# Discrete Choice Modeling using Symbolic Regression with LLM Integration

Rui Jiang[1], Shadi Haj-Yahia[2], Omar Mansour[2], and Tomer Toledo[2]

[1]The Taub Faculty of Computer Science, Technion - Israel Institute of Technology
[2]Faculty of Civil and Environmental Engineering, Technion - Israel Institute of Technology

## Short summary

Discrete Choice Modeling (DCM) serves as a cornerstone in understanding decision-making behaviors across transportation, economics, and marketing. While random utility methods offer interpretability, they often suffer from rigid functional form assumptions. Conversely, machine learning approaches maximize predictive power at the cost of interpretability and the risk of overfitting. This research introduces a novel framework integrating Symbolic Regression (SR) with domain knowledge to automate the discovery of utility functions that are both accurate and interpretable. Within the optimization of this framework, we propose a hybrid evolutionary algorithm where an LLM acts as an intelligent operator for population initialization, mutation, and crossover. While the overarching framework remains a Symbolic Regression search, we incorporate the LLM specifically to enhance the generation of candidate solutions. Results on the Swissmetro and Synthetic datasets demonstrate that our SR-DCM implementation is on par with existing machine learning and hybrid models in predictive performance, while significantly reducing overfitting. Furthermore, it outperforms existing interpretable models, with LLM integration serving to further accelerate convergence.
**Keywords**: Discrete Choice Modeling (DCM), Hybrid Machine Learning Algorithms, Large Language Models (LLMs), Symbolic Regression (SR), Utility Function Discovery

## 1 Introduction

Discrete Choice Modeling (DCM) Ben-Akiva & Lerman (1985) is the mathematical framework used to model and predict the choices individuals make from a finite set of alternatives. Anchored in random utility maximization theory McFadden (1972), the fundamental goal is to identify a utility function $U_{in}$ for individual $i$ and alternative $n$. This framework was expanded through the systematic treatments of Ben-Akiva (1973) and later Revelt & Train (1998), which introduced flexible structures such as mixed logit and simulation-based estimation to address heterogeneity and substitution patterns beyond the multinomial logit. DCM provides a probabilistic framework for modeling decision-making. In this probabilistic framework, A decision-maker $n$ selects an alternative $i$ from a finite choice set $C_n$ based on the probability calculated from the utility model. The utility $U_{in}$ is decomposed into a deterministic component $V_{in}$ and a stochastic error term $\varepsilon_{in}$:

$$U_{in} = V_{in} + \varepsilon_{in} \tag{1}$$

The probability of choosing alternative $i$ is defined as $P(U_{in} > U_{jn}, \forall j \neq i)$. Under the assumption that $\varepsilon$ follows an i.i.d. Gumbel distribution, this yields the standard Multinomial Logit (MNL) probability:

$$P_{in} = \frac{e^{V_{in}}}{\sum_{j \in C_n} e^{V_{jn}}} \tag{2}$$

Traditionally, practitioners have relied heavily on the intuition of domain-knowledge experts to select variables that we enter the specification of utility functions and their functional forms, which guarantees interpretability but often fails to capture complex, non-linear relationships in the data.

In recent years, Machine Learning (ML) techniques have been integrated into DCM to capture these non-linearities Aboutaleb et al. (2021). ML models significantly outperform traditional parametric models in predictive accuracy and can automatically capture complex variable interactions without manual specification. Aboutaleb et al. (2021) show how machine learning can capture

nonlinearities while preserving behavioral structure, and comparative studies such as Rahnasto & Hollestelle (2024); Van Cranenburgh et al. (2022) document the predictive advantages of ML relative to traditional logit models. Hybrid approaches are emerging as well: Sifringer et al. (2020) introduced a two-part utility formulation that combines an interpretable, theory-driven component with a neural network that learns latent representations directly from the data, Han et al. (2020) further demonstrated this integration with TasteNet-MNL, where a compact neural module generates individual-specific taste parameters while the MNL structure preserves behavioral interpretability. However, the "black box" nature of these models makes them difficult to interpret. In policy-sensitive fields like transportation planning, understanding why a model predicts a shift in demand is as crucial as the prediction itself. Furthermore, ML models are prone to overfitting and may violate fundamental economic principles and domain knowledge Chernozhukov et al. (2018); Athey (2018) (e.g., positive elasticity of price).

To bridge the gap between accuracy and interpretability, evolutionary computation, specifically Genetic Algorithms (GA) and Grammatical Evolution (GE) O'Neill & Ryan (2002), has emerged as a promising alternative. Within the paradigm of evolutionary computation, Symbolic Regression (SR) Cranmer (2023); Kommenda et al. (2020) offers a compelling resolution to the tension between predictive accuracy and model interpretability. Symbolic Regression is a non-parametric optimization technique used to discover analytical expressions that describe the relationship between independent variables and a dependent target. Unlike parametric regression, which optimizes coefficients for a pre-specified functional form, SR performs a simultaneous search for both the model structure and its parameters.

In SR, mathematical expressions are represented as *expression trees*. Internal nodes consist of operators from a functional set $\mathcal{F}$ (e.g., $\{+, -, \times, \div, \sqrt{}, \log\}$), while terminal nodes consist of variables from a terminal set $\mathcal{T}$ and constant values. The search process is governed by an evolutionary heuristic:

1. **Population Initialization:** A diverse set of candidate models is generated stochastically to populate the initial search space.
2. **Fitness Assessment:** Each candidate model is evaluated according to a fitness metric.
3. **Selection and Genetic Operators:** To iterate toward better solutions, the algorithm employs:
    - *Tournament Selection:* Choosing the best-performing models to serve as parents for the next generation.
    - *Crossover:* Recombining sub-structures from two parent models to produce offspring that may capture the advantages of both.
    - *Mutation:* Introducing random structural changes to a model to prevent premature convergence and explore novel functional forms.

Udrescu & Tegmark (2020) demonstrated GA-based symbolic regression by rediscovering 100 equations from the Feynman Lectures on Physics, Cranmer et al. (2020) extracted symbolic equations from black-box deep learning models using SR, bridging data-driven and interpretable science. In the context of discrete choice modeling, Haj-Yahia et al. (2025) developed a GA framework with GE, yielding significant improvements in capturing choice behavior. However, pure SR can suffer from slow convergence Virgolin & Pissis (2022); Kantor et al. (2021); Lu et al. (2016) and often produces "bloated" Trujillo et al. (2016) mathematical expressions that are accurate but physically nonsensical Wang et al. (2023).

A growing body of work investigates the synthesis of Large Language Models (LLMs) with GA. FunSearch framework Romera-Paredes et al. (2024) showed the capability of LLM in functional search and algorithmic problem solving, by evolving functions iteratively with updated prompts with past iteration solutions. Recent advancements in neuro-evolutionary hybrids suggest a move away from direct LLM code generation. Instead, several studies have successfully integrated LLMs into the genetic programming workflow to strategically guide initialization and mutation processes Grayeli et al. (2024); Shojaee et al. (2024). Other than the above mentioned methods, Hua et al. (2025) finetuned LLM specifically for SR tasks, LLM-Meta-SR Zhang et al. (2025) explored the option of using LLM to generate selection operators through a Meta-SR process. Furthermore, several studies Merler et al. (2024); Romera-Paredes et al. (2024); Shojaee et al. (2024); Grayeli et al.

(2024); Zhang et al. (2025) leverage the In-Context Learning (ICL) Brown et al. (2020) capabilities of LLMs to facilitate self-evolution, allowing the models to iteratively refine their own performance.

Building upon these developments, this research proposes a novel integrated framework that leverages Symbolic Regression (SR) for the advancement of discrete choice modeling. The primary contribution of this work is threefold.

- we employ SR to autonomously discover functional forms that capture complex choice behaviors without the restrictive assumptions of traditional parametric models.
- we enhance the physical and economic consistency of the discovered models by enforcing domain-specific constraints, ensuring that the mathematical output aligns with established behavioral theory.
- we introduce a hybrid optimization paradigm that utilizes Large Language Models (LLMs) to guide the symbolic search process. Rather than relying on LLMs for the direct generation of utility functions, this approach integrates them as heuristic assistants to navigate the vast expression space more efficiently, thereby bridging the gap between data-driven discovery and expert knowledge.

## 2 Methodology

***Symbolic Regression Formulation for DCM***

The application of Symbolic Regression to DCM requires a change in the objective function compared to standard regression tasks. In typical SR applications (e.g., physics or simple regression), the algorithm minimizes the error (such as MSE) between a predicted value $\hat{y}$ and an observed value $y$. However, in DCM, the utility $U_{in}$ is a latent construct and is not directly observable; the dataset only contains the discrete final choices $y_{in} \in \{0, 1\}$.

Therefore, SR cannot regress the utility directly. Instead, the algorithm is adapted to evolve the functional form of the deterministic utility $V_{in}$ as follows:

1. **Candidate Generation:** The SR engine generates a population of expression trees, where each tree represents a potential specification for $V_{in}$ (e.g., $V = \beta_1 \cdot \text{time} + \beta_2 \cdot \text{cost}^2$).
2. **Probability Mapping:** For a given candidate tree, the utility values are computed and strictly mapped to choice probabilities $P_{in}$ using the MNL formula.
3. **Fitness Evaluation:** The fitness of the tree is not measured by error, but by the Log-Likelihood ($\mathcal{LL}$) of the observed choices under that specific utility formulation:

$$\mathcal{LL} = \sum_{n=1}^{N} \sum_{i \in C_n} y_{in} \ln(P_{in}) \tag{3}$$

The evolutionary process thus searches the space of mathematical expressions to find the functional form of $V$ that maximizes the likelihood of the empirical data.

***Incorporate Domain Knowledge***

Beyond adapting the symbolic regression architecture to the specific requirements of discrete choice modeling, it is essential to incorporate domain-specific constraints derived from behavioral science. By integrating these expert-defined priors, we ensure that the evolved utility functions remain consistent with economic theory and provide the high level of interpretability required for decision-making applications. We now provide a detailed exposition of the specific constraints developed for this framework and their implementation within the evolutionary search.

**Structural Constraints** To ensure interpretability and to keep the solution straightforward, we restrict the utility function to a "linear-in-parameters, non-linear-in-variables" structure:

$$V_{in} = \beta_1 \cdot T_{in1} + \beta_2 \cdot T_{in2} + \cdots + \beta_k \cdot T_{ink} \tag{4}$$

Where each term T is an interaction (multiplication, division) of features or transformed features (e.g., $\log(x)$, $\sqrt{x}$). The model may contain both numerical variables (e.g., *travel_time*) and categorical features (e.g., *age_group*, *income_group*).

**Economic Directions of effects** We enforce economic theory via derivative constraints to ensure behavioral validity. For example, the marginal utility of travel time must be non-positive:

$$\frac{\partial V_{in}}{\partial time_{in}} \leq 0 \tag{5}$$

Solutions violating these constraints are penalized or discarded during the selection process.

***Domain-Knowledge Informed Evolution Operators***

**Structure-Aware Mutation** To preserve the structural integrity of the utility function, we have re-engineered the core stochastic operators of the symbolic regression engine. Rather than merely shielding top-level nodes, we adapted the entire suite of mutation operators—including **node insertion**, **node deletion**, **operator mutation**, and several others—to be intrinsically structure-aware. These operators are designed to accept a valid utility expression and guarantee that the output remains strictly within the defined structural manifold. For instance, in the case of **operator mutation**, swaps are restricted to functionally analogous classes (e.g., swapping $*$ for $\div$, or log for $\sqrt{}$), ensuring that stochastic changes never violate the top-level linear summation required for the model's interpretability, consistently adheres to the form $V = \sum c_k \cdot T_k$.

**Constrained Crossover** To preserve structural integrity during recombination, our crossover operator only permits the exchange of homologous sub-structures. A functional term(s) from one parent can only be swapped with a functional term(s) from another, preventing the formation of mathematically invalid or non-interpretable nested structures.

**Local Coefficient Optimization** Following any structural change via mutation or crossover, the model performs a local optimization of the linear parameters $\beta$ using second-order methods (e.g., BFGS). This ensures that the fitness evaluation reflects the maximum likelihood potential of the newly discovered functional form rather than relying on inherited, sub-optimal constants.

***LLM-Guided Initialization and Evolution***

We also seek to investigate the potential for integrating LLM-informed heuristics into standard genetic operators, moving beyond traditional random perturbations to guide the symbolic search more effectively.

**Initialization** Rather than beginning with a purely random population, we utilize an LLM-driven initialization to seed the evolutionary process with domain-informed utility functions. By prompting the model to generate candidates based on established economic principles, we inject prior knowledge directly into the starting population, providing a robust "head start" that narrows the search space toward physically and theoretically plausible regions.

**Evolution** This semantic guidance continues through our Mutation and Crossover operators, which we transform from mechanical tree-swapping into context-aware evolutions. Instead of performing blind structural changes, the LLM treats parent expressions as conceptual foundations to synthesize offspring that reflect specific economic phenomena, such as diminishing returns or specific elasticity constraints.

- In semantic mutation: the model modifies a single parent to explore targeted behavioral shifts.
- In semantic crossover: it moves beyond the mechanical exchange of expression tree branches to identify and integrate functional motifs from the high-performing parents.

To keep the optimization process effective and diverse, we adopt a hybrid approach. While the LLM provides semantic guidance, we retain a designated fraction of traditional stochastic operators. By injecting "blind" randomness alongside domain-specific reasoning, we prevent the model from overfitting to the LLM's internal biases. This dual-track mechanism ensures that the search remains capable of novel discovery, balancing the sophisticated synthesis of the LLM with the unbiased exploration of traditional genetic programming.

## 3 Results and discussion

***Swissmetro Dataset***

The Swissmetro dataset Bierlaire et al. (2001) consists of survey data collected in Switzerland during March 1998 to analyze the potential impact of a revolutionary maglev underground system—the Swissmetro—against traditional transport modes such as rail and private car. Given that the Swissmetro represented a true modal innovation, researchers utilized Stated Preference (SP) surveys to capture hypothetical choice behavior from 441 rail-based travelers and 750 car-based travelers. Each respondent was presented with nine choice situations involving three alternatives: rail, Swissmetro, and car (available only to car owners). The dataset includes a comprehensive set of variables covering individual demographics (age, income, gender), travel characteristics (purpose, ticket type, baggage), and alternative-specific attributes such as travel time, cost, and headway. This dataset has become a foundational benchmark in discrete choice modeling for studying the acceptance of transportation innovations.

The following table 3 summarizes the key characteristics and descriptive statistics of the Swissmetro survey data.

| **Feature** | **Description / Statistics** |
|---|---|
| **Sample Size** | 1,191 individuals; 10,719 total observations |
| **Alternatives** | Train, Swissmetro (SM), and Car |
| **Individual Covariates** | Age, Gender, Income, GA ownership, Travel Purpose |
| **Alternative Attributes** | Travel Time (TT), Travel Cost (CO), Headway (HE), SM Seats Availability (SM_SEATS) |
| **Avg. Travel Time (min)** | SM: 87.47, Train: 166.63, Car: 123.80 |
| **Avg. Travel Cost (CHF)** | SM: 670.34, Train: 514.34, Car: 78.74 |
| **Availability** | Train/SM (100%), Car (84% of sample) |

Table 1: Summary of the Swissmetro Stated Preference Dataset

***Synthetic Dataset***

While machine learning and neural network-based approaches often achieve high accuracy within training distributions, they frequently fail to generalize when faced with distributional shifts. In the context of Discrete Choice Modeling (DCM), robustness to these shifts is critical; models are often deployed for long-term city planning or economic policy forecasting, where the future data distribution will almost certainly diverge from historical patterns. This makes out-of-distribution (OOD) testing a primary metric of significance.

To evaluate the generalization capabilities and susceptibility to overfitting of various models, we first constructed a synthetic dataset governed by a known ground truth utility function. While the variables and ranges are derived from the Swissmetro dataset, we generated the synthetic choices by stochastically resampling observations based on the probabilities derived from Equation 2, using a set of ground truth utility functions. The synthetic "ground truth" utility function follows the specification established by Haj-Yahia et al. (2025):

$$\begin{aligned} V_j =& \beta_{0_j} + \beta_{time_j} \log(time_j) + \beta_{headway_j} \log(headway_j) + \beta_{cost_j} \log(cost_j) \\ &+ \sum_g \beta_{cost,g_j} \log(cost_j) \delta_{income,g} + \sum_h \beta_{age,h_j} \delta_{age,h} + \beta_{seats,SM} \delta_{seats} \delta_{SM} \end{aligned} \tag{6}$$

Where $time_j$, $headway_j$ and $cost_j$ each represents the time, headway and cost of the individual $j$ using the current alternative. $\delta_{income,g}$ and $\delta_{age,g}$ represent if the individual $j$ is in the income / age group $g$, 1 if it belongs to the group and 0 if not. $\delta_{seats}\delta_{SM}$ term represent only for SM if the airline seats are available. Each $\beta$ is the coefficient for that term.

To simulate a rigorous OOD environment, we created multiple testing sets by inflating the time and cost variables by 20% to 40% before resampling the choices. This setup allows us to train models on a baseline distribution and evaluate their predictive performance on a shifted distribution.

Finally, we benchmark our approach against the state-of-the-art DCM methods, including traditional Expert-Designed Multinomial Logit (MNL, described in Sifringer et al. (2020) as FULL_MNL), pure machine learning (LightGBM) Ke et al. (2017), hybrid models - L-MNL Sifringer et al. (2020), TasteNet Han et al. (2020), and Grammatical DCM Haj-Yahia et al. (2025). The results are in table 2. Notice that here we take the model with the best validation negative log-likelihood for the best OOD performance, and test it on 100 different inflated test datasets. The test results are the average of the statistics on the 100 datasets. Overfit Rate is the increase / decrease of synthetic dataset test negative-log-likelihood nll over real-life dataset test negative-log-likelihood

Table 2: Comparative Performance and Overfitting Analysis

| | **Real-Life NLL** | | | **Synthetic NLL** | | | | |
|---|---|---|---|---|---|---|---|---|
| **Method** | **Train** | **Test** | **Gap (%)** | **Train** | **Test** | **Test$_{OOD}$** | **Gap (%)** | **OF Rate (%)** |
| Ground Truth | 0.706 | 0.745 | 5.52 | 0.706 | 0.733 | 0.683 | 3.82 | −6.82 |
| MNL | 0.754 | 0.795 | **5.44** | 0.790 | 0.800 | 0.778 | **1.27** | −2.75 |
| L-MNL | 0.508 | 0.644 | 26.8 | 0.679 | 0.815 | 0.789 | 20.0 | −3.19 |
| LightGBM | 0.343 | **0.592** | 72.6 | 0.483 | 0.764 | 0.730 | 58.2 | −4.45 |
| TasteNet | 0.564 | 0.674 | 19.5 | 0.706 | 0.750 | 0.725 | 6.23 | −3.33 |
| Gram. DCM | 0.668 | 0.714 | 6.89 | 0.712 | 0.735 | 0.739 | 3.23 | 0.54 |
| SR (No LLM) | 0.665 | 0.713 | 7.22 | 0.709 | 0.738 | 0.694 | 4.09 | **−5.96** |
| SR (With LLM) | 0.655 | 0.707 | 7.94 | 0.706 | **0.733** | **0.690** | 3.82 | **−5.87** |

***Real-life Dataset***

While synthetic datasets provide a controlled environment to verify if a model can recover a known ground truth, real-world data introduces unobserved heterogeneity and measurement noise that are characteristic of behavioral modeling. Alongside the synthetic experiments, we evaluate our proposed framework on the original Swissmetro to demonstrate its ability to capture implicit utility functions from empirical observations. Unlike the synthetic OOD testing, the real-world evaluation focuses on the model's capacity to maintain structural interpretability while achieving competitive predictive performance against black-box benchmarks. The results of these experiments are consolidated in Table 2.

***Result Analysis***

Results reveal a critical trade-off between local predictive fit and global structural robustness across the tested frameworks. On the real-life Swissmetro dataset, LightGBM and neural-hybrid models (L-MNL, TasteNet) achieve the lowest Test NLL (0.592–0.674). However, this performance is highly sensitive; LightGBM's training error (0.343) is nearly half its test error, suggesting it "memorizes" local patterns rather than discovering underlying behavioral laws.

The true strength of the SR framework is evident on the synthetic dataset under 40% out-of-distribution (OOD) shifts in time and cost. While ML models degrade significantly (LightGBM: 0.730; L-MNL: 0.789), our LLM-guided SR remains stable at 0.690, nearly matching the Ground Truth (0.683). This confirms SR recovers the actual functional utility; as explicit mathematical equations rather than dense weight matrices, these models remain valid even when variables move far outside training ranges.

The Train-Test Gap and OF Rate columns quantify this stability. While LightGBM and L-MNL show high local instability (with synthetic train-test gaps of 58.2% and 20.0%, respectively), both the SR (No LLM) and SR (With LLM) frameworks demonstrate minimal gaps (4.09% and 3.82%), closely mirroring traditional MNL (1.27%). Furthermore, because the OOD data inherently yields a lower NLL under the Ground Truth (an OF Rate of -6.82%), a robust model should reflect this exact structural shift. While ML models under-adjust, our SR frameworks (-5.96% and -5.87%) most accurately recover this baseline baseline shift. This highlights a profound advantage: SR maintains the rigorous, mathematical stability of traditional MNL while delivering vastly superior predictive accuracy (0.707 vs. 0.795 on real-world data).

Finally, comparing "SR (With LLM)" to "SR (No LLM)" and Grammatical DCM demonstrates

the clear efficiency of our hybrid approach. While both Grammatical DCM and SR utilize genetic algorithm frameworks, Grammatical DCM fails to capture the distributional shift (exhibiting a positive OF Rate of 0.54%). Furthermore, integrating the LLM acts as a powerful semantic filter, preventing the search process from wasting generations on "bloated" or economically nonsensical equations. This intelligent guidance allows the model to reach a more accurate Real-Life Test NLL (0.707) and superior OOD robustness compared to purely stochastic symbolic methods.

## 4 Conclusions

This research resolves the fundamental tension between the predictive accuracy of machine learning and the theoretical interpretability of traditional discrete choice models by introducing a novel framework that integrates Symbolic Regression (SR) with Large Language Models (LLMs). By treating the utility function as a mathematical expression to be autonomously discovered, and utilizing LLMs as heuristic assistants to guide the evolutionary search, our approach overcomes the mathematical bloat and slow convergence typical of pure symbolic methods. Experimental results on both real-world and systematically shifted synthetic distributions of the Swissmetro dataset demonstrate the framework's superior out-of-distribution (OOD) robustness. While black-box ML models suffered massive train-test gaps (up to 58.2% for LightGBM) under 40% distribution shifts, the LLM-guided SR maintained a stable test negative log-likelihood (0.690) nearly identical to the theoretical ground truth (0.683). Ultimately, the model's ability to mirror the true structural shift (an OF Rate of -5.87%, compared to the ground truth's -6.82%) confirms that this integration of generative AI and symbolic discovery successfully filters empirical noise to capture a generalizable, theoretically sound behavioral signal, providing a highly promising pathway for long-term policy forecasting and future applications in higher-dimensional choice sets.

## References


Aboutaleb, Y. M., Danaf, M., Xie, Y., & Ben-Akiva, M. (2021). Discrete choice analysis with machine learning capabilities. *arXiv preprint arXiv:2101.10261*.

Athey, S. (2018). The impact of machine learning on economics. In *The economics of artificial intelligence: An agenda* (pp. 507–547). University of Chicago Press.

Ben-Akiva, M. E. (1973). *Structure of passenger travel demand models.* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand* (Vol. 9). MIT press.

Bierlaire, M., Axhausen, K., & Abay, G. (2001). The acceptance of modal innovation: The case of swissmetro. In *Swiss transport research conference.*

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... others (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877–1901.

Chernozhukov, V., Demirer, M., Duflo, E., & Fernandez-Val, I. (2018). *Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in india* (Tech. Rep.). National Bureau of Economic Research.

Cranmer, M. (2023). Interpretable machine learning for science with pysr and symbolicregression. jl. *arXiv preprint arXiv:2305.01582*.

Cranmer, M., Sanchez Gonzalez, A., Battaglia, P., Xu, R., Cranmer, K., Spergel, D., & Ho, S. (2020). Discovering symbolic models from deep learning with inductive biases. *Advances in neural information processing systems*, *33*, 17429–17442.

Grayeli, A., Sehgal, A., Costilla Reyes, O., Cranmer, M., & Chaudhuri, S. (2024). Symbolic regression with a learned concept library. *Advances in Neural Information Processing Systems*, *37*, 44678–44709.

Haj-Yahia, S., Mansour, O., & Toledo, T. (2025). Grammar-based approach to data-driven utility specification for discrete choice models. *Available at SSRN 5195530*.

Han, Y., Pereira, F. C., Ben-Akiva, M., & Zegras, C. (2020). A neural-embedded choice model: Tastenet-mnl modeling taste heterogeneity with flexibility and interpretability. *arXiv preprint arXiv:2002.00922*.

Hua, Y., Li, R., Yao, J., Zhuang, G., Tang, S., Liu, B., ... Lu, Y. (2025). Finetuning large language model as an effective symbolic regressor. *arXiv preprint arXiv:2508.09897*.

Kantor, D., Von Zuben, F. J., & de Franca, F. O. (2021). Simulated annealing for symbolic regression. In *proceedings of the genetic and evolutionary computation conference* (pp. 592–599).

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, *30*.

Kommenda, M., Burlacu, B., Kronberger, G., & Affenzeller, M. (2020). Parameter identification for symbolic regression using nonlinear least squares. *Genetic Programming and Evolvable Machines*, *21*(3), 471–501.

Lu, Q., Ren, J., & Wang, Z. (2016). Using genetic programming with prior formula knowledge to solve symbolic regression problem. *Computational intelligence and neuroscience*, *2016*(1), 1021378.

McFadden, D. (1972). Conditional logit analysis of qualitative choice behavior.

Merler, M., Haitsiukevich, K., Dainese, N., & Marttinen, P. (2024). In-context symbolic regression: Leveraging large language models for function discovery. *arXiv preprint arXiv:2404.19094*.

O'Neill, M., & Ryan, C. (2002). Grammatical evolution. *IEEE Transactions on Evolutionary Computation*, *5*(4), 349–358.

Rahnasto, I., & Hollestelle, M. (2024). Comparing discrete choice and machine learning models in predicting destination choice. *European Transport Research Review*, *16*(1), 43.

Revelt, D., & Train, K. (1998). Mixed logit with repeated choices: households' choices of appliance efficiency level. *Review of economics and statistics*, *80*(4), 647–657.

Romera-Paredes, B., Barekatain, M., Novikov, A., Balog, M., Kumar, M. P., Dupont, E., ... others (2024). Mathematical discoveries from program search with large language models. *Nature*, *625*(7995), 468–475.

Shojaee, P., Meidani, K., Gupta, S., Farimani, A. B., & Reddy, C. K. (2024). Llm-sr: Scientific equation discovery via programming with large language models. *arXiv preprint arXiv:2404.18400*.

Sifringer, B., Lurkin, V., & Alahi, A. (2020). Enhancing discrete choice models with representation learning. *Transportation Research Part B: Methodological*, *140*, 236–261.

Trujillo, L., Muñoz, L., Galván-López, E., & Silva, S. (2016). neat genetic programming: Controlling bloat naturally. *Information Sciences*, *333*, 21–43.

Udrescu, S.-M., & Tegmark, M. (2020). Ai feynman: A physics-inspired method for symbolic regression. *Science advances*, *6*(16), eaay2631.

Van Cranenburgh, S., Wang, S., Vij, A., Pereira, F., & Walker, J. (2022). Choice modelling in the age of machine learning-discussion paper. *Journal of choice modelling*, *42*, 100340.

Virgolin, M., & Pissis, S. P. (2022). Symbolic regression is np-hard. *arXiv preprint arXiv:2207.01018*.

Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., ... others (2023). Scientific discovery in the age of artificial intelligence. *Nature*, *620*(7972), 47–60.

Zhang, H., Chen, Q., Xue, B., & Zhang, M. (2025). Llm-meta-sr: Learning to evolve selection operators for symbolic regression. *arXiv preprint arXiv:2505.18602*.