buildings

MDPI

Article

Exploring the Potential of Emerging Digitainability—GPT Reasoning in Energy Management of Kindergartens

Nebojša Jurišević [1,*], Dušan Gordić [1], Danijela Nikolić [1,*], Aleksandar Nešović [2] and Robert Kowalik [3]

[1] Faculty of Engineering, University of Kragujevac, 34000 Kragujevac, Serbia
[2] Institute for Information Technologies, University of Kragujevac, 34000 Kragujevac, Serbia
[3] Faculty of Environmental Engineering, Geodesy and Renewable Energy, Kielce University of Technology, 25-314 Kielce, Poland
* Correspondence: jurisevic@kg.ac.rs (N.J.); danijelan@kg.ac.rs (D.N.)

Abstract: One of the barriers to the rapid transition of societies toward a more sustainable future is a scarcity of field experts. Members of scientific and professional communities believe that this obstacle could be overcome by supplementing the decisions of non-experts with artificial intelligence. To examine this opportunity, this study examines the viability of GPT-3.5 as an expert adviser in the energy management of kindergartens. Thus, field experts investigated the deductive and inductive reasoning potential of GPT-LLM (Large Language Model). The first task was conducted on a sample of kindergartens in the Western Balkans. The LLM was instructed to provide the buildings' specific heat consumption (SHC) by relatively detailed building descriptions and building occupancy. The second task involved kindergartens in various European locations, and the LLM was tasked with estimating energy savings using limited data about the renovation process. The study found deductive reasoning to be insufficient for estimating SHC from the building envelope details, with average accuracy below the least predictive model (R^2 = 0.56; MAPE = 48%). Including the factor of occupancy, the SHC estimates were relatively accurate, wherein the first deductive test proved precise (MAPE = 27%), but it was less so in the opposite case (MAPE = 67%). In terms of inductive reasoning, the LLM assumptions were relatively consistent with practice.

Keywords: GPT; GPT-3.5; energy management; educational buildings; kindergartens; digitainability

check for updates

Citation: Jurišević, N.; Gordić, D.; Nikolić, D.; Nešović, A.; Kowalik, R. Exploring the Potential of Emerging Digitainability—GPT Reasoning in Energy Management of Kindergartens. *Buildings* **2024**, *14*, 4038. https://doi.org/10.3390/buildings14124038

Academic Editors: Antonio Caggiano and Alessandro Cannavale

Received: 1 November 2024
Revised: 1 December 2024
Accepted: 14 December 2024
Published: 19 December 2024

1. Introduction

Due to ever-increasing scientific progress, members of modern societies must adapt to social and technological changes (STCs) faster than previous generations [1]. In the preceding saeculum, the pace of technological change was predetermined by society's ability to automate industries and establish diverse service sectors [2]. In contrast, contemporary saeculum STCs are driven by networking, digitalization, and the ever-increasing presence of artificial intelligence (AI) [3]. The key disparity between the two periods is that the former's dynamic was determined by infrastructure development (e.g., roads, railroads, and the internet), whereas the latter is not. Furthermore, previous technological advancements primarily impacted working-class jobs, while novel technology focuses on decision making, influencing mainly white-collar occupations [4]. As a result, future human progress should be faster than in the past [5], and decisions affecting it will be made with less effort [6]. Ideally, this should enable shared prosperity for humanity, prevent global conflicts, and promote overall well-being [7]. By harnessing the synergy of sustainable ideas and digitalization, defined as digitainability [8], the prospects for a more sustainable future should be brighter than they were previously.

1.1. Subject of Research

One of the relatively significant technological advances in the AI field pertains to the development of Large Language Models (LLMs)—algorithms specifically designed to simulate conversations with human users [9]. Although chatbots have been in use since 1966 [10], they have only recently gained widespread attention due to significant improvements in their usability [11]. These advancements were made possible by progress in natural language processing (NLP) algorithms [12] employing unsupervised learning techniques. Unlike the supervised NLPs utilized by some chatbots before, the latter does not require explicit human instructions or data labeling for LLM chatbot training. This allows prompt learning on large amounts of textual data, and this approach was applied to OpenAI's Generative Pre-trained Transformer (GPT) [13]. To be as exhaustive as possible, GPT was trained on vast amounts of data (Table 1) gathered from different sources (Common Crawl [14], WebText2 [15], Wikipedia [16], and two separate sets of books available on the internet (Books1 and Books2).

Table 1. GPT-3 knowledge base [17].

Dataset	Number of Tokens	Training Mix
Common Crawl (filtered)	490 billion	60%
WebText2	19 billion	22%
Books1	12 billion	8%
Books2	55 billion	8%
Wikipedia	3 billion	3%

The text on which the GPT was trained was divided into smaller units of words or sub-words (tokens). Each of the tokens had embedding that allowed the model to understand the context and the relationship between the words. In this context, the text output LLM provides is based on predictions of the tokens that follow the textual input sequence. To reduce harmful bias and factual inaccuracies, the data LLM was trained on were thoroughly cleaned and filtered. This could include techniques such as identifying and removing harmful stereotypes, flagging potential misinformation, and maintaining data quality standards [17]. Upon the training, the model underwent a fine-tuning process, i.e., adaptation of the pre-trained model to a new task. This can be accomplished by prompt-based fine-tuning, in which the user provides directions for the LLM on how to come to output; or few-shot learning, in which the LLM adapts to a new task following given examples [17].

As a result of the new technology's development, members of the scientific and professional communities began to investigate the opportunities for GPT application in augmenting (non-)experts' knowledge, highlighting the novel technology's strengths and weaknesses. Table 2 provides a brief overview of the studies that have addressed this topic.

Table 2. Short overview of the studies examining GPT usability in a variety of professions.

Field	Ref.	Country	Study Aim	Study Outcome	Stated Concerns/Downsides
Industry	[18]	United Kingdom	To investigate how GPT can be used to reduce waste generation, improve product quality, and achieve sustainability in the textile industry.	By utilizing GPT, companies in the textile industry can improve the customer experience and make their services more efficient, cost-effective, and prompt.	Not stated.
	[19]	United Arab Emirates	To evaluate GPT output by a pool of participants (experts); to gather feedback regarding the overall interaction experience and the quality of the GPT output.	The participants had an overall positive interaction experience and indicated the potential of such a tool in automating many preliminary and time-consuming tasks.	The response is not reliable; generic and boilerplate statements; not connected to real-time internet data.
	[20]	United Kingdom	To explore what users anticipate from AI; to gain insight into GPT's applications and the potential effects they may have soon.	GPT can improve interactive learning, simplify collaborations between students and teachers, and provide a more efficient way to store and access course materials.	Privacy and data security; potential to replace human jobs.
Environment and Sustainable Development	[21]	Brazil	To examine the usability of five LLM models in natural resources management decision making.	In the context of water management, it is possible to support human decisions by the use of conversational agents.	Not stated.
	[22]	Austria	To evaluate contributions and the potential impact of AI on sustainable development in the society domain.	AI has the potential to significantly aid in achieving sustainable development goals.	Lack of transparency concerning AI decisions; bias built into the algorithms; overreliance on automated solutions rather than human intervention.
	[23]	Austria	To investigate the benefits of AI for digitalization, urbanization, globalization, climate change, automation and mobility, global health issues, and the aging population.	GPT-3 provides easily understandable insights into the complex and cross-sectional matters of megatrends.	AI systems can make mistakes or generate wrong output.
	[24]	India	To investigate how GPT can be used to spread the concept and benefits of nearly-zero-energy buildings through the academic community.	GPT can contribute to activities aimed at spreading the benefits of sustainable development.	Not stated.
	[25]	Germany	To investigate the political reasoning, biases, and limitations of GPT.	GPT argues for pro-environmental, left-libertarian ideology. It would impose taxes on flights, restrict rent increases, and legalize abortion.	The study examined just two political orientations, i.e., Germany's Wahl-O-Mat and the Netherlands's Stem Wijzer.

Table 2. *Cont.*

Field	Ref.	Country	Study Aim	Study Outcome	Stated Concerns/Downsides
Education	[4]	Singapore	To discuss the potentials of GPT in education and research; discuss student-facing, teacher-facing, and system-facing applications; and analyze opportunities and threats.	Despite the challenges that GPT poses for traditional assessments, it will not necessarily lead to their extinction. Instead, it will encourage educators to use AI tools to create diverse assessments that evaluate deeper understanding and critical thinking.	Academic dishonesty; superficial understanding; overreliance on chatbots.
	[26]	Kenya	To explore the possibility of implementing a constructivist learning environment using chatbot technology.	Chatbot technology can contribute to education through active and social learning.	Not stated.
	[27]	United Kingdom	To establish an understanding of the ethics of AI applied in educational contexts.	While initial indicators suggest a lack of interest in the ethics of AI in education, the community recognizes its significance. To improve ethical engagement, discussions and frameworks are required to ensure ethical principles for meaningful real-world impact.	Uncertainties in equity, fairness, confidentiality, and anonymity.
	[28]	United States	(Not directly stated) Conversation was aimed to explore complex issues and propose solutions and strategies.	Not directly stated.	(Not directly stated) Limited access to external resources (references).
	[29]	United States	To evaluate the abstracts using an AI output detector, plagiarism detector, and blinded human reviewers trying to distinguish whether abstracts were original or generated.	Most generated abstracts were detected using the AI output detector. Blinded human reviewers correctly identified 68% of generated abstracts as being generated by GPT.	GPT writes believable scientific abstracts, though with completely generated data.
	[30]	India, Zambia	To understand the perceptions and opinions of academicians toward GPT by collecting and analyzing social media comments, and a survey was conducted with library and information science professionals.	While some academicians may not accept GPT-3, most are starting to accept it.	GPT reduces critical thinking and raises ethical concerns.
	[31]	United States	To evaluate the performance of GPT on questions within the scope of the United States Medical Licensing Examination Step 1 and Step 2 exams, as well as to analyze responses for user interpretability.	By performing at a greater than 60% threshold, the model achieved the equivalent of a passing score for a third-year medical student.	GPT training data were not up to date.

Table 2. *Cont.*

Field	Ref.	Country	Study Aim	Study Outcome	Stated Concerns/Downsides
	[32]	China	To evaluate GPT capabilities in open-ended question answering, factual modeling, and following instructions. The study highlights the strengths and weaknesses of the bot in comparison with human experts.	Although GPT demonstrated impressive capabilities, it still cannot replace human experts.	The study findings were based on unbalanced data.
	[33]	Slovakia, UAE, Czech Republic	To provide an up-to-date overview of upcoming changes and advancements in the use of AI in dental education.	GPT can facilitate communication between healthcare providers and patients.	Ethical and legal implications.
	[34]	Germany	To assess the quality of radiology reports simplified by GPT. The evaluation was performed by 15 radiologists.	Most radiologists agreed that the simplified reports were factually correct, complete, and not potentially harmful to the patient.	Instances of incorrect statements; missed key; medical findings.
Computing	[35]	China	To provide an overview of GPT, its features, benefits, and challenges.	GPT is a promising AI technology that can be used to automate conversations and generate more accurate responses.	Security and limited capabilities.
	[36]	United States	To assist researchers and developers in enhancing future language models and chatbots.	Despite its impressive capabilities, GPT improvement is necessary for it to excel in areas such as reasoning, mathematical problem solving, and reducing bias.	Unsatisfactory context comprehension; weak math and arithmetic skills; perception of ethics and morality; difficulty using idioms.
	[37]	United States	(Not directly stated) Highlighting potential limitations of GPT, such as its ability to generate inaccurate or meaningless content as well as raising concerns about the technology's potential harm.	(Not directly stated) GPT has limitations.	Overreliance on AI is harmful.

According to Thurzo et al. [33], ChatGPT can prompt quick decisions with reasonably accurate diagnoses and solutions, resulting in increased operational effectiveness. In terms of sustainable development, Rathore [18] explored the opportunities of ChatGPT utilization in the textile industry, indicating that technology can mitigate waste generation, improve the quality of products, and contribute to sustainability goals. Alves et al. [21] had a similar conclusion, confirming that chatbots can contribute to decision making in natural resource management. Prieto et al. [19] demonstrated that GPT can generate a coherent construction schedule for a simple construction project. According to the authors, the platform used a logical approach to completing the task scope. Other research found that AI platforms can facilitate intelligent traffic management systems [20] and improve the efficiency of supply chains [22]. The Internet of Things and artificial intelligence, in that regard, can be combined to create the AIoT (artificial intelligence of things), improving building and process performance [35]. AI-driven analytics can also be used to identify the impact of climate change on certain communities [23]. Jungwirth and Haluza [22], for example, note that ChatGPT could be useful in addressing social megatrends, though they warn that much work on the platform and its proper use is required before tangible results can be seen. This can be of particular use for both developed and developing countries' educational systems [24,26].

In contrast to just positive aspects, Holmes et al. [27] see the prior opportunities as a threat to humanity, as AI may not always reflect the values of society as a whole. Hartam et al. [25] provided converging evidence on ChatGPT's pro-environmental, left-libertarian orientation. Borji [36] created a categorical archive of ChatGPT failures, referring to false information as bot hallucinations. These errors were observed in other studies as well [4], some of which emphasized the absence of [38] or incorrectly stated references [28] as a particular issue. Marcus and Davis declared GPT to be a "not reliable interpreter of the world" [37], whereas Gao et al. [29] stated the platform can generate realistic scientific abstracts, but the data could be completely made up. Because of all this, Subaveerapandiyan et al. [30] indicate that ChatGPT should aid decisions rather than generate ideas. Consequently, the confidence in ChatGPT as an expert adviser has been examined in several professional and scientific domains.

Guo et al. [32] created a dataset of 40,000 questions and an appropriate mixture of expert and artificially generated answers to test how closely ChatGPT resembles human experts. The question–answer pairs were provided to a pool of experts and non-experts to characterize them. In comparison with expert reports, the study found that the machine writing style was relatively weak, which has also been shown in some other studies [39,40]. Because of this, successfully contrasting different styles was not as difficult a task for experts as it was for non-experts. However, non-experts understood the artificially generated answers better than the expert responses because the former were plainer and simpler. Other studies have proven that ChatGPT has sufficient "knowledge" and adequate reasoning to pass graduate exams in law and business schools, score in the top 10% on a law exam [41], and assist juristic decisions [41]. A study conducted in Turkey showed that ChatGPT performed better than anatomy students [39], while a similar study found that the bot would pass the third year at the faculty of medicine in the US [31]. Even more, Jeblick et al. [34] suggest using ChatGPT in addition to expert opinions. Moving on to more complex intellectual analyses, Borji [36] subjected the bot to a series of challenging logical tests to determine the overall potential of ChatGPT reasoning. He found it to have relatively good physical reasoning skills and particular challenges when dealing with spatial, temporal, psychological, and commonsense tasks. To summarize the reviews: ChatGPT has proven its worth both in the hands of experts (discussing the challenges of modern humanity) and in the hands of non-experts (as an advisor). However, due to the challenges that still exist in terms of AI reliability, governments of countries and regions are treating AI innovations with particular caution [42,43]. Final decisions recommending the use of technology would require years of professional and scientific evaluations to prove the technology is useful and compliant with the ethical principles present in the Data for Humanity Initiative [39].

To contribute to these efforts, this study aims to examine the usability of GPT as an advisor tool in the domain of kindergarten energy management. In this context, experts in the field of energy management evaluated the usability of ChatGPT as an advisor for non-experts. There is no similar study in the available literature. The study findings should fill existing knowledge gaps by answering the following research questions: how successfully GPT can deal with the topic of energy management in kindergartens and how useful the bot could be for energy managers. The novelty of the study lies in the exploration of ChatGPT as an advisory tool in the specific context of energy management in buildings. The study aims to inform and influence AI practice in educational and professional settings.

1.2. Object of Research

The object of the research in this study is a sample of educational buildings, i.e., kindergartens. These buildings were chosen for analysis because they accommodate the youngest population, require strict comfort control, and are prioritized in renovation efforts, making them ideal starting points for research into energy management and comfort in buildings. Depending on the latitude and level of industrial development, buildings in the EU are responsible for 60–80% of countries' final energy consumption [44], and public buildings consume about 50% more specific heat (SHC) (kWh/m^2/a) than residential buildings [45]. Because of this, buildings are the focus of modern initiatives dealing with a more sustainable future and better-organized societies [46]. One of the obstacles to the anticipated level of advancements in the field of public building energy management is a lack of subject matter experts [44]. To address this issue, scientists and professionals in the field developed a variety of simple-to-use models that enable non-experts to monitor and predict building energy consumption. Jurisevic et al. assessed the performance of various predictive models to target energy [47] and water [48] consumption in public preschool buildings, achieving up to 92% accuracy. Similar models were developed in other studies for a variety of building types, including school buildings (86% accuracy) [49], educational buildings (60%) [50], university campuses (89%) [51], banks (up to 69%) [52], and supermarkets (86% to 95%) [51]. Although the models perform relatively well, their limitation is the fact that they were developed on relatively small building samples. Consequently, the models would not accurately describe the energy performance of buildings out of the sample. Apart from this, users of the models need to have at least some field knowledge, as the results of the model are simply numbers representing either the building's energy consumption or its potential for energy savings. On the other hand, LLMs allow building operators to consult the AI platform for advice and potentially receive the right answer. No prior programming or statistical knowledge is required from the operator. Unlike predictive models, GPT answers are generated for a single building, which is an advantage over models that were developed using a sample of multiple buildings. Additionally, the GPT response would not be a number, but rather a clear and concise written response, which is another benefit for non-experts [32]. This could be one of the positive effects that novel technology could have on contemporary challenges that modern humanity encounters in pursuing a more sustainable future. In addition, GPT should execute deductive and inductive reasoning to respond to this specific challenge. This should help the scientific and professional public to gain a better understanding of the platform's reasoning skills. Contributions made by this study are consistent with the Data for Humanity Initiative [39].

2. Materials and Methods

The study used two building samples to assess the effectiveness of GPT (gpt-3.5-turbo) reasoning in managing energy in kindergartens: (1) buildings situated in the same region—a city in the Western Balkans—and (2) buildings distributed across various locations in Europe. The first building sample was used to evaluate GPT's effectiveness in predicting the energy consumption of buildings with different floor areas and different construction periods. The second set of buildings was utilized to assess the ability of GPT to estimate energy savings upon building renovation in various locations (Figure 1). The first building sample was relatively well described (Table 3), whereas the second was not as much (Table 4). Consequently, one will be used to test GPT precision (deductive reasoning), while the other to test the LLM's ability to assess building performance from a relatively subpar building description (inductive reasoning). Figure 1 depicts the locations, images, and basic information about the analyzed buildings, such as the year of construction and heated floor area. Buildings from the first study sample are shown in blue, while those from the second are shown in red squares.

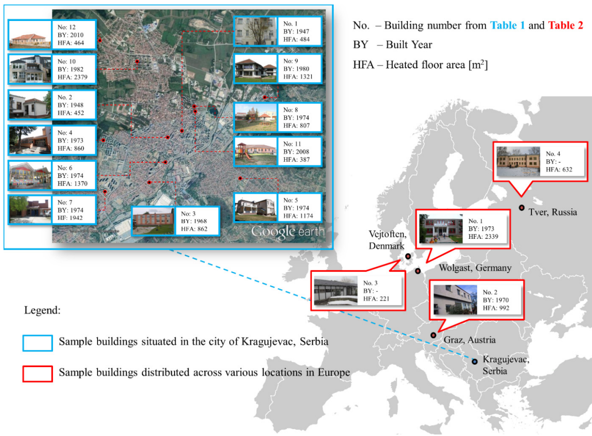

Figure 1. Building samples used for the analysis.

Table 3. Details of the first sample of educational buildings—public kindergartens located in the same city.

		Building Thermal Envelope Details (*i*)									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
		Built Year	**Number of Floors**	**External Walls Gross Area**	**Heated Floor Area**	**Gross Heated Volume**	**Gross Glazing Area**	**External Walls U-Value**	**Glazing Elements U-Value**	**Ceiling U-Value**	**Roof Type**
		[-]	[-]	[m^2]	[m^2]	[m^3]	[m^2]	[W/m^{2}K]	[W/m^{2}K]	[W/m^{2}K]	[-]
Kindergarten No (*kn*)	1	1947	3	468	484	1382	92	1.38	4.01	0.37	Flat
	2	1948	1	740	452	1429	98	1.28	3.68	1.75	Pitched
	3	1968	2	1121	862	2888	548	0.5	1.59	0.52	Flat
	4	1973	2	738	860	2580	236	1.38	3.6	0.25	Flat
	5	1974	3	1036	1174	3745	270	0.46	3.21	0.35	Pitched
	6	1974	1	764	1370	4482	499	2.0	4.26	1.4	Pitched
	7	1974	2	1942	537	5199	453	0.46	3.52	0.34	Pitched
	8	1974	2	685	807	2598	273	1.16	2.88	1.4	Pitched
	9	1980	2	2708	1321	4057	461	1.38	3.52	1.53	Pitched
	10	1982	2	2480	2379	7636	755	0.34	3.11	0.34	Pitched
	11	2008	1	311	387	1136	68	0.16	2.71	0.35	Pitched
	12	2010	1	230	464	1508	80	0.16	2.9	0.35	Pitched

Details describing the first set of buildings (Table 3) were taken from Jurišević's doctoral dissertation [44]. Twelve kindergartens were described in great detail in the building information section. Inputs used in the study were sufficient for accurately estimating the buildings' SHC, achieving performance metrics comparable to those reported in state-of-the-art approaches from the literature (R^2: 0.92; MAPE: 14%) [47]. Henceforth, this study considered the selected inputs sufficient for drawing reliable deductive conclusions when estimating the SHC of the chosen building sample.

A second set of building details to which this study refers was gathered from energy reports and scientific papers. These publications gave different and less thorough descriptions of buildings than they did of energy-saving techniques and energy savings realized. As a result, the available data were unsuitable for drawing deductive conclusions. Nevertheless, these limitations did not hinder the use of inductive reasoning, which involves deriving conclusions from a limited or insufficient set of information. Table 4 lists the available details of four kindergartens in a relatively comparative manner before and after renovation.

Table 4. Details of the second sample of educational buildings—public kindergartens distributed across Europe.

				Building Location (*l*)			
				***l*1**	***l*2**	***l*3**	***l*4**
				Vejtoften, Denmark [53]	**Wolgast, Germany [54]**	**Graz, Austria [55]**	**Tver, Russia [56]**
Before Renovation	Data Label (*k*)	6	Built year	Not stated	1973	1970	Not stated
		5	Heated floor area	221 m^2	2339 m^2	992 m^2	632 m^2
		4	Number of stories	1	2	2	2
		3	Fenestration details	Traditional double-glazed windows	Unknown	Unknown	Wooden frame windows with a total surface of 151 m^2
		2	External walls details	With 95 mm thermal insulation (not stated what type)	Unknown	Unknown	Building brick, plastered and painted, the percent of wear makes 64%
		1	Roof details	Pitched, with 145 mm thermal insulation (not stated what type)	Flat	Pitched	Pitched roof is on rafters and an obreshetka
			Energy consumption	167.4 kWh/m^2/a	158 kWh/m^2/a	Not stated	Not stated
Upon Renovation	Data Label (*j*)	5	Modernization completed in	Before 2015	2009	2010	Before 2014
		4	Fenestration details	Triple-glazed windows	Double glazing with insulating protection (U-value including frame 1.4)	Replacement of windows	Metaplastic-framed windows with a total surface of 151 m^2
		3	External walls details	With 390 mm thermal insulation (not stated what type)	Exterior wall insulation with mineral wool (15 cm, U-value 0.22)	Additional thermal insulation of external walls	Not renovated
		2	Roof details	Pitched, with 145 mm thermal insulation (not stated what type)	Roof insulation (30 cm, U-value 0.12)	Not stated	Not renovated
		1	Additional measures	In order to reduce/remove thermal bridge effects at the uninsulated base/foundation of the building, 200 mm of insulation was added on the outside to a depth of 400 mm.	Not stated	Thermal insulation of heat pipes	Not stated
			Energy consumption	91.7 kWh/m^2/a	116 kWh/m^2/a	Not stated	Not stated
			Energy or CO_2 savings	45.2%	70 t/a	70%	40%

Due to the different nature of the available data, the instructions provided to GPT for the first and second sets of buildings differ. However, to make the GPT responses suitable for fair analysis, the prompt commands were issued in the same way for all buildings from the same set. Commands to the GPT were instructed throughout OpenAI's playground [57] platform, where parameters such as temperature, maximum response length, diversity, wording frequency, and text presence penalties could be set. The parameter values this study utilized are presented in Table 5.

Table 5. OpenAI playground settings.

GPT Parameters	Parameter Value	Parameter Role [17,58]
Model	"gpt-3.5-turbo"	A deep learning model that generates text employing a neural network.
Temperature (ranging from 0 to 1)	1	Determines the randomness of the response. The more closely the temperature approaches 0, the less erratic the result will be.
Maximum length (ranging from 0 to 2048)	200	Caps a number of tokens that are allowed for a response. This varies according to the type of model.
Stop sequences (user input)	-	Makes responses end at the desired point, such as the end of a sentence or list.
Top probabilities/Top P (ranging from 0 to 1)	1	Controls which tokens the model will consider when generating a response. Setting this to 0.9 will consider the top 90% most likely of all possible tokens.
Frequency penalty (ranging from 0 to 1)	0	Controls the repetition of the same tokens in the generated response. The higher the penalty, the lower the probability of seeing the same tokens more than once in the same response.
Presence penalty (ranging from 0 to 2)	0	Reduces the chance of repeating any token that has appeared in the text. It is stricter than the frequency penalty, so it increases the likelihood of introducing new topics in a response.

2.1. GPT-3.5 Deductive Reasoning Test

To examine the usability of GPT as an adviser in kindergarten energy management, a deductive reasoning test was conducted. To evaluate GPT reasoning, the study utilized input-based prompting to initiate the bot's deductive reasoning. In this regard, the prompt instructions included: (a) building description section (D) and (b) questioning sections (Q). The order and content (italic text) of the instructions were as follows:

D1: *The public kindergarten is located in Kragujevac, Serbia. It was built in year i1 (Table 3) and has not been renovated since. The other details of the building are i = from 2 to 11 (all inputs were entered together with their units available in Table 3). The building is heated and naturally ventilated from 6:30 am to 9:30 pm.*

Q1: *How much heat is expected for the building to consume during the heating seasons [kWh/m^2/a] with the following number of heating degree days: (a) 2133 K·Day; (b) 2349 K·Day; and (c) 2510 K·Day.*

Conclusions on the quality of deductive reasoning were drawn from expert judgment based on a comparison of the GPT and mathematically based assessments in [47]. In addition, the study examined the potential of GPT to account for the impact of occupancy (i.e., occupant behavior) on building energy performance. This factor is difficult to quantify and is therefore often overlooked in the field of predictive analytics [44]. Potential advances in novel technologies that can address this challenge could enhance the calibration of predictive models and make predictions more accurate. Because the influence of occupancy on a building's energy performance is better measured in relatively small time steps, this study tested GPT deductive reasoning on a monthly rather than annual time frame. In this context, GPT was provided with the number of HDDs, calculated following Equation (1):

$$\mathrm{HDD} = \sum_{j=1}^{DHS} Tm_j - Tr \quad (1)$$

where HDD [K·Day] is the number of heating degree days, *Tm* is the mean outside temperature [K], *Tr* is the room temperature [K], *j* is the day of a heating season [-], and *DHS* is the duration of a heating season [day]. Room temperature for the examined building was set to 24 °C (297.15 K), while the monthly or seasonal HDD did not include the days with an average daily temperature higher than 12 °C (285.15 K). In addition to HDD, GPT was provided with the number of building monthly visits for two consecutive heating seasons. The task assigned to the prompt was as follows:

Q2: *Having in mind D1, assess the monthly heat consumption of the same building by adding the influence of monthly visits of the building users (children), and the number of heating degree days (HDD). The number of visits nv. How much heat building will consume that month?*

Appendices A and B contain the *hdd* and *nv* values used for each month of the studied period for the buildings analyzed.

2.2. GPT-3.5 Inductive Reasoning Test

In addition to Section 2.1, the study performed an inductive reasoning test to evaluate GPT's usability in energy management tasks with insufficient building details. In this regard, a second set of buildings was used. The GPT was instructed by contextual template-based prompting to answer the questions concerning each of the buildings individually. The order and content (italic text) of the instructions were as follows:

D2: *The public kindergarten is located in: li (Table 4). It was built in: ki (Table 4). The details of the building envelope are k1, . . . , k7 (all inputs were entered together with their units available in Table 4). The building was renovated in the year j6, and considered following improvements of the thermal envelope: j1, . . . , j5.*

Answer the following questions by relying on inductive reasoning:

Q3: *How much specific heat [kWh/m²/a] did the building consume before renovation?*

Q4: *How much specific heat [kWh/m²/a] does the building consume upon renovation?*

3. Results and Discussion

The responses GPT provided to the instructions are presented visually to make them easier to interpret. To measure the accuracy of the assessments, the study used two accuracy indicators: mean absolute error (MAE) [59] (Equation (2)) and mean absolute percentage error (MAPE) [60] (Equation (3)). In addition to MAPE, the study used the coefficient of determination (Equation (4)) [61] to compare the GPT assessments made in this study with the assessments from another study.

$$MAE = \frac{\sum_{i=1}^{n} |(y_i - \hat{y}_l)|}{n} \tag{2}$$

$$MAPE = \frac{\sum_{i=1}^{n} \frac{|(y_i - \hat{y}_l)|}{y_i} \cdot 100\%}{n} \tag{3}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (\hat{y}_l - \overline{y})^2}{\sum_{i=1}^{n} (y_i - \hat{y}_l)^2} \tag{4}$$

where n is the number instances (sample size), y_i the true value of the instance, $\hat{\boldsymbol{y}}_{\boldsymbol{l}}$ the assessed value of the instance, and $\overline{\boldsymbol{y}}$ is the mean value of the sample.

3.1. GPT-3.5 Deductive Reasoning Test

GPT responses to the Q1 set of questions are presented in Figure 2. The actual SHCs for buildings are represented by bars, while the corresponding GPT assessments are represented by dots. The bar and dot colors represent three HDD scenarios. The units used are kWh/m^2/a. As can be seen from the figure, the number of HDDs did not have a decisive influence on the buildings' SHCs. This means that relatively small changes in HDD during the heating season (~200 K·Day) do not necessarily follow seasonal changes in SHC. This could be explained by the fact that variable behavior of building occupants (as determined by the number of monthly visits and activities within the building) has a greater influence on SHC than relatively minor changes in HDD. On the other hand, the order of the GPT-assessed SHCs mainly followed the order of the heating seasons' HDDs, thus neglecting the influence of occupant behavior. This is a shortcoming of deductive reasoning, which was solely based on the data instructors provided to the prompt. On the positive side of deductive reasoning, GPT presented a comprehensive approach by listing the approach segments as bullet points (listing the inputs and calculating the total heat demand for each building and the SHC of each building). The method was systematic and simple to follow. However, the formulas used in the calculation method were oversimplified and inaccurate. The heat consumption was calculated based on the heated floor area rather than the thermal envelope area. The formula did not contain units, but rather dimensional notations. The formula "Heat Demand (kWh/m^2/a) = Heating Degree Days × Gross Heated Floor Area × U-values" used to calculate SHC was oversimplified and incorrect, both dimensionally and formally. In this context, GPT proved unable to replicate the accuracy of traditional calculations, even though the formal approach appeared systematic and logical.

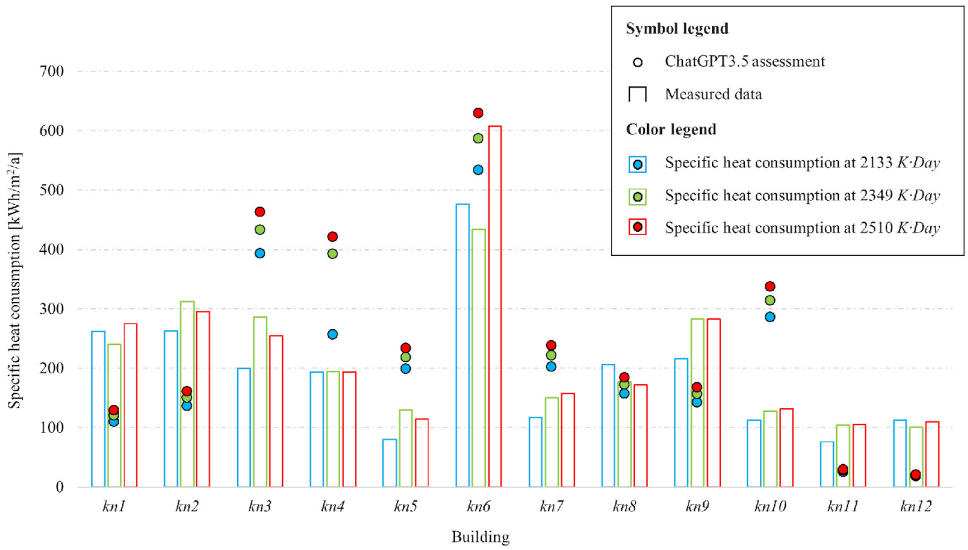

Figure 2. Comparison of kindergartens' real and GPT-assessed SHCs for three heating seasons.

When compared with the actual data, the GPT-assessed SHCs are mainly underestimated (two-thirds of the cases). The greatest underestimation in terms of MAPE was measured in the case of kn12: 469% (MAE: 88.9), and the greatest overestimation in the case of kn10: 60% (MAE: 188.4). Moreover, errors in predicting building SHC were higher when GPT underestimated the value (MAPE: 199%, MAE: 107.3) than when it overestimated it (MAPE: 37%, MAE: 117.4).

Figure 3 depicts the distribution and accuracy of the buildings' actual and GPT-assessed SHCs across different consumption ranges. The x-axis represents the actual SHC, while the y-axis represents the GPT-assessed SHC. Each dot represents the SHC of a building over one heating season. In terms of SHC consumption ranges, the MAE indicators for scenarios with less than 150 kWh/m^2/a and those between 150 and 400 kWh/m^2/a were relatively similar (108.5 and 123.3, respectively). MAPE values for the two same-span categories were 97% and 52%, respectively. Regarding the SHCs greater than 400 kWh/m^2/a, GPT overestimated all the consumptions by 17% on average. The overall coefficient of determination (R^2) between real and GPT-assessed data was 0.38, with a MAPE of 67%. In this context, the most intuitive and least precise statistical model (simple linear regression

(SLR)) developed on the same set of buildings [47] outperformed GPT by around 55% in terms of R^2 and 51% in terms of MAPE. Moreover, SLR required only the HDD and building heated floor areas to provide estimations, whereas the LLM was given five times as many inputs. This performance was significantly lower than the performance of more advanced predictive algorithms developed for the same building sample (multiple linear regression (R^2: 0.88; MAPE: 31%), Decision Tree (R^2: 0.84; MAPE: 25%), and Evolutionary assembled artificial neural network (R^2: 0.92; MAPE: 14%)).

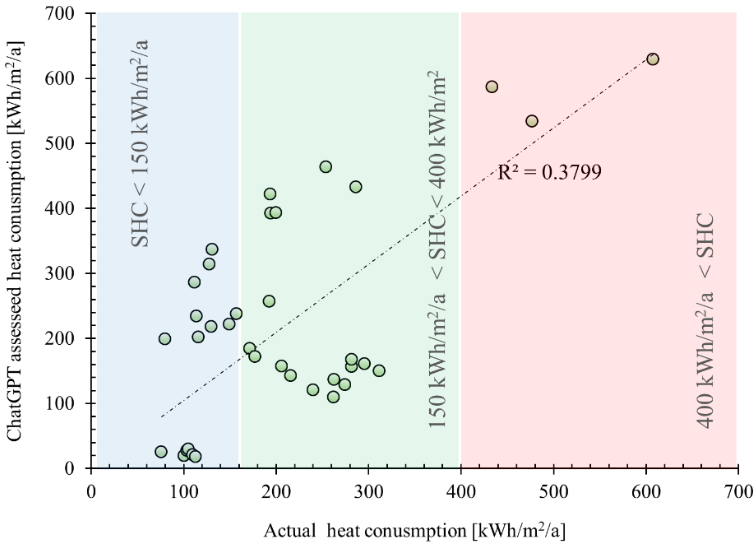

Figure 3. Comparison of buildings' real and GPT-assessed SHCs in different consumption ranges.

To investigate GPT's ability to use occupancy as a factor affecting SHC, this study examined the cases of buildings where the LLM previously assessed the SHC with (1) highest (kn7: MAPE = 13%, MAE = 76.9) and (2) lowest accuracy (kn10). Although most predictive models dealing with energy management in public buildings neglect occupancy as a factor affecting heat consumption, there is no doubt this feature influences the SHC. By Q2, a comparison of two buildings' real and GPT-assessed heat consumption is presented in Figure 4 (due to data availability and data filtering, Figure 4a,b do not represent the same consecutive heating seasons). The bottom axis of both graphs represents the month to which the measurements (SHC, number of visits) relate, while the upper axis shows the number of HDDs for each corresponding month. The data for kn10 and kn7 are available in Appendices A and B, respectively. The blue dots in the graph indicate the real SHC of kindergartens, while the green dots are GPT-assessed SHC. SHC values are shown on the left y-axis, while the number of monthly visits, (represented by red crosses on the graph), is indicated on the right y-axis.

Variations in GPT-assessed heat consumption (HC) relatively fairly followed the variations in the real data. The coefficient of determination between real and LLM-assessed values was the same (0.59), although the assessed values provided a much better fit in the case of kn7 than in the case of kn10, with just two dots being out of the ground truth pattern. As for MAPE, the average error of the GPT estimates for kn10 was 67% (MAE: 39,067), while for kn7 it was 27% (MAE: 730). This suggests that LLM algorithms can reasonably predict the influence of occupancy on HC, but only in kindergartens where they have previously proven to be reliable at predicting SHC. To respond to Q1 and Q2, GPT applied formulas, explaining them step by step. The approach was not entirely correct, nor were the formulas used. In this sense, some of the formulas were dubious and incomplete. Because of this, GPT proved unsuitable for comparison with engineering students. This contradicts the findings of papers dealing with the interpretation of theoretical knowledge such as medicine [31,39] and law [41].

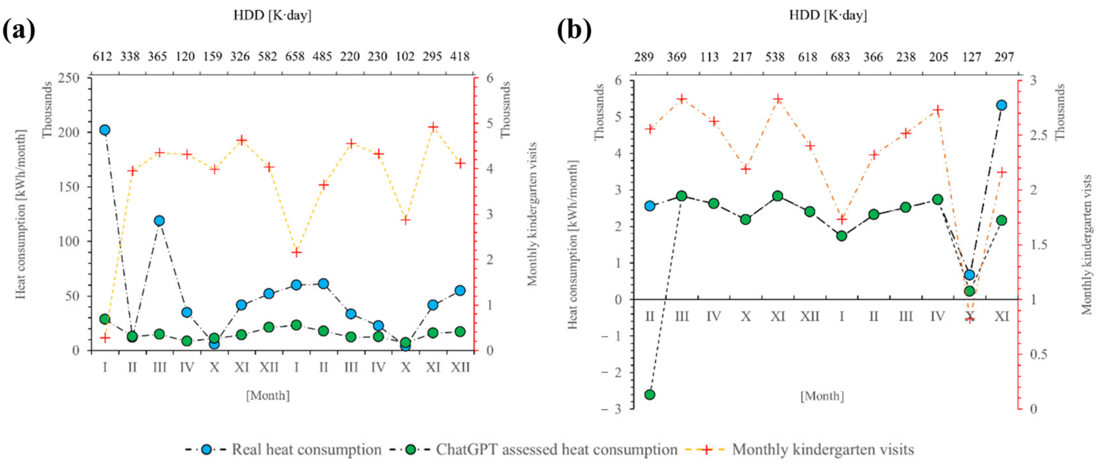

Figure 4. Comparing the influence of building occupancy on a building's real and GPT-assessed heat consumption (**a**) kn10, (**b**) kn7.

3.2. GPT-3.5 Inductive Reasoning Test

The GPT responses to the Q3 set of questions are presented in Figure 5. Figure 5a compares the actual and GPT-assessed SHCs using side-by-side comparable bars, with the actual SHC shown in red and the GPT-assessed SHC in green. Similarly, Figure 5b shows the actual and GPT-assessed savings in SHC. Due to the relatively weak data describing the building and the actions taken, LLM was unable to provide any details before being instructed to rely on inductive reasoning. After this instruction, it began to assume the missing data and the expected energy savings. It was interesting to see that the assumptions were relatively good and in line with practice. When evaluating the building HC before renovation (Figure 5a), LLM overestimated the value by 7% (in the case of the building in Vejtofen, Denmark) and underestimated it by 5% (in the case of the building in Wolgast, Germany). When comparing the energy savings achieved after the renovation (Figure 5b), the errors were higher, between 10% and 40%, when compared with the actual SHC improvements.

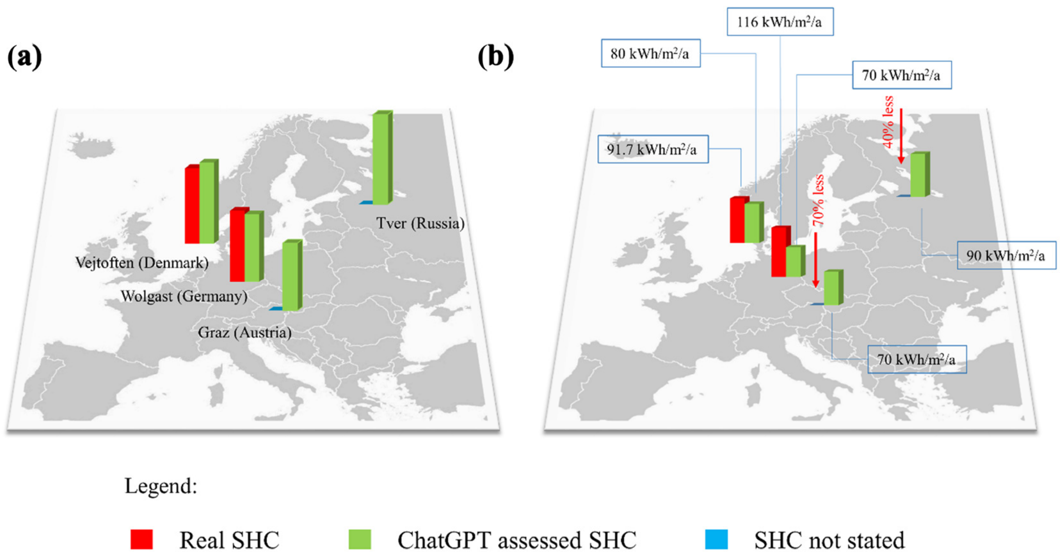

Figure 5. Comparison of buildings' real and GPT-assessed energy savings (**a**) before renovation, (**b**) after renovation.

For the buildings in Graz (Austria) and Tver (Russia), SHC consumption before renovation was not reported in the source literature. However, according to the information provided (Table 4), values were assumed against which energy savings were evaluated. In the case of the kindergarten in Graz, the savings were underestimated by 17%, while in the case of the kindergarten in Tver, they were overestimated by 15% (Table 6).

Table 6. Comparison of buildings' real and GPT-assessed SHCs and buildings' real and GPT-assessed energy savings.

	Vejtofen (Denmark)	Wolgast (Germany)	Graz (Austria)	Tver (Russia)
Real SHC [kWh/m^2/a]	167.4	158	Not stated	Not stated
GPT-assessed SHC [kWh/m^2/a]	180	150	150	200
Real SHC savings [%]	49%	23%	70%	40%
GPT-assessed SHC savings [%]	55%	53%	53%	55%

Assessments based on LLM inductive reasoning were relatively fair, particularly those dealing with SHCs before renovation. This is particularly interesting given the weak data input (Table 4).

3.3. Study Contributions and Directions for Future Research

This was the first study in the field of energy management of public buildings to provide a comprehensive analysis of the applicability and reliability of GPT in real-life scenarios. In addition to the provided results, the study could guide future research by indicating what positive outcomes to expect and what advances to look for. By increasing community evaluation of LLM usability, studies like this contribute to the knowledge base that can provide valuable feedback for future advancements in LLM reasoning.

Future research will assess the reliability of LLM recommendations in shaping decisions related to building renovations. The research will compare the usability of competing technologies in the field. This study will investigate the variety of LLMs' inductive reasoning abilities, emphasizing a thorough analysis of their strengths and limitations. This would encompass assessing the GPT capability to differentiate between construction periods, understand legislation governing building energy efficiency, and recognize changes in building envelope characteristics over time.

4. Conclusions

This study examined the viability of employing GPT as an expert adviser in the field of energy management of kindergartens. The research was conducted on two groups of buildings: (a) 12 public kindergartens in the city of Kragujevac (Serbia) and (b) 4 kindergartens in different cities in Europe. The first group of buildings provided a comprehensive set of data dealing with building physics that facilitated the evaluation of GPT's deductive reasoning potential. The second group of buildings was poorly described, and therefore was used to test GPT's inductive reasoning potential. Concerning deductive reasoning, GPT was tasked to assess the buildings' SHC [kWh/m^2/a]. The response was relatively inaccurate, with an average MAPE of 67%. This outcome can be considered unsatisfactory, especially considering that a simple linear regression, using a single input, outperformed GPT on the same dataset [47]. When dealing with deductive reasoning in assessing the kindergartens' SHCs, GPT proved incapable of performing correct calculations and providing satisfactory accuracy of scores. This aligns with Borji's findings [36], which identified earlier versions of GPT as incapable of math and arithmetic skills. Hence, the success of LLM in this sort of energy management task cannot be compared with that in medicine, where GPT provides the knowledge of a student [31] or even an expert [34]. When dealing with the estimates of monthly heat demand considering the occupancy as an influential factor, LLM proves a promising technology. The average MAPE on this task was 48%. In terms of inductive reasoning, the LLM bot was instructed to assess the building's HC and energy savings by following the renovation procedure. When dealing with missing details in this context, GPT assumptions were in line with practice. As a result, SHC assessments for two of the buildings analyzed indicated MAPE between just 5% and 7%, while energy savings were estimated with poorer performance (15% and 17% error). After analysis, GPT deductive tasks can be considered to be ineligible as an adviser in the field of energy management of kindergartens. This conclusion is based on GPT's weak and unreliable mathematical approach rather than the accuracy of its assessment. Moreover, made-up formulas and false explanations can lead non-experts to make wrong decisions. In the case of inductive reasoning, the technology shows promising potential in augmenting non-experts. Unlike similar studies examining the usability of GPT assessments in other domains, the energy management domain analyzed in this study did not encounter challenges related to the need for real-time internet data (as in [19]), privacy and data security (as in [20]), political bias (as in [25]), or issues of equity and fairness (as in [27]). Therefore, continued advancements in LLM technology could pave the way for practical applications of GPT in addressing energy management challenges in kindergartens.

Author Contributions: Conceptualization, N.J. and D.N.; methodology, N.J. and D.G.; software, N.J.; validation, R.K. and A.N.; formal analysis, N.J., D.N. and D.G.; investigation, N.J.; resources, N.J.; data curation, A.N. and R.K.; writing—original draft preparation, N.J. and D.N.; writing—review and editing, N.J. and D.G.; visualization, N.J.; supervision, N.J., D.N. and D.G.; project administration, D.N.; funding acquisition, N.J., D.N., D.G. and R.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author/s.

Conflicts of Interest: The authors declare no conflicts of interest.

List of Abbreviations Including Units and Nomenclature

HC	Heat consumption [kWh/a]
HDD	Heating degree day [K·Day]
MAE	Mean absolute error
MAPE	Mean absolute percentage error [%]
r	Pearson's correlation coefficient [-]
R^2	Coefficient of determination [-]
SHC	Specific heat consumption [kWh/m^2/annually] i.e., [kWh/m^2/a]
DHS	Duration of a heating season [day]
a	Independent variable
$\bar{a}$	Mean of the values of the a-variable
b	Dependent variable
$\bar{b}$	Mean of the values of the b-variable
bl	Building location
br	Before renovation
D	Description
i	Instance
GPT	Generative pre-trained transformer
j	Day of a heating season
kn	Kindergarten number
LLM	Large Language Model
MLR	Multiple linear regression
n	Number of instances (sample size)
nbv	Number of visits
NLP	Natural language processing
Q	Question
SLR	Simple linear regression
ted	Thermal envelope detail
y	True value of an instance
$\hat{y}$	Predicted value of an instance
$\bar{y}$	Mean value of a sample

Appendix A

Table A1. Describing Details in Figure 4a.

Building: *kn10*		
Month	***HDD***	***nv***
I	612	2778
II	338	3953
III	365	4356
IV	120	4316
X	159	3984
XI	326	4633
XII	582	4031
I	658	2159
II	485	3643
III	220	4558
IV	230	4328
X	102	2877
XI	295	4918
XII	418	4114

Appendix B

Table A2. Describing Details in Figure 4b.

Building: *kn7*		
Month	***HDD***	***nv***
II	289	2557
III	369	2830
IV	113	2626
X	217	2189
XI	538	2830
XII	618	2404
I	683	1733
II	366	2320
III	238	2518
IV	205	2731
X	127	824
XI	297	2162

References

1. Renn, O. How Sustainable Is the Digital World? *Nature* **2023**, *614*, 224–226. [CrossRef]
2. Dunning, S.B. Saeculum. In *Oxford Classical Dictionary*; Oxford University Press: Oxford, UK, 2017. [CrossRef]
3. Knell, M. The Digital Revolution and Digitalized Network Society. *Rev. Evol. Polit. Econ.* **2021**, *2*, 9–25. [CrossRef]
4. Rudolph, J.; Tan, S.; Tan, S. ChatGPT: Bullshit Spewer or the End of Traditional Assessments in Higher Education? *J. Appl. Learn. Teach.* **2023**, *6*, 342–363. [CrossRef]
5. Đukić, P. Just Transition of the Energy Sector in Serbia—Reforms Sustainability in Face of a New Global Crisis. *Energ. Ekon. Ekol.* **2022**, *XXIV*, 53–62. [CrossRef]

6. Cvetanović, A.; Jovičić, M.; Bošković, G.; Jovičić, N. Implementation of Circular Economy and Lean Approaches for a More Competitive and Sustainable Industry. In Proceedings of the 14th International Quality Conference, Kragujevac, Serbia, 24–27 May 2023; Faculty of Engineering, University of Kragujevac: Kragujevac, Serbia, 2023; pp. 1719–1729, ISBN 978-86-6335-104-2.
7. Goh, H.H.; Vinuesa, R. Regulating Artificial-Intelligence Applications to Achieve the Sustainable Development Goals. *Discov. Sustain.* **2021**, *2*, 3–8. [CrossRef]
8. Lichtenthaler, U. Digitainability: The Combined Effects of the Megatrends Digitalization and Sustainability. *J. Innov. Manag.* **2021**, *9*, 64–80. [CrossRef]
9. Adamopoulou, E.; Moussiades, L. An Overview of Chatbot Technology. In *Artificial Intelligence Applications and Innovations, Proceedings of the 6th IFIP WG 12.5 International Conference, AIAI 2020, Neos Marmaras, Greece, 5–7 June 2020*; IFIP Advances in Information and Communication Technology; Springer International Publishing: Cham, Switzerland, 2020; Volume 584, pp. 373–383. [CrossRef]
10. Weizenbaum, J. ELIZA—A Computer Program for the Study of Natural Language Communication between Man and Machine. *Commun. ACM* **1966**, *9*, 36–45. [CrossRef]
11. Gordon, C. ChatGPT Is the Fastest Growing App in the History of Web Applications. Available online: https://www.forbes.com/sites/cindygordon/2023/02/02/chatgpt-is-the-fastest-growing-ap-in-the-history-of-web-applications/?sh=2055a15a678c (accessed on 10 May 2024).
12. Nadkarni, P.M.; Ohno-Machado, L.; Chapman, W.W. Natural Language Processing: An Introduction. *J. Am. Med. Inform. Assoc.* **2011**, *18*, 544–551. [CrossRef]
13. OpenAI. Introducing ChatGPT. Available online: https://openai.com/blog/chatgpt (accessed on 12 May 2024).
14. Common Crawl—Open Repository of Web Crawl Data. Available online: https://commoncrawl.org/ (accessed on 27 January 2024).
15. WebText Background—OpenWebText2. Available online: https://openwebtext2.readthedocs.io/en/latest/background/ (accessed on 27 January 2024).
16. Wikipedia. Available online: https://www.wikipedia.org/ (accessed on 27 January 2024).
17. Alto, V. *Modern Generative AI with ChatGPT and OpenAI Models*; Packt Publishing Ltd.: Birmingham, UK, 2023; ISBN 9781805123330.
18. Rathore, B. Future of Textile: Sustainable Manufacturing & Prediction via ChatGPT. *Eduzone* **2023**, *12*, 52–62. [CrossRef]
19. Prieto, S.A.; Mengiste, E.T.; Soto, B.G. Investigating the Use of ChatGPT for the Scheduling of Construction Projects. *Buildings* **2023**, *13*, 857. [CrossRef]
20. Rathore, D.B. Future of AI & Generation Alpha: ChatGPT beyond Boundaries. *Eduzone* **2023**, *12*, 63–68. [CrossRef]
21. Alves, B.C.; Freitas, L.A.; Aguiar, M.S. Chatbot as Support to Decision-Making in the Context of Natural Resource Management. In Proceedings of the 2021: Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais, Online, 18–23 July 2021; pp. 29–38. [CrossRef]
22. Jungwirth, D.; Haluza, D. Artificial Intelligence and the Sustainable Development Goals: An Exploratory Study in the Context of the Society Domain. *J. Softw. Eng. Appl.* **2023**, *16*, 91–112. [CrossRef]
23. Jungwirth, D.; Haluza, D. Artificial Intelligence and Ten Societal Megatrends: An Exploratory Study Using GPT-3. *Systems* **2023**, *11*, 120. [CrossRef]
24. Rani, P.S.; Rani, K.R.; Daram, S.B.; Angadi, R.V. Is It Feasible to Reduce Academic Stress in Net-Zero Energy Buildings? Reaction from ChatGPT. *Ann. Biomed. Eng.* **2023**, *51*, 2654–2656. [CrossRef]
25. Hartmann, J.; Schwenzow, J.; Witte, M. The Political Ideology of Conversational AI: Converging Evidence on ChatGPT's pro-Environmental, Left-Libertarian Orientation. *arXiv* **2023**, arXiv:2301.01768. [CrossRef]
26. Bii, P. Chatbot Technology: A Possible Means of Unlocking Student Potential to Learn How to Learn. *Educ. Res.* **2013**, *4*, 218–221; ISSN: 2141-5161.
27. Holmes, W.; Porayska-Pomsta, K.; Holstein, K.; Sutherland, E.; Baker, T.; Shum, S.B.; Santos, O.C.; Rodrigo, M.T.; Cukurova, M.; Bittencourt, I.I.; et al. Ethics of AI in Education: Towards a Community-Wide Framework. *Int. J. Artif. Intell. Educ.* **2022**, *32*, 504–526. [CrossRef]
28. King, M.R. A Conversation on Artificial Intelligence, Chatbots, and Plagiarism in Higher Education. *Cell. Mol. Bioeng.* **2023**, *16*, 1–2. [CrossRef]
29. Gao, C.A.; Howard, F.M.; Markov, N.S.; Dyer, E.C.; Ramesh, S.; Luo, Y.; Pearson, A.T. Comparing Scientific Abstracts Generated by ChatGPT to Real Abstracts with Detectors and Blinded Human Reviewers. *NPJ Digit. Med.* **2023**, *6*, 75. [CrossRef]
30. Subaveerapandiyan, A.; Vinoth, A.; Tiwary, N. Netizens, Academicians and Information Professionals' Opinions About AI with Special Reference to ChatGPT. *Libr. Philos. Pract.* **2023**, 1–16. [CrossRef]
31. Gilson, A.; Safranek, C.W.; Huang, T.; Socrates, V.; Chi, L.; Taylor, R.A.; Chartash, D. How Does ChatGPT Perform on the United States Medical Licensing Examination? The Implications of Large Language Models for Medical Education and Knowledge Assessment. *JMIR Med. Educ.* **2023**, *9*, e45312. [CrossRef] [PubMed]
32. Guo, B.; Zhang, X.; Wang, Z.; Jiang, M.; Nie, J.; Ding, Y.; Yue, J.; Wu, Y. How Close Is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection. *arXiv* **2023**, arXiv:2301.07597. [CrossRef]
33. Thurzo, A.; Strunga, M.; Urban, R.; Surovková, J.; Afrashtehfar, K.I. Impact of Artificial Intelligence on Dental Education: A Review and Guide for Curriculum Update. *Educ. Sci.* **2023**, *13*, 150. [CrossRef]

34. Jeblick, K.; Schachtner, B.; Dexl, J.; Mittermeier, A.; Stuber, A.; Topalis, J.; Weber, T.; Wesp, P.; Sabel, B.; Ricke, J.; et al. ChatGPT Makes Medicine Easy to Swallow: An Exploratory Case Study on Simplified Radiology Reports. *arXiv* **2022**, arXiv:2212.14882. [CrossRef]
35. Deng, J.; Lin, Y. The Benefits and Challenges of ChatGPT: An Overview. *Front. Comput. Intell. Syst.* **2022**, *2*, 81–83. [CrossRef]
36. Borji, A. A Categorical Archive of ChatGPT Failures. *arXiv* **2023**, arXiv:2302.03494. [CrossRef]
37. Markus, G.; Davis, E. GPT-3, Bloviator: OpenAI's Language Generator Has No Idea What It's Talking About | MIT Technology Review. Available online: https://www.technologyreview.com/2020/08/22/1007539/gpt3-openai-language-generator-artificial-intelligence-ai-opinion/ (accessed on 27 June 2024).
38. Retief, F.; Bond, A.; Pope, J.; Morrison-Saunders, A.; King, N. Global Megatrends and Their Implications for Environmental Assessment Practice. *Environ. Impact Assess. Rev.* **2016**, *61*, 52–60. [CrossRef]
39. Talan, T.; Kalinkara, Y. The Role of Artificial Intelligence in Higher Education: ChatGPT Assessment for Anatomy Course. *Int. J. Manag. Inf. Syst. Comput. Sci.* **2023**, *7*, 32–40. [CrossRef]
40. Salvagno, M.; Taccone, F.S.; Gerli, A.G. Can Artificial Intelligence Help for Scientific Writing? *Crit. Care* **2023**, *27*, 75. [CrossRef]
41. Li, L.; Ma, Z.; Fan, L.; Lee, S.; Yu, H.; Hemphill, L. ChatGPT in Education: A Discourse Analysis of Worries and Concerns on Social Media. *Educ. Inf. Technol.* **2024**, *29*, 10729–10762. [CrossRef]
42. The Government Offices of Sweden. Lund Declaration on Maximising the Benefits of Research Data. Available online: https://www.government.se/information-material/2023/06/lund-declaration-on-maximising-the-benefits-of-research-data/ (accessed on 27 June 2024).
43. European Commission. The Rome Declaration. Available online: https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_17_767 (accessed on 27 June 2024).
44. Jurišević, N. System for Monitoring and Targeting of Energy and Water Consumption in Public Buildings, University of Kragujevac, Kragujevac, Serbia 2021. Available online: https://nardus.mpn.gov.rs/handle/123456789/18681?locale-attribute=en (accessed on 30 June 2024).
45. Bećirović, S.P.; Vasić, M. Methodology and Results of Serbian Energy-Efficiency Refurbishment Project. *Energy Build.* **2013**, *62*, 258–267. [CrossRef]
46. European Commission. Renovation Wave. Available online: https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficient-buildings/renovation-wave_en (accessed on 30 June 2024).
47. Jurišević, N.; Gordić, D.; Vukićević, A. Assessment of Predictive Models for the Estimation of Heat Consumption in Kindergartens. *Therm. Sci.* **2022**, *26*, 503–516. [CrossRef]
48. Jurišević, N.M.; Gordić, D.R.; Vukašinović, V.; Vukicevic, A.M. Assessment of Predictive Models for Estimation of Water Consumption in Public Preschool Buildings. *J. Eng. Res.* **2021**, *10*, 98–111. [CrossRef]
49. Capozzoli, A.; Grassi, D.; Causone, F. Estimation Models of Heating Energy Consumption in Schools for Local Authorities Planning. *Energy Build.* **2015**, *105*, 302–313. [CrossRef]
50. Beusker, E.; Stoy, C.; Pollalis, S.N. Estimation Model and Benchmarks for Heating Energy Consumption of Schools and Sport Facilities in Germany. *Build. Environ.* **2012**, *49*, 324–335. [CrossRef]
51. Garrido, A.; Hardy, L. *Análisis y Evaluación de las Relaciones Entre el Agua y la Energía en España*; Realigraf, S.A.: Madrid, Spain, 2010; Volume 6, ISBN 9788496655232.
52. Aranda, A.; Ferreira, G.; Mainar-Toledo, M.D.; Scarpellini, S.; Llera Sastresa, E. Multiple Regression Models to Predict the Annual Energy Consumption in the Spanish Banking Sector. *Energy Build.* **2012**, *49*, 380–387. [CrossRef]
53. Rose, J.; Thomsen, K.E. Energy Saving Potential in Retrofitting of Non-Residential Buildings in Denmark. *Energy Procedia* **2015**, *78*, 1009–1014. [CrossRef]
54. Power, A.; Zulaf, M. *Cutting Carbon Costs: Learning from Germany's Energy Saving Program*; London School of Economics: London, UK, 2011.
55. Bleyl-androschin, J.W.; Schinnerl, D. *Comprehensive Refurbishment of Buildings Through Energy Performance Contracting a Guide for Building Owners and ESCos Including Good Practice Examples*; Graz Energy Agency: Graz, Austria, 2010; ISBN 4315861524340.
56. Vatin, N.I.; Nemova, D.V.; Kazimirova, A.S.; Gureev, K.N. Increase of Energy Efficiency of the Building of Kindergarten. *Adv. Mater. Res.* **2014**, *953–954*, 1537–1544. [CrossRef]
57. Playground—OpenAI API. Available online: https://platform.openai.com/playground (accessed on 20 September 2024).
58. Taulli, T. *Generative AI*; Apress: New York, NY, USA, 2023; ISBN 9781484293690.
59. Sammut, C.; Webb, G.I. (Eds.) Mean Absolute Error (MAE). In *Encyclopedia of Machine Learning and Data Mining*; Springer: Boston, MA, USA, 2017; p. 806. [CrossRef]
60. Swamidass, P.M. (Ed.) Mean Absolute Percentage Error (MAPE). In *Encyclopedia of Production and Manufacturing Management*; Springer: New York, NY, USA, 2006; p. 462; ISBN 9781402006128.
61. Chicco, D.; Warrens, M.J.; Jurman, G. The Coefficient of Determination R-Squared Is More Informative than SMAPE, MAE, MAPE, MSE and RMSE in Regression Analysis Evaluation. *PeerJ Comput. Sci.* **2021**, *7*, e623. [CrossRef]