

Em direção a uma Taxonomia de Métricas e Observabilidade para Chatbots no Ensino de Programação

Renato de Souza Garcia¹, Williamson Silva² Pedro Henrique Dias Valle¹

¹Instituto de Matemática e Estatística – Universidade de São Paulo (IME-USP)

²Universidade Federal do Cariri (UFCA)

renato.s.garcia@usp.br

williamson.silva@ufca.edu.br

pedrohenriquevalle@usp.br

A dificuldade no ensino e na aprendizagem de programação é evidenciada pelos baixos índices de aprovação em disciplinas da área [Alves et al. 2019]. Essa complexidade pode estar associada à necessidade de desenvolver habilidades de abstração, resolução de problemas, raciocínio e pensamento lógico [Luxton-Reilly et al. 2018, Turpen et al. 2016]. Além disso, a comunicação insuficiente entre mentores e estudantes pode afetar aspectos motivacionais ao longo do curso [Wijaya and Purwarianti 2024]. Para apoiar esse processo, educadores têm adotado ferramentas tecnológicas, como *chatbots* educacionais e Sistemas Tutores Inteligentes (ITS), que oferecem *feedback* imediato e atuam como agentes pedagógicos com menor dependência de intervenção humana [Wang et al. 2024, Wijaya and Purwarianti 2024].

Mesmo com a adoção de novas tecnologias no ensino de programação, métodos tradicionais de avaliação de aprendizagem frequentemente se mostram inadequados para capturar interações complexas com agentes inteligentes, pois tendem a avaliar apenas o produto final, em vez de considerar processos cognitivos, criatividade e estratégias iterativas de resolução de problemas [Hadyaoui and Cheniti-Belcadhi 2024]. Nesse contexto, evidencia-se a necessidade de sistemas que vão além da verificação de acurácia e consigam registrar e analisar, de forma mais completa, a jornada cognitiva e criativa do estudante [Hadyaoui and Cheniti-Belcadhi 2024].

Nesse contexto, a observabilidade constitui o elemento central desta proposta. A observabilidade educacional permite identificar e monitorar as interações dos alunos com o *chatbot* ou ambiente de programação, estabelecendo a base em tempo real para análises de dados que podem apoiar o aprimoramento do sistema e a identificação da necessidade de intervenções pedagógicas personalizadas [Wang et al. 2024].

Trabalhos recentes têm apresentado diferentes abordagens para coletar e monitorar dados de aprendizagem. No contexto de *Multimodal Learning Analytics* (MMLA), múltiplas fontes de dados, como *logs* do sistema, áudio, vídeo e expressões faciais, podem ser integradas em uma linha do tempo comum, apoiando a análise das interações e das dificuldades dos estudantes [Cohn et al. 2024]. Em outra direção, estudos empregam técnicas de Processamento de Linguagem Natural (NLP) para obter intenções e entidades em interações com *chatbots* e identificar padrões de dúvidas recorrentes, subsidiando o planejamento pedagógico [Mathew et al. 2021].

A observabilidade tem apoiado avaliações automáticas em ambientes *online*,

incluindo a geração de *quizzes* personalizados com base no histórico de submissões e interações [Han and Leung 2023]. Visando identificar e entender os efeitos pedagógicos, a literatura demonstra a aplicação de referenciais como CIE e MELT-OPS [Gottipati et al. 2023] e análises comportamentais quantitativas, por exemplo, comparando o uso do *chatbot* com a participação em fóruns [Ai et al. 2023]. Apesar dos avanços no uso de chatbots no ensino de programação, ainda falta um referencial unificado para organizar dados, métricas e critérios de validação na perspectiva da observabilidade educacional.

Este trabalho compreende a observabilidade na aprendizagem mediada por *chatbots*, como a capacidade do sistema de registrar evidências (eventos, *logs* e dados de interação e de código) que possibilitam compreender aspectos do processo de aprendizagem e, assim, apoiar intervenções pedagógicas e análises de efetividade. A partir dessa perspectiva, investigamos como a literatura tem abordado essa observabilidade em *chatbots* educacionais de programação: quais dados são coletados, quais métricas são implementadas para monitoramento e apoio, e quais procedimentos de avaliação sustentam as conclusões reportadas.

Como base para esta proposta, conduzimos um Mapeamento Sistemático da Literatura (MSL) para organizar o estado da arte sobre métricas aplicadas em chatbots no ensino de programação. O MSL está sendo usado para consolidar e identificar: (i) métricas adotadas e suas definições operacionais; (ii) métodos e instrumentos de avaliação (quantitativos, qualitativos e mistos), como testes de aprendizagem, questionários de percepção/usabilidade e análises estatísticas; (iii) conteúdos de programação, linguagens e níveis educacionais; e (iv) teorias *frameworks* pedagógicos e fundamentos de avaliação associados. A síntese do MSL serve como base para este trabalho, pois consolida evidências que orientam a construção do artefato proposto e definem critérios mínimos para *design* e avaliação de *chatbots* educacionais.

A principal contribuição desta pesquisa é a proposição de uma Taxonomia de Métricas para *chatbots* educacionais em programação, fundamentada em evidências identificadas no MSL e em referenciais de avaliação educacional, usabilidade e aprendizagem. A taxonomia visa organizar as informações encontradas em um *framework* que torne as métricas comparáveis, reutilizáveis e aplicáveis em estudos e sistemas reais. Além de categorizar e relacionar famílias de métricas, segundo dimensões como finalidade (monitoramento, diagnóstico, intervenção/recomendação, avaliação de aprendizagem e avaliação técnica), fonte do dado (diálogo, *logs* de uso, evidências de código e percepção), natureza (quantitativa, qualitativa ou mista) e granularidade temporal (turno/evento, sessão, atividade, disciplina), indicando, quando possível, os fundamentos teóricos e as evidências que comprovam sua efetividade na literatura.

Nesta proposta, a observabilidade complementa a taxonomia: enquanto a taxonomia define o que medir e por quê, a observabilidade orienta como instrumentar e coletar sinais para operacionalizar essas métricas. Como resultado, o trabalho estabelece diretrizes iniciais para a instrumentação e a organização dos dados e métricas observadas.

Link do vídeo: <https://youtu.be/8CMQ6mtUPoQ>

Referências

- Ai, Y., Baveja, M., Girdhar, A., O'Dell, M., and DeOrio, A. (2023). A custom generative ai chatbot as a course resource. In *ASEE Annual Conference*.
- Alves, G., Rebouças, A., and Scaico, P. (2019). Coding dojo como prática de aprendizagem colaborativa para apoiar o ensino introdutório de programação: Um estudo de caso. In *Anais do XXVII Workshop sobre Educação em Computação*, pages 276–290. SBC.
- Cohn, C., Snyder, C., Fonteles, J. H., Ashwin, T. S., Montenegro, J., and Biswas, G. (2024). A multimodal approach to support teacher, researcher and ai collaboration in stem+c learning environments. *British Journal of Educational Technology*, 56:595–620.
- Gottipati, S., Shim, K. J., and Shankararaman, V. (2023). Ai for connectivism learning: Undergraduate students' experiences of chatgpt in advanced programming courses. In *Proceedings of Americas Conference on Information Systems (AMCIS)*.
- Hadyaoui, A. and Cheniti-Belcadhi, L. (2024). Intelliframe: A framework for ai-driven, adaptive, and process-oriented student assessments. In *Proceedings of the 20th International Conference on Web Information Systems and Technologies (WEBIST)*, pages 441–448.
- Han, D. and Leung, M. Y.-Y. (2023). The intelligent cloud teacher. In *Proceedings of the 31st International Conference on Computers in Education (ICCE)*.
- Luxton-Reilly, A., Alblawi, I., Becker, B. A., Giannakos, M., Kumar, A. N., Ott, L., Paterson, J., Scott, M. J., Sheard, J., and Szabo, C. (2018). Introductory programming: a systematic literature review. In *Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education*, pages 55–106.
- Mathew, A. N., V., R., and Paulose, J. (2021). Nlp-based personal learning assistant for school education. *International Journal of Electrical and Computer Engineering (IJECE)*, 11(5):4522–4530.
- Turpen, C., Dancy, M., and Henderson, C. (2016). Perceived affordances and constraints regarding instructors' use of peer instruction: Implications for promoting instructional change. *Physical Review Physics Education Research*, 12(1):010116.
- Wang, J. Y., Wang, W. Y., and Liang, J. C. (2024). Designing an educational chatbot for enhanced learning in programming courses. In *2024 12th International Conference on Information and Education Technology (ICIET)*.
- Wijaya, O. C. and Purwarianti, A. (2024). An interactive question-answering system using large language model and retrieval-augmented generation in an intelligent tutoring system on the programming domain.