Proceedings of the Thirty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-26)
Doctoral Consortium Track

Novel Applications of Large Language Models in Computational Social Choice

Nguyen Thach
University of Nebraska-Lincoln
nate.thach@huskers.unl.edu

Abstract

My research initiates the use of large language models (LLMs) to automate the design of effective algorithms for challenging problems in computational social choice, namely participatory budgeting (PB) and facility location mechanism design. To this end, I propose two LLM-based frameworks that respectively generate interpretable PB rules and mechanisms fulfilling competing objectives, i.e., optimality vs. fairness or strategyproofness. Future directions include extending these frameworks to achieve provable guarantees and addressing other relevant areas such as fair division.

1 Introduction

Computational social choice concerns the problem of aggregating individual or group preferences into collective decisions and has extensive real-world applications, including voting and elections, fair resource allocation, and budget allocation. Let us consider the following two problems.

Example 1. *A city council seeks to allocate a given budget across a set of public projects (e.g., building new parks and libraries, installing public restrooms, and improving street conditions) for serving the agents (i.e., residents of the city), each of whom has their own preferences over the projects. Ideally, the allocation should (i) optimize the overall utility or satisfaction while (ii) ensuring fairness, i.e., each agent has roughly equal influence on the allocation outcome.*

The problem is a central topic in participatory budgeting (PB), a democratic paradigm for deciding the funding of public projects given the agent preferences. Unfortunately, despite the algorithmic progress in recent years, there exists a fundamental tradeoff between notions of utility and fairness—utility-maximizing allocations are generally not fair, and vice versa, fair allocations are likely suboptimal. Given the immense domain expertise and manual efforts typically required to design PB rules—functions mapping agent preferences to an allocation outcome—we ponder the following question: **Can we automate the design of PB rules that achieve (Pareto-)optimal utility-fairness tradeoffs?**

Example 2. *A government agency seeks to locate several public facilities (e.g., hospitals, schools, transit stations) to serve agents within a region. Each agent has their own preferences over the facility locations and hence may have the incentive to misreport preferences to manipulate facility locations, potentially leading to undesirable societal situations. Thus, the goal is to elicit agent preferences truthfully and determine facility locations that optimize some cost objective (e.g., minimizing the social cost—the total distances between agent ideal locations and facility locations).*

The problem describes a standard mechanism design setting in facility location [Moulin, 1980], for which existing studies have proposed to design *strategyproof* mechanisms that ensure no agents are worse off from reporting their preferences truthfully. While a plethora of handcrafted strategyproof mechanisms exist, they often yield poor worst-case guarantees. Furthermore, for more than two facilities under the social cost objective, there is no deterministic strategyproof mechanism with bounded approximation ratios [Fotakis and Tzamos, 2014]. In response, we raise the question: **Can we automate the design of strategyproof mechanisms in facility location that are empirically optimal?**

The common theme across these two problems, as well as many others in computational social choice, is the need of fulfilling competing objectives, which has only been partially addressed via manual algorithm design with rigorous mathematical analyses from human experts. The reliance on human ingenuity despite the rapid advancements in deep learning in the past decades is not without cause as the final collective decision for any given social choice problem may have a profound impact on the daily life of every participating human agent, hence the necessity of transparency and interpretability in reaching such decision. Lately, large language models (LLMs) have shown remarkable capabilities in algorithm design for various combinatorial optimization problems such as traveling salesman problems [Liu *et al.*, 2026], with interpretability attained via code implementation. Thus, my research strives to address the following central question:

> *Can LLMs automate the design of effective algorithms for multi-objective problems in computational social choice?*

2 Designing Participatory Budgeting Rules

To answer the question from Example 1 affirmatively, we propose the LLMRule framework [Thach *et al.*, 2026a] that

maintains a population of heuristic rules. It adopts an evolutionary search procedure, iteratively searching for rules that yield better tradeoffs between utility and fairness objectives. Each rule is evaluated on a set of problem instances and assigned a fitness value that indicates its probability of being selected as reference for generating new rules. There are two main challenges inherent to PB that we specifically address. (**Challenge I**) There is a paucity of quantifiable fairness objectives measuring how fair the allocations returned from PB rules are, whereas existing LLM-based evolutionary frameworks require a fine-grained performance measure that can capture small improvements during search for ensuring convergence at desirable algorithms. (**Challenge II**) Verifying whether a given budget allocation satisfies standard notions of fairness is intractable [Aziz *et al.*, 2017], whereas existing LLM-based frameworks assume the availability of an efficient evaluation procedure such that the performance of the generated algorithms can be quickly computed.

Our Approach. In response to Challenge I, we introduce *Strong-EJR approximation*, a novel fairness objective inspired by the desirable fairness property termed Strong-EJR, to quantify the degree of fairness of the budget allocations returned from a given LLM-generated PB rule. To account for Challenge II, we define this objective based on our provably equivalent definition of Strong-EJR, the verification of which can be done with reduced complexity. To further address Challenge II, we devise techniques to speed up the computation of Strong-EJR approximation, resulting in a linear-time complexity with respect to the number of agents.

Evaluation. Using more than 600 real-world PB instances across Europe and North America, we empirically demonstrate that LLMRule outperforms existing "fair" rules, such as the Method of Equal Shares, in terms of overall utility while maintaining a similar degree of fairness with respect to Strong-EJR approximation on classical PB settings. Each resulting PB rule can be represented as a succinct mathematical expression and is hence fully interpretable.

3 Designing Facility Location Mechanisms

Regarding the question from Example 2, while prior work [Golowich *et al.*, 2018] has proposed automated mechanism design via deep learning, the inherent lack of interpretability from neural networks severely undermines its practicality. Therefore, we develop the LLMMech framework [Thach *et al.*, 2025], in which LLMs are integrated into an evolutionary framework for designing interpretable facility location mechanisms. Different from PB where good utility-fairness tradeoffs suffice, the goal now is to ensure the LLM-generated mechanisms attain (near-)optimality without compromising their (empirical) strategyproofness.

Our Approach. Using the notion of regret (i.e., strategyproof mechanisms always yield zero regret), we design a fitness function that simultaneously optimizes the social cost objective while maintaining empirical strategyproofness (on a finite set of problem instances) by introducing a penalty term for mechanisms with positive regret. By this means, empirically strategyproof mechanisms are more likely to be generated as training progresses.

Evaluation. Our numerical experiments with thousands of synthetic instances from various facility location settings show that LLMMech consistently outperforms all baseline mechanisms, including those learned via neural networks. We also demonstrate that, in addition to interpretability over deep learning approaches, the LLM-generated mechanisms learned from a particular distribution are generalizable to out-of-distribution preferences without retraining.

4 Prospects and Future Directions

My research has demonstrated the algorithmic capabilities of LLMs in multiple areas of computational social choice and hence the central question from Section 1 can be answered affirmatively. Given the effectiveness as well as interpretability of our proposed LLM-based frameworks, stakeholders may utilize them as a decision-aid system before making the final collective decision. We do, however, advocate rigorous testing from human experts before deployment, as these frameworks currently only consider empirical fairness and strategyproofness and may inadvertently misbehave in edge cases.

For ongoing and future work, first, given our prior study [Thach *et al.*, 2026b] on designing strategyproof mechanisms for a structural modification variant of the multi-facility location problem on graphs, which has implications in urban planning, I plan to extend LLMMech to tackle more general settings of such problem. Second, I will develop stronger variants of LLMRule and LLMMech to respectively generate PB rules and mechanisms with provable fairness/strategyproofness guarantees. Finally, I will introduce new LLM-based frameworks for addressing other—potentially more challenging—areas in computational social choice such as fair division and stable matching.

References

[Aziz *et al.*, 2017] Haris Aziz *et al.* Justified representation in approval-based committee voting. *Social Choice and Welfare*, 2017.

[Fotakis and Tzamos, 2014] Dimitris Fotakis and Christos Tzamos. On the power of deterministic mechanisms for facility location games. *ACM TEAC*, 2(4):1–37, 2014.

[Golowich *et al.*, 2018] Noah Golowich, Harikrishna Narasimhan, and David C Parkes. Deep learning for multi-facility location mechanism design. In *IJCAI*, 2018.

[Liu *et al.*, 2026] Fei Liu *et al.* Systematic survey of llms for algorithm design. *ACM Computing Surveys*, 2026.

[Moulin, 1980] Hervé Moulin. On strategy-proofness and single peakedness. *Public Choice*, 35(4):437–455, 1980.

[Thach *et al.*, 2025] Nguyen Thach, Fei Liu, Houyu Zhou, and Hau Chan. Large language models for multi-facility location mechanism design. *arXiv:2503.09533*, 2025.

[Thach *et al.*, 2026a] Nguyen Thach, Xingchen Sha, and Hau Chan. Large language models for designing participatory budgeting rules. In *AAMAS*, 2026. Forthcoming.

[Thach *et al.*, 2026b] Nguyen Thach, C. Wang, and Hau Chan. Improving the location of facilities through network addition modification: a revisit. *Ann. Oper. Res.*, 2026.