International Journal for Quality Research 19(3)
ISSN 1800-6450

Nebojša Jurišević
Robert Kowalik
Dušan Gordić
Aleksandar Novaković
Vladimir Vukašinović
Nikola Rakić
Jelena Nikolić
Arso Vukićević [1]

DOI – 10.18421/IJQR19.03-09

LARGE LANGUAGE MODELS AS TOOLS FOR PUBLIC BUILDING ENERGY MANAGEMENT: AN ASSESSMENT OF POSSIBILITIES AND BARRIERS

Abstract: *This study examins usability of large language model-based (LLM) chat bots, specifically GPT-3.5 and GPT-4, as assisting tools in the energy manageemnt of educational buildings in Serbia and Poland. The assessment is based on the comparison of three key usability aspects: 1) the accuracy of expert opinion replication in classifying building construction periods, 2) familiarity with field-specific legislation; and 3) knowledge of details regarding building thermal envelopes. In replicating constuction periods and sugesting legislation, LLM chat bots performed admirably in Poland, but less so in Serbia. Regarding thermal enveople characteristics, GPT-3.5 indicated U-value spans encompassing actual values, but to broad to be useful. U-value spans provided by GPT-4 were narrower, but they did not generally intersect with the actual U-value range. The study concludes that LLM chat bots indicate great potential for assisting non-experts in public building energy management, with usability varying by country, but still far from experts.*

Keywords: *LLM chat bots, GPT-3.5, GPT-4, energy management, legislation, buildings*

1. Introduction

The rapid evolution of Natural Language Processing (NLP) has resulted in the breakthrough of artificial intelligence (AI)-based large language models (LLMs), such as OpenAI's ChatGPT (Kalyan 2024). These LLM chat bots have demonstrated the capability of providing human-like responses in text-based interfaces (Khennouche et al. 2024). This advancement is due to groundbreaking developments in unsupervised learning, allowing LLM chat bots to learn from vast text datasets without direct human data annotation for specicic tasks. LLMs are trained on extensive text data, utilizing transformer architecture and attention mechanisms to generate contextually relevant and coherent text outputs for a wide range of questions. Contemporary LLM chat bots can learn from provided input prompts, enabling them to adopt and interpret knowledge from various professional and scientific domains. To examine the reasoning potential and reliability of this novel technology, recent studies have conducted a series of assessments in various domanns.

Findings in literature indicate that LLM chat bots have significant potential in education by enhancing exam performance and solving complex problems. Study by Pursnani et al. (2023) highlighted the effectiveness of prompt modifications in improving the model's accuracy on the FE Environmental Exam, suggesting future research into responsible AI integration for optimized learning experiences and improved

[1] Corresponding author: Arso Vukićević
Email: arso_kg@yahoo.com

student outcomes. Related to the responsible use of AI, Chauncey and McKenna (2023) investigated how ethical LLM chat bot can be used in education, proposing a framework to enhance teaching and learning by supporting cognitive flexibility and self-regulation. Generative AI tools have also shown potential to revolutionize management education, balancing technology use with authentic learning. In study by Gupta et al. (2024) SCOT theory and mixed-methods analysis reveal positive impacts and challenges in pedagogy, assessment, and ethics.

Regarding the environmental sciences, Egbemhenghe et al. (2023) explored ChatGPT applications for addressing the global water crisis, highlighting benefits in water management, predictive maintenance, and ethical considerations, while emphasizing the need for responsible AI governance. Besides enhancing building performance (Liao et al. 2024), the generative AI can be also used to improve building design by reusing data, learning from past experiences, and creating new ideas, despite challenges in data representation and design evaluation.

Further advancements, like GPT-4, have shown even greater usability (Rudolph, Tan, and Tan 2023), excelling in theory of mind tests and achieving a 95% accuracy rate on items it can answer correctly, compared to 40% for GPT-3.5 (Campello de Souza, et al. 2023). Case studies of medical training exams reveal GPT-4 achieved an overall score of 90.5%, compared to 63.1% for GPT-3.5 (Passby, et al. 2023). OpenAI's technical report indicates GPT-4 significantly outperforms its predecessor on tests like the Law School Admission Test, Scholastic Assessment Test, Graduate Record Examination, and Advanced Placement Exams (OpenAI 2023). Similar conclusions have been reached in linguistics, sociology (Rahaman et al. 2023). However, there are disparities between different LLM chat bot platforms across various tasks (Ahmed et al. 2023). Webb et al. (2023) performed intelligence tests on different versions of Generative Pretraind Transformers (ChatGPT). They discovered that LLM chat bots (depending on development stage and the test conducted) show above average intelligence when compared to humans (zero-shot reasoning). However, aside from the benefits intelligent technology can provide, its utilization raised a number of questions and concerns: Is it reliable? How should it be used? What constraints does technology have? Is it capable of stealing jobs?

To contribute in answering these questions, this study compared the performance of GPT-3.5, and GPT-4 Turbo, in the field of energy management of public buildings. To make the comparison fair in providing useful information, study does not analyse written quality of the answers, rather just factual data LLM chat bots provide. The comparison is focused on assessing three aspects of LLM chat bot capabilities:

A1) Matching expert opinions when classifying building construction periods (CPs);

A2) Recommending legal acts that govern building thermal performance;

A3) Comprehending period-specific details of building thermal envelopes.

Aside from comparing LLM chat bots' utilization potential, the study emphasizes the feasibility of novel technology in augmenting professional opinions and assisting decision-making in energy management.

2. Materials and methods

To conduct the analysis (second half of April 2024), the study employed GPT-3.5 (Trained on data up until September 2021, Context window 16385 tokens) and GPT-4 Turbo (trained on data up until December 2023, Context window 128000 tokens). To examine how closely the LLM chat bots resemble expert opinions, as well as ground truths within the field, the study utilized contextual prompting by using four lines of

questioning. The responses LLM chat bots provide to these questions, address the study aims. To be more specific, one question (Q1) deals with the first study aim (A1), another question (Q2) addresses the second (A2), and two more questions (Q3, Q4) deal with the third (A3). To ensure objectivity of the findings, research was conducted on the example of two countries: one located in southern Europe (Republic of Serbia), and one located in central Europe (Republic of Poland). The study compared LLM chat bot responses to data from field legislative in two countries, as well as National Typology of Kindergartens in Serbia (Popović et al. 2018), Polish Building Typology (Polish National Energy Conservation Agency 2012) and Typology Approaches for Non-Residential Buildings in Five European Countries (Stein et al. 2012). Given that the studied countries have different geographical, historical, and cultural preferences, building sectors may be considered relatively distinct.

The order (Q/A) and content (italic text) of the questions used in the study were as following:

Q1/A1: *How many discrete construction periods of educational buildings can be identified in Serbia/Poland since 1918? Base your response on factors such are architectural design, regulatory changes and turning points in the advancement of technical and technological capacities in construction!*

Q2/A2: *List the legal acts that govern building thermal performance for each of the construction periods!*

Q3/A3: *Quantify the U-values of the public building's exterior walls for each construction period!*

Q4/A3: *Quantify the U-values of the fenestration elements for each construction period!*

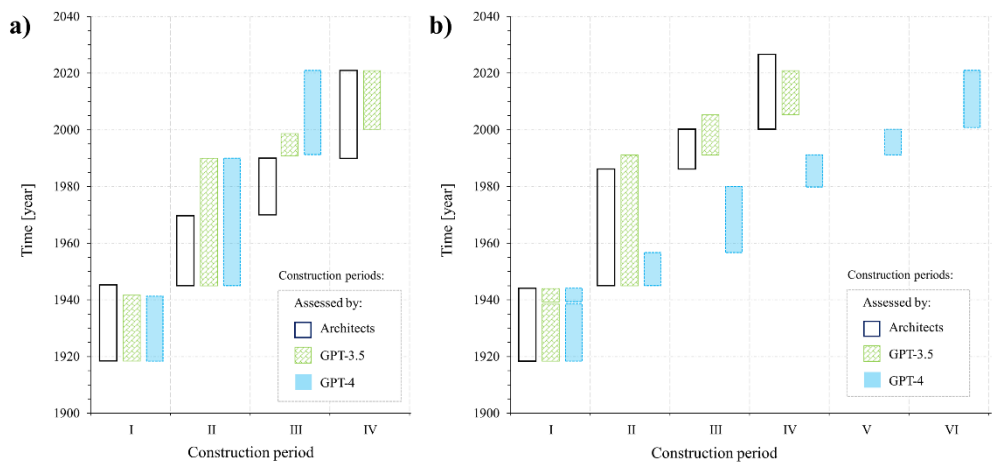

Figure 1. Comparing LLM chat bot performance to architect opinions in classifying CPs: a) Republic of Serbia, b) Republic of Poland

3. Results and discussion

3.1. Matching expert opinions when classifying building construction periods

To avoid unobjective comparisons based on the number of LLM chat bots available tokens, the study examines only factual details of their responses, describing built periods (year spans), legal acts (act titles and issue years), and thermal characteristics of building envelopes (U values of external walls and fenestrations elements). To answer the A1 aim, the study examined LLM chat bot responses when prompted Q1 (Fig. 1). As a result, GPT-3.5 proved capable of accurately resembling the opinions of architects on the number of distinctive CPs (Fig. 1). However, the time spans of the CPs do not entirely match those assessed by architects (Popović et al., 2018). In both countries, only the time span of the first CP (1918-1945: as assessed by architects) was entirely matched by chat bots. Even more, due to the devastating consequences of World War 2, GPT-3.5 and GPT-4 either left a period gap from 1939 to 1945 (Republic of Serbia, Fig. 1a), or identified the gap as a sub-period when the construction and renovation of buildings was carried out below the typical standards of the time (Republic of Poland, Fig. 1b). Regarding the II CP of the Republic of Serbia, GPT chat bots estimated it lasted approximately twice as long as assessed by architects (GPT3.5/GPT-4: 46 years; architects: 24 years). In these estimates, the LLM chat bots combined II and III CPs into one, matching the span of the last period as assessed by architects (GPT-4), or subdividing it into two separate periods (GPT-3.5).

As for the CPs of educational buildings in Poland (Fig. 1b), GPT-3.5 provided the same number of periods as the expert literature (Stein et al. 2012), matching the time spans relatively fairly. The threshold between the post-war CPs was shifted by four years compared to the threshold assessed by the architects. These discrepancies can be considered negligible, particularly as they pertain to period prolongations. In addition, building typology experts emphasize that the thresholds they have adopted are not strict, as there are educational buildings whose construction began in one period and ended in another (Polish National Energy Conservation Agency 2012). Having this in mind, GPT-3.5 can be considered capable of resembling architects' opinions on educational building typology in Poland. On the other hand, GPT-4 provided two CPs more than the architects, subdividing the III CP (four decades long) into three distinct periods: one spanning two decades, and two lasting a decade each. Although this does not resemble architects' assessments, it should not be misinterpreted as a lack of GPT-4 assessment accuracy, but rather a more detailed subclassification within the existing CP.

3.2. Recommending legal acts that govern building thermal performance

To investigate AI factual knowledge in the field's legislative (A2), LLM chat bots were prompted with additional question (Q2). To make the comparison clear, Tab. 1 and Tab. 2 list legal acts valid in each of the CPs in the Republic of Serbia (Gordić et al. 2018; Ministry of Construction Transport and Infrastructure of the Republic of Serbia 2018) and the Republic of Poland (Polish Committee for Standardization 1974; Polish National Energy Conservation Agency 2012), respectively, as well as corresponding suggestions provided by each of the chat bots. The suggestions made were marked: ✓ if true, and: ***x*** if false (the acts was not found in the official gazette archive). If the suggestion was found to be true, a reference was provided. To supplement the list of legal acts used in national typologoies, authors added acts enacted upon the time the typologies was published. When responding, LLM chat bots correctly stated that prior to 1945, there were no legal acts regulating the thermal properties of buildings on the territory of the modern-day Republic of Serbia (Tab. 1). For the

following period (1945-1989), GPT-3.5 claimed that the field was regulated by the *Legal Act #1b* (Tab. 1), which does not appear in the Official Gazette archive. On the plus side, the chat bot pointed out that the proposed law did not specifically address building thermal performance. Going further, the chat bot continued to hallucinate by suggesting *Legal Act #2b* as the law governing the filed from 1989 – 2004, despite the fact that the suggested law did not directly regulate the field and was enacted five years upon the CP ended (Government of the Republic of Serbia 2008). From the beginning of the 2000s, the chat bot suggests more legislation (*Legal Act #3b*, *Legal Act #4b*, *Legal Act #5b*) than before, but none of them successfully resembled the titles of those existing. In this context, *Legal Act #3b* was an incomplete title for the *Legal Act #6a* draft proposed ten years later than suggested, *Legal Act #4b* referred to a document as Technical Regulations rather than Regulation (*Legal Act #4a*) mismatching the enactment date by a year, and *Legal Act #5b* did not appear in the official gazette archive. Regarding GPT-4, it did not recommend legislation before 1990, and those suggested after 1990 did not provide an exact year of legislation enactment rather just period (2000s). As a consequence, the chat bot responses were less precise than they could have been, but they were also less likely to be hallucinations. Of the three proposed documents, the title of *Legal Act #2c* came closest to the title of the actual document (*Legal Act #4a*), mismatching the Rulebook for Regulation. As for the other proposed documents, GPT-4 made similar mistakes as GPT-3.5: *Legal Act #1c* refers to the draft of *Legal Act #4a* instead of the document that entered into force, while *Legal Act #3c* does not exist in offizial gazete archive.

In the case of the Republic of Poland (Tab. 2), both GPT-3.5 and GPT-4 stated correctly that no legal acts governed the thermal performance of buildings prior to 1945. For the period until 1989, GPT-3.5 suggested *Legal Act #1e* – which did not refer to building thermal properties as to construction details. GPT-4 incorrectly stated that the period 1956-1980 was unregulated by legal acts, despite the fact that *Legal Act #2d* and *Legal Act #3d* were in force. In contrast to the scenario in the Republic of Serbia, the legislation referenced by GPT-3.5 regarding the Republic of Poland was in place. Going to the current CP (2000s), GPT-3.5 suggestions became more relevant, as they better matched the domain of building thermal regulations than before. The same can be applied to the GPT-4 suggestions. Although GPT-4 cited more legislation in the field than GPT-3.5 (two instances versus one), GPT-3.5 provided useful insights by indicating the existence of additional regulations and guidelines issued by the ministries, such as *Legal Act #6d*, as well as amendments to the *Legal Act #3e*. The latter suggestion was not as relevant as former as it was not closely related to the domain of thermal regulation.

3.3. Comprehending period-specific details of building thermal envelopes

Aside from examining LLM chat bot potentials to recommend appropriate field legislation, the study looked into opportunities for AI to accurately characterize educational building thermal envelopes. In particular, this refers to the proper description of envelope characteristics (such as external walls U-values and fenestration U-values) over different CPs. This details are of particular importance in assessing building energy consumption (Jurišević, Gordić, and Vukićević 2021). To address the A3 aim, the study examined LLM chat bot responses when prompted Q3 and Q4. The chat bot responses to Q3 are represented in Fig. 2, Tab. 3, and Tab. 4.

Table 1. Actual legislation governing building thermal performance in Serbia, and LLM chat bot outputs (***x*** indicates halicunated legal acts, ✓ indicates correct answer).

Assessed by Architects			Assessed by GPT-3.5		Assessed by GPT-4	
Construction period (ground truth)		Legal Acts	Construction period (predicted)	Legal Acts	Construction period (predicted)	Legal Acts
I	1918-1941	Period without thermal regulations	1918-1939	*Period without thermal regulations* ✓	1918-1939	*Period without thermal regulations* ✓
	1941-1945		1939-1945		1939-1945	
II	1946-1971	***Legal Act #1a*:** Regulation on Technical Measures and Condition-s for Thermal Protection of Buildings (1970) (Government of Socialist Federal Republic of Yugoslavia 1970)	1945-1989	***Legal Act #1b*:** Socialist Federal Republic of Yugoslavia (SFRY) Building Act (1965) ***x*** *not specifically focused on thermal performance* ✓	1945-1956	*There might have been early forms of egulations* ✓ *(regulations not specified)*
III	1971-1990	***Legal Act #2a*:** JUS U.J5.600 - Heat in civil engineering – Requirements for design and manufacturing of buildings (1980) (Institute for Standardization of Yugoslavia 1980)			1956-1980	
					1980-1989	
IV	1991-2011	***Legal Act #3a*:** JUS U.J5.600:1987 SRPSU U.J5.600:1998 Heat in civil engineering – Requirements for design and manufacturing of buildings (Institute for Standardization of Serbia 1998; Institute for Standardization of Yugoslavia 1987)	1989-2004	***Legal Act #2b*:** Law on Planning and Construction (1994) ***x***	1990-2000	***Legal Act #1c*:** Energy Efficiency Act (2000s) ***x*** ***Legal Act #2c*:** Regulation on Energy Efficiency of Buildings (2000s) ***x*** ***Legal Act #3c*:** Building Codes and Standards (2000s) ***x***
	2012 -	***Legal Act #4a*:** Rulebook on Energy Eff. of Buildings (2011) (Government of the Republic of Serbia 2011) ***Legal Act #5a*:** Rulebook on Conditions, Content and Manner of Issuing Energy Performance Certificate Of Buildings (2012) (Government of the Republic of Serbia 2012) ***Legal Act #6a*:** The Law on Energy Efficiency and Rational Use of Energy (2021) (Government of the Republic of Serbia 2021)	2000s	***Legal Act #3b*:** Energy Efficiency Law (2004) ***x*** ***Legal Act #4b*:** Technical Regulation on Energy Efficiency of Buildings (2010) ***x*** ***Legal Act #5b*:** Revised Building Act (2019) ***x***	2000 -	

Table 2. Actual legislation governing building thermal performance in Poland, and LLM chat bot outputs (***x*** indicates halicunated legal acts, ✓ indicates correct answer).

Assessed by Architects			Assessed by GPT-3.5		Assessed by GPT-4	
Construction period		Legal Acts	Construction period	Legal Acts	Construction period	Legal Acts
I	1918-1945	***Legal Act #1d*:** Building Code	1918-1939	*No specific acts focused on building thermal performances. Thermal performances influenced by construction standards.* ✓	1918-1939	*No specific legal acts focusing on the thermal performance of buildings* ✓
			1939-1945	*No legal acts* ✓	1939-1945	
II	1946-1985	***Legal Act #2d*:** PN-64/B-03404: Heat transfer coefficients *k* for building partitions (1964) (Polish Committee for Standardization 1964) ***Legal Act #3d*:** PN-74/B-03404: Heat transfer Coefficient *k* for building partitions (1974) (Polish Committee for Standardization 1974) ***Legal Act #4d*:** PN-82/B-02020: Thermal insulation of buildings. Requirements and calculations (1982) (Polish Committee for Standardization 1982)	1945-1989	***Legal Act #1e*:** Building Law (1961) ✓ (Council of Ministers, Chairman of the Committee of Construction 1961) *Norms and standards issued by state agencies* ✓	1945-1956	
					1956-1980	*No specific legal acts focusing on the thermal performance of buildings* ***x***
III	1986-2000	***Legal Act #5d*:** PN-91/B-20020: Thermal insulation of buildings. Requirements and calculations (1991) (Polish Committee for Standardization 1991)			1980-1989	*Legislation on building energy efficiency and thermal performance was not developed* ***x***
			1989-2004	*Amendments to the Building Law of 1961* ✓ ***Legal Act #2e*:** Building Law (1994) ✓ (Ministry of Spatial Planning and Construction 1994)	1990-2000	***Legal Act #1f*:**: Building Law (1994) ✓ (Ministry of Spatial Planning and Construction 1994) ***Legal Act #2f*:** Energy Efficiency Act (1994) ✓(Ministry of Agriculture and Food Economy 1994)
IV	2000 -	***Legal Act #6d*:** The Decree on technical requirements which buildings and their location should comply with (2002, 2013) (Ministry of Infrastructure 2002; Ministry of Infrastructure Construction and Maritime Economy 2013) ***Legal Act #7d*:** The Act of supporting thermos modernisation and renovation (2008) (The Government of Poland 2008b)	2000s	*Amendments to the Building Law of 1994* ✓ ***Legal Act #3e*:** Energy Performance of Buildings Directive (EPBD) ✓(The European Parlament and the Council of European Union 2010) *Specific regulations and guidelines issued by the Ministry of Infrastructure, Construction, and Spatial Planning* ✓	2000s - present	***Legal Act #3f*:** Regulation on the Technical Conditions to be Met by Buildings and their Location (2002) ✓ (The Government of Poland 2008a) ***Legal Act #4f*:** Energy Performance of Buildings Directive (EPBD) (2010, 2018) ✓ (The European Parlament and the Council of European Union 2010)

In the case of the Republic of Serbia (Fig. 2a), GPT-3.5 indicates poorer thermal characteristics of external walls than actual, except when assessing thermal envelope of the second CP (1945-1990) (Tab. 3). GPT-4 suggested two to three times higher mean U-value for all the CPs. Both LLM caht bots showed a decreasing trend in the mean U-value from the beginning to the end of the period. To create this trend, chat bots significantly overestimated the mean U-value of the first CP (GPT-3.5 by 70%, GPT-4 by 180%), which was actually one-third lower than the U-value of the II CP, and by 16% lower than in the III CP.

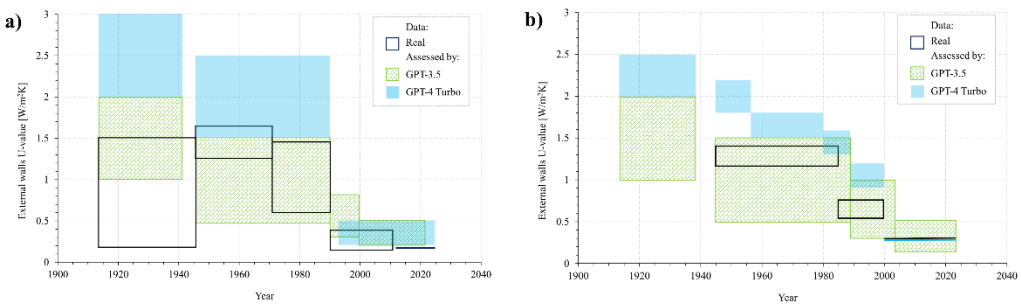

Figure 2. Actual U-values of exterior walls of educational buildings and appropriate LLM chat bot assessments: a) Republic of Serbia, b) Republic of Poland

Table 3. Details of external wall U-values of educational buildings in the Republic of Serbia, real and LLM chat bot assessment comparison

Real data (Gordić et al. 2018; Ministry of Construction Transport and Infrastructure of the Republic of Serbia 2018)			GPT-3.5		GPT-4	
CP		U-value [W/m²K] Mean ± STD	CP	U-value [W/m²K] Mean ± STD	CP	U-value [W/m²K] Mean ± STD
I	1918-1941	0.88±0.85	1918-1941	1.50±0.71	1918-1941	2.5±0.71
	1941-1945		1941-1945		1941-1945	
II	1946-1971	1.43±0.28	1945-1991	1.0±0.71	1945-1991	2.0±0.71
III	1971-1990	1.05±0.63				
IV	1991-2011	0.25±0.14	1991-2000	0.55±0.35	1992 -	0.35±0.21
	2012 -	0.16	2000s	0.35±0.21		

In Polish national typologies this study used (Polish National Energy Conservation Agency, 2012; Stein et al., 2012), there is no data describing the thermal properties of exterior walls in before 1941 (Fig. 2b, Tab. 4). The LLM chat bots correctly identified a decreasing trend of mean U-values from the beginning to the end of the investigated period. As for the assessments made in the case of the Republic of Serbia, GPT-3.5 underestimated the mean U-value of the external walls of Polish educational buildings in II CP (by 22%), while it has it slightly overestimated in III and IV CPs. GPT-4 overestimated the mean U-values in II and III CPs (22-36% in II and 38-55% in III) while it has perfectly match the value of IV CP.

Table 4. Details of external wall U-values of educational buildings in the Republic of Poland, real and LLM chat bot assessment comparison

Real data (Polish National Energy Conservation Agency 2012; Stein et al. 2012)			GPT-3.5		GPT-4	
CP		U-value [W/m²K] Mean ± STD	CP	U-value [W/m²K] Mean ± STD	CP	U-value [W/m²K] Mean ± STD
I	1918-1945	-	1918-1939	2.25±1.06	1918-1939	2.25±0.35
			1939-1945	1.75±0.35	1945-1955	2.05±0.35
II	1946-1985	1.28±0.17	1945-1989	1.75±1.06		
III	1986-2000	0.65±0.14			1956-1989	1.75±0.35
			1990s	1.05±0.64	1990s	0.90±0.42
IV	2000 -	0.30	2000s	0.50±0.42	2000s -	0.23±0.11

The LLM chat bots' responses to Q4 are represented in Fig. 3, Tab. 5. and Tab. 6. In the case of the Republic of Serbia (Fig. 3a, Tab. 5), GPT-3.5 assessments mainly cover the span of actual U-values of the fenestration elements through time, with the exception of the I CP. GPT-4 was unable to match the U-value spans of the I and III CPs. The deviation of the assessed values for the first three periods was ten (GPT-3.5) to thirteen times (GPT-4) greater than the actual (Tab. 5). In terms of the mean U-values of the fenestration elements GPT-3.5 and GPT-4 overestimated the values of the I CP by approximately 40%.

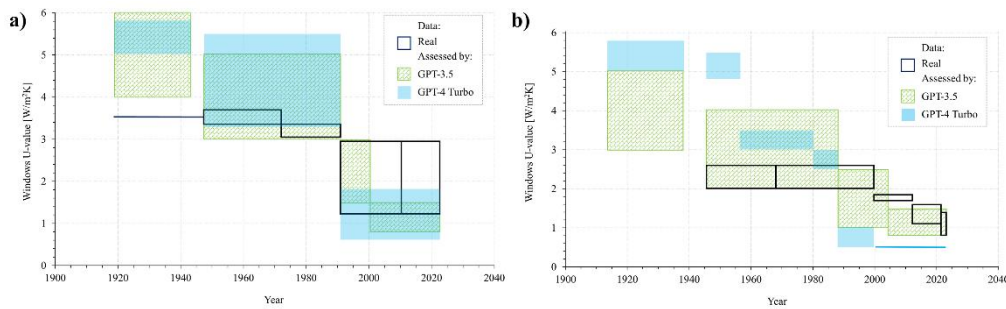

Figure 3. Actual U-values of educational building fenestration and appropriate LLM chat bot assessments: a) Republic of Serbia, b) Republic of Poland

When compared to the actual data, the mean of the overestimated U-values was almost equally far (±2%) from the ground truth: 40% (I CP), 16% (II CP), and 20% (III CP). In the last CP, LLM chat bots assessed the mean to be lower than actual by 52% (GPT-3.5), 41% (GPT-4).

On the example of the Republic of Poland (Fig. 3b, Tab. 6), GPT-3.5 generally covers the span of fenestration U-values through time. Both LLMs exaggerate the value in the II CP (GPT-3.5 by 23%, GPT-4 by 23-55%). Contrary to the results obtained on the example of the Republic of Serbia, GPT-3.5 underestimated the mean U-value by 13%, respectively. In the last CP, GPT-3.5 assessed higher mean U-values than actual (15-34%), as well as GPT-4 (57-71%).

Table 5. Details of fenestration U-values of educational buildings in the Republic of Serbia, real and LLM chat bot assessment comparison

Real data (Gordić et al. 2018; Ministry of Construction Transport and Infrastructure of the Republic of Serbia 2018)			GPT-3.5		GPT-4	
CP		U-value [W/m²K] Mean ± STD	CP	U-value [W/m²K] Mean ± STD	CP	U-value [W/m²K] Mean ± STD
I	1918-1941	3.50	1918-1941	4.0±1.41	1918-1941	2.5±0.71
	1941-1945		1941-1945		1941-1945	
II	1946-1971	3.40±0.14	1945-1990	3.0±1.41	1945-1991	2.0±0.71
III	1971-1990	3.20±0.14				
IV	1991-2011	2.05±1.20	1990s	2.25±1.06	1992 -	0.35±0.21
	2012 -	2.05±1.20	2000s	1.15±0.49	Post-2000	1.75±1.06

Table 6. Details of fenestration U-values of educational buildings in the Republic of Poland, real and LLM chat bot assessment comparison

Real data (Polish National Energy Conservation Agency 2012; Stein et al. 2012) (Kaczorek and Bekierski 2020; Życzyńska and Cholewa 2015)			GPT-3.5		GPT-4	
CP		U-value [W/m²K] Mean ± STD	CP	U-value [W/m²K] Mean ± STD	CP	U-value [W/m²K] Mean ± STD
I	1918-1945	-	1918-1939	4.0±1.41	1918-1939	0.47±0.35
			1939-1945		1939-1945	
II	1946-1985	2.30±0.42	1945-1950	3.25±1.06	1945-1955	4.50±0.71
III	1986-2000	2.30±0.42	1950-1989	2.75±1.06	1956-1989	3.0±0.71
			1990s	2.0±0.71	1990s	2.05±0.64
IV	2000 -	1.75±0.07	2000s	1.15±0.49	2000s -	1.10±0.42
		1.35±0.35				
		1.15±0.35				

4. Conclusion

GPT-3.5 and GPT-4 can resemble expert opinions when classifying construction periods of educational buildings relatively fairly. This task was particularly well assessed by GPT-3.5, which distinguished CPs of countries almost as well as architects. On the other hand, GPT-4 was able to indicate CP subcategories that were of time borders of the CPs assessed by architects. This was the case on the example of the Republic of Poland. In terms of the field legislation assessment, GPT chat bots performed well in detecting periods with no specific acts focusing on building thermal performances (CPs prior to 1955). Following that (until the 1990s), the chat bots started to recommend the laws that do not exist in the Official gazete of the Republic of Serbia. On the example of the Republic of Poland, GPT-3.5 and GPT-4 suggested existing laws in the field. It was interesting that, as opposite to the example of the Republic of Serbia, LLM chat bots did not hallucinate, and GPT-3.5 provided better performance than GPT-4. This is due the fact that former falsely claimed that there was no legislation in the field from 1956 to 1989. It is worth noting that LLM chat bots did not recommend bylaws (standards, regulations)

that contain actual rules in the field. When dealing with thermal envelope characteristics, GPT-3 proved to be the most useful in determining average U-values. GPT-3.5 indicated U-value spans that cover actual, but the span was usually too large to be considered particularly useful. To sum up, studied LLM chat bots inticate a great potential for assisting non-experts in the field of public building energy management. This specifically refers to their capacity to resemble architect opinions in terms of CPs. When it comes to law regulations and details of the building thermal envelopes, LLM chat bots prove to have different potential in different countries. In the case of the Republic of Serbia, LLM chat bots were unable to recommend existing laws, while in the case of republic of Poland this was not the case as all the suggestions were truthful.

References:

Ahmed, Imtiaz, Mashrafi Kajol, Uzma Hasan, and Partha Protim Datta. 2023. "ChatGPT vs. Bard: A Comparative Study." *TechRxiv*. doi: 10.36227/techrxiv.23536290.v2.

Campello de Souza, Bruno, Agostinho Serrano de Andrade Neto, and Antonio Roazzi. 2023. "Are the New AIs Smart Enough to Steal Your Job? IQ Scores for ChatGPT, Microsoft Bing, Google Bard and Quora Poe." *SSRN Electronic Journal*. doi: 10.2139/ssrn.4412505.

Chauncey, Sarah A., and H. Patricia McKenna. 2023. "A Framework and Exemplars for Ethical and Responsible Use of AI Chatbot Technology to Support Teaching and Learning." *Computers and Education: Artificial Intelligence* 5(October):100182. doi: 10.1016/j.caeai.2023.100182.

Council of Ministers, Chairman of the Committee of Construction, Urban Planning and Architecture. 1961. "Building Law." 7(46).

Egbemhenghe, Abel U., Toluwalase Ojeyemi, Kingsley O. Iwuozor, Ebuka Chizitere Emenike, Tolu I. Ogunsanya, Stella Ukamaka Anidiobi, and Adewale George Adeniyi. 2023. "Revolutionizing Water Treatment, Conservation, and Management: Harnessing the Power of AI-Driven ChatGPT Solutions." *Environmental Challenges* 13(October):100782. doi: 10.1016/j.envc.2023.100782.

Gordić, Dušan, Nebojša Jurišević, Dubravka Živković, Vladimir Vukašinović, Dobrica Milovanović, Davor Končalović, and Mladen Josijević. 2018. "Benchmarking of Heat Energy Consumption in Public Buildings in the City of Kragujevac." *ENergetics2018* 139–46.

Government of Socialist Federal Republic of Yugoslavia. 1970. "Regulation on Technical Measures and Conditions for Thermal Protection of Buildings." *Official Gazette of Socialist Federal Republic of Yugoslavia 35/70*.

Government of the Republic of Serbia. 2008. "Planning and Building Act (2008)." *Official Gazette of the Republic of Serbia* (73):1–122.

Government of the Republic of Serbia. 2011. "Rulebook on Energy Efficiency of Buildings." *Official Gazette of the Republic of Serbia No. 61/2011*.

Government of the Republic of Serbia. 2012. "Rulebook on Conditions, Content, and Manner of Issuing Energy Performance Certificate of Buildings." *Official Gazette of the Republic of Serbia 69/2012*.

Government of the Republic of Serbia. 2021. "The Law on Energy Efficiency and Rational Use of Energy." *Official Gazette of the Republic of Serbia* 40.

Gupta, Pragya, Renuka Mahajan, Usha Badhera, and Pooja S. Kushwaha. 2024. "Integrating Generative AI in Management Education: A Mixed-Methods Study Using Social

Construction of Technology Theory." *International Journal of Management Education* 22(3):101017. doi: 10.1016/j.ijme.2024.101017.
Institute for Standardization of Serbia. 1998. "Heat in Civil Engineering - Requirements for Design and Manufacturing of Buildings (SRPS U.J5.600:1998)." *Official Gazette of the Republic of Serbia 11/98*. Retrieved October 16, 2023 (https://iss.rs/sr_Cyrl/project/show/iss:proj:17292).
Institute for Standardization of Yugoslavia. 1980. "Heat in Civil Engineering – Requirements for Design and Manufacturing of Buildings." *Official Gazette of Socialist Federal Republic of Yugoslavia 3/80*.
Institute for Standardization of Yugoslavia. 1987. "Heat in Civil Engineering - Requirements for Design and Manufacturing of Buildings." *Official Gazette of Socialist Federal Republic of Yugoslavia 69/87* 76–78.
Jurišević, Nebojša, Dušan Gordić, and Arso Vukićević. 2021. "Assessment of Predictive Models for the Estimation of Heat Consumption in Kindergartens." *Thermal Science*. doi: https://doi.org/10.2298/TSCI201026084J.
Kaczorek, Dobrosława, and Dominik Bekierski. 2020. *Implementation of the EPBD in Poland – Status in 2020*.
Kalyan, Katikapalli Subramanyam. 2024. "A Survey of GPT-3 Family Large Language Models Including ChatGPT and GPT-4." *Natural Language Processing Journal* 6(October 2023):100048. doi: 10.1016/j.nlp.2023.100048.
Khennouche, Feriel, Youssef Elmir, Yassine Himeur, Nabil Djebari, and Abbes Amira. 2024. "Revolutionizing Generative Pre-Traineds: Insights and Challenges in Deploying ChatGPT and Generative Chatbots for FAQs." *Expert Systems with Applications* 246(October 2023):123224. doi: 10.1016/j.eswa.2024.123224.
Liao, Wenjie, Xinzheng Lu, Yifan Fei, Yi Gu, and Yuli Huang. 2024. "Generative AI Design for Building Structures." *Automation in Construction* 157(October 2023). doi: 10.1016/j.autcon.2023.105187.
Ministry of Agriculture and Food Economy. 1994. "Energy Efficiency Act." *Journal of Laws of the Republic of Poland* 89(414).
Ministry of Construction Transport and Infrastructure of the Republic of Serbia. 2018. "Implementation Status of EPBD in Serbia." Pp. 1–17 in *EECG meeting in Belgrade*. Belgrade.
Ministry of Infrastructure. 2002. "The Decree on Technical Requirements Which Buildings and Their Location Should Comply With." *Journal of Laws of the Republic of Poland* (75).
Ministry of Infrastructure Construction and Maritime Economy. 2013. "Minister of Infrastructure." *Journal of Laws of the Republic of Poland* (32).
Ministry of Spatial Planning and Construction. 1994. "Buildng Law." *Journal of Laws of the Republic of Poland* 89(414).
OpenAI. 2023. "GPT-4 Technical Report." 4:1–100.
Passby, Lauren, Nathan Jenko, and Aaron Wernham. 2023. "Performance of ChatGPT on Dermatology Specialty Certificate Examination Multiple Choice Questions." *Clinical and Experimental Dermatology*. doi: 10.1093/ced/llad197.
Polish Committee for Standardization. 1964. "Heat Transfer Coefficients k for Building Partitions (PN-64/B-03404)." Retrieved October 16, 2023 (https://sklep.pkn.pl/pn-b-03404-1964p.html).
Polish Committee for Standardization. 1974. "Heat Transfer Coefficients K for Building Partitions (PN-B-03404:1974)." Retrieved October 16, 2023 (https://sklep.pkn.pl/pn-b-03404-1974p.html).

Polish Committee for Standardization. 1982. “Thermal Insulation of Buildings. Requirements and Calculations (PN-B-02020).” Retrieved October 16, 2023 (https://sklep.pkn.pl/pn-b-02020-1982p.html).

Polish Committee for Standardization. 1991. “Thermal Insulation of Buildings. Requirements and Calculations (PN-B-02020).” Retrieved October 16, 2023 (https://sklep.pkn.pl/pn-b-02020-1991p.html).

Polish National Energy Conservation Agency. 2012. *Polish Building Typology*. Warsaw.

Popović, Milica Jovanović, Dušan Ignjatović, Aleksandar Rajčić, Ljiljana Đukanović, Miloš Nedić, Bojana Stanković, Nataša Ćuković Ignjatović, Branislav Živković, Aleksandra Sretenović, Željko Đurišić, and Dimitrije Kotur. 2018. *National Typology of Kindergartens in Serbia*. Belgrade: Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH.

Pursnani, Vinay, Yusuf Sermet, Musa Kurt, and Ibrahim Demir. 2023. “Performance of ChatGPT on the US Fundamentals of Engineering Exam: Comprehensive Assessment of Proficiency and Potential Implications for Professional Environmental Engineering Practice.” *Computers and Education: Artificial Intelligence* 5(April):100183. doi: 10.1016/j.caeai.2023.100183.

Rahaman, Md. Saidur, M. M. Tahmid Ahsan, Nishath Anjum, Harold Jan R. Terano, and Md. Mizanur Rahman. 2023. “From ChatGPT-3 to GPT-4: A Significant Advancement in AI-Driven NLP Tools.” *Journal of Engineering and Emerging Technologies* 1(1):50–60. doi: 10.52631/jeet.v1i1.188.

Rudolph, Jürgen, Shannon Tan, and Samson Tan. 2023. “War of the Chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The New AI Gold Rush and Its Impact on Higher Education.” *Journal of Applied Learning and Teaching* 6(1):364–89. doi: 10.37074/jalt.2023.6.1.23.

Stein, Britta, Elena Dascalaki, Malgorzata Popiolek, Jerzy Kwiatkowski, Maria Amtmann, and Zdravko Georgiev. 2012. *Typology Approaches for Non-Residential Buildings in Five European Countries: Existing Information, Concepts and Outlook*.

The European Parlament and the Council of European Union. 2010. “Directive 2010/31/EU on the Energy Performance of Buildings.” *Official Journal of the European Union* 13–35.

The Government of Poland. 2008a. “Regulation on the Technical Conditions to Be Met by Buildings and Their Location.” *Journal of Laws of the Republic of Poland* (201):14.

The Government of Poland. 2008b. “The Act of Supporting Thermomodernisation and Renovation.” *Journal of Laws of the Republic of Poland* 1–1.

Webb, Taylor, Keith J. Holyoak, and Hongjing Lu. 2023. “Emergent Analogical Reasoning in Large Language Models.” *Nature Human Behaviour*. doi: 10.1038/s41562-023-01659-w.

Życzyńska, Anna, and Tomasz Cholewa. 2015. “The Modifications to the Requirements on Energy Savings and Thermal Insulation of Buildings in Poland in the Years 1974-2021.” *Budownictwo i Architektura* 14(1):145–54. doi: 10.35784/bud-arch.1680.

Nebojša Jurišević
Faculty of Engineering,
University of Kragujevac,
Kragujevac, Serbia
jurisevic@kg.ac.rs

Robert Kowalik
Faculty of Environmental Engineering, Geomatics and Renewable Energy, University of Technology,
Kielce, Poland

Dušan Gordić
Faculty of Engineering,
University of Kragujevac,
Kragujevac, Serbia
gordic@kg.ac.rs

rkowalik@tu.kielce.pl

Alekasndar Novaković
School of Mathematics and Physics, Queen's University Belfast, Belfast, United Kingdom
Joint Research Centre in AI for Health and Wellness, Faculty of Business and IT, Ontario Tech University, Ontario, Canada
a.novakovic@qub.ac.uk

Vladimir Vukašinović
Faculty of Engineering, University of Kragujevac, Kragujevac, Serbia
vladimir.vukasinovic@kg.ac.rs

Nikola Rakić
Faculty of Engineering, University of Kragujevac, Kragujevac, Serbia
nikola.rakic@fink.rs

Jelena Nikolić
Faculty of Engineering, University of Kragujevac, Kragujevac, Serbia
jelena.nikolic@fink.rs

Arso Vukićević
Faculty of Engineering, University of Kragujevac, Kragujevac, Serbia
arso_kg@yahoo.com