# Utilizing Large Language Models for Supporting Multi-Criteria Decision Modelling Method DEX

Marko Bohanec
Dept. of Knowledge Technologies
Jožef Stefan Institute
Ljubljana, Slovenia
marko.bohanec@ijs.si

Uroš Rajkovič
Faculty of Organizational Sciences
University of Maribor
Kranj, Slovenia
uros.rajkovic@um.si

Vladislav Rajkovič
Faculty of Organizational Sciences
University of Maribor
Kranj, Slovenia
vladislav.rajkovic@gmail.com

## Abstract

We experimentally assessed the capabilities of two mainstream artificial intelligence chatbots, ChatGPT and DeepSeek, to support the multi-criteria decision-making process. Specifically, we focused on using the method DEX (Decision EXpert) and investigated their performance in all stages of DEX model development and utilization. The results indicate that these tools may substantially contribute in the difficult stages of collecting and structuring decision criteria, and collecting data about decision alternatives. However, at the current stage of development, the support for the whole multi-criteria decision-making process is still lacking, mainly due to occasionally inconsistent and erroneous execution of methodological steps.

## Keywords

Multi-criteria decision-making, decision analysis, large language models, method DEX (Decision EXpert), structuring decision criteria

## 1 Introduction

Multi-criteria decision-making (MCDM) [1] is an established approach to support decision-making in situations where it is necessary to consider multiple interrelated, and possibly conflicting criteria, and select the best solution based on the available alternatives and the preferences of the decision-maker. Traditionally, such models are developed in collaboration with decision makers and domain experts, who define the criteria, acquire decision makers' preferences and formulate the corresponding evaluation rules. The model-development process is demanding, as it includes structuring the problem, formulating all the necessary model components (such as decision preferences or rules) for evaluating decision alternatives, and analyzing the results.

With the development and success of generative artificial intelligence, especially large language models (LLMs) [2], the question arises as to how these models can support or perhaps partially automate decision-making processes. To this end, we explored the capabilities of recent mainstream LLM-based chatbots, specifically ChatGPT and DeepSeek, for supporting the MCDM process. We specifically focused on using the method DEX (Decision EXpert) [3], with which we have extensive experience, spanning multiple decades [4], in the roles of decision makers, decision analysts, and teachers. DEX is a full-aggregation [5] multi-criteria decision modelling method, which proceeds by making an explicit decision model. DEX uses qualitative (symbolic) variables to represent decision criteria, and decision rules to represent decision makers' preferences. Variables (attributes) are structured hierarchically, representing the decomposition of the decision problem into smaller, easier to handle subproblems. Traditionally, DEX models are developed using software such as DEXiWin [6], which helps the user to interactively construct a DEX model and use it to evaluate and analyze decision alternatives.

The reported research is of exploratory nature. We ran ChatGPT and DeepSeek multiple times over the last six months, either individually, as a group or in classrooms with students. Typically, we first formulated some hypothetical decision problem and then guided the chatbot through the main stages of the MCDM process:

A. Model development stages:
   1. Acquiring criteria
   2. Definition of attributes (variables representing criteria)
   3. Structuring attributes
   4. Preference modeling (formulating decision rules)
B. Model utilization stages:
   5. Definition of decision alternatives
   6. Evaluation of alternatives
   7. Explaining the results of evaluation
   8. Analysis of alternatives

In doing this, we observed the responses generated by the LLMs and assessed them from the viewpoint of skilled decision analysts. The main goal was not to solve specific real-life decision problems, but to identify LLMs' strengths and weaknesses that may substantially affect the MCDM process.

Despite focusing on DEX, many of our findings are also applicable to other hierarchical full-aggregation MCDM methods [1, 5], such as AHP, MAUT/MAVT, and MACBETH, which follow the same methodological stages, with slight differences in the representation of model components.

In the following sections, we review the above-mentioned MCDM stages and describe our experience with each of them. Specifically, we illustrate the process with answers generated by ChatGPT-o3 and DeepSeek-V3. We considered a hypothetical personal decision problem of buying an electric-powered vehicle


*Information Society 2025, 6–10 October 2025, Ljubljana, Slovenia*

http://doi.org/10.70314/is.2025.skui.6589

(EV). The chatbots were run in parallel on June 6th, 2025, using similar prompts. Our assessments and comments are somewhat broader, based on some other use-cases, not presented here.

## 2 Acquiring Criteria

A MCDM model includes multiple criteria that capture essential aspects of decision alternatives. These criteria are used to evaluate and compare the alternatives in a structured way. Defining criteria usually requires a good knowledge of the decision problem and the decision maker's goals. This step is usually one of the most difficult and may require consulting domain experts and/or relevant literature.

We asked the chatbots: *I am considering buying a new EV vehicle for up to 35000 EUR in Slovenia. Suggest criteria for evaluating such cars.*

Both ChatGPT and DeepSeek came out with an extensive and structured list of criteria. Table 1 shows the criteria suggested by DeepSeek. ChatGPT's suggestions were similar, though they employed slightly different high-level categories, which also incorporated measurement units associated with each criterion.

**Table 1: EV cars: Criteria suggested by DeepSeek. Items 4–8 shortened due to space limitations.**

**1. Range & Battery Efficiency**
- **Real-world range** (WLTP or EPA estimates) – Aim for **at least 300 km** in mixed conditions.
- **Battery size** (kWh) – Larger batteries offer more range but may increase cost.
- **Energy consumption** (kWh/100 km) – Lower is better (e.g., 14–18 kWh/100 km).

**2. Charging Speed & Compatibility**
- **DC fast-charging capability** (e.g., 50 kW, 100 kW, or higher).
- **AC charging speed** (e.g., 7.4 kW, 11 kW, or 22 kW for home/work charging).
- **Charging port type** (CCS Combo is standard in Europe).

**3. Performance & Driving Experience**
- **Motor power** (kW/HP) – Affects acceleration and highway performance.
- **Drivetrain** (FWD, RWD, AWD) – AWD is better for Slovenia's varied weather.
- **Top speed & acceleration** (0–100 km/h time).

**4. Practicality & Usability:**
- **Cargo space, Seating capacity, Ground clearance**

**5. Price & Incentives**
- **Purchase price, Slovenian EV incentives, Insurance & maintenance costs**

**6. Technology & Features**
- **Infotainment system, Driver assistance, Heat pump availability**

**7. Availability & After-Sales Support**
- **Local dealership & service network, Warranty**

**8. Resale Value & Future-Proofing**
- **Brand reputation, Software updates**

Let us immediately say that we consider this the most important single contribution of LLMs to MCDM modeling. We are not aware of any previous method that would allow identifying and structuring decision criteria in such a depth and detail in literally just a minute. Of course, for "serious" applications getting such a list does not take the burden off the user, who is still responsible for verifying the suggestions and checking the criteria for relevance and correctness. Nevertheless, this is a valuable starting point that can save days or even weeks of work. This stage does not depend on the MCDM method used, so other methods may benefit from using LLMs equally well.

## 3 Definition of Attributes

In this stage, the task is to define variables, called attributes, that represent criteria in a MCDM model. As most MCDM methods use numeric attributes, this stage is specific to DEX, which uses qualitative attributes. Therefore, this and the following stages require LLMs to "understand" the method used. While DEX is less widely known than methods such as AHP or MAUT, it is nonetheless used and valued in various applications. Anyway, we were somewhat surprised to find out that all consulted LLMs were familiar with DEX and reasonably capable of following its main methodological steps. In some steps, however, we had to specify additional requirements to obtain proper DEX model components.

Generally, defining qualitative value scales of attributes was not too difficult for LLMs. Asking *Suggest preferentially ordered value scales* typically gives good suggestions for value scales, for example (DeepSeek):

**Purchase Price**: High (>€34k) → Medium (€30k–34k) → Low (<€30k)
**Government Incentives:** None → Moderate (€1k–3k) → High (>€3k)
**Insurance & Maintenance**: Expensive → Moderate → Cheap

Interestingly, in our first attempts at using LLMs, they were not fully aware of DEX specifics, such as that value scales should have a small number of values, preferentially ordered from "bad" to "good" [3]. After correcting LLMs a few times, they "remembered" and now suggest properly formulated value scales.

```
EV-SUITABILITY                          { excellent, good, adequate, poor }
├─ COST-OWNERSHIP                       { excellent, good, adequate, poor }
│  ├─ ECONOMY                           { very-low-cost, low-cost, medium-cost, high-cost }
│  │  ├─ Purchase-Cost                  { very-low, low, medium, high }
│  │  │  ├─ Net-Price-After-Subsidy     { ≤25 k, 25-30 k, 30-33 k, 33-35 k, >35 k }
│  │  │  └─ Registration-Fee %          { 0.5, 1, 1-2, >2 }
│  │  └─ Running-Cost                   { very-low, low, medium, high }
│  │     ├─ Electricity €/100 km        { ≤3, 3-4, 4-5, >5 }
│  │     └─ Service+Insurance €/yr      { ≤600, 600-800, 800-1 000, >1 000 }
│  ├─ DEPRECIATION-RISK                 { very-low, low, medium, high }
│  │  ├─ 5-yr-Residual-Value %          { ≥65, 60-64, 55-59, <55 }
│  │  └─ Brand-Resale-Index             { strong, above-avg, average, weak }
│  └─ DURABILITY                        { excellent, good, adequate, poor }
│     ├─ Vehicle-Warranty yrs           { ≥7, 5-6, 3-4, <3 }
│     ├─ Battery-Warranty yrs/%         { ≥8/70, 8/60, 5-7/60, <5/60 }
│     └─ Service-Interval km            { ≥30 000, 20-29 000, 15-19 000, <15 000 }
├─ CAPABILITY                           { excellent, good, adequate, poor }
│  ├─ RANGE-&-CHARGING                  { excellent, good, adequate, poor }
│  │  ├─ RANGE                          { very-high, high, medium, low, very-low }
│  │  │  └─ Real-Efficiency kWh/100     { ≤15, 15-17, 17-19, >19 }
│  │  └─ CHARGING                       { excellent, good, adequate, poor }
│  │     ├─ DC-Peak kW                  { ≥150, 100-149, 50-99, <50 }
│  │     ├─ 10-80 %-Time min            { ≤25, 26-35, 36-45, >45 }
│  │     └─ AC-On-board kW              { ≥11 3-ph, 11 1-ph, 7.4, <7 }
│  ├─ PRACTICALITY                      { excellent, good, adequate, poor }
│  │  ├─ Cabin-&-Cargo                  { excellent, good, adequate, poor }
│  │  │  ├─ Boot-Volume L               { ≥450, 400-449, 330-399, <330 }
│  │  │  └─ Rear-Legroom mm             { ≥900, 850-899, 800-849, <800 }
│  │  └─ Versatility                    { excellent, good, adequate, poor }
│  │     ├─ Towing-Capacity kg          { ≥1 000, 750-999, 500-749, <500 }
│  │     └─ Roof-Load kg                { ≥75, 50-74, <50 }
│  └─ PERFORMANCE                       { excellent, good, adequate, poor }
│     ├─ 0-100 km/h s                   { ≤7.5, 7.6-9.0, 9.1-11, >11 }
│     ├─ Top-Speed km/h                 { ≥180, 160-179, 140-159, <140 }
│     └─ Drive-Layout                   { AWD, RWD, FWD }
└─ RESPONSIBILITY                       { excellent, good, adequate, poor }
   ├─ SAFETY-&-TECH                     { excellent, good, adequate, poor }
   │  ├─ SAFETY                         { excellent, good, adequate, poor }
   │  │  ├─ Euro-NCAP-Stars             { 5★, 4★, 3★, ≤2★ }
   │  │  └─ ADAS-Completeness #         { ≥7, 5-6, 3-4, <3 }
   │  └─ TECH-FEATURES                  { excellent, good, adequate, poor }
   │     ├─ Heat-Pump                   { yes, optional, no }
   │     ├─ V2L-Power kW                { ≥3.6, 1.5-3.5, none }
   │     └─ OTA+Nav-Chargers            { full, partial, none }
   └─ SUSTAINABILITY                    { excellent, good, adequate, poor }
      ├─ Battery-Origin-Transparency    { EU+recycle, non-EU+recycle, non-EU, unknown }
      ├─ Recycled-Materials %           { ≥25, 15-24, 5-14, <5 }
      └─ Life-cycle CO₂ t               { ≤15, 16-18, 19-22, >22 }
```

**Figure 1: DEX model structure generated by ChatGPT.**

## 4 Structuring Attributes

After acquiring the criteria and defining attributes, the next step is to structure attributes into a hierarchy (most often an ordinary tree). Following the previous stages, which already resulted in a well-organized criteria structure and proper definition of attributes, this stage looks quite easy for LLMs. Figure 1 shows a full DEX model structure as suggested by ChatGPT without any further instructions. In comparison, the DeepSeek's suggestion (not shown here) was somewhat inferior. Also, it was generated only after we had instructed it that DEX requires "narrow" trees with only two to three descendants for each aggregate attribute. Anyway, we consider both structures appropriate and comparable to the achievements of the best university-level students.

## 5 Decision Preferences and Decision Rules

In DEX, decision maker's preferences are represented by decision rules, organized in decision tables. For each aggregate attribute, the user is asked to investigate all combinations of lower-level attribute values and assess the corresponding outcomes. Depending on the number of aggregate attributes, this might be a laborious task, but can be usually carried out using software such DEXiWin with not too much hassle. LLMs are also capable of suggesting perfectly valid decision tables, as illustrated in Table 2, which suggest the values of *Purchase-Cost* depending on *Net-Price-After-Subsidy* and *Registration-Fee*.

**Table 2: Decision table for Purchase-Cost (ChatGPT).**

| **Net-Price-After-Subsidy ▼ / Registration-Fee ►** | **very-low (0 .5 %)** | **low (1 %)** | **medium (1–2 %)** | **high ( > 2 %)** |
|---|---|---|---|---|
| **very-low** (≤ 25 k €) | very-low | low | low | medium |
| **low** (25–30 k €) | low | low | medium | medium |
| **medium** (30–33 k €) | medium | medium | medium | high |
| **high** (33–35 k €) | high | high | high | high |

From the DEX perspective, it is important to remark that Table 2 is complete (addressing all possible combinations of input values) and preferentially consistent (increasing input values result in increasing outputs). Initially, these requirements were not obvious to LLMs, and we had to request them explicitly. After further use, LLMs now generate appropriate rules by themselves.

However, it is important to emphasize an essential issue. Decision making is considered a subjective process, and decisions should reflect individual decision-makers' preferences. Using LLMs, we only get preferences, generalized from many documents. In order to impose our own preferences, we should (1) carefully check LLMs' suggestions and (2) request modifications, if necessary. According to our experience, this is possible by formulating prompts such as "*increase the importance of* some attribute" or "*alternatives exceeding* some value *of* some attribute *are unacceptable*". Nevertheless, this requires focus, a good understanding of the goals, and perpetual verification of the results.

## 6 Identification of Decision Alternatives

The goal of this stage is to identify decision alternatives and describe them in terms of values, assigned to input attributes of the model. This is another stage in which LLMs excel due to their ability to act like agents: autonomously collecting relevant data from external sources. In this case, the LLMs constructed lists of actual EV cars, consulting sources such as avto.net, citroen.si, reuters.com, and ev-database.org:

**MG 4 Comfort 64** kWh – €31 490 → €24 290 after Eko-sklad
**Citroën ë-C4** 54 kWh – €31 900 → €24 700
**Opel Corsa Electric** 54 kWh – from €28 490 → €21 290
**BYD Dolphin Surf Boost** 43 kWh – €22 990 → €15 790*
**Fiat 500e** 42 kWh – promo €22 290 (price already after bonus)

It is important to understand that alternatives should be defined after developing the model. This allows the LLM to search for and collect the data needed to evaluate the alternatives.

## 7 Evaluation of Alternatives

In this stage, the five cars have been evaluated by the LLMs. All were evaluated as GOOD, except Fiat 500e, which was evaluated as ADEQUATE. Table 3 illustrates the output generated by ChatGPT, showing the evaluation results of the two best-ranked cars. In addition to the overall evaluation (EV-SUITABILITY), the output includes evaluations achieved at the three main branches of the model, together with a brief notification of reasons for such evaluations. This is an appropriate summary of evaluation results.

**Table 3: Evaluation of the two best-ranked EV cars (ChatGPT).**

| **Model** | **MG 4 Comfort 64** | | **Citroën ë-C4** | |
|---|---|---|---|---|
| **Cost-Ownership** | **good** | (low purchase-cost; running-cost low) | **good** | (very-low purchase-cost & running-cost) |
| **Capability** | **good** | (435 km WLTP, 140 kW DC, 363 L boot) | **adequate–good** | (416 km WLTP, 100 kW DC, 380 L boot) |
| **Responsibility** | **good** | (★ 5 Euro-NCAP, heat-pump optional, V2L 3.6 kW) | **good** | (★ 4 Euro-NCAP, heat-pump std., forthcoming V2L) |
| **EV-SUITABILITY** | **GOOD** | | **GOOD** | |
| **Rank** | 1 | | 2 | |

## 8 Explaining the Results of Evaluation

In MCDM, it is essential to understand the evaluation method and explain the reasons underlying evaluation results. This is how ChatGPT explains the evaluation method:

1. Leaf values were taken from the most recent Slovene price-lists or Europe-wide tech sheets (all ≤ €35 000 trims).
2. Running-costs use your 30 000 km/year mileage and the Q1-2025 average household tariff of €0.166 kWh-1.
3. Each quantitative item was mapped to the ordered value-scales we set earlier; the three rule-tables ("purchase-cost", "running-cost", "economy") were then fired, and all higher nodes were combined with simple equal weights (0.4 Cost-Ownership | 0.4 Capability | 0.2 Responsibility).

4. Scores were coded excellent = 4, good = 3, adequate = 2, poor = 1; the root class was chosen by rounding the weighted mean to the nearest integer.

At first, everything seems to make sense. However, careful reading reveals the disturbing fact that different decision rules from those agreed upon in the previous stages were used to evaluate alternatives. Unfortunately, this often happens with LLMs, which tend to "forget" about the previous MCDM stages. It is not uncommon that attributes, their value scales and decision rules change from prompt to prompt. This severely undermines the trust in using LLMs and makes the whole process uneasy: rather than focusing on solving the decision problem, the user is forced to meticulously check each and every step. Also, it is not uncommon to discover logical errors or even basic computational errors (often referred to as "hallucinations" [7]). In one of our sessions with ChatGPT, it displayed the evaluation formula

$$(0.2\times3)+(0.25\times4)+(0.15\times4)+(0.2\times3)+(0.15\times2)+(0.05\times2)=3.15$$

which looked convincing, but gave a hard-to-notice, but wrong result; the correct result is 3.2.

## 9 Analysis of Alternatives

The last stage of the MCDM process is the analysis of alternatives, which is aimed at exploring the decision space using methods such as what-if and sensitivity analysis. Without providing experimental evidence due to space restrictions, we can say that, in principle, LLMs are capable of performing such analyses, giving appropriate answers and explanations to questions such as:

- *Carry out sensitivity analysis for Citroën ë-C3 and MG4 depending on buying price and operating costs.*
- *What would have to change for Fiat 500e 42 to become a good EV vehicle?*

In most cases, results are correct and informative, particularly in cases when an explicit explanation is requested by the user. However, the issues of using inappropriate model components and making logical and computational errors were detected in this stage as well.

## 10 Discussion

LLMs are developing rapidly and becoming increasingly capable. They may evolve under the hood, so that even the same version can behave differently depending on recent updates or user-specific factors. This makes them challenging for conducting a rigorous scientific research. They come without user manuals, requiring their users to explore their capabilities on their own. This study is an experimental attempt to understanding the capabilities of the current (2025) mainstream LLMs for supporting the MCDM process, with special emphasis on the DEX method. On this basis, we could not formulate firm conclusions, but were still able to make observations and formulate recommendations that might help MCDM practitioners.

The single most important contribution of LLMs to MCDM is their ability to formulate a well-structured list of relevant criteria in the first stage (section 2). Nothing nearly as good was available so far for that difficult stage, where LLMs can now substantially boost the process and save a lot of effort and time. The second important contribution is the capability of LLMs to act as agents and collect data about alternatives (section 6) from various external resources.

Considering individual MCDM stages, LLMs performance is quite impressive. They are capable of evaluating and analyzing alternatives, without much instruction. Furthermore, if asked, they can explain the used methods and obtained results quite well. In some cases, however, a seemingly convincing explanation may fall apart, revealing logical and computational errors.

Considering the MCDM process as a whole, the performance of LLMs is not as favorable. In subsequent MCDM stages, LLMs tend to "change their mind" without notice, modifying the already established model components: attributes, value scales, and decision rules. Consequently, this requires a lot of attention from the user's side, who has to check the outputs and perpetually remind the LLMs to remain consistent. This distracts the process and often carries the user away of the main decision-making task. Also, we should warn that in the preference modelling stage (section 5), LLMs suggest generalized decision preferences that might substantially differ from the user's subjective preferences, which need to be enforced explicitly.

In summary, LLMs can substantially contribute to the definition of attributes and alternatives, but are unsuitable for carrying out the whole MCDM process due to possible inconsistent and erroneous executions of the MCDM method. We believe that, given the current state of LLM development, it is more convenient and safer to use specialized and trusted MCDM software, such as DEXiWin. Nevertheless, LLMs evolve fast and we may expect substantial improvements in the future.

## Acknowledgments

The authors acknowledge the financial support from the Slovenian Research and Innovation Agency for the programme Knowledge Technologies (research core funding No. P2-0103 and P5-0018).

## References

[1] Kulkarni, A.J. (Ed.), 2022: Multiple Criteria Decision Making. Studies in Systems, Decision and Control 407, Singapore: Springer, doi: 10.1007/978-981-16-7414-3_3.

[2] Kamath, U., Keenan, K., Somers, G., Sorenson, S., 2024: *Large Language Models: A Deep Dive: Bridging Theory and Practice.* Springer, 506p, ISBN-13 978-3031656460.

[3] Bohanec, M., 2022: DEX (Decision EXpert): A qualitative hierarchical multi-criteria method. In: *Multiple Criteria Decision Making* (ed. Kulkarni, A.J.), Studies in Systems, Decision and Control 407, Singapore: Springer, doi: 10.1007/978-981-16-7414-3_3, 39–78.

[4] Bohanec, M., Rajkovič, V., Bratko, I., Zupan, B., Žnidaršič, M., 2013: DEX methodology: Three decades of qualitative multi-attribute modelling. *Informatica* 37, 49–54.

[5] Ishizaka, A., Nemery, P., 2013: Multi-criteria decision analysis: Methods and software. Chichester: Wiley.

[6] Bohanec, M., 2024: *DEXiWin: DEX Decision Modeling Software, User's Manual, Version 1.2*. Ljubljana: Institut Jožef Stefan, Delovno poročilo IJS DP-14747. Accessible from https://dex.ijs.si/dexisuite/dexiwin.html.

[7] Banerjee, S., Agarwal, A., Singla, S., 2024. LLMs will always hallucinate, and we need to live with this. 10.48550/arXiv.2409.05746.

Zbornik 28. mednarodne multikonference

# INFORMACIJSKA DRUŽBA – IS 2025

Zvezek A

Proceedings of the 28th International Multiconference

# INFORMATION SOCIETY – IS 2025

Volume A

## Slovenska konferenca o umetni inteligenci
## Slovenian Conference on Artificial Intelligence

Uredniki / Editors

Mitja Luštrek, Matjaž Gams, Rok Piltaver

http://is.ijs.si

**8.–9. oktober 2025 / 8–9 October 2025**
**Ljubljana, Slovenia**

Uredniki:

Mitja Luštrek
Odsek za inteligentne sisteme, Institut »Jožef Stefan«, Ljubljana

Matjaž Gams
Odsek za inteligentne sisteme, Institut »Jožef Stefan«, Ljubljana

Rok Piltaver
Outfit7, Ljubljana

Založnik: Institut »Jožef Stefan«, Ljubljana
Priprava zbornika: Mitja Lasič, Vesna Lasič, Lana Zemljak
Oblikovanje naslovnice: Vesna Lasič, uporabljena slika iz Pixabay

Dostop do e-publikacije:
http://library.ijs.si/Stacks/Proceedings/InformationSociety

Ljubljana, oktober 2025

Informacijska družba
ISSN 2630-371X
DOI: https://doi.org/10.70314/is.2025.skui

Kataložni zapis o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani
**COBISS.SI**-ID **255435267**
ISBN 978-961-264-319-5 (PDF)