais e no apoio a interações de equipe. Zhang et al. (2024) demonstram que agentes baseados em LLM podem melhorar clareza e exatidão de histórias de usuário em ambientes industriais, reduzindo ambiguidades no refinamento. Em paralelo, Cabrero-Daniel et al. (2024) avaliam assistentes de reunião baseados em LLM aplicados a reuniões diárias e refinamentos, observando contribuições para síntese de decisões e registro de discussões, ainda que com limitações ligadas ao contexto e ao efeito de novidade. No âmbito da produtividade individual, Peng et al. (2023) mostram que desenvolvedores usando GitHub Copilot concluem tarefas mais rapidamente, reforçando o potencial dos LLMs para acelerar atividades cognitivas e de escrita.

Entretanto, tais estudos apresentam cobertura limitada quanto ao impacto direto dos LLMs em processos coletivos, especialmente na previsibilidade de equipes. Revisões amplas, como a de Hou (2024), apontam que a maioria dos avanços ainda se concentra em tarefas individuais — geração de código, reparo e documentação — e não em interações colaborativas ou rituais ágeis. Trabalhos em previsibilidade utilizando machine learning, como Forouzesh (2025), sugerem que visibilidade de dependências e identificação precoce de riscos favorecem estimativas mais estáveis, oferecendo insumos metodológicos relevantes, embora sem foco em LLMs. Assim, permanece uma lacuna significativa sobre como esses modelos influenciam comunicação coletiva e previsibilidade em equipes, lacuna que este estudo busca investigar por meio de um survey com profissionais e de um experimento controlado planejado. Diferentemente dos estudos anteriores,

este trabalho examina percepções coletivas e práticas reais relacionadas ao uso de LLMs em equipes, um campo ainda pouco explorado.

Questões de pesquisa

Esta pesquisa foi precedida por uma Revisão Sistemática da Literatura (RSL) voltada a identificar evidências sobre o uso de Large Language Models (LLMs) no suporte à comunicação e à previsibilidade em equipes de desenvolvimento. O protocolo seguiu as diretrizes consolidadas de Kitchenham e Charters (2007) e as recomendações PRISMA 2020 para assegurar transparência e rigor na triagem (Page et al., 2021). A RSL foi guiada por três questões principais: **(RQ1)** como LLMs têm sido aplicados para melhorar a comunicação em equipes; **(RQ2)** quais evidências empíricas relacionam LLMs à previsibilidade — incluindo estimativas, riscos e métricas de fluxo; e **(RQ3)** quais métodos, métricas e ameaças à validade são reportados pelos estudos existentes. O escopo foi estruturado pelo modelo PICOC, considerando equipes como população, uso de LLMs como intervenção, práticas sem IA como comparador, comunicação e previsibilidade como outcomes, e organizações de software como contexto.

As buscas foram realizadas em bases consolidadas — ACM Digital Library, IEEE Xplore, Scopus, Web of Science, SpringerLink e repositórios de pré-prints — cobrindo o período de 2018 a 2025, com strings que combinaram termos relacionados a metodologias ágeis, LLMs e comunicação/previsibilidade. Após remoção de duplicatas, a triagem PRISMA incluiu análise por título, resumo e leitura integral, seguida do registro das exclusões. Estudos selecionados atenderam aos critérios de incluir práticas ágeis, uso de LLMs, método empírico e relação explícita com comunicação ou previsibilidade. Para cada trabalho, foram extraídos o tipo de equipe, tarefa investigada, modelo utilizado, desenho metodológico, métricas analisadas e ameaças à validade, formando uma síntese qualitativa agrupada por tipo de prática ágil e impacto reportado, conforme organização sugerida por Dingsøyr, Moe e Ralph (2022).

Metodologia

Este estudo adota uma abordagem empírica em duas fases, iniciando por um **survey exploratório** com profissionais que atuam em equipes e, em etapa futura, um experimento controlado opcional. Embora mencionado para contextualizar o desenho geral da pesquisa, **o experimento não integra o escopo deste artigo**. O foco atual é o survey — ainda não aplicado, vide Apêndice A — destinado a identificar percepções, práticas e dificuldades relacionadas à comunicação, previsibilidade e uso de LLMs em atividades colaborativas. Esta estratégia segue recomendações de Kitchenham e Pfleeger (2008) e Stol e Fitzgerald (2018), que destacam o valor da combinação de métodos para investigar percepções e possíveis efeitos de intervenções tecnológicas.

A população-alvo inclui desenvolvedores, QAs, Scrum Masters e Product Owners que atuam em equipes. A amostragem será não probabilística, por conveniência e “bola de neve”. O questionário contém cinco blocos — perfil, comunicação, previsibilidade, uso e confiança em LLMs, e perguntas abertas — com itens em escala Likert de 1 a 5. A elaboração segue diretrizes de validade de construto (Kitchenham & Pfleeger, 2008), prevendo **pré-teste** com 8 a 10 participantes. Após isso, o formulário será disponibilizado on-line com TCLE garantindo anonimato e uso acadêmico, com coleta prevista entre duas e três semanas.

A análise envolverá estatística descritiva, correlações de Spearman e testes não paramétricos (Mann–Whitney ou Kruskal–Wallis) para comparação entre grupos, além de

categorização temática das respostas abertas. As principais ameaças à validade já foram antecipadas: limitada generalização devido à amostragem por conveniência (mitigada por descrição detalhada da amostra), riscos de construto (reduzidos por avaliação de juízes cegos) e ameaças de conclusão (tratadas pelo uso de testes apropriados e relato de tamanho de efeito). A confiabilidade será reforçada pelo pré-teste e pela análise de consistência interna, como alfa de Cronbach.

Estado atual do trabalho

Até o momento, foram concluídas a definição do problema de pesquisa, a revisão da literatura, a fundamentação teórica e a elaboração do protocolo metodológico. O instrumento do survey também está finalizado, incluindo formulação dos itens, estrutura de análise, termo de consentimento e organização das escalas, embora ainda necessite de validação por pares antes de sua aplicação. A próxima etapa consiste na aplicação do survey a profissionais que atuam com métodos ágeis; após a coleta, serão conduzidas análises descritivas, correlações entre construtos e categorização temática das respostas abertas, dando continuidade ao desenvolvimento da pesquisa.

Resultados esperados

Os resultados esperados deste estudo estão diretamente alinhados às questões de pesquisa propostas. Para a **RQ1**, prevê-se que o survey revele um panorama atualizado das principais dificuldades associadas à comunicação em equipes, incluindo clareza de requisitos, registro de decisões, alinhamento entre áreas e visibilidade de riscos. Também se espera identificar em quais atividades colaborativas profissionais percebem maior potencial para aplicação de LLMs, bem como o grau atual de adoção, confiança e utilidade percebida dessas ferramentas. Em relação à **RQ2**, a análise poderá indicar se experiências prévias com LLMs se correlacionam com percepções mais positivas sobre previsibilidade, estabilidade do fluxo de trabalho e antecipação de riscos, oferecendo indícios sobre possíveis mecanismos de apoio à melhoria dessas dimensões.

No contexto da **RQ3**, a síntese das respostas abertas deve revelar padrões de uso, limitações percebidas e oportunidades onde os LLMs podem contribuir de forma mais efetiva, como documentação de decisões, resumos de reuniões ou identificação preliminar de impedimentos. Esses achados servirão de base para orientar intervenções futuras, incluindo o delineamento de um experimento controlado em etapa posterior da pesquisa. Em conjunto, os resultados esperados contribuem para o avanço do conhecimento sobre aplicação prática de IA generativa em métodos ágeis, além de apoiar a formulação de diretrizes para adoção segura e efetiva de LLMs em rituais colaborativos de desenvolvimento de software.

Este estudo contribui ao caracterizar empiricamente como LLMs podem apoiar comunicação e previsibilidade em equipes, abordando desafios como clareza de requisitos, alinhamento entre áreas e visibilidade de riscos — aspectos ainda pouco explorados na literatura sobre colaboração humano-IA (Stol & Fitzgerald, 2018; Dingsøyr, Moe & Ralph, 2022). No âmbito tecnológico, oferece um instrumento de survey reutilizável, um protocolo experimental reproduzível e diretrizes iniciais para adoção responsável de LLMs em práticas ágeis, fornecendo insumos para pesquisas futuras e para o desenvolvimento de soluções que integrem IA generativa ao trabalho colaborativo em Engenharia de Software.

Conclusão

Este trabalho apresentou uma investigação empírica estruturada para compreender como modelos de linguagem (LLMs) podem apoiar comunicação e previsibilidade em equipes. A revisão da literatura mostrou que, embora a IA generativa tenha avançado em tarefas individuais, faltam evidências sobre seu impacto em atividades colaborativas que exigem alinhamento contínuo, clareza de requisitos e registro sistemático de decisões. Para enfrentar essa lacuna, o estudo propõe um survey exploratório — ainda não aplicado — destinado a mapear percepções, práticas e dificuldades enfrentadas por profissionais, além de orientar etapas posteriores e um experimento controlado opcional que poderá avaliar efeitos potenciais do uso de LLMs em rituais específicos.

Como principal contribuição, o trabalho oferece um **modelo empírico replicável** composto por um instrumento de survey validado, um protocolo experimental reproduzível e diretrizes iniciais para a adoção responsável de LLMs em práticas colaborativas. Espera-se que esses elementos apoiem novas pesquisas e contribuam para a formação de uma base acumulativa de evidências sobre colaboração humano-IA em ambientes de desenvolvimento. Os próximos passos incluem a aplicação do survey, a análise das respostas coletadas e, caso indicado, a realização do experimento, possibilitando refinamentos metodológicos e avanços no entendimento sobre como LLMs podem aprimorar comunicação, previsibilidade e qualidade das entregas em equipes.

Referências

Cabrero-Daniel, B.; Herda, T.; Pichler, V.; Eder, M. (2024). Exploring Human-AI Collaboration in Agile: Customised LLM Meeting Assistants.

Disponível em: <https://doi.org/10.48550/arXiv.2404.14871>

Dingsoyr, T., Moe, N. B., & Ralph, P. (2022). Agile Software Development: Current Research and Future Directions. Foundations and Trends in Information Systems.

Disponível em: <https://doi.org/10.1561/29000000028>

Forouzesh Nejad, Ali Akbar & Arabikhan, Farzad & Gegov, Alexander & Jafari, Raheleh & Ichtev, Alexandar. (2025). Data-Driven Predictive Modelling of Agile Projects Using Explainable Artificial Intelligence. Electronics.

Disponível em: <https://doi.org/14.2609.10.3390/electronics14132609>.

HOU, Xinyi et al. (2024) Large language models for software engineering: A systematic literature review. ACM Transactions on Software Engineering and Methodology, v. 33, n. 8, p. 1-79, 2024. Disponível em: <https://doi.org/10.48550/arXiv.2308.10620>

Kitchenham, B., & Charters, S. (2007). Guidelines for Performing Systematic Literature Reviews in Software Engineering. EBSE Technical Report.

Disponível em: <https://doi.org/10.13140/RG.2.1.1295.4646>

Kitchenham, B.; Pfleeger, S. L. (2008). Personal Opinion Surveys. In Guide to Advanced Empirical Software Engineering. Springer.

Disponível em: https://doi.org/10.1007/978-1-84800-044-5_3

Page, M. J., et al. (2021). PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372:n71.

Disponível em: <https://doi.org/10.1136/bmj.n71>

Peng, S. et al. (2023). The Impact of AI on Developer Productivity. Microsoft Research. <https://www.microsoft.com/en-us/research/publication/the-impact-of-ai-on-developer-productivity/>

Ralph, P. (2021). Toward a Unified Theory of Software Engineering. *ACM/IEEE*.

Disponível em: <https://dl.acm.org/doi/10.1109/ICSE-NIER52604.2021.00022>

Stol, K.-J.; Fitzgerald, B. (2018). The ABC of Software Engineering Research. *ACM Transactions on Software Engineering and Methodology*.

<https://doi.org/10.1145/3236021>

Zhang, Z.; Rayhan, M.; Herda, T.; Goisau, M.; Abrahamsson, P. (2024). LLM-based agents for automating the enhancement of user story quality: An early report.

Disponível em: <https://doi.org/10.48550/arXiv.2403.09442>

Apêndice A – Survey Model

Disponível para consulta em:

<https://github.com/eduardo-andre-de-lima-carneiro/llms-na-gestao-de-equipes/blob/master/complementary-artefacts/survey-model.md>