

Universidade de Pernambuco Programa de Pós-Graduação em Engenharia da Computação (PPGEC)

Proposta de Dissertação de Mestrado

Área: Computação Inteligente

Título: Integração de *In-Context Learning* no Design Automatizado de Algoritmos

Orientador – Fernando Buarque de Lima Neto (fbln@ecomp.poli.br)

Co-orientador – Luís Filipe de Araújo Pessoa (lfap@ecomp.poli)

Contexto

O Design Automatizado de Algoritmos (ADA) visa reduzir o esforço humano e o viés associados ao design manual de heurísticas, metaheurísticas e hiper-heurísticas, automatizando a geração, seleção, configuração e adaptação de algoritmos para classes de problemas [1]. Abordagens clássicas incluem programação genética para hiper-heurísticas, algoritmos de seleção de algoritmos baseados em meta-características e meta-aprendizado para otimização caixa-preta [2]. Mais recentemente, *Large Language Models* (LLMs) emergiram como um novo paradigma para o ADA, atuando como otimizadores (LLMaO), preditores (LLMaP), extratores (LLMaE) e *designers* (LLMaD) de algoritmos, e dando origem a diversos frameworks [1, 3]. Esse paradigma (LLM4AD) explora capacidades de *in-context learning* (ICL), em que o modelo aprende a partir de exemplos no próprio prompt, sem atualização de pesos, para inferir e refinar heurísticas a partir de pares solução-desempenho. Paralelamente, *Tabular Foundation Models* (TabFM), como TabPFN e TabICL [4], demonstraram que ICL pode ser aplicado também a dados tabulares de forma extremamente eficiente, oferecendo previsões calibradas em milissegundos, sem necessidade de treinar modelos. Essa característica é especialmente atrativa para servir como modelos substitutos (*surrogate models*) em otimização cara e para seleção/recomendação de algoritmos a partir de meta-características de instâncias de problemas e dos próprios algoritmos [2, 4].

Problema

Apesar do avanço expressivo de LLM4AD, persistem limitações relevantes: (i) custo computacional e financeiro das chamadas a LLMs; (ii) possíveis vieses de treinamento (generalização limitada); (iii) sensibilidade a prompts; (iv) redução de diversidade; e, (iv) escassa exploração de TabFMs como componentes complementares aos LLMs no ciclo de ADA.

Pergunta Principal

Como conceber, implementar e avaliar um *framework* de Design Automatizado de Algoritmos que combine ICL sobre *foundation models* complementares (LLMs e TabFMs), os quais atuem em papéis distintos do laço evolutivo (geração, predição e seleção), de modo a melhorar a eficiência computacional, a qualidade, a generalização e adaptação de métodos heurísticos?

Perguntas Secundárias

- Quais são, segundo a literatura recente, os papéis canônicos (LLMaO, LLMaP, LLMaE, LLMaD) e as estratégias de ICL mais adequados a cada etapa do laço de ADA, e quais lacunas remanescem?
- Como TabFMs podem ser empregados em ADA como modelos substitutos (e.g., para predição de desempenho de heurísticas e/ou de configurações de algoritmos), comparativamente a LLMs atuando como *surrogates* e a métodos clássicos (e.g., processos gaussianos, *random forest*)?
- Que arquitetura de orquestração, treinamento, política de chamadas e mecanismos de ICL equilibram o *trade-off* multiobjetivo entre custo, qualidade e generalização?

Objetivos

Conceber, implementar e avaliar um *framework* de Design Automatizado de Algoritmos baseado em ICL sobre *foundation models* complementares (i.e., LLMs e TabFMs) com potenciais ganhos em eficiência, qualidade e generalização.

Produtos esperados

- Revisão estruturada da literatura sobre ICL e *foundation models* (LLMs e TabFMs) para o Design Automatizado de Algoritmos.
- Arquitetura de referência de um *framework* híbrido LLM+TabFM para ADA.
- Implementação do *framework*.
- Avaliação empírica comparativa em *benchmarks* otimização combinatória e casos reais de planejamento de políticas públicas, com análises de eficiência, qualidade e generalização.

Referências

- [1] Liu, F., Yao, Y., Guo, P., Yang, Z., Lin, X., Zhao, Z., ... & Zhang, Q. (2026). A systematic survey on large language models for algorithm design. *ACM Computing Surveys*, 58(8), 1-32.
- [2] Zheng, Y., Zhang, L., Li, K., Wang, R., Li, W., Zhang, T., ... & Jin, Y. (2026). A survey on large language models driven meta-optimizers for automated intelligent optimization. *Artificial Intelligence Review*.
- [3] Yu, H., & Liu, J. (2026). Exploring Automated Algorithm Design Synergizing Large Language Models and Evolutionary Algorithms: Survey and Insights. *Evolutionary Computation*, 1-28.
- [4] Qu, J., Holzmüller, D., Varoquaux, G., & Morvan, M. L. (2026). TabICLv2: A better, faster, scalable, and open tabular foundation model. *arXiv preprint arXiv:2602.11139*.