Croatian Journal of Education
Vol.27; No.3/2025, pages: 863-890
Original research paper
Paper submitted: 5th June 2025
Paper accepted: 5th September 2025
https://doi.org/10.15516/cje.v27i3.6596

Appropriate and Inappropriate Use of Artificial Intelligence Language Models by Students: Teachers' (Self-)Perceptions and Experiences Across Educational Levels

Irena Miljković Krecar[1] and Nina Pavlin-Bernardić[2]
[1]VERN' University
[2]University of Zagreb, Faculty of Humanities and Social Sciences

Abstract

This study examined primary, secondary, and higher education teachers' self-efficacy in identifying AI-generated student work, concerns about potential negative effects of AI on learning, perceptions of the frequency of appropriate and inappropriate AI use among students, and actual AI-giarism detection. A total of 511 Croatian teachers completed an online survey comprising newly developed scales. Welch's ANOVA revealed differences between educational levels for most variables, with small to small-to-medium effect sizes. Secondary school teachers perceived the highest rates of inappropriate AI use and had higher confidence in detecting students' AI use than higher education teachers, while primary school teachers reported the lowest rates of appropriate use and detection frequency. Academic AI concern was high across all groups, and self-efficacy scores were slightly above the midpoint of the scale. Hierarchical regression analyses showed that the perceived inappropriate AI use was predicted by the extent of teachers' AI tool use and concerns, whereas the actual AI-giarism detection frequency was predicted by education level, recognized AI tools, extent of AI use, and confidence in detection. The results highlight the need for targeted teacher training, clear institutional guidelines, and AI-resilient assessment design.

Key words: *academic AI concern; AI-giarism detection; appropriate AI use; inappropriate AI use; self-efficacy; teachers' perceptions*

Introduction

The usage of artificial intelligence (AI) by students in academic contexts has attracted increasing attention, especially since OpenAI officially released ChatGPT in November 2022. Educators' perspectives towards the integration of AI tools reveal a tension between the opportunities and concerns. While some instructors recognize the potential of AI for enhancing learning experiences (Aljuaid, 2024; Khalifa & Albadawy, 2024), there is a growing concern that reliance on AI may erode essential cognitive skills among students, leading to a superficial understanding of content and negative impacts on critical thinking and problem-solving abilities (Mai et al., 2024, Miah et al. 2024).

In previous studies that include teachers' estimates of how frequently pupils and students use AI, these estimates typically refer to the overall frequency of AI use for school-related tasks, without distinguishing between the frequency of appropriate and inappropriate uses. For example, in an international study involving teachers across different educational levels (Delello et al., 2025), 8.2 % of teachers estimated that students use AI for their assignments almost always, 41.3 % reported frequent use, 40.1 % occasional use, whereas 10.3 % believed that pupils/students never use AI. At the secondary-school level, teachers' estimates in Norway (Graham et al., 2025) suggest that students use generative AI in writing very rarely, and when they do, teachers most often describe "supportive" functions such as information gathering, improving sentences/grammar, and expanding vocabulary. However, studies based on teachers' estimates remain relatively scarce, and there is a need for research that, in addition to measuring overall frequency of students' AI use, also directly and comparably assesses the frequency of appropriate and inappropriate patterns of AI use across educational levels.

Many concerns were raised regarding the ethical use of such tools by students (Wieczorek et al., 2025). One of the most widely cited definitions of academic integrity describes it as "the commitment, even in the face of adversity, to six fundamental values: honesty, trust, fairness, respect, responsibility, and courage" (International Center for Academic Integrity, 2021, p. 5). These values provide a widely accepted ethical framework for evaluating the use of emerging technologies such as AI in academic contexts, ensuring that innovation does not compromise the principles of responsible academic work. However, recent research indicates that the easy accessibility of AI (e.g., via mobile phones) and its technological capabilities have contributed to more frequent instances of cheating and plagiarism (Chan, 2025; Oravec, 2023), giving rise to the new term "AI-giarism."

Students' motivations for engaging with AI span from the desire to optimize time management to a perceived social norm among peers, belief that "everyone is doing it." Studies also show that many students perceive AI-generated content as a legitimate alternative to traditional research and writing methods, which is supported by the lack of clarity regarding the boundary between AI use and misuse (Chan, 2025; Kalicharan & Butler, 2024).

Overall, research indicates that the use of AI tools in fulfilling academic tasks has become common in recent years. For instance, von Garrel and Mayer (2023), in a large-scale study involving over 6,000 German students, examined the general frequency of AI tools use in academic studies, without distinguishing between ethical and unethical applications. They found that 63.4 % of the respondents reported using AI-based tools in their studies. Approximately one in four students reported using such tools (very) frequently, while nearly a half used them (very) rarely or only occasionally. The 2024 Global AI Student Survey found that 86 % of bachelor, master's, and doctoral students across 16 countries use AI for their studies, with 24 % using it daily and 54 % weekly or more frequently (Digital Education Council, 2024).

Regarding ethical AI use, Miljković Krečar et al. (2024) found that almost one third of Croatian university students in their sample considered ChatGPT as an acceptable tool for getting some initial information or as a source of inspiration. A majority of students (74.8 %) found it unacceptable to take over more than half of the written content from ChatGPT, while the rest found it acceptable. Using ChatGPT to write an entire paper or a task was considered acceptable only by 1 % of the students.

While there is a growing body of research addressing academic dishonesty involving AI in higher education, it seems there have been no empirical studies to date that directly compare inappropriate AI use across different educational levels such as elementary schools, secondary schools, and universities. The study by Štambuk et al. (2015), the only one we found comparing teachers' perceptions of students' academic cheating across different educational levels, showed no differences either in their perceptions of the frequency or acceptability of cheating, nor in their general estimates of cheating prevalence in their own classes. Regarding the perceived prevalence of cheating in their schools or universities, most teachers estimate that it occurs sometimes or often, while a much smaller number of teachers believe it occurs very often, rarely, or never. However, given the year of data collection, this study naturally did not include teachers' estimates of students' use of artificial intelligence.

Thus, the aim of this study was to examine teachers' perceived ability (self-efficacy) to identify AI-generated or AI-assisted student work, their concerns about the potential negative effects of AI on learning, their perceptions of how frequently students use AI in appropriate and inappropriate ways, and the actual detection frequency of inappropriate AI use. Teachers' self-efficacy in recognizing inappropriate AI use is an important construct because it can influence both their classroom practices and their confidence in enforcing academic policies.

The research also focused on teachers at the primary, secondary, and higher education levels to capture possible differences related to the educational context in which they work. The groups of students they teach differ in age, developmental stage, and academic demands, making such a comparison important for identifying context-specific challenges and supporting their needs. In the absence of prior empirical evidence comparing teachers' perceptions of students' appropriate use of AI across

educational levels and based on the findings of Štambuk et al. (2015) regarding inappropriate academic behaviour, we expected no significant differences between elementary, secondary, and higher education teachers in their estimates of students' inappropriate or appropriate use of AI.

Teachers' prior exposure to and familiarity with AI tools is likely to shape both their attitudes toward AI in education and their ability to detect its use by students. According to Bandura's (1997) social-cognitive theory, personal mastery experiences (i.e., hands-on experience) are fundamental sources of self-efficacy beliefs. In the context of AI, this means that teachers with more opportunities to interact with AI tools, as well as more frequent experience in grading student work, may develop stronger detection-related self-efficacy. Research on technology integration in education consistently shows that personal experience with a given technology can influence educators' confidence in managing its classroom use, as well as their perceptions of its benefits and risks (e.g., Ertmer & Ottenbreit-Leftwich, 2010). Therefore, in addition to measuring teachers' self-efficacy in detecting AI-generated work and their concerns about the impact of AI, we also assessed their personal experiences with AI, including the number of AI tools they recognize, the number of AI tools they use, and the frequency of AI tools use.

Differences in the teaching context may also contribute to self-efficacy beliefs. Higher education teachers, for example, not only tend to engage more with AI-related tools due to their academic and research responsibilities, but also use them routinely to grade large numbers of complex written assignments. Secondary school teachers may likewise regularly assess extensive written assignments. In contrast, primary school teachers typically work with younger learners whose developmental stage and assessment formats (e.g., shorter assignments, more in-class work) may reduce both the opportunities and the necessity for detecting AI-generated content. Therefore, we expected higher self-efficacy and higher detection frequency of AI-generated or AI-assisted student work among higher education and secondary school teachers compared to primary school teachers.

While teacher concerns about technology can vary depending on the context, the existing literature does not provide clear or consistent evidence of systematic differences in concerns about the negative educational impact of students' AI use at different levels of education. Therefore, we expect similar levels of concern among teachers across all educational levels. Based on the above, we propose the following two research questions and the corresponding hypotheses:

1. Are there differences between elementary school, secondary school, and university teachers in: (a) self-efficacy in detecting students' inappropriate use of artificial intelligence (AI), (b) concerns about the negative educational impact of students' AI use, (c) the perceived frequency of students' appropriate and inappropriate AI use, and (d) the frequency with which they have confidently detected such inappropriate AI use?

Hypothesis-1a: Secondary school teachers and higher education teachers will have higher self-efficacy in detecting students' inappropriate use of AI than primary school teachers.

Hypothesis-1b: There will be no statistical differences in concerns between the groups of teachers.

Hypothesis-1c: There will be no statistical differences in the perceived frequency of students' appropriate or inappropriate AI use between the groups of teachers.

Hypothesis-1d: Secondary school teachers and higher education teachers will have higher frequency of confident detection of inappropriate AI use (AI-giarism) than primary school teachers.

2. To what extent do the educational level at which teachers teach, the number of AI tools teachers know, the number of AI tools they use, the frequency of their AI use, self-efficacy in detecting students' inappropriate AI use, and concerns about the negative educational impact of students' AI use predict: (a) the perceived frequency of students' inappropriate AI use, and (b) the frequency with which teachers have confidently detected inappropriate AI use?

Hypothesis-2: For both dependent variables, we expect that the background variables (educational level at which teachers teach, the number of AI tools known, the number of AI tools used, and frequency of AI use) will explain a statistically significant proportion of variance. Furthermore, we expect that self-efficacy and concerns will explain a significant additional proportion of variance over and above that accounted for by the background variables.

Methodology

Participants

The study employed a convenience sample of 517 educators working at various educational levels in Croatia: subject teachers in primary schools (i.e., from 5th grade onwards), secondary school teachers, and teachers from higher education institutions. After excluding data from individuals who are not teachers (i.e., secretaries, librarians, administrative staff, professional associates, and classroom teachers), the final sample consisted of 511 educators. 66.1 % of them were females, 32.9 % males, 0.6 % of respondents selected "Other," while 0.4 % selected "I don't want to answer.". The participants' average age was 46.3 (ranging from 24 to 70). Based on the type of institution in which they are employed, 70 participants (13.7 %) worked in elementary schools, 151 (29.5 %) in secondary schools, and 290 (56.8 %) in higher education institutions (across different academic disciplines). Among secondary school teachers, 50.3 % worked in general secondary schools (grammar schools), 37.7 % in four-year vocational schools, 9.2 % in three-year vocational schools, and 2.6 % in art secondary schools. Among university teachers, the largest number were from the field of social sciences (46.6 %), followed by the humanities (17.6 %) and technical sciences (14.1 %). Smaller numbers of teachers were from the natural sciences (7.2 %) and biomedicine and health (6.9 %) fields. Only a small number were from other fields (7.5 % in total).

Procedure

This study employed a cross-sectional, exploratory survey design. Data were collected via an anonymous online questionnaire developed by the research team.

The questionnaire was distributed through the authors' personal and professional networks, social media groups for educators and official institutional emails sourced from the Croatian Ministry of Science and Education. The questionnaire completion time was 10 to 15 minutes and data collection lasted approximately one month.

The participants were informed about the aim, anonymity and voluntary participation in the study, as well as about the fact that the results would be reported exclusively in the aggregate form.

Instruments

In addition to demographic data (age, gender, years of teaching experience, type of institution, educational level), the questionnaire comprised four main sections: familiarity with and use of AI tools, self-efficacy in detecting AI-generated work, concerns about the educational impact of AI, and perceived student use of AI for appropriate and inappropriate purposes. Given that no existing instruments measure teachers' detection of AI use among students, new scales were developed for this study. As this is the first examination of the instruments, exploratory factor analysis (EFA) was conducted to investigate the underlying structures of the scales.

The first group of questions focused on assessing the *number of AI tools recognized and used*. The participants were presented with a list of popular AI-based tools and language models (e.g., ChatGPT, Google Bard, GPT-Neo, BERT) and were asked to indicate all those they had heard of (the results could range from 0 to 12). They were then asked to mark which of the listed tools they personally use, with the option to add any additional tools not included in the list. Finally, the participants indicated how often they use AI tools, choosing one of eight options ranging from "never" to "multiple times a day". The results could range from 0 to 7.

The Self-Efficacy in AI Detection Scale was developed for the purpose of this study. It consisted of 10 items, assessing teachers' self-efficacy related to identifying AI-generated or AI-assisted assignments, plagiarism, and other forms of inappropriate use of AI in the educational context. Responses were given on a 5-point Likert-type scale ranging from 1 ("strongly disagree") to 5 ("strongly agree"). The exploratory factor analysis (principal axes factoring with oblimin rotation) revealed a two-factor structure which jointly explained 55.5 % of the total variance. Two items were excluded from the scale due to high uniqueness values (> .60). The first subscale was named *Confidence in AI detection*, and it consisted of 5 items (e.g., "*I can reliably assess whether a student's assignment or essay was partially or entirely generated by an AI tool*."). The subscale demonstrated excellent internal consistency (McDonald's ω = .88). The second subscale was named *Uncertainty in AI detection*, and it consisted of 3 items (e. g. "*I often doubt my judgment about whether a text is the student's own work or copied*"). The subscale showed good internal consistency (McDonald's ω = .80). The results on the subscales were formed as the arithmetic means of the responses on the corresponding items.

The Academic AI Concern Scale was also developed for the purpose of this study, to measure teachers' concerns about the negative educational impact of students' use of

artificial intelligence tools. It captures apprehension regarding reduced critical thinking, diminished creativity, and potential hindrance to learning outcomes. The responses were given on a 5-point Likert-type scale ranging from 1 ("strongly disagree") to 5 ("strongly agree"). Although the original item pool included 10 statements, after conducting exploratory factor analysis (principal axes factoring with oblimin rotation), only four items were retained for the final version of the scale. These items loaded strongly on a single factor reflecting academic concern about the effects of AI on student learning (e.g., *"I am concerned that students are using AI as a substitute for their own thinking."*). The single-factor solution explained 59.8 % of the total variance. The remaining items were excluded due to weak loadings (< .40; Field, 2018; Tabachnick & Fidell, 2013), high uniqueness (> .60), or conceptual divergence. This shorter, unidimensional scale demonstrated strong internal consistency (McDonald's ω = .86) and clear interpretability. The result on the scale was formed as the arithmetic mean of the responses on the corresponding items.

The Teachers' Perception of Students' (In)Appropriate Use of AI Scale was developed for the purpose of this study, and included items related to students' use of AI in learning, completing homework, and taking exams. The participants were asked to estimate the percentage of their students who use AI tools for those purposes. The scale consisted of 5 items, and via exploratory factor analysis (principal axes factoring with oblimin rotation), we determined a two-factor structure which explained 56.5 % of the total variance. The subscales were: *Appropriate Use of AI* (two items, e.g. *"To assist with learning, e.g., finding information, explaining a process*"; McDonald's ω = .70), and *Inappropriate use of AI* (three items, e.g., "*To help with homework in an excessive or inappropriate manner, i.e., use that violates the rules set by your institution or yourself";* McDonald's ω = .74). The results on the subscales were formed as the arithmetic means of the responses on the corresponding items.

One item assessed *Detection frequency of inappropriate AI use*, i.e., whether teachers had ever confidently identified inappropriate use of AI by their students, and if so, how often. The participants chose one of the seven options: *(a) I have not, and I don't know how to recognize it; (b) I have not, and I believe my students don't do it; (c) I have not, and I believe AI cannot be used in my subject; (d) Yes, once; (e) Yes, two to four times; (f) Yes, five to seven times; (g) Yes, eight or more times*. For analysis, the responses were recoded into an ordinal variable from 0 to 4, reflecting the increasing frequency of detection. Responses (a) and (c) were treated as missing values, as they do not reflect judgments about students' actual behavior.

Results

Skewness and kurtosis values were examined to assess the suitability of parametric analyses. For most variables, skewness and kurtosis fell within acceptable ranges (±2; Field, 2018), indicating approximately normal distributions. An exception was the variable representing the number of AI tools used, which showed substantial positive

skewness (2.01) and leptokurtosis (6.36), suggesting non-normality. For this reason, the variable was not included in the subsequent parametric analyses (Mdn = 1, IQR = 1.00). The descriptive statistics for all analysed constructs, grouped by teachers' educational level, is shown in Table 1.

Table 1

Descriptive statistics for study variables by educational level (N = 551)

	Primary school teachers				Secondary school teachers			
	N	M	SD	Observed range	N	M	SD	Observed range
Recognized AI tools	70	3.86	2.13	1.00 - 10.00	151	3.90	2.13	0.00 - 9.00
Extent of AI tools use	70	3.21	2.59	0.00 - 7.00	151	3.24	2.33	0.00 - 7.00
Self-efficacy in AI use detection (Confidence)	70	3.41	0.85	1.00 - 5.00	151	3.47	0.81	1.00 - 5.00
Self-efficacy in AI use detection (Uncertainty)	70	3.14	1.00	1.00 - 5.00	151	3.18	1.02	1.00 - 5.00
Academic AI concern	70	3.93	0.90	1.00 - 5.00	151	3.92	0.85	1.00 - 5.00
Perceived percentage of appropriate AI use	70	35.50	25.20	0.00 - 85.00	151	51.40	25.60	0.00 - 95.00
Perceived percentage of inappropriate AI use	70	24.20	19.60	0.00 - 73.30	151	36.80	21.50	0.00 - 93.30
Detection frequency of inappropriate AI use	50	1.96	1.19	0.00 - 4.00	130	2.64	1.14	0.00 - 4.00

	Higher education teachers			
	N	M	SD	Observed range
Recognized AI tools	290	4.59	2.35	0.00 - 12.00
Extent of AI tools use	290	4.10	2.20	0.00 - 7.00
Self-efficacy in AI use detection (Confidence)	290	3.16	0.88	1.00 - 5.00
Self-efficacy in AI use detection (Uncertainty)	290	3.20	1.02	1.00 - 5.00
Academic AI concern	290	3.73	0.94	1.00 - 5.00
Perceived percentage of appropriate AI use	290	54.20	25.10	0.00 - 100.00
Perceived percentage of inappropriate AI use	290	27.50	19.30	0.00 – 100.00
Detection frequency of inappropriate AI use	228	2.61	1.08	0.00 - 4.00

To address the first research problem, we performed a one-way ANOVA. Given that the assumption of homogeneity of variances was violated, Welch's ANOVA was conducted, followed by Games–Howell post hoc tests, which are appropriate for comparisons between groups with unequal variances and sample sizes (Field, 2018). The results are presented in Table 2.

Table 2
One-way ANOVA results examining differences between primary school, secondary school, and higher education teachers in the study variables

	F	*df*	*p*	η^2	Games-Howell
Self-efficacy in AI use detection (Confidence)	7.72	2.184	< .001	0.028	3<2
Self-efficacy in AI use detection (Uncertainty)	0.11	2.182	.895	0.000	/
Academic AI concern	3.04	2.184	.051	0.011	/
Perceived percentage of appropriate AI use	15.55	2.180	< .001	0.057	1<2, 1<3
Perceived percentage of inappropriate AI use	12.84	2.179	< .001	0.052	1<2, 2>3
Detection frequency of inappropriate AI use	6.86	2.128	.001	0.037	1<2, 1<3

Note. 1 – primary school, 2 – secondary school, 3 – higher education

ANOVA revealed statistically significant differences between educational levels for most variables, with no significant differences found only for *Uncertainty* dimension of *Self-efficacy in AI use detection* and *Academic AI concern*. Post hoc comparisons showed that higher education teachers reported lower confidence in AI use detection than secondary school teachers. Primary school teachers reported lower perceived percentages of appropriate AI use compared to both secondary school and higher education teachers. For the perceived percentages of inappropriate AI use, secondary school teachers reported the highest values, while primary school teachers reported the lowest. Primary school teachers also reported a lower detection frequency of inappropriate AI use compared to both secondary school and higher education teachers. According to Cohen's (1988) guidelines, most effect sizes were in the small or small-to-medium range, with the largest effects observed for the perceived percentages of both appropriate and inappropriate AI use.

Prior to conducting the hierarchical multiple regression analyses, Pearson correlation coefficients between the study variables were calculated (Table 3). The correlation coefficients and variance inflation factors (VIFs; ranging from 1.07 to 1.40) indicated no problems with multicollinearity (Field, 2018).

Table 3
Pearson correlations for all study variables (N = 511)

	1	2	3	4	5	6	7	8	9
1. Educational level	—								
2. Recognized AI tools	.140**	—							
3. Extent of AI tools use	.169***	.361***	—						
4. Self-efficacy in AI use detection (Confidence)	-.143**	.107*	. 011	—					
5. Self-efficacy in AI use detection (Uncertainty)	.020	-.331***	-.203***	-.387***	—				

	1	2	3	4	5	6	7	8	9
6. Academic AI concern	-.100*	-.078	-.214***	.129**	.194***	—			
7. Perceived percentage of appropriate AI use	.215***	.070	.162***	-.088*	.023	.009	—		
8. Perceived percentage of inappropriate AI use	-.037	.007	.112**	.054	.070	.233***	.299***	—	
9. Detection frequency of inappropriate AI use	.141**	.223***	.196***	.224***	-.072	.068	.083	.302***	—

Note. * p < .05, ** p < .01, *** p < .001. N = 511 for all variables, except for Detection frequency of inappropriate AI use where N = 408. Educational level: 1 – primary school, 2 – secondary school, 3 – higher education

Table 4 presents the results of the hierarchical multiple regression analyses. Two dependent variables were examined: *Perceived percentage of students' inappropriate AI use* and *Detection frequency of inappropriate AI use*. In Step 1, educational level, the number of recognized AI tools, and the extent of AI tools use were entered as predictors. In Step 2, three additional predictors were added: self-efficacy in AI use detection (Confidence), self-efficacy in AI use detection (Uncertainty), and academic AI Concern.

Table 4

Results of hierarchical regression analyses with perceived percentage of students' inappropriate AI use and detection frequency of inappropriate AI use as dependent variables (N = 511)

	Perceived percentage of students' inappropriate AI use		Detection frequency of inappropriate AI use	
	Step 1 B	Step 2 β	Step 1 β	Step 2 β
Educational level	-.056	-.038	.095	.137**
Recognized AI tools	-.003	-.002	.170**	.158**
Extent of AI tools use	.133**	.193***	.115*	.144**
Self-efficacy in AI use detection (Confidence)		-.044		.256***
Self-efficacy in AI use detection (Uncertainty)		.073		.081
Academic AI concern		.249***		.083
R	.129	.297	.271	.374
R^2	.017	.088	.074	.140
adj. R^2	.011	.080	.067	.127
F (df)	2.88* (3,507)	8.12*** (6,504)	10.7*** (3,404)	10.9*** (6,401)
Δ R^2		.071		.067
*F***Δ** *(df)*		13.20*** (3,504)		10.3*** (3,401)

Note. * *p* < .05, ** *p* < .01, *** *p* < .001. Educational level: 1 – primary school, 2 – secondary school, 3 – higher education

For *Perceived percentage of students' inappropriate AI use*, Step 1 accounted for 1.7 % of the variance ($p < .05$), with the extent of AI tools use emerging as the only significant predictor. In Step 2, the explained variance increased to 8.8 %, with the extent of AI tools use and academic AI concern being significant positive predictors. For *Detection frequency of inappropriate AI use*, Step 1 accounted for 7.4 % of the variance ($p < .001$), with recognized AI tools and the extent of AI tools use as significant predictors. In Step 2, the explained variance increased to 14.0 %, with recognized AI tools, the extent of AI tools use, and self-efficacy in AI use detection (Confidence) as significant predictors.

Discussion

The aim of this study was to examine teachers' self-efficacy in identifying AI-generated or AI-assisted student work, their concerns about the potential negative effects of AI on learning, their perceptions of how frequently students use AI in appropriate and inappropriate ways, and the actual detection frequency of inappropriate AI use, across different educational levels. The factor analysis revealed that the self-efficacy measure comprises two factors – *Confidence and uncertainty in detecting AI-generated student work*. This suggests that teachers' self-perceptions are not simply placed on a single continuum ranging from low to high self-efficacy. Rather, teachers may simultaneously experience both confidence and uncertainty; for example, they may feel confident when assessing assignments from students they know well, while still doubting their judgment in more ambiguous cases. This duality reflects the complex and context-dependent nature of teachers' self-efficacy in AI detection.

We expected that secondary school and higher education teachers would have a higher self-efficacy level in detecting students' inappropriate use of AI than primary school teachers. However, the results showed that on the confidence factor, secondary school teachers scored higher than higher education teachers. This difference could be linked to differences in the types of assignments produced at different educational levels, as well as the extent to which teachers know their students. Secondary school teachers typically assign more homework and administer more short tests. This may provide them with more opportunities to encounter and practice identifying AI-generated content, thereby strengthening their confidence. In contrast, university teachers often teach larger groups of students and do not know them well personally, while the absence of a difference between primary school teachers and the other two groups may be due to the nature of work at the primary level, where assignments are typically shorter, more structured, and closely monitored, making potential AI use easier to detect. In such contexts, even limited exposure to AI-generated work might be sufficient to foster a comparable level of confidence.

In contrast, no significant differences were found for the other self-efficacy factor – *Uncertainty in AI use detection*. Given the short timeframe since the introduction of widely available generative AI tools, teachers at all educational levels are likely relying on similar sources of information – such as media reports, professional networks, and

informal peer discussions—rather than on systematic, level-specific training. This shared exposure may contribute to a relatively uniform sense of uncertainty. This is consistent with the findings obtained by Miljković Krečar et al. (2024), who reported that university teachers expressed only moderate certainty as well as limited accuracy (53 %) in detecting AI-generated work.

Also, as expected, for academic AI concern no significant differences were found between teachers, suggesting that apprehension regarding the potential negative impact of AI on student learning is shared across primary, secondary, and higher education contexts. The arithmetic means, close to 4 on a 5-point scale, indicate a generally high level of concern. Such scores reflect widespread uncertainty about how AI will influence students' critical thinking, creativity, and independent problem-solving skills, as highlighted in previous studies (e.g., Mai et al., 2024; Miah et al., 2024).

For appropriate AI use, we initially hypothesized that there would be no significant differences between educational levels, based on the absence of prior empirical evidence suggesting such variation. Our results indicated that primary school teachers reported the lowest rates, and the effect size was small-to-medium. A possible reason for that might be that younger students typically have fewer academic tasks that lend themselves to AI-assisted work and less developed digital skills, contributing to the perception that appropriate AI use is less frequent among them. It is also possible – though more speculative – that greater parental supervision and stronger adherence to classroom rules among younger students play a role. At the same time, one should not exclude the possibility that some parents themselves may assist their children by using AI tools to find answers or explanations.

Although we had postulated no differences for inappropriate AI use, the results revealed them, with a small-to-medium effect size. Secondary school teachers perceived the highest rates of inappropriate AI use among their students. One possible explanation is that secondary school students may have sufficient digital skills and access to AI tools to use them in ways that violate institutional or teacher-set rules, while still lacking the academic integrity standards expected at the higher education level. The academic demands in secondary schools, such as frequent homework, written assignments, and tests, may also create more opportunities or incentives for inappropriate AI use compared to primary school.

When considering the absolute mean values, the results suggest that, across all educational levels, teachers perceive appropriate use of AI to be more frequent than inappropriate use, though mean estimates for the latter were also substantial. However, studies relying on students' self-reports (Digital Education Council, 2024; Miljković Krečar et al., 2024; von Garrel & Mayer, 2023) indicate even higher rates of AI use. Therefore, it is possible that teachers' estimates underestimate both students' appropriate and inappropriate AI use.

With respect to our second research problem, for both the perceived percentage of students' inappropriate AI use and detection frequency of inappropriate AI use, the

results were consistent with the hypothesis: in both steps of the hierarchical regression analyses, a significant proportion of variance was explained, with the second step accounting for additional variance beyond the first.

For the first dependent variable, significant positive predictors were – the extent of AI tools use and concerns. This suggests that attitudinal and affective factors may shape such perceptions and teachers who are more apprehensive about the potential negative consequences of AI on learning may be more likely to interpret student behaviour as inappropriate or to assume that such behaviour is frequent.

In contrast, for detection frequency of inappropriate AI use, significant predictors were educational level, recognized AI tools, the extent of AI tools use, and self-efficacy in AI use detection (confidence factor), indicating that teachers who believe they can effectively identify AI-generated or AI-assisted work are also more likely to report having detected it. The role of educational level, recognized AI tools, and the extent of AI tools use suggests that familiarity and experience with AI, both personally and within the teaching context, may provide teachers with more opportunities or cues for detection. For example, teachers at higher educational levels might be exposed to more complex students' assignments where AI-assisted work is more feasible, and those who know and use more AI tools themselves may recognize AI use more easily in student work. It should be noted however, that the predictors used in this study explained only a modest proportion of variance in the dependent variables (8 % and 12.7 %), indicating that future research should consider additional potential predictors.

This study focused on a novel and rapidly evolving topic – teachers' perceptions of AI use in education – addressing both appropriate and inappropriate use, as well as teachers' self-efficacy and concerns, in a single framework. The research included teachers from primary, secondary, and higher education, allowing for comparisons across educational levels that have rarely been examined in previous research. However, the study has several limitations. The sample is not representative. For instance, the participation rate might have been higher among teachers with greater concerns regarding AI use. The number of primary school teachers in the sample was smaller (n = 70) than in the other groups. It is also possible that teachers' estimates of the frequency of students' AI use are underestimated; therefore, future research in Croatia should also examine students' self-reports on this matter. Also, we did not consider the type of secondary school or field of study, although these factors may also influence the results. For example, von Garrel and Mayer (2023) reported the highest AI usage rates among students in engineering, mathematics, and natural sciences, while Miljković Krečar et al. (2024) found that science and technology students reported using AI more frequently than students in the social sciences.

Conclusion

This study provided novel insights into teachers' perceptions, self-efficacy, concerns and experiences regarding students' use of AI tools across primary, secondary, and

higher education. One of these insights is the duality of the self-efficacy measure, which reflects the complex and context-dependent nature of teachers' self-perceptions regarding their capacity to detect AI-generated work. While uncertainty in detecting AI-generated work was similar across all educational levels, secondary school teachers reported the highest levels of confidence in their capacity for such detection.

Additionally, although concerns about AI's potential negative impact on learning were uniformly high across all educational levels, differences emerged in perceptions of both appropriate and inappropriate AI use. Primary school teachers reported the lowest rates of appropriate AI use, whereas secondary school teachers perceived the highest rates of inappropriate use.

Hierarchical regression analyses further elaborated these differences. For the perceived percentage of students' inappropriate AI use, significant predictors were the extent of AI tools use and concern about the educational impact of AI. Concern emerged as a positive predictor, indicating that teachers with higher levels of concern perceived higher rates of inappropriate AI use among students. For detection frequency of inappropriate AI use, significant predictors were educational level, recognized AI tools, the extent of AI tools use, and confidence in AI use detection.

Our results, e.g., teachers' medium self-efficacy in students' AI use detection, indicate the need for clearer guidelines and better training in this area for academic staff. Also, given the teachers' concerns and the perceived frequency of students' AI use, there is a growing need to rethink assessment strategies and design AI-resilient tasks, for example, with greater emphasis on oral examinations and demonstrations of conceptual understanding. In line with this, Lye & Lim (2024) note that integration of AI in education should be considered through Fit-for-Purpose Assessment: The Against, Avoid and Adopt Principle. The principle encourages educators to clearly define what uses of AI are prohibited, to redesign or avoid assignments that are easily compromised by AI, and to thoughtfully adopt AI where it can enhance learning, all while explicitly communicating these expectations to students.

Acknowledgement

During the research and writing of this paper, artificial intelligence tools were used. Specifically, ChatGPT (OpenAI, 2025) was employed to assist with questionnaire item development, translation, language refinement, and summarizing ideas, while Perplexity AI (2025) was used as a supplementary tool for gathering and exploring the relevant literature. All content generated with the help of these tools was thoroughly and critically reviewed and edited by the authors to ensure academic integrity.

References

Aljuaid, H. (2024). The impact of artificial intelligence tools on academic writing instruction in higher education: A systematic review. *Arab World English Journal (AWEJ), Special Issue on ChatGPT*, 26–55. https://dx.doi.org/10.24093/awej/ChatGPT

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Freeman.

Chan, C. K. Y. (2025). Students' perceptions of 'AI-giarism': investigating changes in understandings of academic misconduct. *Education and Information Technologies*, *30*, 8087–8108. https://doi.org/10.1007/s10639-024-13151-7

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Delello, J. A., Sung, W., Mokhtari, K., Hebert, J., Bronson, A., & De Giuseppe, T. (2025). AI in the Classroom: Insights from Educators on Usage, Challenges, and Mental Health. *Education Sciences*, *15*(2), 113. https://doi.org/10.3390/educsci15020113

Digital Education Council. (2024). *Global AI Student Survey*. https://campustechnology.com/articles/2024/08/28/survey-86-of-students-already-use-ai-in-their-studies.aspx?utm_source=chatgpt.com

Ertmer, P. A., & Ottenbreit-Leftwich, A. T. (2010). Teacher technology change: How knowledge, confidence, beliefs, and culture intersect. *Journal of Research on Technology in Education*, *42*(3), 255–284. https://doi.org/10.1080/15391523.2010.10782551

Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics* (5th ed.). Sage.

Graham, S., Skar, G. B., Kvistad, A. H., & Jeffery, J. V. (2025). Secondary teachers' use and beliefs about generative artificial intelligence and writing in school. *Reading and Writing*. https://doi.org/10.1007/s11145-025-10699-9

International Center for Academic Integrity. (2021). *The fundamental values of academic integrity* (3rd ed.). International Center for Academic Integrity. https://academicintegrity.org/the-fundamental-values-of-academic-integrity

Kalicharan, L., & Butler, C. (2024). Tertiary education students' perceptions of plagiarism in academic writing. *Journal of Applied Research in Higher Education*, *17*(1), 510-525. https://doi.org/10.1108/jarhe-01-2024-0015

Khalifa, M., & Albadawy, M. (2024). Using artificial intelligence in academic writing and research: An essential productivity tool. *Computer Methods and Programs in Biomedicine Update*, *5*, Article 100145. https://doi.org/10.1016/j.cmpbup.2024.100145

Lye, C., & Lim, L. (2024). Generative artificial intelligence in tertiary education: assessment redesign principles and considerations. *Education Sciences*, *14*(6), 569. https://doi.org/10.3390/educsci14060569

Mai, D. T. T., Da, C. V., & Hanh, N. V. (2024). The use of ChatGPT in teaching and learning: A systematic review through SWOT analysis approach. *Frontiers in Education*, *9*, 1328769. https://doi.org/10.3389/feduc.2024.1328769

Miah, A. S. M., Tusher, M. M., Hossain, M. M., Hossain, M. M., Rahim, M. A., Hamid, M. E., Islam, M. S., & Shin, J. (2024). *ChatGPT in Research and Education: Exploring benefits and threats* (arXiv preprint, arXiv:2411.02816). arXiv. https://doi.org/10.48550/arXiv.2411.02816

Miljković Krečar, I., Kolega, M., & Jurčec, L. (2024). Perception of ChatGPT usage for homework assignments: Students' and professors' perspectives. *IAFOR Journal of Education*, *12*(2), 33-60. https://doi.org/10.22492/ije.12.2.02

Oravec, J. A. (2023). Artificial intelligence implications for academic cheating: Expanding the dimensions of responsible human-AI collaboration with ChatGPT. *Journal of Interactive Learning Research*, *34*(2), 213–237.

Štambuk, M., Maričić, A., & Hanzec, I. (2015). Cheating is unacceptable, but... Teachers' perceptions of and reactions to students' cheating at schools and universities. *Croatian Journal of Education*, *17*(4), 259–288. https://doi.org/10.15516/cje.v17i0.1258

von Garrel, J., & Mayer, J. (2023). Artificial intelligence in studies – Use of ChatGPT and AI-based tools among students in Germany. *Humanities and Social Sciences Communications*, *10*, 799. https://doi.org/10.1057/s41599-023-02304-7

Wieczorek, M., Hosseini, M., & Gordijn, B. (2025). Unpacking the ethics of using AI in primary and secondary education: A systematic literature review. *AI and Ethics*. https://doi.org/10.1007/s43681-025-00770-0

Irena Miljković Krecar
VERN' University
Palmotićeva 82/1, 10000 Zagreb, Croatia
irena.miljkovic@vern.hr

Nina Pavlin-Bernardić
University of Zagreb
Faculty of Humanities and Social Sciences
Ivana Lučića 3, 10000 Zagreb, Croatia
nbernardi@m.ffzg.hr

Prikladna i neprikladna uporaba jezičnih modela umjetne inteligencije među učenicima: (samo) percepcije i iskustva nastavnika u različitim obrazovnim razinama

Sažetak

U ovom istraživanju nastavnika osnovnoškolske, srednjoškolske i visokoškolske razine ispitani su samoefikasnost u prepoznavanju UI generiranih učeničkih radova, zabrinutost zbog mogućih negativnih učinaka UI-ja na učenje, percepcija učestalosti primjerenih i neprimjerenih učeničkih korištenja UI-ja, te realna detekcija UI plagijata. Ukupno 511 hrvatskih nastavnika ispunilo je mrežni upitnik koji je sadržavao novorazvijene skale. Welchova ANOVA pokazala je razlike među obrazovnim razinama za većinu varijabli, s malim do malim-srednjim veličinama učinka. Srednjoškolski nastavnici percipirali su najveće stope neprimjerene uporabe UI-ja te iskazali veću sigurnost u prepoznavanju učeničke uporabe UI-ja nego sveučilišni nastavnici, dok su učitelji osnovnih škola percipirali najniže stope primjerene uporabe UI alata i učestalosti prepoznavanja. Zabrinutost zbog akademske uporabe UI-ja bila je visoka u svim skupinama, a rezultati samoefikasnosti bili su nešto iznad sredine skale. Hijerarhijske regresijske analize pokazale su da se percipirana neprimjerena uporaba UI-ja može predvidjeti opsegom nastavničke uporabe UI alata i njihovom zabrinutošću, dok se realna učestalost prepoznavanja UI-plagijata mogla predvidjeti obrazovnom razinom, poznavanjem UI alata, opsegom osobne uporabe i sigurnošću u prepoznavanje. Nalazi naglašavaju potrebu za ciljanom edukacijom nastavnika, jasnim institucionalnim smjernicama i osmišljavanjem zadataka za provjeru znanja koji su otporni na UI.

Ključne riječi: *detekcija; UI plagijata neprimjerena uporaba UI-ja; percepcije nastavnika; primjerena uporaba UI-ja; samoefikasnost; zabrinutost zbog akademske uporabe UI-ja;*

Uvod

Učenička[1] uporaba umjetne inteligencije (UI) u obrazovnom kontekstu privlači sve veću pozornost, osobito otkako je OpenAI učinio javno dostupnim ChatGPT u studenome 2022. godine. Percepcije nastavnika prema integraciji UI alata variraju između prilika i prijetnji: dok neki nastavnici prepoznaju potencijal UI-ja za unaprjeđenje iskustava učenja (Aljuaid, 2024; Khalifa i Albadawy, 2024), raste zabrinutost da bi oslanjanje na UI moglo narušiti ključne kognitivne vještine učenika, dovodeći do površnoga razumijevanja sadržaja i negativnoga utjecaja na kritičko mišljenje i sposobnosti rješavanja problema (Mai i sur., 2024; Miah i sur., 2024).

U dosadašnjim istraživanjima koja uključuju procjene nastavnika o učeničkoj i studentskoj učestalosti uporabe UI-ja, ove se procjene obično odnose na opću učestalost korištenja UI-ja za školske zadatke, bez razdvajanja čestine primjerenih i neprimjerenih načina korištenja. Tako je u međunarodnom istraživanju s nastavnicima s različitih obrazovnih razina (Dellelo i sur., 2025) 8,2 % nastavnika procijenilo da učenici/studenti koriste UI za svoje zadatke gotovo uvijek, 41,3 % da je koriste često, 40,1 % ponekad, dok 10,3 % smatra da učenici/studenti nikad ne koriste UI. Na srednjoškolskoj razini, procjene nastavnika u Norveškoj (Graham i sur., 2025) sugeriraju da učenici generativnu UI u pisanju koriste rijetko, a kada je koriste, nastavnici najčešće navode „podržavajuće" funkcije poput prikupljanja informacija, poboljšanja rečenica/gramatike i obogaćivanja vokabulara. No, ovakva su istraživanja s nastavničkim procjenama još uvijek rijetka, a potrebna su i ona koja, uz opću učestalost učeničkoga ili studentskoga korištenja UI-ja, izravno i usporedivo mjere i učestalost primjerenih i neprimjerenih obrazaca uporabe UI-ja u različitim obrazovnim razinama.

U literaturi su izražene brojne zabrinutosti vezane uz učeničke etičke uporabe takvih alata (Wieczorek i sur., 2025). Jedna od najcitiranijih definicija akademskoga integriteta opisuje ga kao „predanost, čak i u suočavanju s poteškoćama, prema šest temeljnih vrijednosti: poštenju, povjerenju, pravednosti, poštovanju, odgovornosti i hrabrosti" (International Center for Academic Integrity, 2021, str. 5). Ove vrijednosti pružaju široko prihvaćen etički okvir za vrednovanje uporabe novih tehnologija poput UI u akademskom kontekstu, osiguravajući da inovacije ne ugroze načela odgovornoga akademskog rada. Međutim, nedavna istraživanja ukazuju na to da su laka dostupnost (posredovanjem mobilnih telefona) i tehnološke mogućnosti umjetne inteligencije pridonijele povećanoj učestalosti varanja i plagiranja (Chan, 2025; Oravec, 2023), što je dovelo do novoga pojma „AI-giarism".

Motivacije učenika za uporabu UI mogu se kretati od želje za optimizacijom upravljanja vremenom do percipirane društvene norme među vršnjacima, odnosno vjerovanja da „svi to rade". Istraživanja također pokazuju da mnogi učenici doživljavaju sadržaje generirane pomoću UI-ja kao legitimnu alternativu tradicionalnim metodama istraživanja i pisanja, što dodatno potkrepljuje nejasnoća oko granice između uporabe i zloporabe UI-ja (Chan, 2025; Kalicharan i Butler, 2024).

[1]U tekstu se, radi jednostavnosti, termini „nastavnik" i „učenik" koriste za sve razine obrazovanja, te spolne i rodne identitete, osim ako se govori specifično o visokoškolskoj razini, kada se koristi termin „student".

Općenito, istraživanja koja uključuju procjene studenata pokazuju da je uporaba UI alata u izvršavanju akademskih zadataka posljednjih godina postala uobičajena. Primjerice, von Garrel i Mayer (2023), u opsežnom istraživanju koje je obuhvatilo više od 6000 njemačkih studenata, ispitali su opću učestalost uporabe UI alata na studiju, ne razlikujući etičku i neetičku primjenu. Autori nalaze da 63,4 % ispitanika koristi UI alate, pri čemu ih otprilike svaki četvrti student koristi (vrlo) često, a gotovo polovica (vrlo) rijetko ili samo povremeno. *Global AI Student Survey* iz 2024. pokazuje da 86 % studenata preddiplomskih, diplomskih i poslijediplomskih studija u 16 zemalja koristi UI na studiju, pri čemu 24 % koristi UI svakodnevno, a 54 % barem jednom tjedno (Digital Education Council, 2024).

Ispitujući percepciju etičnosti uporabe UI alata, Miljković Krečar i sur. (2024) nalaze da gotovo trećina ispitanih hrvatskih sveučilišnih studenata smatra korištenje ChatGPT-a prihvatljivim za dobivanje početnih informacija ili kao izvor inspiracije. Većina (74,8 %) smatra neprihvatljivim da više od polovice pisanoga sadržaja bude preuzeto, dok je preostalom dijelu to prihvatljivo. Potpuno pisanje rada ili zadatka pomoću ChatGPT-a smatralo je prihvatljivim svega 1 % studenata.

Iako se sve više istraživanja bavi akademskim nepoštenjem povezanim s UI-jem u visokom obrazovanju, čini se da do sada nema empirijskih istraživanja koja bi izravno uspoređivala neprimjerenu uporabu UI-ja na različitim obrazovnim razinama, poput osnovnih škola, srednjih škola i sveučilišta. Studija Štambuk i sur. (2015), jedina koju smo pronašli da je uspoređivala percepcije nastavnika o akademskom varanju učenika na različitim obrazovnim razinama, nije pokazala razlike u njihovim percepcijama učestalosti ili prihvatljivosti varanja, ni u općim procjenama raširenosti varanja u vlastitim razredima. Što se tiče percipirane raširenosti varanja u školama ili sveučilištima u cjelini, većina nastavnika procijenila je da se ono događa ponekad ili često, dok je manji dio smatrao da se događa vrlo često, rijetko ili nikada. Međutim, s obzirom na godinu prikupljanja podataka, istraživanje nije uključivalo procjene nastavnika o učeničkoj uporabi umjetne inteligencije.

Stoga je cilj ovoga istraživanja bio ispitati nastavničku percipiranu sposobnost (samoefikasnost) u prepoznavanju učeničkih radova generiranih ili asistiranih umjetnom inteligencijom, njihovu zabrinutost zbog mogućih negativnih učinaka UI-ja na učenje, percepciju učestalosti učeničkih/studentskih primjerenih i neprimjerenih korištenja UI-ja te učestalost pouzdane detekcije UI plagijata. Samoefikasnost nastavnika u prepoznavanju neprimjerene uporabe UI-ja važan je konstrukt ovdje jer može utjecati i na njihove nastavne prakse i na sigurnost u provođenju akademskih pravila.

Također, ovo je istraživanje usmjereno na nastavnike na osnovnoškolskoj, srednjoškolskoj i visokoškolskoj razini, kako bi se uočile moguće razlike povezane s obrazovnim kontekstom u kojem nastavnici rade. Grupe učenika koje poučavaju razlikuju se prema dobi, razvojnoj fazi i akademskim zahtjevima, što ovu usporedbu čini važnom za identificiranje specifičnih izazova i potreba za podrškom. S obzirom na nedostatak ranijih empirijskih dokaza koji bi uspoređivali percepcije nastavnika o primjerenoj uporabi UI-ja na različitim obrazovnim razinama, te na temelju nalaza

Štambuk i sur. (2015) o neprimjerenom akademskom ponašanju, očekujemo da neće biti značajnih razlika između nastavnika osnovnih, srednjih i visokoškolskih ustanova u procjenama učeničkih primjerenih i neprimjerenih korištenja UI-ja.

Prethodna izloženost nastavnika i njihova upoznatost s UI alatima vjerojatno oblikuju i njihove stavove o UI-ju u obrazovanju i njihovu sposobnost detekcije učeničke uporabe UI-ja. Prema socijalno-kognitivnoj teoriji Bandure (1997), osobna iskustva ovladavanja (tj. iskustva „iz prve ruke") predstavljaju temeljne izvore samoefikasnosti; u kontekstu UI-ja, to znači da nastavnici s više prilika za interakciju s UI alatima, kao i s većim iskustvom u ocjenjivanju učeničkih radova, mogu razviti snažnija vjerovanja o vlastitoj samoefikasnosti u detekciji UI-generiranih radova. Istraživanja o integraciji tehnologije u obrazovanju dosljedno pokazuju da osobno iskustvo s određenom tehnologijom može utjecati na sigurnost nastavnika u upravljanju njezinom uporabom u učionici, kao i na njihove percepcije o koristima i rizicima (npr. Ertmer i Ottenbreit-Leftwich, 2010). Stoga smo, uz mjerenje samoefikasnosti nastavnika u detekciji UI-generiranih učeničkih radova i njihove zabrinutosti zbog utjecaja UI-ja, također procijenili njihovo osobno iskustvo s UI-jem, uključujući broj UI alata koje prepoznaju, broj UI alata koje koriste i učestalost njihova korištenja.

Razlike u nastavnom kontekstu također mogu pridonijeti razlikama u samoefikasnosti. Nastavnici u visokom obrazovanju, primjerice, ne samo da se češće koriste alatima povezanim s UI-jem zbog svojih akademskih i istraživačkih odgovornosti, već i rutinski ocjenjuju velik broj složenih pisanih radova. Srednjoškolski nastavnici također redovito vrednuju opsežnije pisane zadatke. Nasuprot tome, nastavnici u osnovnoj školi tipično rade s mlađim učenicima, čija razvojna faza i oblici vrednovanja (npr. kraći zadatci, više rada u učionici) smanjuju i prilike i potrebu za detekcijom UI generiranih sadržaja. Stoga od nastavnika u visokom obrazovanju i srednoj školi očekujemo višu samoefikasnost i češću detekciju prepoznavanja učeničkih radova generiranih ili asistiranih pomoću UI-ja u usporedbi s nastavnicima u osnovnoj školi.

Iako se zabrinutosti nastavnika zbog uporabe tehnologija mogu razlikovati ovisno o kontekstu, postojeća literatura ne pruža jasne ni dosljedne dokaze o sustavnim razlikama u zabrinutosti zbog negativnih obrazovnih posljedica učeničke uporabe UI-ja na različitim obrazovnim razinama. Stoga očekujemo slične razine zabrinutosti među nastavnicima na svim obrazovnim razinama. Na temelju svega navedenoga, postavljamo sljedeća dva istraživačka pitanja i pripadajuće hipoteze:

1. Postoje li razlike među nastavnicima osnovnih škola, srednjih škola i sveučilišta u: (a) samoefikasnosti u detekciji neprimjerene učeničke uporabe UI-ja, (b) zabrinutosti zbog negativnih obrazovnih posljedica uporabe UI-ja, (c) percipiranoj učestalosti učeničkoga primjerenoga i neprimjerenoga korištenja UI-ja te (d) učestalosti pouzdane detekcije neprimjerene uporabe UI-ja?

Hipoteza-1a: Srednjoškolski nastavnici i nastavnici u visokom obrazovanju imat će višu samoefikasnost u detekciji neprimjerene učeničke uporabe UI-ja nego nastavnici u osnovnim školama.

Hipoteza-1b: Neće biti statistički značajnih razlika u zabrinutosti među skupinama nastavnika.

Hipoteza-1c: Neće biti statistički značajnih razlika u percipiranoj učestalosti primjerenoga i neprimjerenoga korištenja UI-ja među skupinama nastavnika.

Hipoteza-1d: Srednjoškolski nastavnici i nastavnici u visokom obrazovanju imat će veću učestalost pouzdane detekcije neprimjerene uporabe UI-ja (UI plagijarizam) nego nastavnici u osnovnim školama.

2. U kojoj mjeri obrazovna razina na kojoj nastavnici predaju, broj UI alata koje poznaju, broj alata kojima se koriste, učestalost njihove uporabe, samoefikasnost u detekciji neprimjerene učeničke uporabe UI-ja i zabrinutost zbog negativnih obrazovnih posljedica učeničkih uporaba UI-ja predviđaju: (a) percipiranu učestalost neprimjerenih učeničkih uporaba UI-ja, (b) frekvenciju pouzdane detekcije neprimjerene uporabe UI-ja?

Hipoteza-2: Za obje zavisne varijable očekujemo da će pozadinske varijable (obrazovna razina na kojoj nastavnici predaju, broj poznatih UI alata, broj korištenih UI alata i učestalost uporabe UI-ja) objasniti statistički značajan udio varijance. Nadalje, očekujemo da će samoefikasnost i zabrinutost dodatno objasniti značajan udio varijance iznad onoga koji objašnjavaju pozadinske varijable.

Metodologija

Sudionici

U istraživanju je sudjelovao prigodni uzorak od 517 hrvatskih nastavnika zaposlenih na različitim obrazovnim razinama: nastavnici predmetne nastave u osnovnim školama (od 5. razreda nadalje), nastavnici u srednjim školama te nastavnici u ustanovama visokoga obrazovanja. Nakon isključenja sudionika koji nisu nastavnici (npr. tajnik, knjižničar, administrativno osoblje, stručni suradnici i učitelji razredne nastave u nižim razredima), konačni uzorak činilo je 511 nastavnika. Od toga je 66,1 % bilo žena, 32,9 % muškaraca, 0,6 % ispitanika odabralo je opciju „Ostalo", a 0,4 % „Ne želim odgovoriti". Prosječna dob sudionika bila je 46,3 godine (raspon 24 – 70). Prema vrsti institucije u kojoj su zaposleni, 70 sudionika (13,7 %) radi u osnovnim školama, 151 (29,5 %) u srednjim školama, a 290 (56,8 %) u ustanovama visokoga obrazovanja (različitih znanstvenih područja). Među srednjoškolskim nastavnicima 50,3 % radi u gimnazijama, 37,7 % u četverogodišnjim strukovnim školama, 9,2 % u trogodišnjim strukovnim školama, a 2,6 % u umjetničkim srednjim školama. Među sveučilišnim nastavnicima najveći udio je iz društvenih znanosti (46,6 %), potom humanističkih znanosti (17,6 %) i tehničkih znanosti (14,1 %). Manji udjeli su iz prirodnih znanosti (7,2 %) te biomedicine i zdravstva (6,9 %). Samo mali broj sudionika je iz ostalih područja (ukupno 7, 5 %).

Postupak

U istraživanju je korišten presječni, eksploratorni nacrt. Podatci su prikupljeni anonimnim mrežnim upitnikom koji je razvila istraživačka skupina. Upitnik je distribuiran putem

osobnih i profesionalnih mreža autora, grupa na društvenim mrežama za nastavnike te službenih institucionalnih e-adresa objavljenih na mrežnim stranicama Ministarstva znanosti i obrazovanja Republike Hrvatske. Vrijeme ispunjavanja upitnika bilo je od 10 do 15 minuta, a prikupljanje podataka trajalo je približno jedan mjesec. Sudionici su bili informirani o cilju istraživanja, anonimnosti i dobrovoljnosti sudjelovanja te da će rezultati biti prikazani isključivo u zbirnom obliku.

Instrument

Uz demografske podatke (dob, rod, godine nastavnoga iskustva, vrsta institucije, obrazovna razina), upitnik je sadržavao četiri glavna dijela: poznavanje i uporaba alata UI-ja, samoefikasnost u detekciji radova generiranih pomoću UI-ja, zabrinutost zbog obrazovnih posljedica UI-ja te percepcija učeničke uporaba UI alata u primjerene i neprimjerene svrhe. Budući da ne postoje instrumenti koji mjere nastavničku detekciju uporabe UI-ja među učenicima, za potrebe ovoga istraživanja razvijene su nove skale. Kako je riječ o prvoj primjeni instrumenta, provedena je eksploratorna faktorska analiza (EFA) radi provjere latentne strukture njegovih skala.

Prva skupina pitanja odnosila se na procjenu *broja UI alata koje nastavnici poznaju i kojima se koriste*. Sudionicima je ponuđen popis popularnih alata temeljenih na UI-ju i jezičnih modela (npr. ChatGPT, Google Bard, GPT-Neo, BERT) na kojem su trebali označiti one za koje su čuli (rezultati su se mogli kretati od 0 do 12). Zatim su trebali označiti one alate kojima se osobno koriste, s mogućnošću dodavanja alata koji nisu bili na popisu. Na kraju su označavali koliko često koriste UI alate, u rasponu od „nikada" do „više puta dnevno" (rezultati od 0 do 7).

Skala *samoefikasnosti u detekciji UI-ja* razvijena je za potrebe ovoga istraživanja. Sadržavala je 10 tvrdnji kojima se procjenjivala samoefikasnost nastavnika u prepoznavanju UI-generiranih zadataka, plagijata i drugih oblika neprimjerene učeničke uporabe UI-ja u obrazovnom kontekstu. Tvrdnje se procjenjuju na Likertovoj skali od 5 stupnjeva (1 = „uopće se ne slažem" do 5 = „u potpunosti se slažem"). Eksploratorna faktorska analiza (metoda glavnih osi s oblimin rotacijom) otkrila je dva faktora koji su zajedno objasnili 55,5 % ukupne varijance. Dvije tvrdnje su isključene zbog visokoga unikviteta (> ,60). Prva podskala nazvana je *Pouzdanje* u točnu detekciju UI-generiranih učeničkih radova i sastoji se od 5 tvrdnji (npr. „Mogu pouzdano procijeniti je li studentski zadatak ili esej djelomično ili u potpunosti generiran alatom UI-ja."). Podskala pokazuje izvrsnu unutarnju konzistenciju (McDonaldovo ω = ,88). Druga podskala nazvana je *Nesigurnost* u detekciji UI-generiranih učeničkih radova te sadrži 3 tvrdnje (npr. „Često sumnjam u svoju procjenu je li tekst vlastiti rad učenika ili prepisan"). Podskala pokazuje dobru unutarnju konzistenciju (McDonaldovo ω = ,80). Rezultati podskala formirani su kao aritmetičke sredine odgovora na pripadajućim tvrdnjama.

Skala *zabrinutosti zbog akademskoga utjecaja UI-ja* također je razvijena za potrebe istraživanja, s ciljem mjerenja zabrinutosti nastavnika u vezi s negativnim obrazovnim posljedicama učeničke uporabe alata UI-ja. Skala obuhvaća bojazan vezanu uz smanjeno

kritičko mišljenje, umanjenu kreativnost i potencijalno narušavanje obrazovnih ishoda. Tvrdnje se procjenjuju na Likertovoj skali od 5 stupnjeva (1 = „uopće se ne slažem" do 5 = „u potpunosti se slažem"). Iako je izvorno sadržavala 10 tvrdnji, nakon provedene EFA-e (metoda glavnih osi s oblimin rotacijom), u konačnoj verziji zadržane su samo 4. One su se snažno saturirale na jednom faktoru koji odražava akademsku zabrinutost zbog utjecaja UI-ja na učenje učenika (npr. „Zabrinut/a sam da učenici koriste UI kao zamjenu za vlastito razmišljanje."). Jednofaktorsko rješenje objasnilo je 59,8 % ukupne varijance. Preostale tvrdnje isključene su zbog slabih faktorskih opterećenja (< ,40; Field, 2018; Tabachnick i Fidell, 2013), visoke jedinstvenosti (> ,60) ili konceptualnoga odstupanja. Ova kraća, jednofaktorska skala pokazuje visoku unutarnju konzistenciju (McDonaldov ω = ,86) i jasnu interpretabilnost. Rezultat na skali formiran je kao aritmetička sredina odgovora na pripadajućim tvrdnjama.

Skala *percepcija nastavnika o (ne)primjerenoj uporabi UI-ja* razvijena je za potrebe istraživanja i uključivala je tvrdnje povezane s uporabom UI-ja u učenju, rješavanjem domaćih zadaća i polaganjem ispita. Sudionici su procjenjivali postotak učenika koji koriste UI alate u navedene svrhe. Skala je sadržavala 5 tvrdnji, a EFA je polučila dva faktora koji su objasnili 56,5 % ukupne varijance. Podskale su: *primjerena uporaba UI-ja* (dvije tvrdnje, npr. „Za pomoć u učenju, npr. pronalaženje informacija, objašnjavanje procesa"; McDonaldovo ω = ,70) i *neprimjerena uporaba UI-ja* (tri tvrdnje, npr. „Za pomoć u domaćoj zadaći na pretjeran ili neprimjeren način, tj. uporaba koja krši pravila koja je postavila institucija ili nastavnik"; McDonaldovo ω = ,74). Rezultati pdskala formirani su kao aritmetičke sredine odgovora na pripadajuće tvrdnje.

Jedno pitanje ispitivalo je jesu li nastavnici ikada *pouzdano detektirali neprimjerenu ili neovlaštenu uporabu UI*–ja svojih učenika i, ako jesu, koliko često. Sudionici su birali među sedam opcija: (a) Nisam, i ne znam kako to prepoznati; (b) Nisam, i vjerujem da to moji učenici ne rade; (c) Nisam, i vjerujem da se UI ne može koristiti u mojem predmetu; (d) Da, jednom; (e) Da, dva do četiri puta; (f) Da, pet do sedam puta; (g) Da, osam ili više puta. Za analizu su odgovori rekodirani u ordinalnu varijablu od 0 do 4, koja odražava rastuću učestalost detekcije. Odgovori (a) i (c) tretirani su kao nedostajući podatci jer ne odražavaju procjene stvarnoga ponašanja učenika.

Rezultati

Vrijednosti asimetrije (*skewness*) i spljoštenosti (*kurtosis*) ispitane su radi procjene prikladnosti parametrijskih analiza. Za većinu varijabli asimetričnost i spljoštenost bile su unutar prihvatljivih raspona (± 2; Field, 2018), što ukazuje na približno normalne distribucije. Iznimka je bila varijabla koja predstavlja broj korištenih UI alata, koja je pokazala značajnu pozitivnu asimetriju (2,01) i leptokurtiju (6,36), što upućuje na odstupanje od normalnosti. Iz toga razloga, varijabla nije uključena u daljnje parametrijske analize (Mdn = 1, IQR = 1,00). Deskriptivna statistika za sve analizirane konstrukte, grupirane prema obrazovnoj razini nastavnika, prikazana je u Tablici 1.

Tablica 1

Kako bismo odgovorili na prvi istraživački problem, proveli smo jednosmjernu analizu varijance (ANOVA). Budući da je pretpostavka homogenosti varijanci bila narušena, provedena je Welchova ANOVA, a potom i Games–Howellovi post hoc testovi, koji su primjereni za usporedbe između skupina s nejednakim varijancama i veličinama uzoraka (Field, 2018). Rezultati su prikazani u Tablici 2.

Tablica 2

ANOVA je pokazala statistički značajne razlike između obrazovnih razina za većinu varijabli, pri čemu značajne razlike nisu pronađene samo za varijable nesigurnost (dimenzija samoefikasnosti u detekciji uporabe UI-ja) i zabrinutost zbog uporabe UI-ja u obrazovanju. Post hoc usporedbe pokazale su da nastavnici u visokom obrazovanju iskazuju nižu razinu pouzdanja u detekciju uporabe UI-ja nego srednjoškolski nastavnici. Nastavnici u osnovnoj školi iskazuju niže percipirane postotke primjerenoga korištenja UI-ja u usporedbi i sa srednjoškolskim i s visokoškolskim nastavnicima. Za percipirane postotke neprimjerenoga korištenja UI-ja, srednjoškolski nastavnici iskazuju najviše vrijednosti, a nastavnici u osnovnoj školi najniže. Nastavnici u osnovnim školama također iskazuju nižu učestalost detekcije neprimjerene uporabe UI-ja u usporedbi i sa srednjoškolskim i s visokoškolskim nastavnicima. Prema Cohenovim (1988) smjernicama, većina veličina učinaka bila je u rasponu od malih do malih-srednjih, s najvećim učincima uočenim za percipirane postotke i primjerenoga i neprimjerenoga korištenja UI-ja.

Prije provođenja višestrukih hijerarhijskih regresijskih analiza, izračunati su Pearsonovi koeficijenti korelacije između varijabli u istraživanju (Tablica 3). Koeficijenti korelacije i faktori inflacije varijance (VIF; u rasponu od 1,07 do 1,40) nisu ukazivali na probleme s multikolinearnošću (Field, 2018).

Tablica 3

Tablica 4 prikazuje rezultate višestrukih hijerarhijskih regresijskih analiza. Analizirane su dvije zavisne varijable: percipirani postotak učenika koji neprimjereno koriste UI i učestalost detekcije neprimjerene uporabe UIj-a. U prvom koraku prediktori su bili obrazovna razina, broj prepoznatih alata UI-ja i opseg uporabe UI alata kojim se koriste nastavnici. U drugom koraku dodana su tri dodatna prediktora: samoefikasnost u detekciji uporabe UI-ja (dimenzija pouzdanje), samoefikasnost u detekciji uporabe UI-ja (dimenzija nesigurnost) i zabrinutost zbog uporabe UI-ja u obrazovanju.

Tablica 4

Za percipirani postotak učenika koji se neprimjereno koriste UI-jem prvom koraku objašnjeno je 1,7 % varijance ($p < ,05$), pri čemu se opseg uporabe alata UI-ja (kod nastavnika) pokazao kao jedini značajan prediktor. U drugom koraku objašnjena varijanca porasla je na 8,8 %, pri čemu su opseg uporabe alata UI-ja i zabrinutost zbog uporabe UI-ja u obrazovanju bili značajni pozitivni prediktori.

Za učestalost detekcije neprimjerene uporabe UI-ja, u prvom koraku objašnjeno je 7,4 % varijance ($p < ,001$), pri čemu su broj prepoznatih UI alata i opseg uporabe UI alata bili značajni prediktori. U drugom koraku objašnjena varijanca porasla je na 14,0 %, pri čemu su broj prepoznatih UI alata, opseg uporabe UI alata i samoefikasnost u detekciji uporabe UI-ja (skala pouzdanje) bili značajni prediktori.

Diskusija

Cilj ovoga istraživanja bio je ispitati nastavničku samoefikasnost u prepoznavanju učeničkih radova generiranih ili asistiranih umjetnom inteligencijom, njihovu zabrinutost zbog mogućih negativnih učinaka UI-ja na učenje, percepciju učestalosti učeničkih primjerenih i neprimjerenih korištenja UI-jem te realnu detekciju AI plagijata. Faktorska analiza pokazala je da se mjera samoefikasnosti u detekciji UI-generiranih učeničkih radova sastoji od dva faktora – pouzdanje i nesigurnost. To sugerira da se samopercepcije nastavnika ne nalaze na jednom kontinuumu koji se proteže od niske do visoke samoefikasnosti. Naprotiv, nastavnici istovremeno mogu doživljavati i pouzdanje i nesigurnost; primjerice, mogu se osjećati sigurnima kada ocjenjuju zadatke učenika koje dobro poznaju, dok i dalje sumnjaju u vlastite prosudbe u nejasnijim slučajevima. Ova dvoznačnost odražava složenu i kontekstualno uvjetovanu prirodu samoefikasnosti nastavnika u detekciji UI-generiranih radova.

Očekivali smo da će srednjoškolski i visokoškolski nastavnici imati višu samoefikasnost u detekciji neprimjerene učeničke uporabe UI-ja nego nastavnici u osnovnim školama. Međutim, rezultati su pokazali da su na faktoru pouzdanja srednjoškolski nastavnici postigli više rezultate od visokoškolskih. Ova se razlika može povezati s razlikama u vrstama zadataka na različitim obrazovnim razinama, kao i s razinom poznavanja učenika. Srednjoškolski nastavnici tipično zadaju više domaćih zadaća i kraćih provjera znanja. To im pruža više prilika za susret s UI-generiranim sadržajem i vježbanjem njegove identifikacije, čime jačaju svoje pouzdanje. Nasuprot tome, sveučilišni nastavnici često predaju većim skupinama studenata i osobno ih slabije poznaju, dok odsutnost razlike između osnovnoškolskih nastavnika i druge dvije skupine može biti posljedica prirode rada na osnovnoškolskoj razini, pri čemu su zadatci kraći, strukturiraniji i pod većim nadzorom, što potencijalnu uporabu UI-ja čini lakše uočljivom. U takvom kontekstu čak i ograničeno izlaganje UI-generiranim radovima može biti dovoljno da nastavnici razviju usporedivu razinu samoefikasnosti.

S druge strane, nisu utvrđene značajne razlike za drugi faktor samoefikasnosti u detekciji uporabe UI-ja – nesigurnost. S obzirom na kratko razdoblje od uvođenja široko dostupnih generativnih UI alata, moguće je da se nastavnici na svim obrazovnim razinama oslanjaju na slične izvore informacija poput - medijskih izvještaja, profesionalnih mreža i neformalnih razgovora s kolegama - a ne na sustavnu ili specificiranu obuku. To zajedničko izlaganje može pridonijeti relativno ujednačenom osjećaju nesigurnosti. To je u skladu s nalazima Miljković Krečar i sur. (2024), koje nalaze da sveučilišni nastavnici iskazuju tek umjerenu sigurnost te ograničenu točnost (53 %) u detekciji UI-generiranih radova.

Također, kao što se i očekivalo, za akademsku zabrinutost zbog UI-ja nisu utvrđene značajne razlike među nastavnicima, što sugerira da je zabrinutost zbog potencijalno negativnoga utjecaja UI-ja na učenje učenika slična na osnovnoškolskoj, srednjoškolskoj i visokoškolskoj razini. Aritmetičke sredine blizu vrijednosti 4 (na skali od 5 stupnjeva) upućuju na općenito visoku razinu zabrinutosti. Takvi rezultati odražavaju široko rasprostranjenu nesigurnost u vezi s time kako će UI utjecati na kritičko mišljenje, učenička kreativnost i samostalnost u rješavanju problema , što su već istaknula i prethodna istraživanja (npr. Mai i sur., 2024; Miah i sur., 2024).

Nadalje, za primjerenu uporabu UI-ja u početku smo pretpostavili da neće postojati značajne razlike među obrazovnim razinama, temeljem nedostatka prethodnih empirijskih dokaza koji bi ukazivali na takve varijacije. No naši rezultati pokazuju da nastavnici u osnovnim školama iskazuju najniže stope, s malom do srednjom veličinom učinka. Razlog može biti taj što mlađi učenici obično imaju manje zadataka primjerenih za rad pomoću UI-ja te slabije razvijene digitalne vještine, što doprinosi percepciji da je (primjerena) uporaba UI-ja među njima rijeđa. Također je moguće – iako spekulativnije – da ulogu ima i veći roditeljski nadzor te strože pridržavanje učioničkih pravila kod mlađih učenika. Istodobno, ne treba isključiti ni mogućnost da neki roditelji sami pomažu djeci koristeći alate UI-ja za pronalazak odgovora ili objašnjenja.

Za neprimjerenu uporabu UI-ja, iako je nismo postulirali, rezultati pokazuju razlike (ponovno uz malu do srednju veličinu učinka). Srednjoškolski nastavnici percipiraju najviše stope neprimjerene uporabe UI-ja među učenicima. Jedno moguće objašnjenje jest da srednjoškolci imaju dovoljno razvijene digitalne vještine i pristup alatima UI-ja da ih koriste tako da krše institucionalna ili nastavnikova pravila, pri čemu još uvijek nemaju razvijene standarde akademskoga integriteta koji se očekuju na visokoškolskoj razini. Akademski zahtjevi u srednjim školama – poput čestih domaćih zadaća, pisanih zadataka i testova – također mogu stvarati više prilika ili poticaja za neprimjerenu uporabu UI-ja u usporedbi s osnovnom školom.

Kada se razmotre apsolutne vrijednosti aritmetičkih sredina, rezultati sugeriraju da nastavnici na svim obrazovnim razinama percipiraju primjerenu uporabu UI-ja češćom od neprimjerene, iako su procijenjene srednje vrijednosti i za neprimjerenu uporabu značajne. Međutim, istraživanja koja se oslanjaju na studentske samoprocjene (Digital Education Council, 2024; Miljković Krečar i sur., 2024; von Garrel i Mayer, 2023) ukazuju na još više stope uporabe UI-ja. Stoga je moguće pretpostaviti da nastavnici podcjenjuju i primjerenu i neprimjerenu uporabu UI-ja među učenicima / studentima.

Što se tiče drugoga istraživačkog problema, za obje varijable – percipirani postotak učenika koji se neprimjereno koriste UI-jem i učestalost detekcije neprimjerene uporabe UI-ja – rezultati su bili u skladu s hipotezom: u oba koraka hijerarhijskih regresijskih analiza objašnjen je značajan udio varijance, pri čemu je drugi korak objasnio dodatnu varijancu povrh prvoga. Za prvu zavisnu varijablu značajni pozitivni prediktori bili su opseg uporabe UI alata kojima se koriste nastavnici te zabrinutost. To sugerira da

stavovni i afektivni čimbenici mogu oblikovati takve percepcije, a nastavnici koji su zabrinutiji zbog potencijalnih negativnih posljedica UI-ja na učenje vjerojatnije će interpretirati ponašanje učenika kao neprimjereno ili pretpostaviti da se ono često pojavljuje.

Nasuprot tome, za učestalost detekcije neprimjerene uporabe UI-ja značajni prediktori bili su obrazovna razina, broj prepoznatih UI alata, opseg uporabe UI alata i samoefikasnost u detekciji uporabe UI-ja (dimenzija pouzdanje), što ukazuje da su nastavnici koji vjeruju da mogu učinkovito prepoznati radove generirane ili asistirane pomoću UI-ja također skloniji izvještavanju da su ih doista detektirali. Uloga obrazovne razine, prepoznatih alata i opsega uporabe UI-ja sugerira da upoznatost i iskustvo s UI-jem – i osobno i u nastavnom kontekstu – mogu nastavnicima pružiti više prilika ili signala za detekciju. Primjerice, nastavnici na višim obrazovnim razinama mogu biti izloženi složenijim studentskim zadatcima u kojima je asistirana uporaba UI-ja izvedivija, a oni koji sami poznaju i koriste više alata UI-ja mogu lakše prepoznati njihovu uporabu u studentskim radovima. No treba napomenuti da su prediktori korišteni u ovom istraživanju objasnili samo skroman udio varijance zavisnih varijabli (8 % i 12,7 %), što upućuje na to da bi buduća istraživanja trebala razmotriti dodatne potencijalne prediktore.

Ovo istraživanje usmjereno je na novu i aktualnu temu – percepcije nastavnika o uporabi UI-ja u obrazovanju – obuhvaćajući i primjerene i neprimjerene oblike uporabe, kao i samoefikasnost i zabrinutost nastavnika, u istom okviru. U istraživanje su bili uključeni nastavnici osnovnih, srednjih i visokoškolskih ustanova, što je omogućilo usporedbe između obrazovnih razina koje su rijetko bile ispitivane u ranijim radovima. Ipak, istraživanje ima nekoliko ograničenja. Uzorak nije reprezentativan. Moguće je, primjerice, da su u istraživanju u većoj mjeri sudjelovali nastavnici s izraženijom zabrinutošću zbog uporabe UI-ja. Broj nastavnika osnovnih škola u uzorku bio je manji (n = 70) nego u drugim skupinama. Također je moguće da nastavnici podcjenjuju učestalost studentske uporabe UI-ja, stoga bi buduća istraživanja u Hrvatskoj trebala ispitati i samoprocjene učenika o ovoj temi. Nadalje, nismo uzeli u obzir vrstu srednje škole ili područje studija, iako ti čimbenici također mogu utjecati na rezultate. Primjerice, von Garrel i Mayer (2023) izvijestili su o najvišim stopama uporabe UI-ja među studentima inženjerstva, matematike i prirodnih znanosti, dok su Miljković Krečar i sur. (2024) utvrdili da studenti prirodnih i tehničkih područja češće koriste UI nego studenti društvenih znanosti.

Zaključak

U ovome istraživanju doneseni su novi uvidi u percepcije, samoefikasnost, zabrinutost i iskustva nastavnika u vezi s učeničkom uporabom UI alata na osnovnoškolskoj, srednjoškolskoj i visokoškolskoj razini. Jedan od uvida odnosi se na dualnost mjere samoefikasnosti, što odražava složenu i kontekstualno uvjetovanu prirodu samoprocjena nastavnika o vlastitoj sposobnosti detekcije UI-generiranih radova. Iako je nesigurnost

u detekciji bila slična na svim obrazovnim razinama, srednjoškolski nastavnici izvijestili su o najvišim razinama pouzdanja u sposobnosti takve detekcije.

Pored toga, iako su zabrinutosti o potencijalno negativnom utjecaju UI-ja na učenje bile ujednačeno visoke na svim obrazovnim razinama, uočene su razlike u percepcijama primjerene i neprimjerene uporabe UI-ja. Nastavnici u osnovnoj školi izvještavali su o najnižim stopama primjerene uporabe, dok su srednjoškolski nastavnici percipirali najviše stope neprimjerene uporabe.

Hijerarhijske regresijske analize dodatno su razjasnile te razlike. Za percipirani postotak učenika koji neprimjereno koriste UI, značajni prediktori bili su opseg nastavničke uporabe UI alata i zabrinutost zbog UI-ja. Zabrinutost se pokazala kao pozitivan prediktor, što upućuje na to da nastavnici s višom razinom zabrinutosti percipiraju više stope neprimjerene učeničke uporabe UI-ja. Za učestalost detekcije neprimjerene uporabe UI-ja, značajni prediktori bili su obrazovna razina, prepoznati UI alati, opseg uporabe UI alata i samoefikasnost u detekciji uporabe UI-ja (faktor pouzdanje).

Naši nalazi, poput umjerene samoefikasnosti nastavnika u detekciji učeničke uporabe UI-ja, ukazuju na potrebu za jasnijim smjernicama i boljom obukom nastavnoga osoblja u ovom području. Također, s obzirom na zabrinutosti nastavnika i percipiranu učestalost učeničke uporabe UI-ja, sve je izraženija potreba za promišljanjem strategija vrednovanja i oblikovanjem zadataka otpornijih na UI, primjerice s većim naglaskom na usmene ispite i demonstracije konceptualnoga razumijevanja. U skladu s tim, Lye i Lim (2024) napominju da se integracija UI-ja u obrazovanje treba razmatrati u pristupu *Fit-for-Purpose Assessment: The Against, Avoid and Adopt Principle*. Ovaj princip potiče nastavnike da jasno definiraju koje su uporabe UI-ja zabranjene, da redizajniraju ili izbjegavaju zadatke koji se lako mogu kompromitirati pomoću UI-ja, te da promišljeno usvoje UI ondje gdje on može unaprijediti učenje, pri čemu se o tim očekivanjima treba eksplicitno komunicirati sa studentima.

Napomena

Tijekom istraživanja i pisanja ovoga rada korišteni su alati umjetne inteligencije. Konkretno, ChatGPT (OpenAI, 2025) korišten je za pomoć u kreiranju tvrdnji upitnika, prevođenju, jezičnom usavršavanju i sažimanju ideja, dok je Perplexity AI (2025) korišten kao dopunski alat za prikupljanje i istraživanje relevantne literature. Svaki navod generiran pomoću ovih alata autor je detaljno i kritički pregledao i uredio, kako bi se osigurala akademska čestitost.