UDC 004.8
DOI: 10.15587/2706-5448.2025.333726

Oleh Lukianykhin,
Vira Shendryk

USE OF GENERATIVE ARTIFICIAL INTELLIGENCE TO IMPROVE OUTPUT MESSAGE EFFECTIVENESS IN DECISION SUPPORT SYSTEMS FOR PROSUMERS

The object of this study is the use of generative artificial intelligence (GenAI) to create output messages in a decision support system (DSS) for prosumers. The research addresses the challenge of improving user experience (UX) by enhancing the effectiveness of DSS messages. A prototype DSS was developed for a specific private household equipped with solar panels. A rule-based message generation system was created as a baseline for comparison. An evaluation was conducted through surveys in Ukrainian and English. GenAI models from OpenAI and Anthropic were compared. Messages were assessed along two key dimensions of UX quality: usefulness and ease of comprehension.

The results indicate that GenAI can enhance the effectiveness of DSS recommendations for specific user groups without adverse effects. The Sonnet 3.5 model (Anthropic) generated messages that were rated as statistically more useful ($p < 0.05$) by female users in Ukrainian. Users preferred shorter messages in English, and Sonnet 3.5 outperformed GPT-4 (OpenAI) in terms of usefulness in both languages ($p < 0.05$).

The higher usefulness ratings can be attributed to more detailed recommendations while maintaining natural language. The English-language results were likely influenced by the fact that respondents were not native speakers. Differences between the models are associated with the specifics of their integration into the DSS.

The results prove the hypothesis that GenAI can improve the efficiency of DSSs by generating more useful but not more complex messages. These results also indicate that GenAI's main advantage is in tailoring the DSS output to the needs of different user groups. The difference in results between the models highlights the need for proper testing of the developed AI solutions in specific contexts. The results will be used to develop a more efficient DSS for electricity prosumers.

***Keywords**: generative artificial intelligence, decision support system, prosumers, user experience, photovoltaics.*

Received: 02.04.2025
Received in revised form: 02.06.2025
Accepted: 23.06.2025
Published: 29.08.2025

How to cite
Lukianykhin, O., Shendryk, V. (2025). Use of generative artificial intelligence to improve output message effectiveness in decision support systems for prosumers. Technology Audit and Production Reserves, 4 (2 (84)), 13–23. https://doi.org/10.15587/2706-5448.2025.333726

1. Introduction

Contemporary power systems (PS) have an increasing number of components, and the complexity of their structure is growing. The contributing factors include smart grid participants, private households with generation capacity (prosumers), and the integration of microgrids [1–3]. This trend increases the amount of automated and manual decisions that have to be made in the grid at every point in time to manage different parts of the system [1, 4]. While automated solutions can often scale and rise to the challenge, human operators at different levels struggle with making so many management decisions [5]. Moreover, some control processes can be automated on a macro level of the grid, but they face additional challenges when adapting to the micro level. For instance, users may doubt fully automated AI-based decisions because of a lack of trust, transparency, and explainability [6, 7]. Because of that, some fully automated solutions are adapted as semi-automated or manual decision-making at the micro level of the grid. These factors increase the need for decision support systems (DSSs) in PS [1, 5, 8–10].

The main challenges human operators face in decision-making are the amount of information and cognitive load required to make each decision. The number of decisions to make and the significant amount of available information contribute to the first component and make the operator's workload higher [11, 12]. The cognitive load on a human operator reduces efficiency, as even when a reasonable amount of data is provided, the operator should often do some intermediate calculations or reasoning and find insights in the data [13, 14]. These factors reduce the number of decisions one operator can make at a desired quality without additional support [15, 16].

An efficient DSS targets all these challenges. First, it gathers, processes, enhances, and structures the information to present it to a user. This helps to avoid information overload [15, 17]. Second, it reduces the cognitive load on the operator by preparing intermediate calculations and extracting patterns and insights from the data volumes [17–21]. Third, DSS provides better transparency into all steps of information processing and thus can increase trust and allow for broader adoption of semi-automated solutions [6, 22], for example, AI-based solutions in prosumers. However, the described benefits depend on how well the information is delivered to an operator, for example, if correct information was chosen to be presented and if it was in an optimal format [17–19]. Moreover, an optimal format depends on a specific domain and application [20–24].

Therefore, it is a potent research direction to focus on improving the efficiency of how DSS delivers information to human operators, specifically in prosumer's decision-making, at the micro level of a power grid where many decisions are made.

As mentioned before, there is a naturally growing need for DSS in the PS domain. Many applications exist, but solution design often resorts to manually constructing the format of the DSS outputs. For example, in [10] authors developed a DSS for managing and optimizing the flexibility of a power grid enhanced with renewable energy sources and energy storage. In [8], a DSS is applied to support wind energy generation by providing forecasts to the operator [9] demonstrates that DSS can be employed to identify urban areas with the potential for improving energy efficiency and renewable energy sources. In [25], the authors present a DSS that helps analyze the feasibility of renewable energy developments in England.

In all these examples, no primary focus was on the output format of the DSS, only on its task-solving effectiveness. Moreover, the related literature review revealed that a significant proportion of the DSS studies in the PS domain focus on centralized, macro-level operations, while micro-level subjects and DSS for individual prosumers are comparatively underrepresented. Some examples of DSS specifically for prosumers include [26], where authors suggested a DSS for prosumer communities, and [27], where a rule-based DSS was introduced to reduce the electricity cost for a microgrid with renewable energy sources. Also, in [28], the authors introduce the DSS, which minimizes electricity costs while maintaining comfort in terms of temperature. As outlined in [29], DSS for prosumers can enable complex mechanisms, like a reaction to dynamic price changes, energy exchange with other facilities, supplier change, etc. Such actions require complex decision-making, while prosumers are often unable or unwilling to do that. It's common for the operators of prosumer facilities to rely on intuition and resort to common behavior patterns based on experience. This doesn't always lead to optimal decisions and may even lead to errors due to human factors, highlighting the importance and potential impact of the DSSs for prosumers.

Although the optimality of the output format of a DSS is not a common subject for research, specifically in PS applications, there are some attempts to explore this research direction. For example, combining human-computer interfaces research with the PS domain in [11] leads to the development of a framework that aims to improve the efficiency of AI assistants by applying vision aids and bidirectional interaction. Authors in [16] focused on providing recommendations for redesigning control rooms for better operator efficiency in centralized power grid operators. The recommendations were constructed by investigating operators' behavior and aiming to reduce cognitive load. [12] considered the operators' efficiency in nuclear power plants, in particular, based on the amount of information provided during the situation diagnosis process. It was demonstrated that an excessive amount of information impairs the decision-making process. Nevertheless, such studies on the efficient format of the DSS output are quite rare in the PS domain and mostly focus on centralized grid operation, not microgrid level.

At the same time, this aspect is studied much more in other domains where DSSs are successfully applied – both in specific use cases and in identifying general patterns and dependencies. In a significant amount of studies, researchers aim to improve the efficiency of the output format of a specific DSS. For instance, [18] presents exploratory research trying to improve the efficiency of the user interface (UI) of a DSS by adopting augmented reality technologies. However, other studies also aim to find general dependencies and provide guidance for future researchers and applied solutions developers. For example, in [20], authors try to find the theoretical underpinnings of DSS for efficient cognition. This provides an essential contribution to relevant research directions, such as developing better support for cognition with DSSs. In [17], the authors offer a comprehensive review of DSSs in different domains and analyze, among other things, the aspect of efficient output delivery to a DSS user. [30] provides another example of such general dependency: authors demonstrate that there seems to be a trade-off between the increased situation awareness of an operator and the task performance. They observed that the increased amount of information provided to a user increased situation awareness significantly but negatively impacted the downstream task performance. These findings highlight the need to develop a DSS with a user interface fine-tuned for a specific task.

The number of studies focused on efficient information delivery to the user is extensive, and it is not possible to cover all of them here. However, it demonstrates the importance of optimizing the DSS interaction with a user. Moreover, most of these works highlight the need to research optimal output formats for specific domains and applications, as significant differences exist among them.

Therefore, the review continued with the search for effective methods to improve user experience (UX) in DSSs with optimal output. GenAI approaches have attracted attention, as in recent years, they – especially those based on large language models (LLMs) – have become increasingly popular and widely adopted, as demonstrated in [31]. Naturally, they were also applied in various domains to enhance decision-making processes. For example, [32] looks into how users accept recommendations when collaborating with GenAI in decision-making, particularly in dimensions of trust and perceived benefits. In [33], authors focused on the medical domain and explored potential GenAI applications there, including the impact of LLMs application on decision-making in medical applications. In particular, there were some attempts to improve the UX and efficiency of DSS using LLMs. For example, in [34], techniques to develop adaptive UX based on the LLMs. Common ideas among different studies include attempts to influence how useful the information is to a user and decrease the complexity of comprehension, similar to non-LLM powered DSS [35].

At the same time, there is a very limited number of studies using GenAI in DSS in PS. For example, in [36], authors suggest using an LLM as a core component of the DSS for the control rooms. However, this technical report doesn't focus specifically on the output format or other grid participants, like prosumers.

Based on related research works, it is evident that there is a research gap in the application of GenAI in DSS in PS with a focus on improving the DSS output format and DSS efficiency of information delivery to a human operator, especially for applications in prosumers.

In this study, let's focus on transforming technical suggestions of a control module of a DSS into a human-readable format, i. e., natural language, using Generative AI. An intelligent DSS that gathers and processes data of the prosumer facility usually has two essential steps: generating a control suggestion (CS) from the technical standpoint and converting that into a human-readable format.

More formally, given a set of control suggestions S for the time period t_i and data d_i available to the system, a natural language transformation NL is applied to generate a human-readable message m_i

$$NL(d_i;t_i;S_i) \rightarrow m_i. \quad (1)$$

This is a 2-step process, where the control suggestion S_i is generated by a control suggestion function CS based on d_i and t_i

$$CS(d_i;t_i) \rightarrow S_i. \quad (2)$$

So, the whole process can be written as

$$NL(d_i;t_i;CS(d_i;t_i)) \rightarrow m_i. \quad (3)$$

It is possible to suppose the control suggestion (CS) is the same for several NL functions. In that case, it is possible to evaluate and compare different NL functions directly by comparing their outputs, as any extra variance of different control suggestions is excluded.

Focusing on the human-facing interface and considering recent advances in the GenAI approaches, it is possible to formulate the main research hypothesis: utilization of the GenAI can improve the user experience of a DSS for prosumers by increasing the efficiency of converting technical output from a DSS's control suggestion module (CSM) into a human-readable message.

In particular, related work indicates the opportunities for increasing the usefulness of the messages and making it easy for humans to read them. Hence, the following research questions were formulated:

1. How can GenAI be utilized in a prosumer DSS to generate output messages efficiently based on CSM output?
2. How will it change the difficulty of perceiving the message?
3. How will it change how useful the message feels to the user?
4. What are the differences between such GenAI applications in the Ukrainian and English target languages?

That is, *the aim of research* is to increase the efficiency of output messages of a prosumer DSS with GenAI. Therefore, the first step is to develop a DSS with several NL transformations integrated as natural language modules (NLM) of the DSS, including GenAI-based ones. The second step is to compare different NLMs by measuring the difficulty of perception and informativeness of their outputs via human evaluation.

2. Materials and Methods

2.1. Object and design of research

The object of this research is the use of generative artificial intelligence (GenAI) to create output messages in a decision support system (DSS) for prosumers. To achieve the goals of this study, DSS has to be developed with different NL transformations integrated into it as natural language modules (NLMs). Let's aim to compare rule-based and GenAI-based approaches. A specific practical use case was considered to evaluate different approaches to creating an output message of a DSS.

2.2. Data

This study is conducted based on prosumer data from a private household with photovoltaic installations, where data on generation and consumption was gathered for several years. It has daily information on the generated electricity amount, amounts of energy used from a battery and bought from the grid, and temperature at 6:00 and 18:00 on the given day. Due to data particularities, such as missing values, data from 01-01-2020 to 10-11-2023 was utilized for experiments in this study. The data in detail was discussed in [37], so only essential particularities and preprocessing steps are listed here:

– Data has strong yearly seasonality.
– Outliers were corrected using the IQR approach, and the resulting minimum and maximum clipping values for the generation column are 0 and 14.
– Missing values in generation and consumption from battery columns were filled using a median value.
– Negative (corrupted) values of power used from the battery were replaced with a median value.
– The power used from the battery column was converted to kWh for consistency with other columns.
– Several corrupted data points were fixed based on the domain knowledge, e. g., setting consumption from the grid and from the battery to 0 on 31/07/2023.

Date-related features were added to enhance the modeling and decision-making process: day of the week, weekend flag, day of the month, month, sin and cos of the day of the year. A weather-related feature was added as well: temperature change during the day.

As complete household consumption was not available in the data, an estimation was performed using Formula (4) to model the household behavior required for developing and evaluating a DSS. The scaling factor for a generation was based on the industry average for prosumers with energy storage systems [38]. This assumption was necessary due to uncertainty regarding the proportion of generated energy consumed by the household. In particular, the following formula was used

$$HC = 0.75 \cdot G + BU + GS, \tag{4}$$

where HC – the estimated household consumption, G – the generated energy, BU – the energy used from batteries, and GS – the energy supplied from the main grid.

Outliers for the calculated HC values were corrected using the IQR approach. The resulting minimum and maximum clipping values are 0 and 13, respectively. Missing values were filled using a median value.

2.3. Decision support system

DSS for a prosumer focuses on providing relevant recommendations for operating the prosumer facility in the next time period. It should consider all available information and generate a human-understandable output presented to an operator. The operator should review the human-readable output before they decide what to do tomorrow.

The DSS designed for this study relies on the estimates of generation and consumption for tomorrow, as well as relevant information based on historical and current data. The designed system (Fig. 1) consists of:

1. *Forecasting modules* – responsible for estimating generation and consumption for tomorrow with ML models.
2. *Control suggestion module (CSM)* – responsible for suggesting a relevant course of action for the next time period based on the available inputs.
3. *Natural language module (NLM)* – responsible for converting the technical representation of DSS suggestions into a human-readable format.

All modules of the system are discussed in detail further in this section.

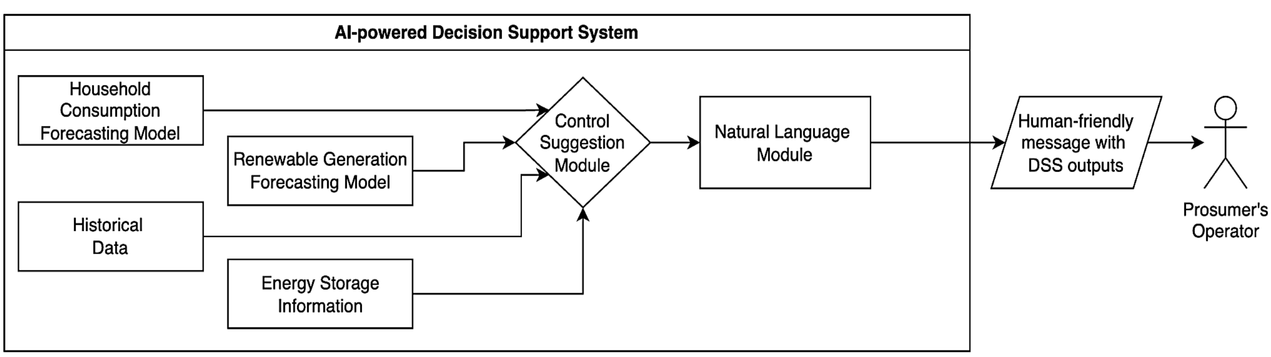


Fig. 1. Information flow diagram of the developed DSS

2.4. Forecasting models

Estimates of consumption and generation are required to provide the required information for decision-making to CSM and operators. To this end, forecasting models were built using ML approaches. LightGBM regressor was chosen due to its significant capacity, which allows learning complex patterns in the data while keeping models lightweight and allowing for quick fine-tuning.

Both generation and forecasting models were trained using a similar approach discussed in detail in [37]. The approach was as follows:

1. Select target – generation or HC.
2. Conduct final data preprocessing for a model:

a) Fill missing values with quadratic interpolation for any numeric feature column if necessary.

b) Add lags – values of the target variable in specified past steps, as features for the given data point. Lags of 1–7, and 14 were used.

3. Use classic train-test split to have hold-out test data for final evaluation. 5% chronologically latest data was used for the test.
4. Specify hyperparameter ranges (Table 1).
5. Tune for 2000 iterations of the hyperparameters tuning process using RandomSearch.
6. In each tuning iteration, use time-series-specific cross-validation (TimeSeriesSplit, [39]) with four folds to get more robust model performance estimates during tuning and avoid data leakage from the future.

The final best model hyperparameters for both models and hyperparameter ranges used for tuning are given in Table 1. All values are rounded to four decimal places for clarity.

Table 1

Hyperparameters of final LightGBM models

Hyperparameter	Range for tuning	Generation model	HC model
n_estimators	[20, 1000]	470	652
learning_rate	[0.01, 0.3]	0.0419	0.0176
num_leaves	[20, 200]	121	99
max_depth	[–1, 15]	1	1
min_child_samples	[5, 50]	48	8
subsample	[0.7, 0.3]	0.9975	0.8397
colsample_bytree	[0.7, 0.3]	0.8036	0.8103
reg_alpha	[0, 1.0]	0.2695	0.8849
reg_lambda	[0, 1.0]	0.2427	0.6995

2.5. Control suggestion module

The control suggestion module (CSM) is responsible for generating suggestions on how to operate the prosumer facility in the next time period. In the considered use case, the next time period is one day, i. e., recommendations are made for tomorrow.

The recommendations focus on three aspects: energy-intensive activities, usual daily activities, and battery usage. Each aspect has a defined set of possible actions, that are as follows:

– Daily activities: Keep them as usual or decrease them.

– Energy-intensive activities: keep them as usual, postpone them to other days (decrease), or schedule them for tomorrow (increase).

Battery: consume energy from the batteries, charge it, or preserve energy *for later usage.*

As this study focuses on the efficiency of the final messages, designing the optimal CSM was out of scope. As long as the input to the next module (NLM) is the same and reasonable, the comparison of NLMs is valid and sufficient for this study. To make the comparison more illustrative, five common prosumer scenarios were identified. Suggestions for these scenarios were formulated based on the domain knowledge and standards of the industry.

The scenarios are as follows:

1. Generation (G) is greater than Consumption (C), and Battery (B) is empty.
2. G is greater than C, and B is fully charged.
3. G is equal to C, and B is fully charged.
4. G is less than C, and B is empty.
5. G is less than C, and B has enough energy to cover the difference.

The algorithm of the CSM is based on estimating the battery charge and comparing estimates of upcoming generation and consumption. The battery charge is estimated using historical data for the last 7 days before tomorrow:

1. Start with zero charge – pessimistic assumption in terms of battery, should balance other optimistic assumptions.
2. For each day, calculate the difference between generation and consumption:

a) if $G - C > 0$ – add it to the charge;

b) if $G - C <= 0$ – subtract it from the battery, but if the resulting charge is negative set it to zero;

c) if the battery charge exceeds the maximum battery capacity – set it to battery capacity.

The algorithm can be presented using formula

$$\begin{aligned} & B=0 \\ & \text{for } i \text{ from 7 to 1:} \\ & \quad B=B+(G_{-i}-C_{-i}), \\ & \quad \text{if } B<0: \\ & \qquad B=0; \\ & \quad \text{if } B>B_{\max}: \\ & \qquad B=B_{\max}. \end{aligned} \tag{5}$$

Then, generation and consumption estimates are compared with the battery charge, and if one of the five scenarios is identified – the corresponding suggestion is given. Otherwise, no specific suggestion is given. The proposed suggestions for each of the five scenarios are presented in Table 2. In addition, specific examples of the scenarios from the historical data were chosen, one per scenario, to evaluate the NL modules (Table 2).

Table 2

Five common prosumer scenarios: date and proposed CSM suggestions

Scenario	Date	Daily	Energy intensive	Battery
Scenario 1	23/05/2022	keep	keep	charge
Scenario 2	11/05/2020	keep	increase	preserve
Scenario 3	12/09/2020	keep	keep	consume
Scenario 4	30/10/2023	decrease	decrease	preserve
Scenario 5	10/05/2022	keep	decrease	consume

The program implementation is parametrized with maximum battery capacity, the threshold for considering the battery empty or full, and generation and consumption equal. Hence, the approach is designed to be flexible and can be adapted to different setups while being justified by domain knowledge. This makes developed CSM relevant for various applications despite the straightforward decision-making logic. To accommodate for human-friendly output creation by an NLM, CSM provides its output in two main parts:

1. Dictionary with control suggestions, 1 per each aspect, e. g.:

```
{
  "battery_decision": "charge_battery",
  "enrg_intns_actvt_decision": "enint_keep_as_usual",
  "daily_consumption_decision": "daily_keep_as_usual"
}
```

2. Dictionary with the information directly used to provide the suggestions, as demonstrated in the following example (all values are rounded to four decimal places for clarity):

{
"generation_prediction": 5.8729,
"household_consumption_prediction": 5.4218,
"battery_estimate": 0,
"all_week_batteries": [0, 0, 0, 0, 0, 0.077, 0, 0],
"max_battery_capacity": 2.5
}

2.6. Natural language module

The Natural Language Module is a module that integrates natural language transformations (rule-based and GenAI-based) that are being evaluated in this study. All other parts of the DSS remain the same while NLM is being changed, and its outputs are evaluated to answer research questions set in Section 1. The general design is as follows:

1. Information gathered by previous modules is taken as input. The inputs include:

a) CSM suggestions.

b) Information that the CSM used to generate the suggestions.

c) Data gathered for tomorrow, including estimates of generation and consumption and features used by the forecasting module to generate these estimates.

2. Parse the inputs and generate a human-readable message.

The ultimate goal of this module is to convert a CSM output from a technical format, which may or may not be readable by the non-specialist audience, to something that can convey the meaning efficiently using a natural language. Given the assumption of having efficient control suggestions, the goal of the message is to maximize the probability of the operator complying with it.

2.6.1. Rule-based suggestion message generation

Rule-based message generation works in a straightforward but meaningful way to provide a baseline for comparison with GenAI-based alternatives. In the rule-based NLM (RB-NLM), the generated message structure is predefined:

1. It starts with a header with a date, e. g., "*Suggestion for tomorrow (Friday, May 01):*".

2. The main part provides recommendations in a natural language, one sentence per aspect of recommendations. Messages are predefined for each control action (Table 3), e. g., if a suggestion is that batteries will be charged tomorrow, a sentence stating *"Extra energy will be used to charge the batteries."* will be added to a generated output message.

3. At the end, a more technical part is added, which explains why these particular suggestions are made by providing values for expected generation, consumption, current battery capacity, and maximum battery capacity, e. g., *"This suggestion is based on the following information: expected generation: 4.96 kWh, expected consumption: 4.78 kWh, battery reserve: 0.00 kWh, maximum battery capacity: 2.50 kWh."*

An overview of messages used for each action and aspect is given in Table 3.

An output message example produced with an RB NLM for 01/05/2020 is as follows:

"Suggestion for tomorrow (Friday, May 01):

There is no need to limit usual daily electricity consumption.
There is no need to avoid energy intensive activities tomorrow.
Extra energy will be used to charge the batteries.

This suggestion is based on the following information:
expected generation: 4.96 kWh, expected consumption: 4.78 kWh,
battery reserve: 0.00 kWh, maximum battery capacity: 2.50 kWh."

Table 3

Message used in an RB NLM for each CSM suggestion

CSM Suggestion	English	Ukrainian
Battery – charge	Extra energy will be used to charge the batteries	Надлишок згенерованої енергії буде використано для заряджання батарей
Battery – consume	Energy from the batteries will be used to meet demand	Буде використано енергію з батарей для задоволення потреб
Battery – preserve	Energy from the batteries will not be used tomorrow	Енергія з батарей не буде використовуватись
Energy Intensive – increase	Consider scheduling more energy-intensive activities for tomorrow	Варто запланувати енергоємні активності на завтра
Energy Intensive – decrease	Consider postponing energy-intensive activities to another day	Варто відкласти енергоємні активності на пізніше
Energy Intensive – keep	There is no need to avoid energy-intensive activities tomorrow	Немає потреби уникати енергоємних активностей
Daily – keep	There is no need to limit usual daily electricity consumption	Немає потреби обмежувати звичне щоденне споживання електроенергії
Daily – decrease	Reduce usual electricity consumption. Please try to limit the use of household appliances	Потрібно зменшити звичне споживання електроенергії. Будь ласка, намагайтесь обмежити щоденне використання електроприладів

2.6.2. GenAI-powered suggestion message generation

GenAI approaches were applied to improve the efficiency of message generation. In this study, models by OpenAI and Anthropic were utilized: GPT-4 [40] and Sonnet 3.5 [41], respectively. The inputs of the process are the same as for the rule-based NLM, but there is no deterministically predefined structure of the message output, as LLMs produce response messages based on input messages from the user – prompts.

In the case of a specialized application, it's essential to contextualize the model using a system prompt [42, 43]. In this solution, the system prompt was constructed separately for OpenAI and Anthropic models due to the difference in how they handle provided instructions, e. g., Sonnet tended to infer details and worked fine with more abstract instructions. At the same time, GPT-4 required more explicit highlights in the system prompt. Given a model instructed with an effective system prompt, further interactions are much simpler and require minimal instructions in the following user prompts.

Thus, the suggested GenAI-powered NLM creates a chat with a model on initialization of the module and provides a system prompt once to set the context and clarify the requirements for further message generation. Then, NLM output messages are generated by repeated interaction with the same contextualized model, i. e., sending simple formatted prompts with input data. In addition, this reduces the budget spent for API calls, as the most significant number of tokens – system prompt – is sent only once, as shown in Fig. 2.

System prompts were constructed following current best practices identified by researchers and industry experts [42, 43]. In particular, the following approaches were applied:

1. Examples were provided for both the final desired output and specific instructions in the prompt.

2. Instructions focused on what is correct and desired instead of listing what the model shouldn't do.

3. The system prompt was clearly structured with sections and lists.

4. The system prompt explained the input the model receives, its format, and its meaning without mentioning all the detailed variables or column names.

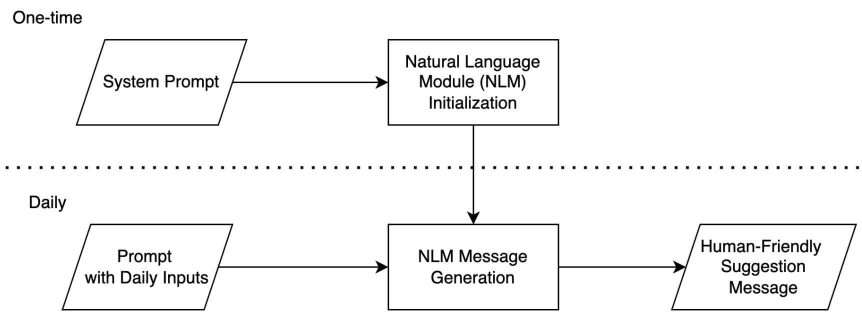


Fig. 2. Natural language module workflow: one-time vs daily prompts

Next, the prompt that delivers daily inputs to the LLM was designed to be structured and straightforward but extensible. This way, the GenAI-based NLM implementation is not bound to the specific fields or parameters in the input, unlike the rule-based NLM. This prompt is the same for both English models:

*"Suggestion from the decision support system ('suggestions'): {suggestion}.
Information that was used to make the suggestions ('used_info'): {used_info}.
Full data available for tomorrow ('df_str'): {df_str}."*

Where *suggestion*, *used_info,* and *df_str* are the inputs of the NLP module, namely:

– *suggestion* – a dictionary, the output of the CSM – control suggestions;
– *used_info* – a dictionary, the output of the CSM – variables used by CSM to make the suggestion;
– *df_str* – a string representation of a data frame with all relevant info for tomorrow, e. g., generation and consumption estimates, and input features of forecasting models.

The same approach was applied to designing prompts for the Ukrainian target language. Constructing a system prompt in the language of the desired output is a better option than straightforward prompt translation or keeping an English prompt with instructions to generate further answers in Ukrainian. This is due to the requirement for natural language flow in the result message, which is very important for UX [44]. This way, an efficient GenAI-powered NLM was also constructed for the Ukrainian language.

To account for the impact of message length on user experience, a system prompt with an explicit message size limitation was designed for both models. The final NLMs chosen for evaluation in both languages are rule-based, GPT-4-based, and Sonnet 3.5-based. Only the Sonnet-based NLM has a message size limitation in the system prompt.

2.7. Survey

There was a need for human evaluation to compare the quality of NLM outputs. To this end, a survey containing examples of each NLM's outputs was designed. To accommodate the audience and research question 4, the survey was prepared in Ukrainian and English.

To improve the statistical properties of the sample, a paired design is utilized: instead of random samples from different groups, the survey contains balanced groups, i. e., each respondent evaluates the same items (scenarios) with various treatments (NLMs applied to generate a message). For example, the survey contains three messages for an example of scenario 1 (23/05/2022) considered in the study, i. e., one per evaluated NLM.

This approach allows for directly evaluating the difference between groups' scores instead of comparing average group scores. It also excludes intersubject variability and significantly enhances statistical properties. For instance, it reduces the number of samples needed to achieve the same statistical significance. To minimize order bias in respondents' answers, messages generated for the same example by different modules for the same scenario are presented in varying order across examples. To this end, permutations of the three elements were used to define the presentation order of examples. With five scenarios considered in this study, five out of six possible permutations were used. Scenario representation was balanced: each scenario identified for the Control Module was considered via one example per NLM. Thus, with five scenarios and three NLMs, the survey contains 15 message examples for a given language.

Considering the limited time and the respondents being volunteers, the number of survey questions was deliberately kept low to maximize the response rate.

The final form design was as follows:

1. Language choice question – depending on the answer, all the following steps are presented in the corresponding language.
2. Instructions for the survey, explaining what is the context and what kind of messages and questions will be shown for evaluation.
3. 15 messages are shown – one at a time, with two questions each. Each question is mandatory and allows answers with an integer score from 1 to 10:

a) How easy was it to understand the message?
b) How useful and informative is the message?

Before launching the survey, the parameters of the experiment were set to define the number of respondents to achieve statistically significant results: significance $\alpha = 0.05$, $\beta = 0.2$. Then, a rough estimate of the required sample size can be calculated using the formulas (6) and (7) [45]:

$$N = \left[\frac{\sigma(z_{1-\beta} + z_{1-\alpha/2})}{\delta} \right]^2, \tag{6}$$

$$N = \frac{\sigma^2 (1.96 + 0.84)^2}{\delta^2} = 7.84 \frac{\sigma^2}{\delta^2} = 7.84 \cdot d^2, \tag{7}$$

where δ – practical significance, σ – variance, d – effect size, calculated as $d = \delta / \sigma$. This leads to the following sample sizes: for small effect size ($d = 0.2$) – 196, for medium effect size ($d = 0.5$) – 32, and for large effect size ($d = 0.8$) – 13.

A paired t-test can be utilized to check if the average score for the two groups is significantly different. To apply it, the average score per group is calculated for each user, i. e., average over five examples per NLM. Next, the difference between the two NLMs' average scores is calculated. This way, a list of differences is obtained per pair of NLMs – one difference per user. This sample can be tested with a paired t-test using the following formula

$$T_{n-1} = \frac{\bar{d}}{SE}, SE = \frac{std(d)}{\sqrt{n}}. \tag{8}$$

A corresponding p-value in a t-distribution with $n - 1$ degrees of freedom is then compared to the parameter of experiment α to check if the result is statistically significant. One tail test is used as the goal of comparison is to detect approach superiority.

3. Results and Discussion

3.1. Scenarios and survey metadata

Historical examples of five scenarios used for the survey are visualized in Fig. 3, *a–e*. Historical data is enhanced with generation and consumption estimates for tomorrow and battery charge estimates used by the CSM to decide on the relevant suggestions.

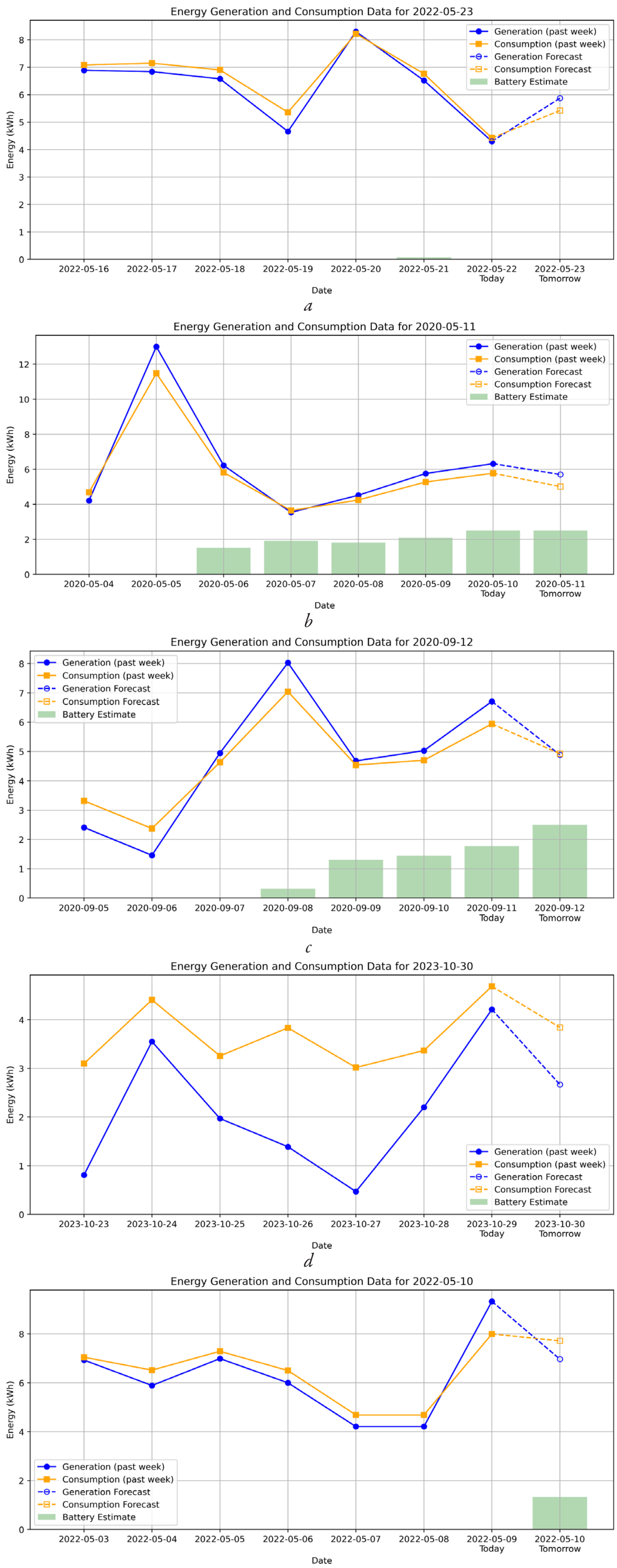


Fig. 3. Examples of scenarios: *a* – scenario 1 – 23/05/2022 – generation surplus with an empty battery; *b* – scenario 2 – 11/05/2020 – generation surplus with full battery; *c* – scenario 3 – 12/09/2020 – generation and consumption balance with full battery; *d* – scenario 4 – 30/10/2023 – lack of generation; *e* – scenario 5 – 10/05/2022 – lack of generation covered by battery reserve

The survey was run online for 2 weeks: from 18 April to 03 May 2025. The number of responses is given in Table 4.

Table 4

Number of responses per language and gender

Language	Female	Male	No answer	Total
English	11	18	1	30
Ukrainian	43	20	1	64
Total	54	38	2	94

The number of English-language responses was insufficient for gender-specific analysis, and because of that, they were analyzed together. As none of the respondents were native English speakers, it's also reasonable to combine them to get corresponding results for that group. In contrast, the number of Ukrainian-language responses allows for analysis per gender.

Answers distribution per age group is presented in Table 5. The number of responses for different age groups doesn't allow for robust age group comparison.

Table 5

Number of responses per language and age group

Language	18–24	25–34	35–44	45–54	55–64	No answer	Total
English	4	14	9	2	0	1	30
Ukrainian	46	7	4	6	1	0	64
Total	50	21	13	8	1	1	94

3.2. Analysis of survey responses

Tables 6–8 demonstrate the results of statistical tests [46] comparing the scores of different NLMs in two dimensions – usefulness and ease of understanding. In these tables, RB is rule-based NLM, S is NLM using Sonnet 3.5, and G is NLM using GPT-4. In particular, a summary of statistical tests for the English language is in Table 6. It can be observed that for English, only the usefulness score of Sonnet-based NLM is statistically significantly higher than for GPT-4-based NLM.

Table 7 and Table 8 demonstrate the results of the analysis of the Ukrainian responses for male and female respondents, respectively. In Ukrainian, Sonnet-based NLM is considered more useful than both rule-based systems and GPT-4-based NLM among the female audience. These results indicate that for this user group, using GenAI-based NLM is indeed efficient in generating optimal DSS messages. The corresponding NLM was designed for a general audience, without targeting a specific group, e. g., female users, which allows expecting higher improvement in case of such targeting. Considering the smaller adoption of such systems by female users and the increasing need for that, it may be beneficial to introduce GenAI-powered NLM to improve UX in such applications.

Table 6

Summary of statistical tests for the English language

Dimension	RB vs. G	G vs. S	RB vs. S
Usefulness	$RB > G$, $m = 0.41$, $std = 1.908$, $p = 0.1225$	$S > G$, $m = 0.19$, $std = 0.538$, $p = 0.0337$	$RB > S$, $m = 0.23$, $std = 1.822$, $p = 0.2505$
Ease of understanding	$RB > G$, $m = 0.46$, $std = 1.752$, $p = 0.0805$	$S > G$, $m = 0.11$, $std = 0.506$, $p = 0.1147$	$RB > S$, $m = 0.35$, $std = 1.546$, $p = 0.1146$

ISSN-L 2664-9969; E-ISSN 2706-5448

Table 7

Summary of statistical tests for the Ukrainian language, male

Dimension	RB vs GPT	GPT vs Sonnet	RB vs Sonnet
Usefulness	$G > RB$, $m = 0.32$, $std = 0.930$, $p = 0.0702$	$S > G$, $m = 0.01$, $std = 0.505$, $p = 0.4652$	$S > RB$, $m = 0.33$, $std = 1.020$, $p = 0.0822$
Ease of understanding	$RB > G$, $m = 0.04$, $std = 0.692$, $p = 0.3993$	$G > S$, $m = 0.12$, $std = 0.814$, $p = 0.2588$	$RB > S$, $m = 0.16$, $std = 0.772$, $p = 0.1829$

Table 8

Summary of statistical tests for the Ukrainian language, female

Dimension	RB vs GPT	GPT vs Sonnet	RB vs Sonnet
Usefulness	$G > RB$, $m = 0.02$, $std = 0.958$, $p = 0.4372$	$S > G$, $m = 0.31$, $std = 0.996$, $p = 0.0233$	$S > RB$, $m = 0.33$, $std = 1.231$, $p = 0.0408$
Ease of understanding	$RB > G$, $m = 0.01$, $std = 1.062$, $p = 0.4659$	$S > G$, $m = 0.13$, $std = 1.054$, $p = 0.2111$	$S > RB$, $m = 0.12$, $std = 1.192$, $p = 0.2630$

In addition, no expected adverse effects, such as increased complexity due to long or less structured messages, were observed. It is worth noting that several users submitted comments in the survey's optional notes field. They explicitly stated that long messages are difficult to read and should be avoided in both English and Ukrainian. However, this was not reflected in aggregated survey results for both languages. On one side, it reflects how personal experience may differ from reality. On the other side, this indicates that there may be some dependency of scores (especially ease of understanding) on the length of the message. This dependency is discussed further in this section.

Across all groups, the Sonnet-based NLM has a statistically significantly higher score in usefulness than NLM based on GPT-4. It is not clear why exactly this occurs. It may be due to a better integration into the DSS pipeline, a more effective system prompt, internal model particularities combined with these factors, or other aspects not considered here. This underscores the importance of evaluating AI model-based solutions in both offline and online modes to find an optimal approach.

Another notable difference related to the Sonnet 3.5-based NLM is that ease and usefulness scores are not strongly correlated, unlike for the RB and GPT-4-based NLMs, as shown in Table 9. While the overall correlation between ease and usefulness is 0.82, this correlation for Sonnet is only 0.019. Moreover, the ease score is negatively correlated with the scores of all other NLMs, indicating disagreement in user evaluation between these models. This suggests the presence of some structural or internal characteristics unique to Sonnet-based NLM.

Table 9

Scores correlation between models and evaluation dimensions: ease of understanding (E) and usefulness (U)

Model	GPT-4 E	GPT-4 U	RB E	RB U	Sonnet 3.5 E	Sonnet 3.5 U
GPT-4 E	–	0.694	0.784	0.808	–0.266	0.746
GPT-4 U	0.694	–	0.594	0.525	–0.74	0.486
RB E	0.784	0.594	–	0.921	–0.232	0.957
RB U	0.808	0.525	0.921	–	–0.356	0.797
Sonnet 3.5 E	–0.266	–0.74	–0.232	–0.356	–	0.019
Sonnet 3.5 U	0.746	0.486	0.957	0.797	0.019	–

Nevertheless, statistical test results and average scores per language per category (all higher than 7.5) clearly indicate that the developed DSS prototype is effective and efficient in delivering output to users in a human-friendly manner. Moreover, applying GenAI allows for better results than the baseline approach. Its main advantage is increasing how useful messages feel to the audience by adapting the format to their needs. In addition, possible adverse effects of increasing message complexity were not observed. However, developing an optimal GenAI-based NLM may require a deeper understanding of existing dependencies among user groups. Therefore, the analysis was extended to explore patterns related to languages, genders, and message length.

The difference between languages can be attributed to the survey audience, which is mostly native Ukrainian speakers who use English as a foreign language. There is no evidence of a correlation between answers for the two languages: average ease answers have a Pearson correlation of –0.036, and for usefulness, it is 0.002.

The relationship between message length and score was also investigated. The average message length for each model is presented in Table 10.

Table 10

Average message length per model and language

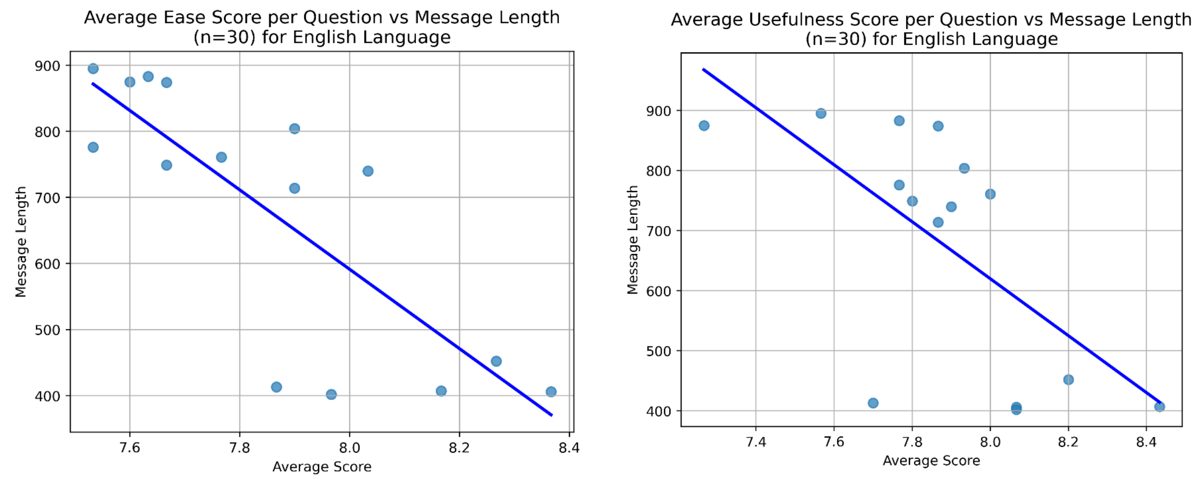
Language	RB	GPT-4	Sonnet 3.5
English	416	866	748
Ukrainian	450	564	744

While Sonnet is considered superior to GPT in usefulness for both languages, its average message length is longer than GPT's in Ukrainian but shorter in English. This suggests that message length is not the only factor impacting the UX, and likely not the main one, provided the message length remains reasonable for the applications. Nevertheless, this aspect deserves further investigation.

Fig. 4 presents linear dependencies between average score and message length for English-language responses. Both Pearson coefficients for ease (–0.798) and usefulness (–0.648) are negative, and their difference from zero is statistically significant (p-values of 0.0003 and 0.0089, respectively). Such an effect is not observed for Ukrainian-language responses – both Pearson correlations are positive but not statistically different from zero. This is likely due to the difference in respondents' native and survey languages, with most Ukrainian respondents being native speakers, while English respondents are mostly non-native speakers.

The correlation with length also supports previous findings, i. e., longer texts are not desired by foreign language users. Because of that, the impact of the GenAI approach in DSS is more clearly observed in the Ukrainian responses, where the factor of foreign language complexity is not present. This finding also led to the conclusion that different target languages require adaptation of the general approach and flexibility in the DSS to accommodate users with diverse language backgrounds.

The relationship between message length and responses by gender demonstrates a slightly different pattern that aligns with previous findings. As shown in Table 11, answers from both genders in English have a strong negative correlation between message length and ease scores. This pattern is plausible if to assume the main difference was caused by the difference between native and non-native speakers' comprehension. However, for Ukrainian responses, both correlations are small and point in different directions: male answers still indicate a negative impact of the length on scores, while female answers indicate the opposite. It is aligned with p-values for ease correlation being closer to the critical value for male responses than for female ones.



a *b*

Fig. 4. Relationship between average score and message length, modeled linearly for English-language responses:
a – ease of understanding; *b* – usefulness

Table 11

Correlations between message length and average ease score for different languages and genders

Language	Gender	Pearson Correlation	*p*-value
English	Male	–0.73	0.002
English	Female	–0.648	0.0089
Ukrainian	Male	–0.247	0.3742
Ukrainian	Female	0.173	0.537

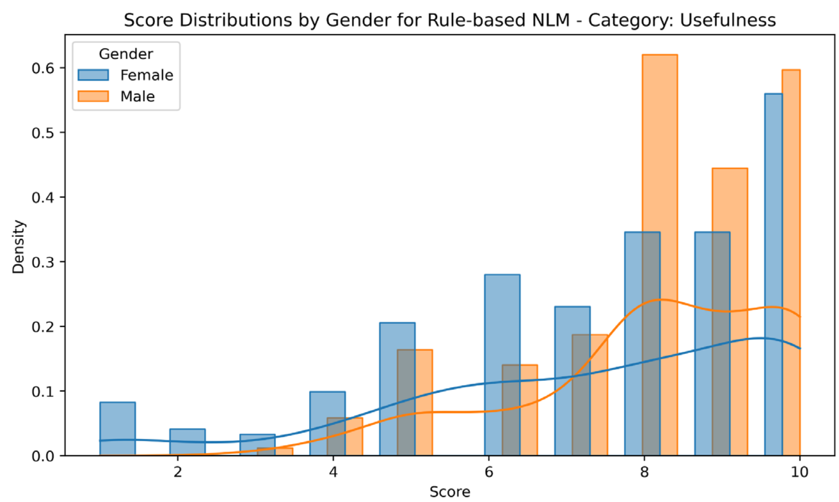


Fig. 5. Distribution of usefulness scores for a rule-based NLM by gender

In addition, the average usefulness score is strongly negatively correlated with message length for male English users, with a correlation coefficient of –0.736 and a *p*-value of 0.0018. These results highlight that while the message length may impact the ease of understanding and consequently influence UX, it's essential to understand underlying reasons and target specific groups influenced by such dependencies.

For example, UX may be improved by simplifying messages for non-native speakers or providing shorter, more explicitly structured messages with optional explanations to male users.

Another important relationship in this study is between ease of understanding and usefulness.

These two aspects are strongly correlated, with a Pearson correlation of 0.823. However, for male users, this correlation is just 0.407 compared to 0.927 in female users. This illustrates further how different groups of users are different and should be accounted for flexibly by the DSS.

There are other examples of male and female users disagreeing on scores. One of the most illustrative ones is the usefulness of rule-based NLM.

Fig. 5 shows the distribution of scores per gender.

Male users not only rate the usefulness of RB NLM systems higher, but their answers are also negatively correlated with female ones – Pearson correlation of –0.589 (*p*-value 0.296). Another example is Sonnet-based NLM. Its usefulness scores by male and female users are strongly negatively correlated, with a correlation of –0.971 and a *p*-value of 0.006, making it statistically significant.

These examples are well aligned with the results of statistical tests, i. e., not only do different groups evaluate models based on their preferences and needs, as can be seen for RB NLM, but they also react differently to a new solution – GenAI-based NLM.

3.3. Limitations and future work directions

These results and analyses should be interpreted with certain limitations in mind. First, the number of English-language responses could be higher for a more robust analysis. Second, the GenAI-powered NLM did not explicitly target specific parts of the audience, while its main potential is a more individual approach to delivering information. Third, Ukrainian responses were from native Ukrainian speakers, while English responses were from non-native English speakers. Fourth, the study wasn't run on a production system in the form of an A/B test. Finally, in this study, statistical tests were conducted independently, thus excluding family-wise error rates from the scope. Acknowledging these limitations is crucial for researchers conducting further studies on related topics, as well as for industry practitioners aiming to apply the findings in practical contexts.

Future work directions include both extending this research and overcoming its limitations. First, the number of responses should be increased to validate some conclusions and improve the statistical properties of the results, e. g., the lack of adverse effects of GenAI application on perceived message complexity. This also includes more strict thresholds for the statistical significance of multiple tests, e. g., with Holm-Bonferotti correction-like approaches. Second, the development of a GenAI-powered or hybrid system can be done with an explicit focus on different user groups and languages. For example, this may include targeting female users with clear explanations or male users with short, structured reports. Another possible adjustment is the simplification of message formats for non-native speakers. Third, it would be highly beneficial to introduce live testing in the form of an A/B test in real prosumer DSS. A considered production system would also allow for investigating other aspects of the decision-making process in combination with text messages, such as visual information aids.

4. Conclusions

This study focused on investigating the influence of GenAI approaches on DSSs in prosumers. To answer the formulated research questions, a specific use case was considered with a private household with a solar generation capacity. A prototype DSS was developed with several variations of NLM, transforming technical DSS output into human-friendly text messages. In particular, a rule-based NLM was developed as a baseline, along with several variations of GenAI-powered NLM, e. g., using GPT-4 and Sonnet 3.5 models by OpenAI and Anthropic, respectively. NLMs were evaluated based on historical examples of five common prosumer scenarios in a survey with human evaluation. NLMs were evaluated in terms of ease of understanding and subjective usefulness. The survey conducted in English and Ukrainian indicated the benefits of utilizing the GenAI model, which is properly integrated into a DSS, while there is no significant evidence of disadvantages.

Statistically significant moderate improvement was detected for female audiences in Ukrainian in the aspect of usefulness when Sonnet 3.5-generated messages were evaluated as more useful than rule-based or GPT-4 alternatives. Moreover, it was observed that male and female respondents disagreed on the usefulness of messages from different NLMs, while answers were not correlated between the two languages. For the English language, respondents preferred shorter messages, although there were no statistically significant score differences. This observation can be explained by respondents being non-native speakers of English. These findings highlight the need for more flexible output generation in prosumer DSSs, satisfying the specific needs of different user groups, such as caused by gender differences, while demonstrating GenAI's capability to cater to those needs.

There was also evidence across different user groups and languages that Sonnet-generated messages were more useful than those generated with GPT-4. Although the exact reason is unclear, it highlights the importance of thoroughly testing specific GenAI-based solutions before using them in production.

These results have both theoretical and practical significance. First, further research and applied developments can benefit from the validated hypothesis that GenAI can be effectively used to prepare optimal DSS outputs for different user groups. Second, specific improvements can be made in adapting DSS outputs for different genders and languages based on the presented results. Considering the limited amount of research on optimal outputs of prosumers' DSSs, this study is an essential step toward closing the research gap and improving the efficiency of the related applied solutions.

Acknowledgments

The authors would like to express their sincere gratitude to all survey participants. Their time, thoughtful responses, and engagement made this research possible.

Conflict of interest

The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, authorship or otherwise, that could affect the research, and its results presented in this paper.

Financing

The work is supported by the state budget scientific research project of Sumy State University, "Intelligent information technology for proactive management of energy infrastructure in conditions of risk and uncertainty" (state registration number 0123U101852).

Data availability

Data will be made available on reasonable request.

Use of artificial intelligence

The authors have used artificial intelligence technologies within acceptable limits to provide their own verified data, which is described in the research methodology section.

References

1. Lopes, J. A. P., Madureira, A. G., Matos, M., Bessa, R. J., Monteiro, V., Afonso, J. L. et al. (2020). The future of power systems: Challenges, trends, and upcoming paradigms. *WIREs Energy and Environment, 9 (3).* https://doi.org/10.1002/wene.368
2. Alotaibi, I., Abido, M. A., Khalid, M., Savkin, A. V. (2020). A Comprehensive Review of Recent Advances in Smart Grids: A Sustainable Future with Renewable Energy Resources. *Energies, 13 (23)*, 6269. https://doi.org/10.3390/en13236269
3. Peng, F. Z., Liu, C.-C., Li, Y., Jain, A. K., Vinnikov, D. (2024). Envisioning the Future Renewable and Resilient Energy Grids – A Power Grid Revolution Enabled by Renewables, Energy Storage, and Energy Electronics. *IEEE Journal of Emerging and Selected Topics in Industrial Electronics, 5 (1)*, 8–26. https://doi.org/10.1109/jestie.2023.3343291
4. Kröger, W., Nan, C.; Büscher, C., Schippl, J., Sumpf, P. (Eds.) (2018). Power systems in transition. *Energy as a Sociotechnical Problem.* London: Routledge, 41–78. https://doi.org/10.4324/9781315186313-3
5. Marot, A., Kelly, A., Naglic, M., Barbesant, V., Cremer, J., Stefanov, A. et al. (2022). Perspectives on Future Power System Control Centers for Energy Transition. *Journal of Modern Power Systems and Clean Energy, 10 (2)*, 328–344. https://doi.org/10.35833/mpce.2021.000673
6. Machlev, R., Heistrene, L., Perl, M., Levy, K. Y., Belikov, J., Mannor, S. et al. (2022). Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities. *Energy and AI, 9*, 100169. https://doi.org/10.1016/j.egyai.2022.100169
7. Shin, D. (2020). User Perceptions of Algorithmic Decisions in the Personalized AI System:Perceptual Evaluation of Fairness, Accountability, Transparency, and Explainability. *Journal of Broadcasting & Electronic Media, 64 (4)*, 541–565. https://doi.org/10.1080/08838151.2020.1843357
8. Donida Labati, R., Genovese, A., Piuri, V., Scotti, F., Sforza, G. (2020). A Decision Support System for Wind Power Production. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 50 (1)*, 290–304. https://doi.org/10.1109/tsmc.2017.2783681
9. Sztubecka, M., Skiba, M., Mrówczyńska, M., Bazan-Krzywoszańska, A. (2020). An Innovative Decision Support System to Improve the Energy Efficiency of Buildings in Urban Areas. *Remote Sensing, 12 (2)*, 259. https://doi.org/10.3390/rs12020259
10. Badami, M., Fambri, G., Mancò, S., Martino, M., Damousis, I. G., Agtzidis, D. et al. (2019). A Decision Support System Tool to Manage the Flexibility in Renewable Energy-Based Power Systems. *Energies, 13 (1)*, 153. https://doi.org/10.3390/en13010153
11. Marot, A., Rozier, A., Dussartre, M., Crochepierre, L., Donnot, B. (2022). Towards an AI Assistant for Power Grid Operators. *HHAI2022: Augmenting Human Intellect.* IOS Press, 354, 79–95. https://doi.org/10.3233/faia220191
12. Lyu, X., Li, Z., Ma, Q., She, M. (2023). Effects of accessible information amount and judgment times on human diagnostic performance of nuclear power plant faults. *Ergonomics, 66 (7)*, 927–938. https://doi.org/10.1080/00140139.2022.2118836
13. Fehrenbacher, D. D., Djamasbi, S. (2017). Information systems and task demand: An exploratory pupillometry study of computerized decision making. *Decision Support Systems, 97*, 1–11. https://doi.org/10.1016/j.dss.2017.02.007
14. Allen, P. M., Edwards, J. A., Snyder, F. J., Makinson, K. A., Hamby, D. M. (2014). The Effect of Cognitive Load on Decision Making with Graphically Displayed Uncertainty Information. *Risk Analysis, 34 (8)*, 1495–1505. https://doi.org/10.1111/risa.12161
15. Phillips-Wren, G., Adya, M. (2020). Decision making under stress: the role of information overload, time pressure, complexity, and uncertainty. *Journal of Decision Systems, 29 (1)*, 213–225. https://doi.org/10.1080/12460125.2020.1768680

16. Afzal, U., Prouzeau, A., Lawrence, L., Dwyer, T., Bichinepally, S., Liebman, A. et al. (2022). Investigating Cognitive Load in Energy Network Control Rooms: Recommendations for Future Designs. *Frontiers in Psychology, 13.* https://doi.org/10.3389/fpsyg.2022.812677
17. Yacob, A., Roslim, A. D., Abdul Wahab, S. F., Abdul Halim, I. I., Baharum, Z. W., Hamzah, W. M. A. F. et al. (2024). Exploring the Landscape of Decision Support Systems: A Comprehensive Review of Implementations and Key Characteristics. *Journal of Advanced Research in Applied Sciences and Engineering Technology, 51 (1)*, 141–149. https://doi.org/10.37934/araset.51.1.141149
18. Moncur, B., Galvez Trigo, M. J., Mortara, L.; Schmorrow, D. D., Fidopiastis, C. M. (Eds.) (2023). Augmented Reality to Reduce Cognitive Load in Operational Decision-Making. *Augmented cognition. HCII 2023. Lecture Notes in Computer Science.* Cham: Springer, 328–346. https://doi.org/10.1007/978-3-031-35017-7_21
19. Perry, N. C., Wiggins, M. W., Childs, M., Fogarty, G. (2013). The Application of Reduced-Processing Decision Support Systems to Facilitate the Acquisition of Decision-Making Skills. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 55 (3)*, 535–544. https://doi.org/10.1177/0018720812467367
20. Phillips-Wren, G., Daly, M., Burstein, F. (2022). Support for cognition in decision support systems: an exploratory historical review. *Journal of Decision Systems, 31 (1)*, 18–30. https://doi.org/10.1080/12460125.2022.2070946
21. Stemmet, L., Ahmed, M. D.; Khosrow-Pour, M. (Ed.) (2014). The Decision Maker's Cognitive Load. *Encyclopedia of Information Science and Technology.* IGI Global, 6466–6474. https://doi.org/10.4018/978-1-4666-5888-2.ch635
22. Morrison, B. W., Kelson, J. N., Morrison, N. M. V., Innes, J. M., Zelic, G., Al-Saggaf, Y. et al. (2023). You're Not the Boss of me, Algorithm: Increased User Control and Positive Implicit Attitudes Are Related to Greater Adherence to an Algorithmic Aid. *Interacting with Computers, 35 (3)*, 452–460. https://doi.org/10.1093/iwc/iwad028
23. Kanojiya, A., Nagori, V. (2018). Analysis of Architecture and Forms of Outputs of Decision Support Systems implemented for different domains. *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*. Coimbatore: IEEE, 346–350. https://doi.org/10.1109/icicct.2018.8472981
24. Kelton, A. S., Pennington, R. R., Tuttle, B. M. (2010). The Effects of Information Presentation Format on Judgment and Decision Making: A Review of the Information Systems Research. *Journal of Information Systems, 24 (2)*, 79–105. https://doi.org/10.2308/jis.2010.24.2.79
25. Beriro, D., Nathanail, J., Salazar, J., Kingdon, A., Marchant, A., Richardson, S. et al. (2022). A Decision Support System to Assess the Feasibility of Onshore Renewable Energy Infrastructure. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.3986807
26. Taik, A., Nour, B., Cherkaoui, S. (2021). Empowering Prosumer Communities in Smart Grid with Wireless Communications and Federated Edge Learning. *IEEE Wireless Communications, 28 (6)*, 26–33. https://doi.org/10.1109/mwc.017.2100187
27. Stamatescu, I., Arghira, N., Făgărăşan, I., Stamatescu, G., Iliescu, S., Calofir, V. (2017). Decision Support System for a Low Voltage Renewable Energy System. *Energies, 10 (1)*, 118. https://doi.org/10.3390/en10010118
28. Ponraj, P., Murugesan, S. (2024). A Bi-level Decision Support System for Home Energy Management in Smart Homes. *2024 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES).* Mangalore: IEEE, 1–5. https://doi.org/10.1109/pedes61459.2024.10961886
29. Gržanić, M., Capuder, T., Zhang, N., Huang, W. (2022). Prosumers as active market participants: A systematic review of evolution of opportunities, models and challenges. *Renewable and Sustainable Energy Reviews, 154*, 111859. https://doi.org/10.1016/j.rser.2021.111859
30. Vachon, F., Lafond, D., Vallieres, B. R., Rousseau, R., Tremblay, S. (2011). Supporting situation awareness: A tradeoff between benefits and overhead. *2011 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA).* Miami Beach: IEEE. https://doi.org/10.1109/cogsima.2011.5753460
31. Lukianykhin, O., Shendryk, V., Shendryk, S., Malekian, R.; Karabegovic, I., Kovačević, A., Mandzuka, S. (Eds.) (2024). Promising AI Applications in Power Systems: Explainable AI (XAI), Transformers, LLMs. *New Technologies, Development and Application VII.* Cham: Springer, 66–76. https://doi.org/10.1007/978-3-031-66271-3_8
32. Huynh, M.-T. (2024). Using generative AI as decision-support tools: unraveling users' trust and AI appreciation. *Journal of Decision Systems*, 1–32. https://doi.org/10.1080/12460125.2024.2428166
33. Yu, P., Xu, H., Hu, X., Deng, C. (2023). Leveraging Generative AI and Large Language Models: A Comprehensive Roadmap for Healthcare Integration. *Healthcare, 11 (20)*, 2776. https://doi.org/10.3390/healthcare11202776
34. Huang, Y., Kanij, T., Madugalla, A., Mahajan, S., Arora, C., Grundy, J. (2024). Unlocking Adaptive User Experience with Generative AI. *Proceedings of the 19th International Conference on Evaluation of Novel Approaches to Software Engineering.* Angers: SciTePress, 760–768. https://doi.org/10.5220/0012741000003687
35. Iovane, G., Chinnici, M. (2024). Decision Support System Driven by Thermo-Complexity: Scenario Analysis and Data Visualization. *Applied Sciences, 14 (6)*, 2387. https://doi.org/10.3390/app14062387
36. Choi, S., Jain, R., Emami, P., Wadsack, K., Ding, F., Sun, H. et al. (2024). eGridGPT: Trustworthy AI in the Control Room. *Office of Scientific and Technical Information (OSTI).* https://doi.org/10.2172/2352232
37. Lukianykhin, O., Shendryk, V. (2025). Machine learning-driven photovoltaic generation forecasting for prosumer decision support. *Artificial Intelligence, 30 (1)*, 107–119. https://doi.org/10.15407/jai2025.01.107
38. Huld, T., Ruf, H., Heilscher, G. (2014). Self-Consumption of Electricity by Households, Effects of PV System Size and Battery Storage. *29th European Photovoltaic Solar Energy Conference and Exhibition.* Amsterdam, 4014–4017. https://doi.org/10.4229/EUPVSEC20142014-7AV.6.8
39. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O. et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.
40. *GPT-4 is OpenAI's most advanced system, producing safer and more useful responses.* OpenAI. Available at: https://openai.com/gpt-4
41. Claude 3.5 Sonnet (2024). *Anthropic.* Available at: https://www.anthropic.com/news/claude-3-5-sonnet
42. Fagbohun, O., Harrison, R. M., Dereventsov, A. (2023). An Empirical Categorization of Prompting Techniques for Large Language Models: A Practitioner's Guide. *Journal of Artificial Intelligence, Machine Learning and Data Science, 1 (4)*, 1–11. https://doi.org/10.51219/jaimld/oluwole-fagbohun/15
43. Beri, G., Srivastava, V. (2024). Advanced Techniques in Prompt Engineering for Large Language Models: A Comprehensive Study. *2024 IEEE 4th International Conference on ICT in Business Industry & Government (ICTBIG).* Indore: IEEE, 1–4. https://doi.org/10.1109/ictbig64922.2024.10911672
44. Moore, R. J., Arar, R. (2019). *Conversational UX Design: A Practitioner's Guide to the Natural Conversation Framework.* ACM Books. https://doi.org/10.1145/3304087
45. van Belle, G. (2008). Sample Size. *Statistical Rules of Thumb.* John Wiley & Sons, 27–51. https://doi.org/10.1002/9780470377963.ch2
46. Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D. et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods, 17 (3)*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

✉***Oleh Lukianykhin,*** *PhD Student, Department of Information Technologies, Sumy State University, Sumy, Ukraine, e-mail: oleh.lukianykhin.a@gmail.com, ORCID: https://orcid.org/0000-0002-4211-2401*

Vira Shendryk, *PhD, Associate Professor, Head of Department of Information Technologies, Sumy State University, Sumy, Ukraine, ORCID: https://orcid.org/0000-0001-8325-3115*

✉*Corresponding author*