European Research Studies Journal
Volume XXIX, Issue 4, 2026

pp. 89-106

Use of Publicly Available Artificial Intelligence Tools to Support Prosumer Decision-Making: The Case of Households with Photovoltaic Installations

Submitted 03/09/26, 1st revision 25/09/26, 2nd revision 16/10/26, accepted 01/11/26

Mariusz Chudzicki[1]

Abstract:

Purpose: *The aim of this article is to assess the usefulness of publicly available generative artificial intelligence tools in supporting prosumers in decisions related to energy and cost management. Particular attention is paid to the ability of the models to consider the structure and timing of energy consumption, individualize recommendations, and recognize limitations caused by incomplete input data.*
Methodology/Approach: *The study was conducted as a comparative case study of two real households equipped with photovoltaic micro-installations with similar production parameters but different structures of major loads and different usage profiles. ChatGPT, Gemini, and Microsoft Copilot were examined in their publicly available versions, without paid plans, APIs, agents, or additional tools. Each system received identically structured datasets and the same prompt. The responses were assessed qualitatively in terms of data use, individualization, factual correctness, feasibility and prioritization of recommendations, recognition of missing information, and quality of justification.*
Findings: *All tools correctly recognized the importance of the time mismatch between production and consumption and produced different recommendations for the two households. The largest differences concerned the prioritization of actions and the way the tools responded to incomplete data. Gemini more often filled data gaps with additional assumptions, which led, among other things, to an unjustified estimate of PHEV energy consumption and an imprecise interpretation of the settlement system. Copilot more often made its conclusions conditional on obtaining additional parameters. ChatGPT took an intermediate position, providing recommendations that could be made from the available data while also indicating the limits of their accuracy. None of the tools considered an additional investment to be clearly justified on the basis of the available information.*
Practical implications: *Publicly available generative AI can support prosumers at the stage of initial analysis and identification of possible actions without the need to build a formal optimization model. However, recommendations that require precise technical, economic, or investment assessment should be verified using more detailed data and specialist methods.*
Originality/Value: *The study does not assess a specialized energy system. Instead, it evaluates the practical usefulness of freely available conversational tools in the same clearly defined real-world decision problem. The value of the study also lies in comparing the*

[1]*Dr., Czestochowa University of Technology, Poland, ORCID: 0000-0003-0686-5438, e-mail: mariusz.chudzicki@pcz.pl;*

flexibility of generative AI with the transparency and repeatability of classical multi-criteria methods.

Keywords: *Self-consumption, large language models, energy flexibility, personal finance, demand-side management.*

JEL codes: *C44, D14, Q42.*

Paper type: *Research article.*

Declaration: *Generative artificial intelligence tools were employed during the preparation of this work for language proofreading and style improvement. The author remains fully responsible for the final text and its scientific integrity, having thoroughly reviewed and revised the generated content afterward.*

1. Introduction

The development of prosumer energy systems changes the nature of decisions made by households. Before a micro-installation is commissioned, the main issue is the choice of technology and the assessment of investment profitability. After commissioning, the focus shifts to the ongoing management of energy production and consumption.

The economic effect of the installation then depends not only on annual generation, but also on when energy is produced, when it is consumed, which loads account for the largest share of demand, and to what extent their operation can be shifted in time. These decisions become more important as households become increasingly electrified, because heat pumps, electric domestic hot water systems, and BEV and PHEV vehicles may account for a significant share of total electricity consumption.

Managing such a household is a multi-criteria problem. A prosumer may simultaneously aim to reduce the cost of electricity purchased from the grid, increase self-consumption, maintain user comfort, avoid unnecessary investment expenditure, and reduce the need for manual control of loads.

Classical multi-criteria decision-support methods can formally describe such a problem, but they require the definition of criteria, methods of measurement, and often also weights. For an average installation user, this means building a model whose preparation and later updating require specialist knowledge.

The widespread availability of large language models has created an alternative way to obtain decision support. A user can describe a household in natural language,

provide quantitative and qualitative data, and then receive an interpretation of the situation, possible courses of action, and their justification. This does not mean that the recommendation is automatically correct. A model may misinterpret data, create unjustified assumptions, give estimates a false appearance of precision, or prioritize criteria in different ways even though their importance is not visible to the user.

Therefore, the main research problem is not simply whether AI can generate advice, but whether the advice is of sufficient quality under conditions similar to the real use of publicly available services.

The aim of the study is to assess the usefulness of publicly available generative artificial intelligence tools in supporting prosumer decisions related to energy and cost management. The main research question is: to what extent can such tools use data from a specific household, differentiate recommendations for different consumption profiles, and recognize situations in which the available information is insufficient for a reliable recommendation? The study is a comparative analysis of two real households with similar photovoltaic installations but different load structures and usage profiles.

The same procedure was applied to ChatGPT, Gemini, and Microsoft Copilot using their publicly available free versions. The originality of the study therefore lies in evaluating the practical usefulness of non-specialized conversational tools in the same clearly defined prosumer decision problem, rather than analyzing a dedicated energy management system.

2. Literature review

2.1 The Household as a Prosumer and an Economic Decision-Making Entity

A household is an entity that makes interrelated decisions concerning consumption, saving, investment, and financing. Household finance therefore includes both the ongoing management of income and expenditure and decisions concerning savings, the use of financial surpluses, and the financing of future needs (Campbell, 2006; Bywalec, 2009).

These decisions are made by households that differ considerably in their resources, number of residents, structure of needs, and financial competences. Therefore, there is no single universal model of behavior that is appropriate for all households (Hoffmeyer-Zlotnik and Warner, 2009; Świecka and Musiał, 2016).

Investment in a photovoltaic micro-installation extends the scope of these decisions. The household is no longer only an energy consumer and becomes a prosumer, combining the roles of consumer and producer. Photovoltaics can also be analyzed as an element of household financial management because the economic effect of the investment is mainly reflected in lower future expenditure on energy (Chudzicki,

2020). After the installation is commissioned, the problem is no longer limited to assessing the return on investment but also includes how the generated energy is used.

The same annual production may produce different effects in households with different demand structures. Therefore, not only the amount of generation but also the ability to match consumption to periods of own production becomes important.

An additional condition in the analyzed cases is the quantity-based energy discount settlement scheme. For micro-installations up to 10 kWp, each 1 kWh exported to the grid allows the prosumer to later settle 0.8 kWh of electricity drawn from the grid. Direct consumption of 1 kWh at the time of production avoids the purchase of the full 1 kWh from the grid.

This mechanism means that increasing self-consumption may have economic value even when the grid provides an indirect energy-balancing function. In this study, the settlement system is a condition common to both case studies and is not treated as a separate subject of regulatory analysis.

2.2 Matching Production and Consumption, Self-Consumption, and Load Flexibility

An annual comparison of production and consumption is a useful indicator of scale, but it does not show the actual level of on-site energy use. Research on matching local generation and load shows that even an annual balance close to zero may coexist with frequent energy flows to and from the grid if production and demand occur at different hours or during different seasons (Salom *et al.,* 2014).

The temporal resolution of the data is also important: excessive aggregation of production and load profiles may lead to an incorrect assessment of self-consumption and the sizing of solutions used together with PV systems (Beck *et al.,* 2016). In photovoltaic systems, the problem has a clear daily and seasonal character: production is concentrated during daytime hours and in months with higher solar irradiation, while some loads operate mainly in the morning, evening, or heating season.

Self-consumption means the share of PV energy used directly by the owner of the installation. The literature identifies energy storage and demand-side management, meaning the shifting of some loads to periods of own production, as the two main ways to increase self-consumption (Luthander *et al.,* 2015; Widén, 2014).

Existing energy storage systems are not analyzed in this study because neither household has one. Storage may appear only as a potential investment recommendation made by a model. Therefore, the flexibility of already used loads is more important in this analysis.

Flexibility is not only a technical property of a device. The same load may be easy to shift in one household and practically inflexible in another. This is especially relevant to electric vehicles. A BEV or PHEV battery creates a large periodic electricity demand, but the possibility of moving charging to periods of PV production depends on actual mileage, the energy level required before the next trip, charging power, and, above all, whether the vehicle is present at home. Reviews of smart EV charging confirm that the benefits of integrating PV and electromobility depend largely on coordinating charging time with production and with the profile of other building loads (Fachrizal *et al.,* 2020; Munkhammar *et al.,* 2013).

A similar mechanism applies to domestic hot water preparation and space heating. A DHW tank, the thermal inertia of the building, or a flexibly controlled heat pump may allow part of the electricity demand to be shifted to periods of photovoltaic production. Power-to-heat analyses identify heat pumps and passive heat storage as important sources of flexibility in the integration of renewable energy sources (Bloess *et al.,* 2018; Fischer and Madani, 2017).

However, resident comfort and the technical operating conditions of the installation remain the limits of such actions. The development of building energy management systems shows that load flexibility can also be used by specialized control algorithms (Beaudin and Zareipour, 2015). Pinthurat, Surinkaew and Hredzak (2024) indicate the broad use of reinforcement learning methods in home energy management systems that operate with energy storage, while Kelm, Mieński and Wasiak (2025) analyze control of an energy storage system in a prosumer installation under dynamic tariff conditions.

These solutions confirm the potential of algorithmic optimization, but they require a dedicated model, data, and control infrastructure (Kaschub et al., 2016). In the present study, the focus is shifted to a different question: can a prosumer who uses only a publicly available conversational tool also obtain useful support?

From the perspective of decision support, the quality of input information is therefore particularly important. Annual production and annual consumption are not sufficient to reliably determine device operating schedules. At a minimum, relevant information includes the temporal availability of a load, its operating power, programmability, the PV production profile, seasonality, and user-related constraints. Missing data should not lead to arbitrary values being presented as facts. A proper decision-support tool should be able to distinguish between a recommendation that can be made from the available data and a conclusion that requires additional measurement or information.

2.3 Multi-criteria decision support and generative artificial intelligence

Energy problems are a classical field of application for multi-criteria methods. Literature reviews show the broad use of MCDM/MCDA in energy planning,

assessment of renewable technologies, and comparison of alternatives that differ in economic, technical, and environmental terms (Pohekar and Ramachandran, 2004; Løken, 2007; Bhowmik *et al.,* 2017).

Later studies confirm the continuing importance of multi-criteria methods both in planning renewable energy development and in managing energy problems (Kumar *et al.,* 2017; Mardani *et al.,* 2017). Their main advantage is procedural transparency: the decision-maker defines the alternatives and criteria, specifies how assessments are aggregated, and, with unchanged data and parameters, the result can be reproduced and checked.

At the same time, this formality is a limitation from the perspective of a typical prosumer. The model must first be built, and the importance of individual criteria is often expressed through weights. The choice of weights may reflect the user's preferences, but it may also introduce subjectivity and strongly affect the result. In addition, the structure of the problem changes over time: a new vehicle may appear, the hours when residents are at home may change, the heating method may change, or economic conditions may change. A classical model can be updated, but this requires the data, constraints, or weights to be defined again.

Large language models offer a different mechanism of decision support. Their practical advantage is the ability to combine numerical data with descriptive information and conduct analysis in natural language (Feuerriegel *et al.,* 2024). Handler, Larsen and Hackathorn (2024) indicate that LLMs can support successive stages of the decision-making process, from obtaining and interpreting information to generating alternatives.

In the prosumer context, Oprea and Bâra (2024) proposed a recommendation system that uses an LLM to generate recommendations from data on production, demand, and market conditions. These studies confirm the potential of personalized recommendations, but they concern systems prepared specifically for a particular application.

In contrast, this article evaluates general-purpose tools that are available without building a dedicated system. This difference is important. The user does not control how information is internally aggregated, and the importance assigned to individual criteria is not explicit. A model may therefore correctly identify all relevant factors but rank actions differently from another tool that receives identical data. Flexibility in interpreting the problem is an advantage, but it reduces the formal transparency of the result.

A second limitation is the possibility of generating content that is not supported by the data. Hallucination in natural language generation is well documented for both NLG systems and modern large language models (Ji *et al.,* 2023; Huang *et al.,* 2025; Farquhar *et al.,* 2024).

In a prosumer problem, the consequence may be, for example, assuming that each vehicle charging session uses the full battery capacity, assuming an unspecified charger power, or giving an exact PV operating schedule without information about the orientation of the installation. Therefore, assessment of AI usefulness should consider not only the number and detail of recommendations but also how the model handles uncertainty and missing data.

3. Methodology

3.1 Study Design and Household Characteristics

The study was designed as a qualitative comparative analysis of the use of publicly available generative AI tools to support prosumer decisions. It did not assess the maximum capabilities of the models or systems using APIs, agents, external automation workflows, or locally installed models. The adopted perspective was that of a user who has basic information about their own household and uses a standard conversational interface.

The analysis covered two real households with PV micro-installations below 10 kWp, located under comparable regional conditions. Typical annual production of both installations is about 10 MWh, while the highest observed values reach about 12 MWh. The households differ in electricity consumption, the structure of major loads, and the temporal usage profile. This design makes it possible to examine whether recommendations are based on the data of a specific case or are limited to general advice for PV owners.

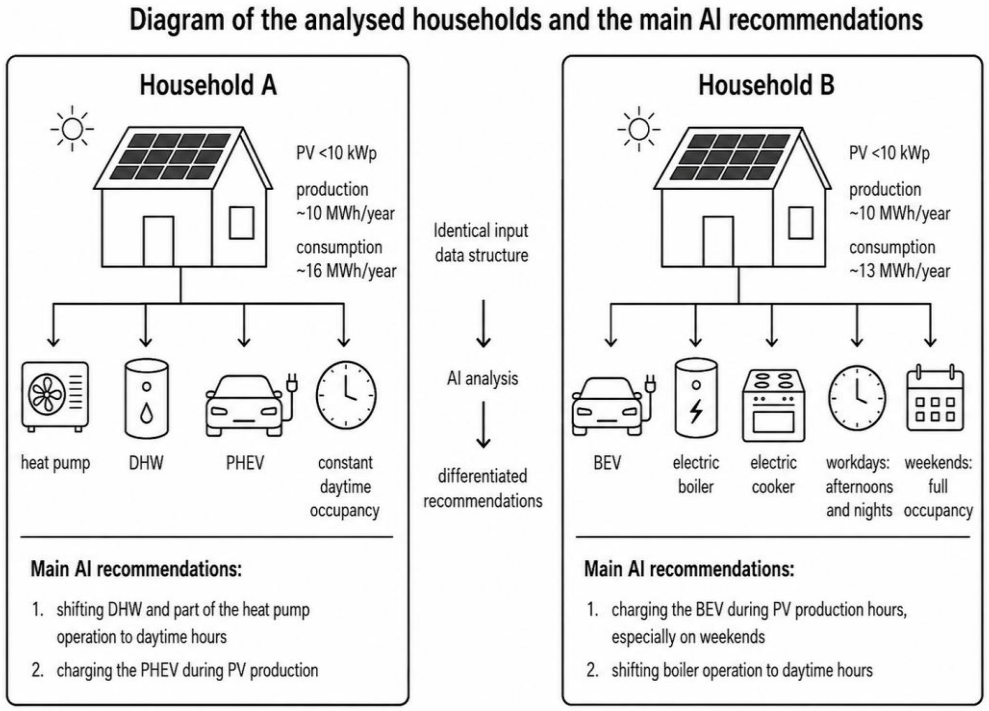
Table 1. *Characteristics of the analyzed households*

Parameter	**Household A**	**Household B**
PV installation capacity	<10 kWp	<10 kWp
Typical / maximum annual PV production	~10 / ~12 MWh	~10 / ~12 MWh
Annual electricity consumption	~16 MWh	~13 MWh
Settlement system	discount scheme 1:0.8	discount scheme 1:0.8
Main loads	heat pump, PHEV	BEV, electric water heater, electric cooker
Heating and DHW	heat pump; 200 l DHW tank	120 l water heater, 3 kW
Annual heat pump electricity consumption	14-15 MWh including DHW	not applicable
Vehicle	PHEV, 18 kWh	BEV, 88 kWh
Vehicle energy consumption	~21 kWh/100 km	~21 kWh/100 km

Parameter	Household A	Household B
Annual mileage	<10,000 km	≥15,000 km
Building usage profile	continuous daytime use	weekdays: mainly afternoons and nights; weekends: full use

Source: *Author's own elaboration based on data from the analyzed households.*

Figure 1. *Diagram of the analyzed households*



Source: *Author's own elaboration using AI.*

3.2 Tool Selection and Procedure Standardization

ChatGPT, Google Gemini, and Microsoft Copilot were selected for the study. The selection criterion was their wide availability and the possibility of conversational data analysis without specialist infrastructure.

The tests were conducted in the second half of August 2026 using newly created free accounts or anonymous access where the service allowed it. No commercial plans were used. This assumption was intended to reflect the situation of a potential user who wants to obtain support without paying for access to the model.

Each tool received the same range of data categories and an identically structured instruction. For household A, all systems received exactly the same dataset A, and for household B, exactly the same dataset B. The order of information was kept constant. Each test was conducted in a new session so that earlier context would not influence the response. After the first complete response was obtained, no corrective questions were asked and no data were clarified for only one model.

Standardization is particularly important in studies of generative AI because the result may depend on how the prompt is formulated. The adopted procedure does not eliminate the internal variability of the models, but it reduces the influence of the author on differences between the three tools. The model or service variant available during each test and the session start time were also recorded, which is reasonable because generative AI services are updated frequently.

3.3 Prompt Construction and Response Analysis Method

The prompt was designed for practical use and did not employ advanced prompt-engineering techniques. The same structure of input information was used for both households, with only the values and characteristics specific to each case being changed. The general structure of the data is presented in Table 2, and the full text of the instruction given to each analyzed tool is provided below.

Table 2. *General structure of input data provided to the AI tools*

Data category	Scope of information
Location	town or region relevant to PV production conditions
Photovoltaic installation	installation capacity and typical and maximum annual energy production
Energy balance	annual household electricity consumption
Settlement system	mechanism for settling energy exported to and imported from the grid
Main loads	devices responsible for a significant share of electricity consumption
Heating and DHW	type of devices, tank capacities, and known operating parameters
Vehicle	vehicle type, battery capacity, energy consumption, mileage, range, and charging method or frequency, if this information was available
Usage profile	presence of residents and typical periods of use of the building and loads

Source: *Author's own elaboration.*

3.4 Original Polish Prompt Used in the Study

Posiadam gospodarstwo domowe z instalacją fotowoltaiczną i chciałbym lepiej wykorzystać produkowaną energię oraz ograniczyć koszty zakupu energii z sieci. Zależy mi przede wszystkim na rozwiązaniach możliwych do zastosowania bez pogorszenia komfortu użytkowania gospodarstwa.

Poniżej podaję dostępne dane: [DANE GOSPODARSTWA]

Na podstawie tych informacji:

1. Oceń obecną sytuację energetyczną gospodarstwa i wskaż najważniejsze problemy lub obszary, w których sposób wykorzystania energii można poprawić.
2. Zaproponuj działania, które można wdrożyć bez ponoszenia dodatkowych kosztów inwestycyjnych. Wskaż w szczególności, czy sposób lub czas użytkowania któregoś z odbiorników można zmienić tak, aby zwiększyć wykorzystanie energii produkowanej przez instalację fotowoltaiczną.
3. Jeżeli rekomendujesz zmianę czasu pracy określonych urządzeń, podaj praktyczne, orientacyjne przedziały godzinowe, w których ich użytkowanie byłoby najbardziej korzystne. Uwzględnij lokalizację gospodarstwa oraz różnice występujące pomiędzy poszczególnymi porami roku. Jeżeli dostępne informacje nie pozwalają na wiarygodne określenie takich przedziałów, wyraźnie to zaznacz.
4. Uszereguj proponowane działania od najbardziej do najmniej zasadnego z punktu widzenia spodziewanych korzyści ekonomicznych i wyjaśnij krótko przyczynę takiej kolejności.
5. Oceń, czy na podstawie przekazanych danych można wiarygodnie wskazać zasadność wykonania dodatkowej inwestycji związanej z poprawą wykorzystania energii produkowanej przez instalację PV. Jeżeli tak, wskaż możliwe rozwiązanie i je uzasadnij. Jeżeli dostępne dane są niewystarczające do takiej oceny, wskaż, jakich informacji brakuje.

Nie przyjmuj samodzielnie wartości parametrów, których nie podałem. Jeżeli do sformułowania określonej rekomendacji potrzebujesz dodatkowych informacji, wskaż je zamiast zastępować brakujące dane własnymi założeniami. Oddziel wnioski wynikające bezpośrednio z przekazanych danych od przybliżeń wynikających z ogólnej wiedzy.

3.4a English Translation for the Reader

I have a household with a photovoltaic installation and I would like to make better use of the energy it produces and reduce the cost of purchasing electricity from the grid. I am primarily interested in solutions that can be implemented without reducing household comfort.

Below I provide the available data: [HOUSEHOLD DATA]

Based on this information:
1. Assess the current energy situation of the household and identify the main problems or areas in which energy use could be improved.

2. Propose actions that can be implemented without additional investment costs. In particular, indicate whether the way or timing of using any load can be changed to increase the use of energy produced by the photovoltaic installation.
3. If you recommend changing the operating time of specific devices, provide practical approximate time ranges in which their use would be most beneficial. Consider the location of the household and differences between seasons. If the available information is insufficient to determine such time ranges reliably, state this clearly.
4. Rank the proposed actions from the most to the least justified in terms of expected economic benefits and briefly explain the reason for this order.
5. Assess whether the provided data are sufficient to reliably determine whether an additional investment related to improving the use of energy produced by the PV installation is justified. If so, indicate a possible solution and justify it. If the available data are insufficient for such an assessment, indicate what information is missing.

Do not independently assume values for parameters that I have not provided. If additional information is needed to formulate a specific recommendation, indicate what information is needed instead of replacing missing data with your own assumptions. Separate conclusions that follow directly from the provided data from approximations based on general knowledge.

The responses were analyzed qualitatively without assigning scores. Seven aspects were assessed: use of the provided data, individualization of responses for households A and B, factual correctness, feasibility of recommendations, prioritization of actions, recognition of missing information, and quality of justification. The aim was not to create a ranking of the models but to identify characteristic differences in how they solved the same problem. The analysis included six responses: one response from each of the three tools for each household.

4. Results and Discussion

4.1 Recommendations for households A and B

All analyzed tools correctly recognized the basic problem in both households: annual PV production alone does not determine how effectively the energy is used because the timing of production and consumption must also be matched. Each model identified the possibility of increasing direct use of PV production by changing the operating schedule of some loads.

At the same time, the recommendations differed depending on the structure of devices and the way the building was used, which confirms that the models did not limit themselves to the same list of general recommendations.

In household A, all tools noted the dominant share of the heat pump in electricity consumption and the seasonal mismatch between heating demand and PV production. A common recommendation was to shift DHW preparation and, where possible, part of the heat pump operation to daytime hours. All models also identified the PHEV as a potentially flexible load.

The largest difference concerned the order of actions. Gemini ranked changing the DHW preparation time first, followed by PHEV charging. Copilot identified changing the heat pump operating schedule as the most important action, linking the priority to its very high share of consumption. ChatGPT ranked shifting PHEV charging first, followed by DHW preparation, and only then changes to the heating operation of the heat pump. Its justification placed greater emphasis on how easily consumption could be shifted without reducing comfort.

This difference shows a key feature of generative decision support: the same parameters may be recognized as important, but an implicit weighting process may lead to a different hierarchy. Copilot prioritized the scale of electricity consumption, whereas ChatGPT gave greater importance to flexibility and the organizational cost of change. Neither reasoning process was based on an explicitly defined set of weights, as would be the case in a classical MCDA method.

For household B, the level of agreement was higher. All three tools identified the BEV and electric water heater as the main loads that could be shifted in time, and they gave the highest priority to vehicle charging. Weekends were treated as particularly important because the vehicle may remain at home during PV production hours. Copilot and ChatGPT correctly linked annual mileage of at least 15,000 km and consumption of about 21 kWh/100 km with a minimum annual vehicle energy demand of about 3.15 MWh. For a typical mileage of about 300 km per week, they also indicated demand of about 63 kWh/week.

The second common recommendation was to change the operating time of the water heater. The models interpreted the 120-liter tank as a simple form of energy storage in the form of heat. Gemini went furthest by suggesting that the temperature could be increased to as much as 60-65°C, subject to technical suitability. Copilot and ChatGPT focused mainly on shifting the heating time without changing the set temperature. The difference is important in practical terms: both approaches may increase daytime energy use, but the first interferes more strongly with the way the device is operated.

4.2 Use of Incomplete Data and Correctness of Recommendations

The clearest error resulting from an unjustified assumption appeared in Gemini's response for household A. The model estimated annual PHEV energy consumption at about 3-3.5 MWh by multiplying about 14 charging days per month by the full battery capacity of 18 kWh.

However, the input data did not state that each charging session starts with an empty battery and uses the full 18 kWh. Moreover, the result was difficult to reconcile with the stated annual mileage below 10,000 km and consumption of about 21 kWh/100 km. In the same case, ChatGPT calculated only a theoretical upper limit based on mileage and unit consumption, while noting that actual electricity consumption must be lower because the vehicle is a PHEV.

Gemini's second error concerned household B and the terminology used for the settlement system. The 1:0.8 discount mechanism was described as 'net-billing in the form of a discount'. This did not change the general direction of the recommendation to increase self-consumption, but it shows the risk of incorrectly interpreting a concept even when the subsequent reasoning remains approximately correct. In practical use, such errors are important because a user may confuse fluent wording with reliable economic or regulatory knowledge.

All tools provided approximate time windows of increased PV production, most often covering the middle of the day in spring and summer and a shorter window in winter. Copilot and ChatGPT more clearly stated that the exact hours depend on orientation, tilt, shading, and the actual profile of the installation. ChatGPT additionally noted that more precise values without these data would create false precision. This is desirable behavior from a decision-support perspective: the model still provides useful approximate information but does not present it as a measurement from the specific installation.

With regard to additional investments, all systems showed relative caution. Energy storage appeared spontaneously as a possible option in the responses, but none of the models considered its purchase to be clearly justified from the available data. The tools mainly identified the actual level of self-consumption, energy exported to and imported from the grid, hourly production and consumption profiles, and the cost of a potential investment as necessary information.

Copilot extended this list to include information about the possibility of programming the water heater and BEV charging. ChatGPT additionally pointed to the missing vehicle charging power, without which it is difficult to assess what share of weekly demand can be covered during the weekend PV production window.

Table 3. *Main differences between the analyzed tools*

Area	**Gemini**	**Copilot**	**ChatGPT**
Use of data	broad, but in some places supplemented with assumptions	consistent, often conditional on additional data	broad, with separation of data from estimates
Priority - household A	DHW / PHEV	heat pump	PHEV / DHW

Area	Gemini	Copilot	ChatGPT
Priority - household B	BEV	BEV	BEV
Response to missing data	some gaps replaced with assumptions	strong tendency to identify required parameters	recommendation based on available data + indication of accuracy limits
Time ranges	practical, but sometimes overly certain	approximate with qualifications	approximate with clear indication of precision limits
Additional investment	no clear recommendation	no clear recommendation	no clear recommendation

Source: *Author's own elaboration based on the study results.*

4.3 Discussion: AI as Decision Support Versus Classical MCDA

The results confirm that publicly available generative models can provide practical decision support at the stage of an initial analysis of a prosumer household. All three systems used both numerical and descriptive data, recognized differences between the households, and identified actions that could be implemented without additional expenditure. It is particularly important that the study used free versions of the services. This means that basic support does not require specialist software or the construction of a mathematical model.

However, this does not mean that generative AI is equivalent to a classical optimization model. In MCDA, the criteria and their importance are explicit. The user can check why a given alternative received a higher score and can reproduce the calculation. In generative AI, the equivalents of weights are in practice hidden in the way the model interprets the context.

The different prioritization of the heat pump, DHW, and PHEV in household A directly illustrates this feature. All tools identified the same loads but assigned them different ranks. From the user's perspective, the result is easier to obtain but less transparent.

A second difference concerns the response to incomplete data. A classical model usually cannot operate without a required value or must explicitly use a scenario. A language model, however, may fill a missing value with its own assumption and then present the result in a convincing form. Gemini's estimate of PHEV energy consumption shows that this is a real rather than only theoretical risk. Therefore, the desirable feature of a system is not maximum detail in the response but the ability to show the boundary between information supported by the data and an assumption.

The differences between the tools can be described as different strategies for supporting the user. Gemini more often aimed to provide a complete and detailed

response by filling data gaps with assumptions. This increased the fluency of the recommendations but also the risk of error. Copilot used a more cautious approach: it more often made further conclusions dependent on obtaining additional parameters. This increased formal caution but also increased the length of the response and the number of conditions.

ChatGPT took an intermediate position: it more often formulated actions that could be applied with the available data while indicating which elements were only approximate. These differences do not justify a universal ranking, but they are practically important for how the tools are used.

The results are consistent with the direction indicated in the literature on LLMs as decision-support tools: the models are flexible, can combine heterogeneous information, and can generate alternative actions, but they introduce new problems related to reliability, interpretability, and control of the decision-making process (Handler *et al.,* 2024).

For prosumers, this means that generative AI may be particularly useful as an interactive 'first level' of analysis: it can help define the problem, identify the most important loads, recognize missing data, and propose low-cost actions. However, it should not independently determine investment decisions that require precise economic or technical assessment.

The analysis was qualitative and was deliberately conducted from the perspective of a typical user of publicly available AI tools. Its purpose was not to provide a technical benchmark of the models, but to examine whether, with the same structure of input data and an identical instruction, they can interpret the situation of a specific prosumer, differentiate recommendations, and identify the information needed for further analysis.

5. Conclusions, Proposals, and Recommendations

The study indicates that publicly available generative artificial intelligence tools can provide useful support to prosumers in analyzing how energy is used. ChatGPT, Gemini, and Microsoft Copilot were able to use data from two real households and differentiate their recommendations despite similar PV installation parameters.

In particular, they recognized the importance of the temporal consumption profile, the flexibility of vehicle charging, DHW preparation, and heat pump operation. The result provides a positive but conditional answer to the main research question: AI can support decision-making if its response is treated as supporting analysis rather than as an autonomous decision.

The most important difference between generative AI and classical MCDA concerns how importance is assigned to individual criteria. In a multi-criteria model, the

criteria and weights are explicit, whereas in a generative tool their relative importance results from the model's internal interpretation of the context. The study demonstrated this in practice: the same data led to different rankings of actions for household A.

This does not reduce the usefulness of AI, but it shows that the two approaches are complementary. AI supports rapid identification of the problem, relationships, and possible actions, while MCDA provides a formal way to compare alternatives when the user needs an explicit structure of criteria and weights.

For the user, a practical procedure may therefore have two stages. First, an AI tool can be used to identify the largest loads, possible opportunities to shift demand, and missing information. Second, decisions that require financial expenditure should be verified using data on actual energy flows, temporal profiles, prices, and technical parameters. In the analyzed households, particularly important data would include hourly or at least daily production, import and export values, the actual vehicle charging power, the programmability of devices, and the current level of self-consumption.

A natural direction for further research is the direct combination of generative AI with a formal multi-criteria approach. In the first variant, the same set of criteria and weights could be defined for a classical MCDA model and then provided to the analyzed AI tools. The resulting recommendation rankings and their response to changes in weights could then be compared. In the second variant, AI models could propose weights on the basis of household data and stated user preferences, and the usefulness of these weights could then be verified in a formal MCDA model.

This would make it possible to assess whether generative AI can serve not only as a source of recommendations but also as a tool supporting the construction stage of a multi-criteria model. An additional extension could use hourly profiles to quantitatively verify the economic effect of the recommended actions.

The most useful feature of generative AI in the analyzed application is therefore not the automatic identification of an 'optimal' decision. Its value lies in lowering the entry barrier to the analysis of a multi-criteria problem: the user can start with the data that are actually available, obtain a meaningful interpretation, and learn what information is needed for the next stage. Responsible use, however, requires verification of numerical values, caution toward false precision, and keeping the final decision with the human user.

References:

Beck, T., Kondziella, H., Huard, G., Bruckner, T. 2016. Assessing the Influence of the Temporal Resolution of Electrical Load and PV Generation Profiles on Self-Consumption and Sizing of PV-Battery Systems. Applied Energy, 173, 331-342.

Beaudin, M., Zareipour, H. 2015. Home Energy Management Systems: A Review of Modelling and Complexity. Renewable and Sustainable Energy Reviews, 45, 318-335.

Bhowmik, C., Bhowmik, S., Ray, A., Pandey, K.M. 2017. Optimal Green Energy Planning for Sustainable Development: A Review. Renewable and Sustainable Energy Reviews, 71, 796-813.

Bloess, A., Schill, W.P., Zerrahn, A. 2018. Power-to-Heat for Renewable Energy Integration: A Review of Technologies, Modeling Approaches, and Flexibility Potentials. Applied Energy, 212, 1611-1626.

Bywalec, C. 2009. Ekonomika i finanse gospodarstw domowych. Warszawa: Wydawnictwo Naukowe PWN.

Campbell, J.Y. 2006. Household Finance. The Journal of Finance, 61(4), 1553-1604.

Chudzicki, M. 2020. Photovoltaics as a Tool for Managing Household Finances. Proceedings of the 35th International Business Information Management Association Conference (IBIMA), Seville, Spain, 8386-8394.

Fachrizal, R., Shepero, M., van der Meer, D., Munkhammar, J., Widén, J. 2020. Smart Charging of Electric Vehicles Considering Photovoltaic Power Production and Electricity Consumption: A Review. ETransportation, 4, 100056.

Farquhar, S., Kossen, J., Kuhn, L., Gal, Y. 2024. Detecting Hallucinations in Large Language Models Using Semantic Entropy. Nature, 630, 625-630.

Feuerriegel, S., Hartmann, J., Janiesch, C., Zschech, P. 2024. Generative AI. Business & Information Systems Engineering, 66, 111-126.

Fischer, D., Madani, H. 2017. On Heat Pumps in Smart Grids: A Review. Renewable and Sustainable Energy Reviews, 70, 342-357.

Handler, A., Larsen, K.R., Hackathorn, R. 2024. Large Language Models Present New Questions for Decision Support. International Journal of Information Management, 79, 102811.

Hoffmeyer-Zlotnik, J.H.P., Warner, U. 2009. Private Household Concepts and Their Operationalisation in Cross-National Social Surveys. Metodološki zvezki, 6(1), 1-26.

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., Liu, T. 2025. A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. ACM Transactions on Information Systems, 43(2), Article 42, 1-55.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y.J., Madotto, A., Fung, P. 2023. Survey of Hallucination in Natural Language Generation. ACM Computing Surveys, 55(12), Article 248, 1-38.

Kaschub, T., Jochem, P., Fichtner, W. 2016. Solar Energy Storage in German Households: Profitability, Load Changes and Flexibility. Energy Policy, 98, 520-532.

Kelm, P., Mieński, R., Wasiak, I. 2025. Control of an Energy Storage System in the Prosumer's Installation Under Dynamic Tariff Conditions. Energies, 18(23), 6313.

Kumar, A., Sah, B., Singh, A.R., Deng, Y., He, X., Kumar, P., Bansal, R.C. 2017. A Review of Multi Criteria Decision Making (MCDM) Towards Sustainable Renewable Energy Development. Renewable and Sustainable Energy Reviews, 69, 596-609.

Løken, E. 2007. Use of Multicriteria Decision Analysis Methods for Energy Planning Problems. Renewable and Sustainable Energy Reviews, 11(7), 1584-1595.

Luthander, R., Widén, J., Nilsson, D., Palm, J. 2015. Photovoltaic Self-Consumption in Buildings: A Review. Applied Energy, 142, 80-94.

Mardani, A., Zavadskas, E.K., Khalifah, Z., Zakuan, N., Jusoh, A., Nor, K.M., Khoshnoudi, M. 2017. A Review of Multi-Criteria Decision-Making Applications to Solve Energy Management Problems: Two Decades from 1995 to 2015. Renewable and Sustainable Energy Reviews, 71, 216-256.

Munkhammar, J., Grahn, P., Widén, J. 2013. Quantifying Self-Consumption of On-Site Photovoltaic Power Generation in Households with Electric Vehicle Home Charging. Solar Energy, 97, 208-216.

Oprea, S.V., Bâra, A. 2024. A Recommendation System for Prosumers Based on Large Language Models. Sensors, 24(11), 3530.

Pinthurat, W., Surinkaew, T., Hredzak, B. 2024. An Overview of Reinforcement Learning-Based Approaches for Smart Home Energy Management Systems with Energy Storages. Renewable and Sustainable Energy Reviews, 202, 114648.

Pohekar, S.D., Ramachandran, M. 2004. Application of Multi-Criteria Decision Making to Sustainable Energy Planning-A Review. Renewable and Sustainable Energy Reviews, 8(4), 365-381.

Salom, J., Marszal, A.J., Widén, J., Candanedo, J., Lindberg, K.B. 2014. Analysis of Load Match and Grid Interaction Indicators in Net Zero Energy Buildings with Simulated and Monitored Data. Applied Energy, 136, 119-131.

Świecka, B., Musiał, M. 2016. Diagnoza finansów gospodarstw domowych w największych miastach Polski. Finanse, Rynki Finansowe, Ubezpieczenia, 4(82/2), 821-832.

Widén, J. 2014. Improved Photovoltaic Self-Consumption with Appliance Scheduling in 200 Single-Family Buildings. Applied Energy, 126, 199-212.