An Empirical Study Into Higher Education Students' Critical Thinking Towards AI-generated Output

Master Thesis

Burgemeester Oudlaan 50, 3062PA Rotterdam, Nederland

July 6th, 2025

Luuk Oskam [589603]

Department of Social and Behavioural Sciences | Erasmus University Rotterdam

MSc Educational Sciences

Dr. Kim van Broekhoven [vanbroekhoven@essb.eur.nl]

Prof. Dr. Sabine Severiens [sabine.severiens@eur.nl]

Word count: 7766

Samenvatting

Eerder onderzoek toont een afname in het kritisch denkvermogen van nieuwe studenten in het hoger onderwijs. Daarnaast hebben empirische studies negatieve gevolgen gerapporteerd van kunstmatige intelligentie (AI) voor de cognitie van studenten. Vergelijking van onderzoeksresultaten over de relaties tussen kritisch denken, AI-gebruik en de intelligentiedomeinen ‘fluïde en gekristalliseerde intelligence’ onthult tegenstrijdige bevindingen. Deze bevindingen vergen nader onderzoek naar hun samenhang. Deze studie beoogt de associatie te onderzoeken tussen de intelligentiedomeinen (fluïde en gekristalliseerd) van Duitse studenten uit het hoger onderwijs (n = 353) en hun vermogen om AI-gegenereerde output voor hun studieproces kritisch te beoordelen op nauwkeurigheid, betrouwbaarheid en toepasbaarheid (genaamd AI-CT). De verzamelde data maken deel uit van de grootschalige online enquête genaamd ‘Fachkraft 2030’, waarin de economische, onderwijs- en algemene leefomstandigheden worden bevraagd van studenten en recent afgestudeerden die in Duitsland wonen. Daarnaast worden participanten in de enquête getoetst op hun cognitieve vaardigheden. Hiërarchische *weighted least squares* regressieanalyses zijn uitgevoerd om zowel fluïde als gekristalliseerde intelligente te onderzoeken in verband met AI-CT. Invloedrijke factoren op AI-gebruik – leeftijd, hoogste academische graad, beoogd studeerdoeleinde van GenAI LLMs, en de *Big Five* persoonlijkheidskenmerken – zijn als controle variabelen meegenomen. Na controle op deze factoren bleef gekristalliseerde intelligentie significant positief geassocieerd met AI-CT (B = .21; p < .001), terwijl fluïde intelligentie niet meer significant geassocieerd was met AI-CT (B = .09; p = .069). De onderzoeksbevindingen dragen bij aan het algemeen begrip over kritisch denkvermogen van studenten en brengen implicaties voor de onderwijspraktijk teweeg.

Trefwoorden: Generatieve Kunstmatige Intelligentie | Fluïde Intelligentie | Gekristalliseerde Intelligentie | Kritisch Denkvermogen | Hoger Onderwijs | Studenten.

Abstract

Previous research has revealed a decrease in newly enrolled higher education student's critical thinking ability. In addition, studies have reported on the negative implications of artificial intelligence (AI) for students' cognitive abilities. When comparing studies that have investigated the relationships between critical thinking, AI-usage and the intelligence domains 'fluid and crystallised intelligence', contradicting findings arise. These findings provoke further examination of their interplay. This study aims to bridge this gap by investigating the relationship between German higher education students' (n = 353) intelligence domains (fluid and crystallised) and their ability to critically assess AI-generated output for their studies on its accuracy, reliability, and applicability (called AI-CT). The collected data is part of a large-scale online survey called 'Fachkraft 2030', in which the economic, educational, and general living situations of students and young graduates residing in Germany are asked and additional cognitive tests are conducted. Hierarchical weighted least squares regression analyses have been conducted to investigate both fluid and crystallised intelligence in relation to AI-CT. Influential factors on AI-usage – age, highest academic accomplishment, intended study purpose of GenAI LLMs, and the Big Five personality traits – have been included as control variables. The results showed that after controlling for these factors, crystallised intelligence remained significantly positively associated with AI-CT (B = .21; p < .001), whereas fluid intelligence was no longer significantly positively associated with AI-CT (B = .09; p = .069). The study findings contribute to the understanding of critical thinking among students and bring about implications for educational practices.

Keywords: Generative Artificial Intelligence | Fluid Intelligence | Crystallised Intelligence | Critical Thinking | Higher Education | Students.

Introduction

The latter decades of the 20th century and first decades of the 21st century have introduced a variety of novel and perpetually developing technological advancements; from the introduction of the internet, which in turn has led to the public accessibility of artificial intelligence (AI) (Binkley et al., 2012). Although the integrity of information on the internet has always been brought into question (United Nations, 2023), generative AI large language models (GenAI LLMs) have further elevated the risk of encountering falsified information (Verma, 2023). GenAI LLMs are designed to supply AI-generated content – based on complex algorithms and prior human-constructed content (World Economic Forum, 2025) – to specific user-prompts in a human-like manner (Databricks, 2024).

These technological developments were cause for a paradigm shift in which skillset students must be taught in order to become a functioning member of society. Although development of all 21st Century Skills – collaborative, sociocultural, digital literacy, creative, communicative, and critical thinking skills (Binkley et al., 2012; Stichting Leerplan Ontwikkeling, 2023) – is considered essential, a focus on critical thinking (CT) seems to be most crucial for thriving in the 'digital information age' (Markauskaite et al., 2022; Tewari, 2022). This is because in a world with AI, learners need 'individual resourcefulness': self-regulated learning, creative problem-solving, and making sense of AI-generated output (Markauskaite et al., 2022). Dwyer et al. (2014) – even before AI had developed into the cutting-edge technology it is today – underscored the importance of CT in regard to people's ability and willingness to engage with ever-evolving information in a critical manner.

Among higher education (HE) students a trend has been reported; there is a decrease in newly enrolled HE student's CT-abilities (Westrick et al., 2024). HE faculty members emphasize that this growing deficit – along with lower reading and writing skills – brings about a steady decline in these students' academic readiness. Although the COVID-19 pandemic is rightfully found culpable for this decrease, it is important to note that students' usage of AI for their studies saw a significant increase during this period (Pantelimon et al., 2023). Various 'strength-weaknesses-opportunities-threats' (SWOT) analyses have underscored AI's negative influences on students' academic attitudes and abilities (Farrokhnia et al., 2024; Humble & Mozelius, 2022; Krstiç et al., 2022). For instance, Krstić et al. (2022) found that AI-usage reduced the cognitive barrier essential for deep cognitive engagement. In other words, it heavily fosters dependency on its generative guidance, impeding the development of students' higher-order thinking skills. Among which, critical thinking.

Theoretical Framework

Critical Thinking Framework

The 21st Century Skill 'critical thinking' (CT) is defined by Rivas et al. (2022, p.2-3) as "a knowledge-seeking process via reasoning skills to solve problems and make decisions". Although there are many varying definitions of CT, a consensus in its research field is that it involves the 'self'. Paul and Elder (2019) argue the core definition to be self-directed, self-disciplined, self-monitored, and self-corrective thinking. Research has revealed that people who show a more developed CT-ability reap many significant benefits, such as academic achievements, an enhanced self-regulation in their learning process, and more effectiveness in achieving their goals (Bagheri & Ghanizadeh, 2016; Ghanizadeh, 2017; Rivas et al., 2022).

Several studies have defined two distinct components of CT which are both essential for acquiring and developing this skill: the cognitive component – CT-skills – and the non-cognitive component – CT-disposition (Bagheri & Ghanizadeh, 2016; Nieto & Saiz, 2011; Rivas et al., 2022). CT-skills refer to knowledge acquisition mechanisms. Such mechanisms involve knowledge perception, learning, and memory processes. CT-disposition refers to individual propensity (i.e., willingness and motivation) to apply critical thinking (Rivas et al., 2022). In other words, CT-skills concern capability – whether a person *can* think critically – whereas CT-disposition concerns attitudes – whether a person *will choose* to think critically.

Intelligence Domains Framework

The widely cited Cattell-Horn-Caroll (CHC) theory (Horn & Cattell, 1966) proposes a cognitive framework in which intelligence is divided into two fundamental cognitive dimensions: fluid intelligence and crystallised intelligence. Fluid intelligence concerns people's cognitive ability to adapt to novel information and to solve unfamiliar problems. This dimension thus regards problem-solving, abstract thinking and knowledge adaptation. Crystallised intelligence is comprised of people's prior knowledge, which improves and builds up through experience and education. This dimension pertains to verbal general knowledge and domain-specific expertise. Whereas fluid intelligence provides someone with an adaptive cognitive skillset, their crystallised intelligence serves as the knowledge base to consult during these cognitive processes. Thus, both dimensions complement one another. There is various research which indicates a relationship between these two fundamental cognitive intelligence domains and CT, described below.

Critical Thinking, Fluid Intelligence, and Crystallised Intelligence

In their study into HE students' cognition, Liu et al. (2024) found that fluid intelligence and CT are grounded in significantly similar cognitive mechanisms (executive functions).

While their study showed that CT and fluid intelligence are significantly differing constructs, it was found that both CT and fluid intelligence shared the same cognitive demands for ‘updating’ and ‘inhibition’. Updating is seen as the brains ability to actively monitor and adjust information in the working memory. Inhibition refers to the cognitive system that suppresses people’s biases and tendencies to jump to conclusions, allowing them to objectively evaluate things. The study findings showed that higher fluid intelligence is a significant predictor of a higher CT-capability, likely due to the shared foundational executive functions. Liu et al.’s (2024) study does not, however, take into account the influence of AI. Further examination of the relationships between fluid intelligence, CT, and AI is required.

Crystallised intelligence supplies people with the knowledge necessary to make informed problem-solving decisions (Horn & Cattell, 1966). Stanovich and Stanovich (2010) argue that without this prior knowledge, people lack the necessary foundation to critically evaluate things. Heijltjes et al.’s (2014) empirical study findings support this claim. They investigated the efficacy of differing instructional conditions for enhancing students’ CT-abilities. They found that explicit instruction on CT combined with practice of CT yielded the most effective results. More notably, they found that students in this condition – who expanded their foundational knowledge (crystallised intelligence) – achieved better CT(-skills) than their peers in other instructional conditions, while simultaneously showing equivalent efforts to apply critical thinking (CT-disposition). Heijltjes et al.’s (2014) study did not, however, take into account the influence of AI. Further examination of the relationships between crystallised intelligence, CT, and AI is required.

The Intelligence Domains, Critical Thinking, and Artificial Intelligence

While prior research has examined the relationship between the intelligence domains (fluid and crystallised) and CT, the role of AI in their relationship remains underinvestigated. Therefore, it is scientifically relevant to not only examine the association between the intelligence domains and CT, but to factor AI into this relation as well. Below, the findings of several empirical studies into the association of AI with both the intelligence domains and CT have been summarised.

Artificial Intelligence, Fluid Intelligence, and Crystallised Intelligence

Vázquez-Cano et al. (2021) investigated the didactic functionality of GenAI LLMs on HE students’ learning outcomes. The focus was on learning activities for acquiring Spanish language punctuation skills. The study results showed that utilising GenAI LLMs, as opposed to traditional learning methods, significantly boosted HE students’ information acquisition. Vázquez-Cano et al.'s (2021) findings introduce interesting insights when aligned with the intelligence domains, as GenAI LLMs were found to facilitate adaptive cognitive processes (fluid intelligence) by enabling HE students to practice problem-solving and applying the study material in various interactive contexts. Additionally, Sheth et al. (2019) emphasised how GenAI LLMs are capable of collecting and providing information in a rich, contextualised and personalised manner. HE students can now, in essence, actively converse with their study material. GenAI LLMs are thus found to significantly further leverage people's foundational knowledge (crystallised intelligence).

Artificial Intelligence and Critical Thinking

A recent study by Lee et al. (2025) investigated the influence of AI-usage on knowledge workers' cognitive abilities. They found that the usage of GenAI for one's work significantly lessened the cognitive effort they employed and, moreover, changed how users utilised their CT-skills. Participants who implemented GenAI spent more time (1) verifying information rather than gathering it, (2) integrating AI-generated output in their work rather than employing their own problem-solving skills, and (3) supervising tasks rather than executing tasks (Lee et al., 2025). The study findings suggest that AI-usage is significantly negatively associated with people's cognitive abilities. Additionally, this negative relationship between GenAI LLMs and CT has been found among HE students as well, because over-reliance on GenAI LLMs diminished their cognitive abilities (Zhai et al., 2024).

Contradiction and Cause for Investigation

When combining the research findings on the relationships between AI and the intelligence domains and between AI and CT, a notable contradiction about their interplay emerges. On the one hand, empirical study findings into the relationship between AI and the intelligence domains have shown a positive association between AI-usage (for one's studies) and both fluid and crystallised intelligence (Sheth et al., 2019; Vázquez-Cano et al., 2021). In turn, this should be beneficial for HE students' CT-ability as other studies have found positive associations between the intelligence domains and CT (Heijltjes et al., 2014; Liu et al., 2024). On the other hand, however, empirical studies into the relationship between AI and CT have found that AI-usage significantly diminishes HE students' CT-ability (Krstić et al., 2022; Lee et al., 2025; Zhai et al., 2024). Thus, when grouping these three concepts (AI, the intelligence domains, and CT), a complex and underinvestigated empirical area arises. This empirical gap is addressed in the current study.

Influential Factors on AI-usage

In light of the influence of AI, it is important to control for factors that mediate HE students' utilisation of GenAI LLMs (age, highest academic accomplishment, intended study purpose of GenAI LLMs, and the Big Five personality traits). First, research has revealed that younger people are more dependent on AI, are more susceptible to AI-utilisation, and hold less critical attitudes towards AI (Neaves & Nagar, 2025; Gerlich, 2025; Persson et al., 2021; Zhang & Dafoe, 2019). Therefore, I will control for age. Additionally, Klarin et al. (2024) have found that a higher educational level is indicative of higher propensity to use AI for studying. Other studies have also found that educational level is positively associated with AI-familiarity (Hamad et al., 2024; Persson et al., 2021). Therefore, I will control for highest academic accomplishments. Furthermore, Klarin et al. (2024) also investigated students' intended study purposes of GenAI LLMs. They found that students who use GenAI LLMs for task-initiation and task-structuring thrive significantly better in the long run than those who use it for task-completion. Therefore, I will control for intended study purpose of GenAI LLMs. Lastly, Kaya et al. (2024) investigated people's attitudes towards AI-usage, based on each of their 'Big Five' personality traits: (1) openness to experience (i.e., open-mindedness and preference for novelty and variety), (2) conscientiousness (i.e., self-discipline and thoughtfulness), (3) extraversion (i.e., assertiveness and sociability), (4) agreeableness (i.e., tendency towards cooperation), and (5) neuroticism (i.e., tendency towards negative emotions) (McCrae & Costa, 2008). Kaya et al.'s (2024) study results suggest that people with higher openness to experience and extraversion showed more positive attitudes towards AI-usage; people with higher conscientiousness were more aware of negative aspects of AI-usage; and people with higher agreeableness and neuroticism were found to be more tolerant towards these negative aspects of AI-usage and therefore more likely to use AI. Moreover, study by Nieto and Saiz (2011) has revealed that people's personality traits are also significantly correlated to their CT-disposition; with conscientiousness, openness to experience, and neuroticism being the most significant predictors of CT. Only neuroticism was found to be negatively correlated with CT, whereas the other personality traits were positively associated. Therefore, I will control for the Big Five personality traits.

Current Study

The empirical insights on GenAI LLMs in relation to the intelligence domains and critical thinking provoke further examination of their interplay. This current study aims to bridge this empirical gap by investigating each of the intelligence domains (fluid and crystallised intelligence) in relation to HE students' ability to critically assess AI-generated output for their studies on its accuracy, reliability, and applicability (from now on 'AI-CT'). Both fluid and crystallised intelligence are thus investigated in relation to AI-CT. For this, two research questions and subsequent hypotheses have been formulated (see Figure 1):

RQ1: *To what extent is fluid intelligence associated with AI-CT?*

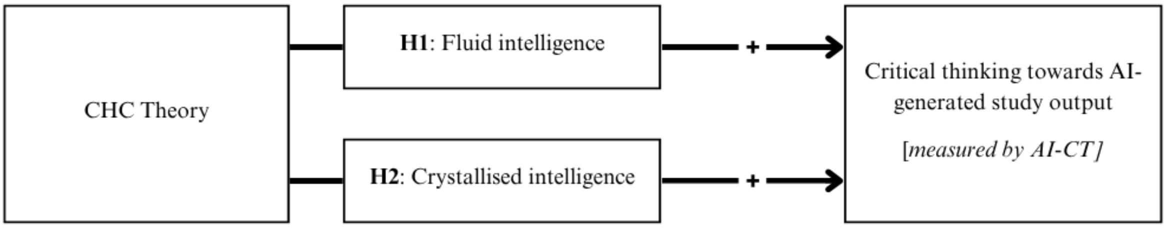
H1: *There is a positive association between fluid intelligence and AI-CT.*

RQ2: *To what extent is crystallised intelligence associated with AI-CT?*

H2: *There is a positive association between crystallised intelligence and AI-CT.*

Figure 1

Conceptual Model of Research Hypotheses



Method

Design

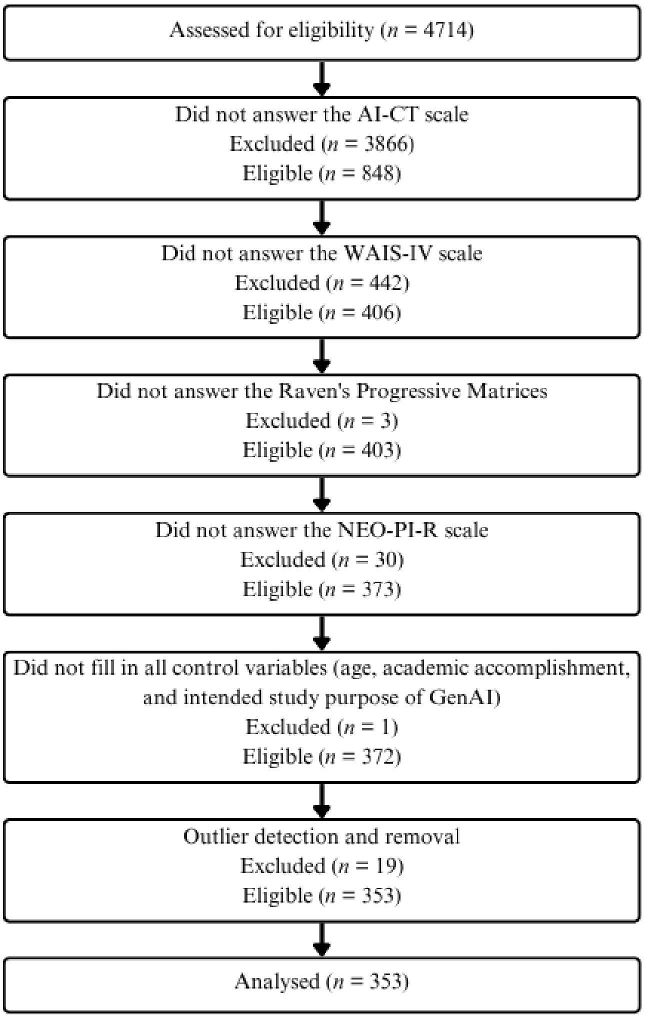
The current study employs a cross-sectional design with quantitative measures. The collected data is part of a large-scale online survey called ‘Fachkraft 2030’. It is constructed and distributed by the Constata research institute and the department of labour economics at Maastricht University (Constata, 2023; Maastricht University, 2023). In the survey, the economic, educational, and general living situations of students and young graduates residing in Germany are asked additional cognitive tests are conducted. The data-collection process for Fachkraft 2030 has been approved by the data privacy officer of the Constata research institute and all German and European legal requirements for this data-collection have been met (Constata, 2023; Maastricht University, 2023).

Participants

The original sample included 4714 participants. The participants who did not answer the scales relevant to the current study and also did not fill in the information for the control variables were excluded in the study sample (see Figure 2 for the exclusion process). Due to participants answering the AI-CT survey section at their own volition, only 848 answered these AI-CT-items. Only 373 of these participants answered all of the survey sections that measured their fluid intelligence, crystallised intelligence and Big Five personality traits. Furthermore, one of these remaining participants had missing values on a control variable used in this study (highest academic accomplishment). Lastly, after outlier detection and removal, another 19 participants were removed from the sample. This leaves 353 participants in the sample of this current study.

Figure 2

Visual Representation of the Participant Exclusion Process



The socio-demographic characteristics of this sample are as follows: the average age of the sample was 25 years old (M = 25.01; SD = 3.59), ranging from 18 to 34; the male-female gender-ratio was approximately evenly distributed (male = 54.7%; female = 43.6%; neither = 1.7%); and most of the participants had no migration background (childhood in Germany = 56.1%; childhood not in Germany = 7.6%; missing = 36.3%; see Table 1).

Table 1

Socio-demographic Characteristics of Participants

Characteristic	Sample					
	N	*M*	%	*SD*	*Min.*	*Max.*
Gender, %					1	3
Male	193	-	54.7%	-		
Female	154	-	43.6%	-		
Other	6	-	1.7%	-		
Age (in years), *M* (*SD*)	353	25.01	-	3.59	18	34
Migration background, %					0	1
Childhood in Germany	198	-	56.1%	-		
Childhood not in Germany	27	-	7.6%	-		
Missing value	128	-	36.3%	-		

Note. $n = 353$. *M* = mean (for numeric variables) or % = percentage (for categorical variables); *SD* = Standard Deviation; *Min.* = Minimum; *Max.* = Maximum. 'Fachkraft 2030' shows no systematic deviations from German student population (Seegers et al., 2019).

Procedure

Fachkraft 2030 was launched through an online German job platform, on which the participants have been recruited (JobValley, 2023). Participation in the study was voluntary. Data-collection was thus done through voluntary sampling. Upon participation, people could win vouchers with a monetary total of €1950,- divided into one €500,- voucher and 29 €50,- vouchers. The survey takes approximately 30 minutes to complete (Seegers et al., 2019). Participants have been able to biannually (in March and September) sign up for the survey since 2012. Fachkraft 2030 is not part of a longitudinal study design, because participants are recruited anew for each survey. For this current study, data has been utilised from the survey taken in September of 2024.

Variables and Instruments

Independent Variables

Fluid intelligence is measured with the Raven's Progressive Matrices (PRM, Burke, 1985). These are a set of non-verbal items used to measure people's abstract reasoning and problem-solving ability. Test-takers are presented with a three-by-three grid of spaces with shapes that show logical progression (e.g., rotating lines, changing colours) and one empty space. Their task is to choose the correct shape from a variety of options that fits the pattern (see Appendix I for an example). The PRM is a validated and reliable (α = .95, see Burke, 1985) instrument. The current study included 27 PRM-items, of which each participant was randomly presented with 6. The total number of correct responses forms the fluid intelligence score.

Crystallised intelligence is measured with the Wechsler's Adult Intelligence Scale (WAIS-IV, Seegers et al., 2019). This test measures participants' cognitive ability through a series of verbal questions. The WAIS-IV is a validated and reliable scale (α = .91, see Abdelhamid et al., 2019). An example of a WAIS-IV-item is: 'Which number fits spot X in the following sequence: 1, 2, 3, 5, X, 13' (see Appendix II for all 40 included items). Each participant has been randomly presented 12 out of the 40 WAIS-IV-items. The total number of correct responses forms the crystallised intelligence score.

Dependent Variable

The dependent variable is the degree in which AI-generated output for one's studies is evaluated on its accuracy, reliability, and applicability (AI-CT). This is measured with the newly-constructed 'AI-CT scale', derived from Lee & Park's (2024) 'ChatGPT literacy critical thinking subscale' (see Appendix IV for a comparison). Participants could agree with 6 given statements through a 5-point Likert-scale ranging from 'strongly disagree' (score = 1) to 'strongly agree' (score = 5). The sum score of all answers forms the AI-CT score.

The 'AI-CT-items' could be answered for GenAI LLMs in general and for specific GenAI LLMs where people could fill in the name of the AI-tool they used. This study, however, does not differentiate between specific GenAI LLMs. Therefore, the specific GenAI LLM item-variants have been computed into the general item-variants. An example of an AI-CT-item is: '*I check the accuracy of information provided by GenAI LLMs*'. After factor analysis, two reverse-coded items that loaded on a different factor were removed: '*using LLM reduces my motivation to put text into my own words*' and '*I assume the text that LLM generates is accurate*' (see Appendix IV for the remaining items). This resulted in a Cronbach's alpha of .85, which is similar to the Cronbach's alpha of the original subscale (α = .86) by Lee and Park (2024) (for the SPSS output tables, see Appendix V).

Control Variables

The following control variables are included in this study: age, highest academic accomplishment, the intended study purpose of GenAI LLM, and the Big Five personality traits (see Appendix VII for the variable descriptives). The control variable 'intended study purpose of GenAI LLMs' is comprised of multiple dichotomous variables, each representing a study purpose for which participants could answer whether they did or did not use GenAI LLMs. Although it is customary for nominal variables that one dichotomous variable – a dummy variable – is used as the reference group, this is not possible here. The reason for this is that participants could choose multiple purposes, which negates the possibility of appointing a reference group. All the study purposes (of GenAI LLMs) are thus included as standalone dichotomous control variables. The nominal control variable 'highest academic accomplishment' originally consisted of thee more categories (diplom, magister, and PhD) which each included too few participants and were therefore recoded into 'other degree'. Additionally, for the same issue, the category 'master with state exam' is recoded into 'master'. The category 'no degree' will be included as the reference group. The control variables 'Big Five personality traits' were measured with the NEO-PI-R scale (Costa & McCrae, 2008). For each personality trait (openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism), ten statements are provided with which participants could agree on a 5-point Likert-scale ranging from 'very inaccurate' (score = 1) to 'very accurate' (score = 5). The sum score of the answers forms the score for each personality trait. An example of a NEO-PI-R-item for Extraversion is: 'I am the life of the party' (see Appendix III for all 50 included NEO-PI-R-items). In this current study, the scale has a Cronbach's alpha of .87 (see Appendix III for the output table).

Analyses

SPSS (version 28) will be used in this study. During the preparation of the dataset, a confirmatory factor analysis and a reliability (Cronbach's alpha) analysis are conducted on the AI-CT scale to measure its internal consistency. Principal axis factoring (promax rotation) is used, because this method accounts for interrelated constructs that are associated with psychological scales. After this preparation, the data is examined for outliers by computing Cook's distances – which measure the influence a datapoint has on the skewness of the overall data. Any datapoints with a Cook's distance value greater than 1 ($D\,i > 1$) are labelled as outliers. Additionally, there is indication of outliers when approximately 5% of the dataset consists of standardised residual values greater than 1.96 ($Z < -1.96$; $Z > 1.96$). These datapoints are investigated to find plausible cause for the influence they have and are either adjusted in case of input-errors or otherwise removed from the sample. In the aforementioned sample exclusion process, listwise deletion of missing values is used.

Once the data is ready to be analysed, hierarchical regressions are conducted to investigate each hypothesis, taking into account the control variables. For this type of statistical analysis, a couple of assumptions have to be met. The first is the assumption of normally distributed residuals. This is checked by computing a histogram and the P-P normality plot for the standardised residuals. There is also the assumption of homoscedasticity, meaning the presence of homogeneity of error-variance across the data. This is checked by computing a scatterplot of the standardised residuals on the y-axis and the standardised predicted values on the x-axis and then checking for equal variance across the y-axis. Furthermore, there needs to be linearity between the independent and the dependent variables. To test for this assumption, the aforementioned scatterplot (including standardised residuals and predicted values) can be checked for linear distribution. Lastly, there must be an absence of multicollinearity ($r > .80$) between all the independent and control variables. Too strong of a correlation would indicate that both variables measure the same thing, thus negating any individual effects they might have. Additionally, Tolerance and VIF-values are examined to check for the assumption of no multicollinearity. The assumption is met when the Tolerance values are greater than .10 and the VIF-values are smaller than 10. Lastly, to draw conclusions regarding the hypotheses, statistical findings are significant at a p-value smaller than 5% ($p < 0.05$) and standardised regression coefficients (*Beta* values) are reported to indicate the direction of the relationships between variables. Positive values correspond with a positive direction, whereas negative values correspond with a negative direction. Standardised regression coefficients also serve to measure the effect sizes of the associations (Nieminen, 2022), with values between .10 and .29 being small, between .30 and .49 being medium, and greater than .50 being large.

Results

The aim of the current study was to investigate both fluid and crystallised intelligence in relation to HE students' ability to critically assess AI-generated output for their studies (AI-CT). Therefore, this study posed the following two research questions: (RQ1) *To what extent is fluid intelligence associated with AI-CT?* and (RQ2) *To what extent is crystallised intelligence associated with AI-CT?* Subsequent hypotheses were formulated and investigated. In this section, the prerequisite assumptions are first checked for and then the regression analyses are conducted and reported on.

Data Inspection

First, outlier detection, inspection and removal were conducted. The Cook's Distance values for the regression for both RQ1 and RQ2 did not exceed the value of 1. However, the standardised residual data of RQ1 and RQ2 together showed a total of 18 cases with values greater than 1.96 ($Z < -1.96$; $Z > 1.96$). Upon closer inspection, these participants showed acquiescence bias (i.e., saying yes to all questions without reading the questions). This was deducted from the fact that their values on the reverse-coded questions differed extremely from the other non-reverse-coded questions, thus breaking a consistent logical pattern in their answers. Additionally, it was found that one remaining participant had not filled in all the AI-

CT scale questions. All these datapoints were deleted listwise from the sample, resulting in a sample size of 353 participants. The descriptive statistics of this sample on all the study variables are summarised in Table 2.

Table 2

Descriptive Statistics

Characteristic	Sample					
	N	*M*	%	*SD*	*Min.*	*Max.*
AI-CT, *M (SD)*	353	15.74		3.21	7	20
Fluid intelligence, *M (SD)*	353	3.04		1.34	0	6
Crystallised intelligence, *M (SD)*	353	3.92		2.41	0	12
Age (in years), *M* (*SD*)	353	25.01		3.59	18	34
Highest academic accomplishment, %					1	4
No degree	177		50.1%			
Other degree	16		4.5%			
Bachelor	142		40.2%			
Master	1		1.6%			
Intended study purpose of GenAI LLMs [a], %						
Providing definitions	231		65.4%		0	1
Providing feedback on writing	261		73.9%		0	1
Generating ideas or solutions	222		62.9%		0	1
Evaluating ideas or solutions	254		72.0%		0	1
Answering practice assignments	294		83.3%		0	1
Generating practice assignments	307		87.0%		0	1
Assisting programming tasks	271		76.8%		0	1
Rewriting or paraphrasing texts	238		67.8%		0	1

Generating text summaries	248	70.3%		0	1
Big Five personality traits, *M (SD)*					
Openness to experience	353	36.58	5.86	18	50
Conscientiousness	353	34.89	6.20	17	50
Extraversion	353	29.79	8.02	11	50
Agreeableness	353	38.83	6.07	18	50
Neuroticism	353	30.54	7.79	10	48

Note. n = 353. *M* = mean (for numeric variables) or % = percentage (for categorical variables); *SD* = Standard Deviation; *Min.* = Minimum; *Max.* = Maximum. 'Fachkraft 2030' shows no systematic deviations from German student population (Seegers et al., 2019).

[a] Reflects the number and percentage of participants answering applying this purpose, with value 0 representing 'no' and 1 representing 'yes'. Multiple purposes can be answered by each participant. This means that the % scores do not accumulate to 100%.

Assumption Checks

The assumption of homoscedasticity was not met for either of the regression analyses (RQ1 nor RQ2). This was deducted from the scatterplots of the standardised residuals on the y-axis and the standardised predicted values on the x-axis, which showed heteroscedasticity (i.e., the absence of equal variance across the y-axis). To make the regression analyses robust, weighted least squares (WLS) have been computed for the regression analyses of both RQ1 and RQ2. By adding the WLS, each datapoint was adjusted for the amount of prediction error they were associated with. Thus, hierarchical WLS regression analyses have been conducted for RQ1 and RQ2. Furthermore, the residuals are normally distributed and there is linearity between the dependent and the independent variables for both the RQ1 and RQ2 regression analyses. There is also no multicollinearity between the independent variables, between the control variables, nor between the independent and the control variables. This is because there are no Pearson's correlation values greater than .80, no Tolerance values smaller than .10, and no VIF-values greater than 10 (see Appendix VI for the SPSS output of the assumption checks).

Research Question 1

In Hypothesis 1, it is expected that higher scores on fluid intelligence are associated with higher critical thinking towards AI-generated output for one's studies (i.e., evaluation on its accuracy, reliability, and applicability). A hierarchical WLS regression analysis was conducted to examine the association between fluid intelligence and AI-CT, controlling for age, highest academic accomplishment, intended study purpose of GenAI LLMs, and Big Five personality traits (see Tables 3, 4, and 5). Model 1 (M1) reports the relationship between the independent variable (fluid intelligence) and the dependent variable (AI-CT). The control variables were added in Model 2 (M2). For highest academic accomplishment, the 'no degree' category is used as the reference group.

In Model 1, fluid intelligence was found to significantly explain 2.6% of the variance in AI-CT ($R_{squared}$ = .026; $F_{\text{change}}(1, 351) = 9.28$; p_{change} = .002). Once the control variables were added, Model 2 grew more significant and could explain an additional 29.8% of the variance in AI-CT (R_{change} = .298; $F_{\text{change}}(18, 333) = 8.14$; $p_{\text{change}} < .001$).

Table 3

Model Statistics of the Hierarchical (WLS) Regression Analysis for RQ1

	M1					*M2*				
	Sum of Squares	*df*	*Mean Square*	*F*	*p*	*Sum of Squares*	*df*	*Mean Square*	*F*	*p*
Regression	19.09	1	19.09	9.28**	.002	239.92	19	12.63	8.38***	<.001
Residual	722.61	351	2.06			501.78	333	1.51		
Total	741.70	352				741.70	352			

Note: $n = 353$. *df* = degrees of freedom; *p* = significance.

* Significant at $p < .05$; ** Significant at $p < .01$; *** Significant at $p < .001$.

Table 4

Model Summary of the Hierarchical (WLS) Regression Analysis for RQ1

M1						*M2*					
R_{squared}	R_{change}	F_{change}	*df1*	*df2*	p_{change}	R_{squared}	R_{change}	F_{change}	*df1*	*df2*	p_{change}
.026	.026	9.28**	1	351	.002	.323	.298	8.14***	18	333	<.001

Note: $n = 353$. *R* = explained variance; R_{change} = changed explained variance; *df1*= numerator degrees of freedom; *df2* = denominator degrees of freedom; p_{change} = change significance.

* Significant at $p < .05$; ** Significant at $p < .01$; *** Significant at $p < .001$.

Fluid intelligence was no longer found to be significantly associated with AI-CT, once the control variables were added (*Beta* = .09; *p* = .069). This finding contradicts Hypothesis 1. In other words, HE students who show higher scores on fluid intelligence do not necessarily show higher scores on critical thinking towards AI-generated output for their studies.

Additionally, for the control variable 'highest academic accomplishment', it was found that having a degree (category 'other degree') was significantly more negatively associated with AI-CT than having no degree (*Beta* = -.11; p = .033). Categories 'bachelor' and 'master' did not significantly differ from 'no degree'. For the control variable 'intended study purpose of GenAI LLMs' there were four purposes found to be significantly associated with AI-CT. The following three purposes were found to be positively associated: providing definitions and explanations (*Beta* = .22; p < .001), providing feedback on own writing (*Beta* = .21; p < .001), and assisting with coding and programming (*Beta* = .24; p < .001). Answering practice questions or assignments was found to be negatively associated (*Beta* = -.14; p = .018). For the control variable 'Big Five personality traits', only 'openness to experience' was found to be significantly positively associated with AI-CT (*Beta* = .16; p = .006). None of the other personality traits were found to be significantly associated.

Table 5

Hierarchical (WLS) Regression Analysis Coefficients for RQ1

Variable	*M1*						*M2*					
	Unstd. B	*Std. Beta*	*SE*	*95% CI*		*p*	*Unstd. B*	*Std. Beta*	*SE*	*95% CI*		*p*
				LL	*UL*					*LL*	*UL*	
Intercept	15.29***		.37	14.56	16.01	<.001	10.75***		1.65	7.50	13.99	<.001
Fluid intelligence	.35**	.16**	.12	.12	.57	.002	.19	.09	.10	-.02	.39	.069
Age							-.01	-.01	.05	-.10	.08	.877
Highest academic accomplishment (ref. No degree)												
Other degree							-1.35*	-.11*	.63	-2.58	-.11	.033
Bachelor							-.02	-.003	.31	-.63	.60	.959
Master							-.72	-.04	.96	-2.62	1.17	.453
Intended study purpose												
Definitions / explanations							1.30***	.22***	.39	.53	2.06	<.001
Feedback on writing							1.23***	.21***	.37	.51	1.95	<.001
Generate ideas or solutions							.29	.05	.39	-.48	1.05	.462

Evaluate own ideas or solutions		.05	.01	.36	-.66	.75	.900
Answer questions or assignments		-.96*	-.18*	.41	-1.75	-.16	.018
Generate questions or assignments		-.69	-.08	.46	-1.59	.22	.137
Assist coding / programming		1.42***	.24***	.34	.75	2.09	<.001
Rewrite own text		.24	.04	.36	-.47	.95	.513
Creating summaries of texts		-.12	-.02	.35	-.81	.56	.730
Big Five personality traits							
Openness to experience		.08**	.16**	.03	.02	.13	.006
Conscien-tiousness		-.004	-.01	.02	-.05	.04	.864
Extraversion		-.02	-.05	.02	-.06	.02	.400
Agree-ableness		.04	.09	.02	-.01	.09	.080
Neuroticism		-.003	-.01	.02	-.04	.04	.884

Note: n = 353. R = explained variance; *Unstd. B* = unstandardised coefficient; *Std. Beta* = standardised coefficient; *SE* = Standard Error; *CI* = confidence interval; *LL* = lower limit; *UL* = upper limit; p = significance. M1 (Model 1) portrays the relationship between the independent variable (*fluid intelligence*) and the dependent variable (*AI-CT*). M2 (Model 2) adds the control variables (*age, highest academic accomplishment, the intended study purpose of GenAI LLMs, and the Big Five personality traits*).
* Significant at $p < .05$; ** Significant at $p < .01$; *** Significant at $p < .001$.

Research Question 2

In Hypothesis 2, it is expected that higher scores on crystallised intelligence are associated with higher critical thinking towards AI-generated output for one's studies (i.e., evaluation on its accuracy, reliability, and applicability). A hierarchical WLS regression analysis (see Tables 6, 7, and 8) was conducted to examine the association between crystallised intelligence and AI-CT, when controlling for age, highest academic accomplishment, intended study purpose of GenAI LLMs, and Big Five personality traits. Model 1 (M1) reports the relationship between the independent variable (crystallised intelligence) and the dependent variable (AI-CT). The control variables were added in Model 2 (M2). For highest academic accomplishment, the 'no degree' category is used as the reference group.

In Model 1, crystallised intelligence was found to significantly explain 5.2% of the variance in AI-CT ($R_{squared}$ = .052; $F_{\text{change}}(1, 351) = 19.08$; $p_{\text{change}} < .001$). Once the control variables were added, Model 2 remained significant and could explain an additional 31.1% of the variance in AI-CT ($R_{\text{change}} = .311$; $F_{\text{change}}(18, 333) = 9.04$; $p_{\text{change}} < .001$).

Table 6

Model Statistics of the Hierarchical (WLS) Regression Analysis for RQ2

	M1					*M2*				
	Sum of Squares	*df*	*Mean Square*	*F*	*p*	*Sum of Squares*	*df*	*Mean Square*	*F*	*p*
Regression	41.39	1	41.39	19.08***	<.001	291.33	19	15.33	9.98***	<.001
Residual	761.35	351	2.17			511.41	333	1.54		
Total	802.73	352				802.73	352			

Note: $n = 353$. *df* = degrees of freedom; p = significance.
* Significant at $p < .05$; ** Significant at $p < .01$; *** Significant at $p < .001$.

Table 7

Model Summary of the Hierarchical (WLS) Regression Analysis for RQ2

M1						*M2*					
R_{squared}	R_{change}	F_{change}	*df1*	*df2*	p_{change}	R_{squared}	R_{change}	F_{change}	*df1*	*df2*	p_{change}
.052	.052	19.08***	1	351	<.001	.363	.311	9.04***	18	333	<.001

Note: $n = 353$. R = explained variance; R_{change} = changed explained variance; *df1*= numerator degrees of freedom; *df2* = denominator degrees of freedom; p_{change} = change significance.
* Significant at $p < .05$; ** Significant at $p < .01$; *** Significant at $p < .001$.

Crystallised intelligence was still found to be significantly positively associated with AI-CT once the control variables were added (*Beta* = .21; $p < .001$). This is a small to medium effect size. This finding is in line with Hypothesis 2. In other words, HE students who show higher scores on crystallised intelligence indeed also show higher scores on critical thinking towards AI-generated output for their studies.

Additionally, for the control variable 'intended study purpose of GenAI LLMs' there were four purposes found to be significantly associated with AI-CT. The following three purposes were found to be positively associated: providing definitions and explanations (*Beta* = .21; p = .001), providing feedback on own writing (*Beta* = .21; p < .001), and assisting with coding and programming (*Beta* = .26; p <.001). Answering practice questions or assignments was found to be negatively associated (*Beta* = -.13; p = .023). For the control variable 'Big Five personality traits', only 'openness to experience' was found to be significantly positively associated with AI-CT (*Beta* = .13; p = .021). None of the other personality traits were found to be significantly associated.

Table 8

Hierarchical (WLS) Regression Analysis Statistics for RQ2

Variable	*M1*						*M2*					
	Unstd. B	*Std. Beta*	*SE*	*95% CI LL*	*95% CI UL*	*p*	*Unstd. B*	*Std. Beta*	*SE*	*95% CI LL*	*95% CI UL*	*p*
Intercept	15.38***		.29	14.82	15.95	<.001	10.77***		1.62	7.57	13.96	<.001
Crystallised intelligence	.27***	.23***	.06	.15	.40	<.001	.25***	.21***	.06	.14	.36	<.001
Age							.02	.02	.04	-.07	.11	.662
Highest academic accomplishment (ref. No degree)												
Other degree							-1.09	-.10	.56	-2.19	.003	.051
Bachelor							-.06	-.01	.30	-.66	.53	.833
Master							-.73	-.04	.91	-2.51	1.06	.425
Intended study purpose												
Definitions / explanations							1.20**	.21**	.37	.47	1.94	.001
Feedback on writing							1.23***	.21***	.35	.55	1.92	<.001
Generate ideas or solutions							.16	.03	.38	-.60	.91	.684
Evaluate own ideas or solutions							.26	.05	.34	-.41	.93	.440

Answer questions or assignments	-.89*	-.13*	.39	-1.66	-.12	.023
Generate questions or assignments	-.80	-.09	.45	-1.68	.08	.075
Assist coding / programming	1.53***	.26***	.33	.88	2.17	<.001
Rewrite own text	.29	.05	.34	-.39	.96	.406
Creating summaries of texts	-.20	-.04	.33	-.86	.46	.551
Big Five personality traits						
Openness to experience	.06*	.13*	.03	.01	.12	.021
Conscien-tiousness	-.001	-.003	.02	-.05	.04	.949
Extraversion	-.03	-.07	.02	-.06	.01	.191
Agree-ableness	.03	.06	.02	-.02	.08	.218
Neuroticism	<.001	<.001	.02	-.04	.04	.995

Note: n = 353. *Unstd. B* = unstandardised regression coefficient; *Std. Beta* = standardised regression coefficient; *SE* = Standard Error; *CI* = confidence interval; *LL* = lower limit; *UL* = upper limit; p = significance. M1 (Model 1) portrays the relationship between the independent variable (*crystallised intelligence*) and the dependent variable (*AI-CT*). M2 (Model 2) adds the control variables (*age, highest academic accomplishment, the intended study purpose of GenAI LLMs, and the Big Five personality traits*).

* Significant at $p < .05$; ** Significant at $p < .01$; *** Significant at $p < .001$.

Discussion

The aim of this current study was to investigate the intelligence domains in relation to higher education (HE) students' ability to critically assess AI-generated output for their studies (AI-CT), after controlling for age, highest academic accomplishment, study purpose of GenAI LLMs, and the Big Five personality traits. For this, the study posed two research questions: (RQ1) *To what extent is fluid intelligence associated with AI-CT?* and (RQ2) *To what extent is crystallised intelligence associated with AI-CT?*

Fluid Intelligence, Artificial Intelligence, and Critical Thinking

No substantial support was found in favour of a positive association between fluid intelligence and AI-CT (H1). The results showed that, after adding the control variables, HE students who show higher scores on fluid intelligence do not necessarily show higher scores on critical thinking towards AI-generated output for their studies (i.e., evaluation on its accuracy, reliability, and applicability). Therefore, the answer to RQ1 is that there is no substantial association between fluid intelligence and AI-CT.

This finding furthers our understanding of the relationship between fluid intelligence and critical thinking (CT). In their study, Liu et al. (2024) had found a positive association between these concepts. However, their study did not account for the influence of AI. The current study shows that once GenAI LLMs get involved, the association between fluid intelligence and CT is mitigated. This aligns with the aforementioned empirical studies that reported a negative association between AI-usage (for one's studies) and CT (Krstić et al., 2022; Lee et al., 2025; Zhai et al., 2024). It is likely that the usage of GenAI LLMs for one's studies excessively impedes HE students' cognitive abilities, consequentially diminishing their ability to critically assess AI-generated output.

Additionally, the current study results add an important note to the findings by Vázquez-Cano et al. (2021). They suggested that AI-usage is positively associated with fluid intelligence, because GenAI LLMs were proposed to have the ability to facilitate adaptive cognitive processes. In turn, this association should yield higher scores on CT (Heijltjes et al., 2014; Liu et al., 2024). The current study, however, indicates that the utilisation of GenAI LLMs does not result in higher scores on CT when applied to AI-generated output for studying.

Crystallised Intelligence, Artificial Intelligence, and Critical Thinking

Substantial support was found in favour of a positive association between crystallised intelligence and AI-CT (H2). The results showed that, after adding the control variables, HE students who show higher scores on crystallised intelligence will (to moderate extent) show higher scores on critical thinking towards AI-generated output for their studies (i.e., evaluation on its accuracy, reliability, and applicability). Therefore, the answer to RQ2 is that there is a moderate positive association between crystallised intelligence and AI-CT.

In their study, Heijltjes et al. (2014), had reported a positive association between crystallised intelligence and critical thinking (CT). They found that when HE students' foundational knowledge was expanded, their CT-ability improved as well. This current study shows that – in contrast to the relationship between fluid intelligence and CT – when involving GenAI LLMs, the association between crystallised intelligence and CT persists. The studies that report on the negative association between AI-usage and people's CT-ability mostly focus on CT-skills. The persistence of this relationship is likely due to the facilitating nature of GenAI LLMs in building one's foundational knowledge, and thus improvement of crystallised intelligence (Sheth et al., 2019).

Control Variables

This study controlled for factors that have been proven to mediate HE students' utilisation of GenAI LLMs (age, highest academic accomplishment, intended study purpose of GenAI LLMs, and the Big Five personality traits).

First, it was found that HE students' highest academic accomplishment was negatively associated with AI-CT. Having an academic degree (category 'other') was found to be indicative of lower AI-CT than having no academic degree. This finding was expected, because earlier research showed academic level is positively associated with both AI-familiarity and AI-utilisation (Gerlich, 2025; Persson et al., 2021; Zhang & Dafoe, 2019). Consequentially, the usage of GenAI LLMs for one's studies diminishes HE students' CT-ability (Krstić et al., 2022; Lee et al., 2025; Zhai et al., 2024).

Second, the following three study purposes of GenAI LLMs were found to be positively associated with AI-CT: providing definitions and explanations, providing feedback on own writing, and assisting with coding and programming. Another purpose, answering practice questions or assignments, was found to be negatively associated with AI-CT. These findings can be explained by Klarin et al.'s (2024) study insights. They reported that students who use GenAI LLMs for task-initiation and task-structuring thrived better in the long run than those who use it for task-completion. The three positively associated purposes all support the learning process. Contrarily, the purpose found to be negatively associated with AI-CT aligns with task-completion.

Lastly, for the Big Five personality traits, only 'openness to experience' was found to be significantly positively associated with AI-CT. Openness to experience refers to the extent in which a person is open-minded and shows curiosity towards novelty. This finding supports the claim by Nieto and Saiz (2011) that openness to experience is positively associated with CT. Additional research into CT-disposition serves to explain why only this personality trait was found to be significantly associated with AI-CT. Namely, 'open-mindedness' is one of the CT-dispositions (Facione et al., 1994). The other personality traits do not show equally strong connections to the CT-dispositions as openness to experience does.

Practical Implications

In addition to the theoretical implications above, the current study findings also highlight two important practical implications. First, educational tasks must focus on stimulating the usage of GenAI LLMs for task-initiation and task-structuring, rather than task-completion. This can be done by, e.g., requiring students to hand in additional 'CT-analyses' of AI-generated output that has been used. Through this additional assignment the focus would shift more towards developing and applying critical thinking and reasoning skills when utilising GenAI LLMs for one's studies. Second, the study findings underscore the importance for educational institutes to embed AI-literacy trainings into their study programs, specifically focused on acquiring and developing CT-skills towards AI-generated output. The tools to implement these trainings are already available. For example, Schüller (2022) has proposed a framework focused on enabling students to consciously and ethically interact with AI-generated output for their studies. AI-CT-abilities can and should be trained.

Limitations and Future Research

This study has several limitations to take into account. First, the current study has a cross-sectional design, meaning that the presence of causal pathways cannot be established. Future longitudinal research would be necessary to draw conclusions about causal relationships between the factors investigated in this empirical study (fluid intelligence, crystallised intelligence, and AI-CT). Even so, the cross-sectional design was suited in attaining the current research objectives: investigating associations, not causations. Second, the data-collection process was executed with unsupervised self-reports by the participants, which in turn introduces the risk of weakening the data validity. Namely, self-reports allow for more subjectivity when interpreting the survey items. To mitigate this limitation, only well-validated instruments were employed for data-collection and the new AI-CT scale was based on another prior existing (well-validated) scale. Third, this study did not have a control group that did not use GenAI LLMs for their studies. This was due to 'Fachkraft 2030' (i.e., the data-collection survey) not posing the AI-CT items to participants who answered to not use GenAI LLMs at all. For investigating the current study objectives, this did not form a restriction. However, a comparison on CT-abilities between students who do and those who do not use GenAI LLMs for their studies could result in more distinct insights on the relationships between CT, AI, and the intelligence domains. Another limitation of this study is that participants could choose multiple study purposes of GenAI LLMs, which negated the possibility of appointing a reference group. Due to this, the valid interpretation of isolated associations is limited. A last limitation was that the assumption of homoscedasticity could initially not be met. However, each datapoint was adjusted for the amount of prediction error they were associated with by computing weighted least squares (WLS) and integrating them into the hierarchical regression analyses, in turn making the analyses robust against violation of the assumption of homoscedasticity.

Conclusion

In conclusion, the introduction of GenAI LLMs has complicated the relationship between the intelligence domains 'fluid and crystallised intelligence' and critical thinking. The current study shows that while crystallised intelligence robustly (i.e., controlling for age, highest academic accomplishment, intended study purpose of GenAI LLMs, and the Big Five personality traits) supports HE students' ability to critically evaluate AI-generated output for their studies, fluid intelligence shows no association with critical thinking once GenAI LLMs get involved. This finding emphasises the disruptive influence of AI-usage on the underlying cognitive processes. The persistent nature of the positive association between crystallised intelligence and AI-CT suggests that HE students' foundational knowledge remains a crucial drive for their CT-ability. Ultimately, generative artificial intelligence remains a dual-edged sword for education, introducing both new learning opportunities and creating challenging issues. Nevertheless, we can retain control. The key is human agency. Through focus on developing our fluid intelligence, crystallised intelligence, and critical thinking skills, we can influence our learning journeys and persevere in the face of educational adversity.

References

Abdelhamid, G. S. M., Gómez-Benito, J., Abdeltawwab, A. T. M., Abu Bakr, M. H. S., & Kazem, A. M. (2019). Hierarchical structure of the wechsler adult intelligence scale–fourth edition with an Egyptian sample. *Journal of Psychoeducational Assessment*, *37*(3), 395–404. https://doi.org/10.1177/0734282917732857

Bagheri, F., & Ghanizadeh, A. (2016). Critical thinking and gender differences in academic self-regulation in higher education. *Journal of Applied Linguistics and Language Research*. https://www.semanticscholar.org/paper/Critical-Thinking-and-Gender-Differences-in-in-Bagheri-Ghanizadeh/35cea1e0065cad7d38151a03cbde6663255671e6

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and Teaching of 21st Century Skills* (pp. 17–66). Springer Netherlands. https://doi.org/10.1007/978-94-007-2324-5_2

Burke, H. R. (1985). Raven's progressive matrices: More on norms, reliability, and validity. *Journal of Clinical Psychology*, *41*(2), 231–235. https://doi.org/10.1002/1097-4679(198503)41:2<231::AID-JCLP2270410216>3.0.CO;2-Z

Constata. (2023). *Fachkraft 2030*. https://www.constata.de/#beratung

Costa, P. T., & McCrae, R. R. (2008). The revised NEO personality inventory (NEO-PI-R). In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment: personality measurement and testing* (Vol. 2, Issue 2, pp. 179–198). Sage Publications Inc. https://doi.org/10.4135/9781849200479.n9

Databricks. (2024). *A compact guide to large language models*. Databricks. https://scalexm.ai/wp-content/uploads/2023/10/White-Paper-LLM.pdf

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2014). An integrated critical thinking framework for the 21st century. *Thinking Skills and Creativity*, *12*, 43–52. https://doi.org/10.1016/j.tsc.2013.12.004

Eaves, S., & Nagar, D. (2025, February 18). *AI's efficiency wars have begun*. Foreign Policy. https://foreignpolicy.com/2025/02/05/deep-seek-china-us-artificial-intelligence-ai-arms-race/

Facione, N. C., Facione, P. A., & Sanchez, C. A. (1994). Critical thinking disposition as a measure of competent clinical judgment: The development of the California critical thinking disposition inventory. *Journal of Nursing Education*, *33*(8), 345–350. https://doi.org/10.3928/0148-4834-19941001-05

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International, 61*(3), 460–474. https://doi.org/10.1080/14703297.2023.2195846

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, *15*(6), 1–28. https://doi.org/10.2139/ssrn.5082524

Ghanizadeh, A. (2017). The interplay between reflective thinking, critical thinking, self-monitoring, and academic achievement in higher education. *Higher Education*, *74*(1), 101–114. https://doi.org/10.1007/s10734-016-0031-y

Hamad, N. I., El-Ashry, A. M., Ibrahim, I. M., & Hassan, E. A. (2024). Embracing the future: An insight into nursing students' attitude and perception towards the usability of artificial intelligence in healthcare. *Teaching and Learning in Nursing*, *[accepted for publication]*, 1–9. https://doi.org/10.1016/j.teln.2024.12.017

Heijltjes, A., van Gog, T., & Paas, F. (2014). Improving students' critical thinking: Empirical support for explicit instructions combined with practice. *Applied Cognitive Psychology*, *28*(4), 518–530. https://doi.org/10.1002/acp.3025

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, *57*(5), 253–270. https://doi.org/10.1037/h0023816

Humble, N., & Mozelius, P. (2022). The threat, hype, and promise of artificial intelligence in education. *Discover Artificial Intelligence, 2*(1), 1–13. https://doi.org/10.1007/s44163-022-00039-z

JobValley. (2023). *Fachkraft 2030*. https://jobvalley.com/en-de/fachkraft-2030/

Kaya, F., Aydin, F., Schepman, A., Rodway, P., Yetişensoy, O., & Kaya, M. D. (2024). The roles of personality traits, AI anxiety, and demographic factors in attitudes toward artificial intelligence. *International Journal of Human–Computer Interaction*, *40*(2), 497–514. https://doi.org/10.1080/10447318.2022.2151730

Klarin, J., Hoff, E., Larsson, A., & Daukantaitė, D. (2024). Adolescents' use and perceived usefulness of generative AI for schoolwork: Exploring their relationships with executive functioning and academic achievement. *Frontiers in Artificial Intelligence*, *7*, 1–13. https://doi.org/10.3389/frai.2024.1415782

Krstić, L., Aleksić, V., & Krstić, M. (2022). Artificial intelligence in education: A review. *IT Education and Practice*, 223–228. https://doi.org/10.46793/TIE22.223K

Lee, H. P., Drosos, I., Sarkar, A., Rintel, S., Wilson, N., Tankelevitch, L., & Banks, R. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. *Association of Computing Machinery*. https://doi.org/10.1145/3706598.3713778

Lee, S., & Park, G. (2024). Development and validation of ChatGPT literacy scale. *Current Psychology*, *43*(21), 18992–19004. https://doi.org/10.1007/s12144-024-05723-0

Liu, X., Han, X., Wang, T., & Ren, X. (2024). An account of the relationship between critical thinking and fluid intelligence in considering executive functions. *Thinking Skills and Creativity*, *52*, 101538. https://doi.org/10.1016/j.tsc.2024.101538

Maastricht University. (2023, April 11). *Study 'Fachkraft 2030': German university research made in Maastricht*. https://www.maastrichtuniversity.nl/news/study-fachkraft-2030-german-university-research-made-maastricht

Markauskaite, L., Marrone, R., Poquet, O., Knight, S., Martinez-Maldonado, R., Howard, S., Tondeur, J., De Laat, M., Buckingham Shum, S., Gašević, D., & Siemens, G. (2022). Rethinking the entwinement between artificial intelligence and human learning: What capabilities do learners need for a world with AI? *Computers and Education: Artificial Intelligence*, *3*, 100056. https://doi.org/10.1016/j.caeai.2022.100056

McCrae, R. R., & Costa, P. T. (2008). The five-factor theory of personality. In *Handbook of personality: Theory and research* (3rd ed., pp. 159–181). The Guilford Press.

Nieminen, P. (2022). Application of standardized regression coefficients in meta-analysis. *BioMedInformatics, 2*(3). https://doi.org/10.3390/biomedinformatics2030028

Nieto, A. M., & Saiz, C. (2011). Skills and dispositions of critical thinking: Are they sufficient? *Anales de Psicología*, *27*(1), 202–209. https://doi.org/10.6018/analesps

Pantelimon, F., Bologa, R., Toma, A., & Posedaru, B.-S. (2023). The evolution of AI-driven educational systems during the Covid-19 pandemic. *Sustainability*, *13*(23), 13501. https://doi.org/10.3390/su132313501

Paul, R., & Elder, L. (2019). *The miniature guide to critical thinking: Concepts and tools* (8th ed.). Rowman & Littlefield Publishers.

Persson, A., Laaksoharju, M., & Koga, H. (2021). We mostly think alike: Individual differences in attitude towards AI in Sweden and Japan. *The Review of Socionetwork Strategies*, *15*(1), 123–142. https://doi.org/10.1007/s12626-021-00071-y

Rivas, S. F., Saiz, C., & Ossa, C. (2022). Metacognitive strategies and development of critical thinking in higher education. *Frontiers in Psychology*, *13*, 913219. https://doi.org/10.3389/fpsyg.2022.913219

Schüller, K. (2022). Data and AI literacy for everyone. *Statistical Journal of the AIOS, 38*(2), 477–490. https://doi.org/10.3233/SJI-220941

Seegers, P. K., Bergerhoff, J., Hartmann, S., & Knappe, A. (2019). *Fachkraft 2030: Erhebung zur wirtschaftlichen und allgemeinen lebenssituation der studierenden in Deutschland*. Studitemps GmbH. https://jobvalley.com/pdf/fachkraft-13-und-14-erhebung.pdf

Sheth, A., Yip, H. Y., Iyengar, A., & Tepper, P. (2019). Cognitive services and intelligent chatbots: Current perspectives and special issue introduction. *IEEE Internet Computing*, *23*(2), 6–12. https://doi.org/10.1109/MIC.2018.2889231

Stanovich, K. E., & Stanovich, P. J. (2010). A framework for critical thinking, rational thinking, and intelligence. In *Innovations in educational psychology: Perspectives on learning, teaching, and human development* (pp. 195–237). Springer Publishing Company.

Stichting Leerplan Ontwikkeling. (2023, October 9). *21e-eeuwse vaardigheden: Kritisch denken*. SLO. https://www.slo.nl/thema/meer/21e-eeuwsevaardigheden/kritisch-denken/

Tewari, G. (2022, November 8). *AI: The apex technology of the information age*. Forbes. https://www.forbes.com/councils/forbesbusinesscouncil/2022/11/08/ai-the-apex-technology-of-the-information-age/

United Nations. (2023). *United Nations global principles for information integrity: Recommendations for multi-stakeholder action.* United Nations. https://www.un.org/sites/un2.un.org/files/un-global-principles-for-information-integrity-en.pdf

Vázquez-Cano, E., Mengual-Andrés, S., & López-Meneses, E. (2021). Chatbot to improve learning punctuation in Spanish and to enhance open and flexible learning environments. *International Journal of Educational Technology in Higher Education*, *18*, 1–20. https://doi.org/10.1186/s41239-021-00269-8

Verma, P. (2023, December 17). The rise of AI fake news is creating a misinformation superspreader. *The Washington Post*. https://www.washingtonpost.com/technology/2023/12/17/ai-fake-news-misinformation/

Westrick, P. A., Angehr, E. L., Shaw, E. J., & Marini, J. P. (2024). *Recent trends in college readiness and subsequent college performance: With faculty perspectives on student readiness* (pp. 1–52). CollegeBoard. https://research.collegeboard.org/reports/college-access/recent-trends-college-readiness

World Economic Forum. (2025). *Generative artificial intelligence*. World Economic Forum4. https://intelligence.weforum.org

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, *11*(1), 28. https://doi.org/10.1186/s40561-024-00316-7

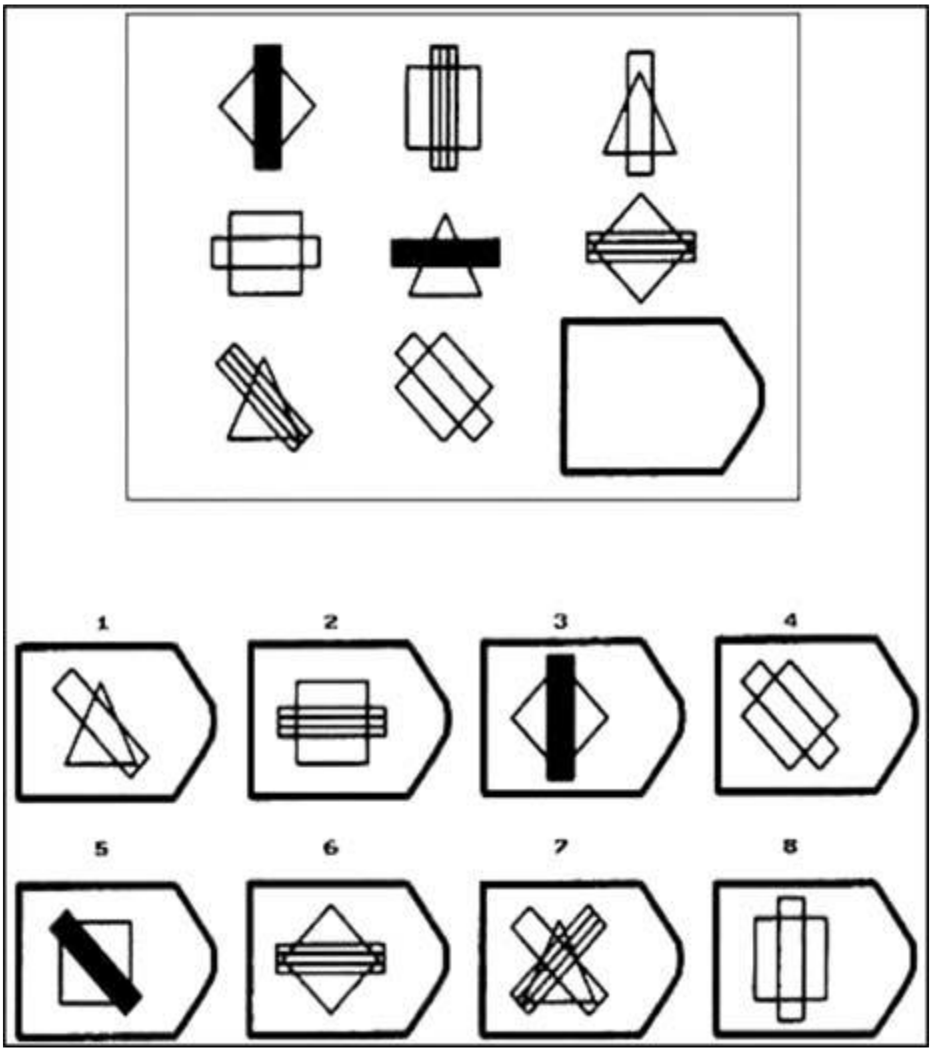
Zhang, B., & Dafoe, A. (2019). *Artificial intelligence: American attitudes and trends*. Social Science Research Network. https://doi.org/10.2139/ssrn.3312874

Appendices

Appendix I – Raven's Progressive Matrix example

Figure 3

Example of a Raven's Progressive Matrix



Appendix II – WAIS-IV Scale Items | Crystallised Intelligence

Table 9

Wechsler's Adult Intelligence Scale Items Translated from German to English

Item	German	English
1.	7 Mitarbeiter können in 7 Tagen eine bestimmte Arbeit verrichten. Wie viele Mitarbeiter braucht man um dieselbe Arbeit in 1 Tag bei halber Arbeitszeit zu erledigen?	7 employees can complete a certain job in 7 days. How many employees are needed to finish the same job in 1 day with half the working hours?
2.	Ein Eimer enthält von jeder Münzenart (1cent, 2cent, 5cent, 10cent, 20cent, 50cent, 1€, 2€) 100 Münzen. Wie viel Euro sind insgesamt im Eimer?	A bucket contains 100 coins of each type (1c, 2c, 5c, 10c, 20c, 50c, €1, €2). How many euros are there in total?
3.	Der Logarithmus von 100 zur Basis 10 ist ...?	The logarithm of 100 to the base 10 is...?
4.	Ein Brötchen kostet 7 Cent. Ein Brot kostet 1,13 Euro mehr als ein Brötchen. Wie viel kosten 3 Brote und 2 Brötchen zusammen?	A bread slice costs 7 cents. A loaf of bread costs €1.13 more. How much do 3 loafs and 2 slices cost in total?
5.	Welche Zahl X passt in folgende Reihe: 1, 2, 3, 5, X, 13?	Which number fits the spot X in the following sequence: 1, 2, 3, 5, X, 13?
6.	Welche Zahl X passt in folgende Reihe: 19, 17, 13, X, 7, 5, 3, 2?	Which number fits the spot X in the following sequence: 19, 17, 13, X, 7, 5, 3, 2?
7.	Ein Hektar sind 100 Meter mal 100 Meter. Wie viele Quadratkilometer sind 1.000 Hektar?	One hectare is 100 meters by 100 meters. How many square kilometres are 1.000 hectares?
8.	Es ist 11:53 Uhr. Zwei Züge sind 250 km voneinander entfernt. Zug 'A' fährt mit 120 km/h los, Zug 'B' mit 100 km/h zehn Minuten später. Wann treffen sie sich?	It is 11:53 a.m. Two trains are 250 km apart. Train 'A' departs at 120 km/h, train 'B' at 100 km/h ten minutes later. When will they meet?

9.	‘A’ ist zweimal so groß wie ‘B’. ‘C’ ist genau halb so groß wie ‘B’. Was ist die Summe von ‘A’, ‘B’ und ‘C’?	‘A’ is twice as large as ‘B’. ‘C’ is exactly half as large as ‘B’. What is the sum of ‘A’, ‘B’, and ‘C’?
10.	Am Jahresanfang steht ein Wertpapier bei 20 Euro. Am Jahresende steht es bei 24,20 Euro. Wie viel Prozent Gewinn ergibt das?	At the beginning of the year, a stock is worth €20. At year-end it is worth €24,20. What is the percentage gain?
11.	Was ist Dihydrogenmonoxid?	What is dihydrogen monoxide?
12.	Wie viele Liter Wasser passen ungefähr in einen Kubikmeter?	Approximately, how many litres of water fit into a cubic meter?
13.	Wie groß ist der Umfang der Erde am Äquator in Kilometern?	What is the Earth's circumference at the equator in kilometres?
14.	Wie viel Prozent Sauerstoff enthält die Erdatmosphäre?	What percentage of the Earth's atmosphere is oxygen?
15.	Welchen Wert hat die übliche Haushaltsspannung in Deutschland (Steckdosenspannung)?	What is the standard household voltage in Germany (outlet voltage)?
16.	Aus welchen drei Bestandteilen besteht ein Atom?	What are the three components of an atom?
17.	Welcher Planet des Sonnensystems ist der Sonne am nächsten?	Which planet in the solar system is closest to the Sun?
18.	Wie hoch ist die Flughöhe der Internationalen Raumstation (ISS) in Kilometern?	What is the altitude of the International Space Station (ISS) in kilometres?
19.	Durch welche Stadt läuft der Nullmeridian?	Through which city does the Prime Meridian pass?
20.	Welche Stadt auf der Südhalbkugel hat die meisten Einwohner?	Which city in the Southern Hemisphere has the most inhabitants?
21.	Wer war kein Bundespräsident von Deutschland?	Who was not a President of Germany?
22.	Wie viele Städte mit mehr als 1,2 Millionen Einwohnern gibt es in Deutschland?	How many cities in Germany have more than 1.2 million inhabitants?

23.	In welchem Jahr begann der 1. Weltkrieg?	In which year did World War I start?
24.	An welchem Tag fiel die Berliner Mauer?	On what day did the Berlin Wall fall?
25.	Welches Land hat keine gemeinsame Grenze mit Deutschland?	Which country does not share a border with Germany?
26.	Wie lang ist der Rhein in Kilometern?	How long is the Rhine River in kilometres?
27.	Wie viele deutsche Bundeskanzler gab es vor Angela Merkel?	How many German chancellors were there before Angela Merkel?
28.	Wer war der erste Mensch im All?	Who was the first human in space?
29.	Welches Land ist seit 2013 Mitglied der Europäischen Union?	Which country has been a member of the European Union since 2013?
30.	Wie heißt der höchste Berg der Alpen?	What is the highest mountain in the Alps?
31.	Huhn verhält sich zu Vogel wie Tisch zu ...?	Chicken is to bird as table is to ...?
32.	Was bedeutet "kumulieren"?	What does "to cumulate" mean?
33.	Wie viele Silben hat das Wort "Monumentalepos"?	How many syllables does the word "Monumentalepos" have?
34.	Welches der folgenden Wörter passt inhaltlich und strukturell nicht zu den anderen: "Beispiel", "Vorbild", "Fehler"?	Which of the following words does not fit with the others: "example", "model", "mistake"?
35.	Konvergieren verhält sich zu divergieren wie sammeln zu ...?	Converge is to diverge as gather is to ...?
36.	Was bedeutet "affektiert"?	What does "affected" mean?
37.	Welches Wort passt nicht zu den anderen: “Quadrat”, “Rechteck”, “Kreis”, “Dreieck”?	Which word does not fit with: “square”, “rectangle”, “circle”, “triangle"?
38.	Welches der Wörter hat den gleichen logischen Aufbau wie "Regenwurm"	Which word has the same logical structure as "earthworm" or

	oder "Schneeflocke": Sonnenaufgang, Baumhaus, Wandfarbe, oder Tisch?	"snowflake": sunrise, treehouse, wall-paint, or table?
39.	Wenn morgen Montag ist, welcher Tag war vorgestern?	If tomorrow is Monday, what day was the day before yesterday?
40.	Was bedeutet "eminent"?	What does "eminent" mean?

Appendix III – NEO-PI-R Test Items & Reliability Analysis

Table 10

NEO-PI-R Test Items Translated from German to English

Personality Domain & Item	German	English
Openness to experience		
1.	Ich habe einen reichen Wortschatz.	I have a rich vocabulary.
2.	Ich habe eine lebhafte Vorstellungskraft.	I have a vivid imagination.
3.	Ich habe hervorragende Ideen.	I have excellent ideas.
4.	Ich verstehe Dinge schnell.	I understand things quickly.
5.	Ich benutze schwierige Worte.	I use difficult words.
6.	Ich verbringe Zeit damit, Dinge zu reflektieren.	I spend time reflecting on things.
7.	Ich bin voller Ideen.	I am full of ideas.
8.	Ich habe Schwierigkeiten abstrakte Ideen zu verstehen.	I have difficulty understanding abstract ideas.
9.	Ich interessiere mich nicht für abstrakte Ideen.	I am not interested in abstract ideas.
10.	Ich habe keine gute Vorstellungskraft.	I do not have a good imagination.
Conscientiousness		
11.	Ich bin immer vorbereitet.	I am always prepared.
12.	Ich achte auf Details.	I pay attention to details.
13.	Ich erledige Hausarbeit sofort.	I get chores done right away.
14.	Ich mag Ordnung.	I like order.
15.	Ich folge einem Plan.	I follow a schedule.
16.	Bei der Arbeit bin ich exakt.	I am exacting in my work.
17.	Ich lasse meine Sachen herumliegen.	I leave my belongings around.
18.	Ich setze Dinge in den Sand.	I make a mess of things.

	19.	Ich vergesse oft, Dinge wieder an den richtigen Platz zurück zu legen.	I often forget to put things back in their proper place.
	20.	Ich drücke mich vor meinen Pflichten.	I shirk my duties.
Extraversion			
	21.	Ich bringe Leben in eine Party.	I bring life to a party.
	22.	Ich fühle mich in Gesellschaft anderer wohl.	I feel comfortable around people.
	23.	Ich beginne Unterhaltungen.	I start conversations.
	24.	Auf Partys spreche ich mit vielen verschiedenen Leuten.	I talk to a lot of different people at parties.
	25.	Es stört mich nicht, im Mittelpunkt der Aufmerksamkeit zu stehen.	I do not mind being the centre of attention.
	26.	Ich rede nicht viel.	I do not talk a lot.
	27.	Ich halte mich im Hintergrund.	I keep in the background.
	28.	Ich habe wenig zu sagen.	I have little to say.
	29.	Ich ziehe nicht gern Aufmerksamkeit auf mich.	I do not like to draw attention to myself.
	30.	Ich bin schweigsam unter fremden Menschen.	I am quiet around strangers.
Agreeableness			
	31.	Ich interessiere mich für Leute.	I am interested in people.
	32.	Ich kann die Gefühle anderer verstehen.	I understand others' feelings.
	33.	Ich habe ein weiches Herz.	I have a soft heart.
	34.	Ich nehme mir Zeit für andere.	I take time out for others.
	35.	Ich kann die Gefühle anderer nachfühlen.	I feel others' emotions.
	36.	In meiner Gegenwart fühlen sich andere wohl.	People feel comfortable around me.

37.	Ich interessiere mich nicht wirklich für andere.	I am not really interested in others.
38.	Ich beleidige Leute.	I insult people.
39.	Ich interessiere mich nicht für die Probleme anderer Leute.	I am not interested in other people's problems.
40.	Andere Menschen kümmern mich wenig.	I do not care about other people.
Neuroticism		
41.	Ich bin die meiste Zeit entspannt.	I am relaxed most of the time.
42.	Ich fühle mich selten traurig.	I seldom feel blue.
43.	Ich rege mich leicht auf.	I get upset easily.
44.	Ich mache mir Sorgen um Dinge.	I worry about things.
45.	Ich fühle mich schnell gestört.	I get irritated easily.
46.	Ich gerate leicht aus der Fassung.	I get easily disturbed.
47.	Meine Laune ändert sich häufig.	My mood changes frequently.
48.	Ich habe häufig Stimmungsschwankungen.	I have frequent mood swings.
49.	Ich lasse mich leicht irritieren.	I get easily annoyed.
50.	Ich fühle mich oft traurig.	I often feel saddened.

Table 11

Reliability Statistics

Cronbach's Alpha	N of Items	N of Cases
.87	50	2327

Appendix IV – AI-CT & ChatGPT Literacy Scale Items

Table 12

Item-comparison of the Critical Thinking Towards AI Scales

New AI-CT scale item	Original adjacent ChatGPT Literacy scale item (Lee & Park, 2024)
1. I check the accuracy of the information provided by GenAI LLMs.	I can evaluate the accuracy of ChatGPT's responses.
2. I carefully consider which GenAI LLM-generated text I add to my own assignments.	I can evaluate the reliability of ChatGPT's responses.
	I can evaluate the completeness of ChatGPT's responses.
3. I assume the text that GenAI LLMs generate is accurate. [a]	I can evaluate the accuracy of ChatGPT's responses.
	I can determine whether ChatGPT's responses are true.
	I can identify errors in ChatGPT's responses.
4. Using GenAI LLMs encourages me to critically evaluate the generated text.	I can recognize and explain bias in ChatGPT's responses.
5. Using GenAI LLMs reduces my motivation to put text into my own words. [a]	No corresponding item(s).
6. I reflect on the text that GenAI LLMs produce before integrating it into my own work.	I can determine whether ChatGPT's responses are true
	I can identify errors in ChatGPT's responses.
	I can evaluate the completeness of ChatGPT's responses.

Note. Multiple AI-CT scale items are adjacent with several of the ChatGPT Literacy scale items by Lee and Park (2024). One AI-CT scale item does not correspond with any of the original ChatGPT Literacy scale items.

[a] These reverse-coded items loaded on a different factor and have been deleted from the AI-CT scale.

Appendix V – AI-CT Scale Factor Analysis & Reliability Analysis

Graph 1

Factor Analysis Scree Plot

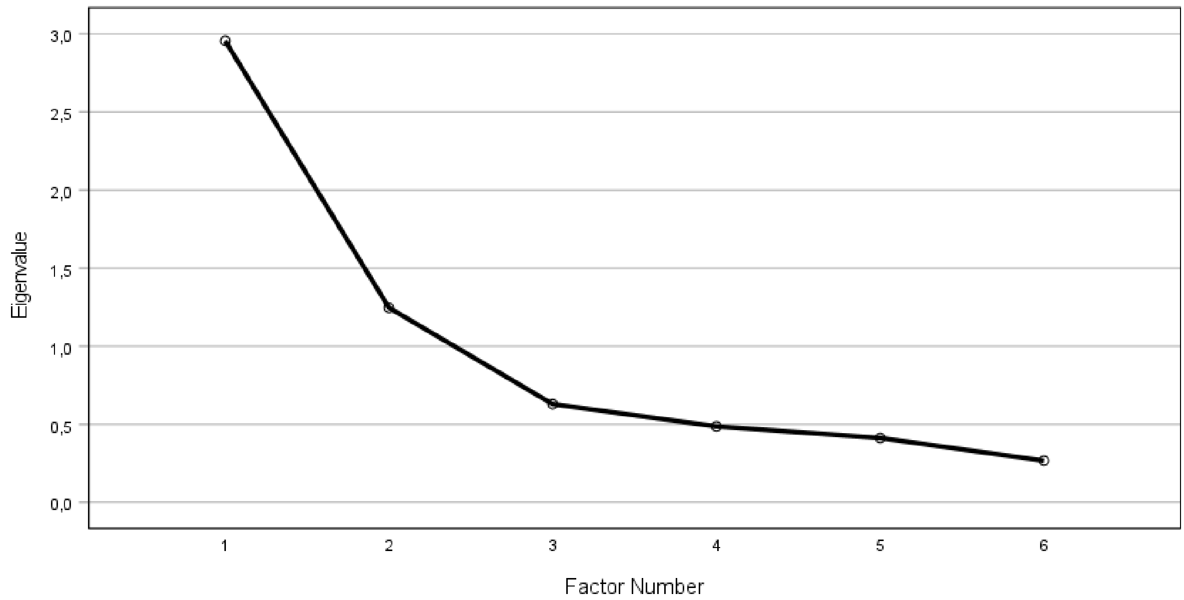


Table 13

Factor Analysis Pattern Matrix[a]

	Factor	
	1	2
I carefully consider which GenAI LLM-generated text I add to my own assignments	.89	
I check the accuracy of the information provided by GenAI LLMs	.79	
I reflect on the text that GenAI LLMs produce before integrating it into my own work	.72	
Using GenAI LLMs encourages me to critically evaluate the generated text	.65	

Unreversed: I assume the text that GenAI LLMs generate is accurate	.78
Unreversed: Using GenAI LLMs reduces my motivation to put text into my own words	.47

Note. Extraction Method: Principal Axis Factoring. Rotation Method: Promax with Kaiser Normalization. Values <.30 are suppressed.

a. Rotation converged in 3 iterations.

Table 14

Inter-Item Correlation Matrix

	1	2	3	4	5	6
I check the accuracy of the information provided by GenAI LLMs	1.000					
I carefully consider which GenAI LLM-generated text I add to my own assignments	.68	1.000				
Unreversed: I assume the text that GenAI LLMs generate is accurate	.001	-.11	1.000			
Using GenAI LLMs encourages me to critically evaluate the generated text	.49	.59	-.18	1.000		
Unreversed: Using GenAI LLMs reduces my motivation to put text into my own words	-.19	-.31	.38	-.26	1.000	

I reflect on the text that GenAI LLMs produce before integrating it into my own work	.53	.66	-.20	.57	-.26	1.000

Table 15

Reliability Statistics: Unreversed items included

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items	N of Cases
.54	.54	6	848

Table 16

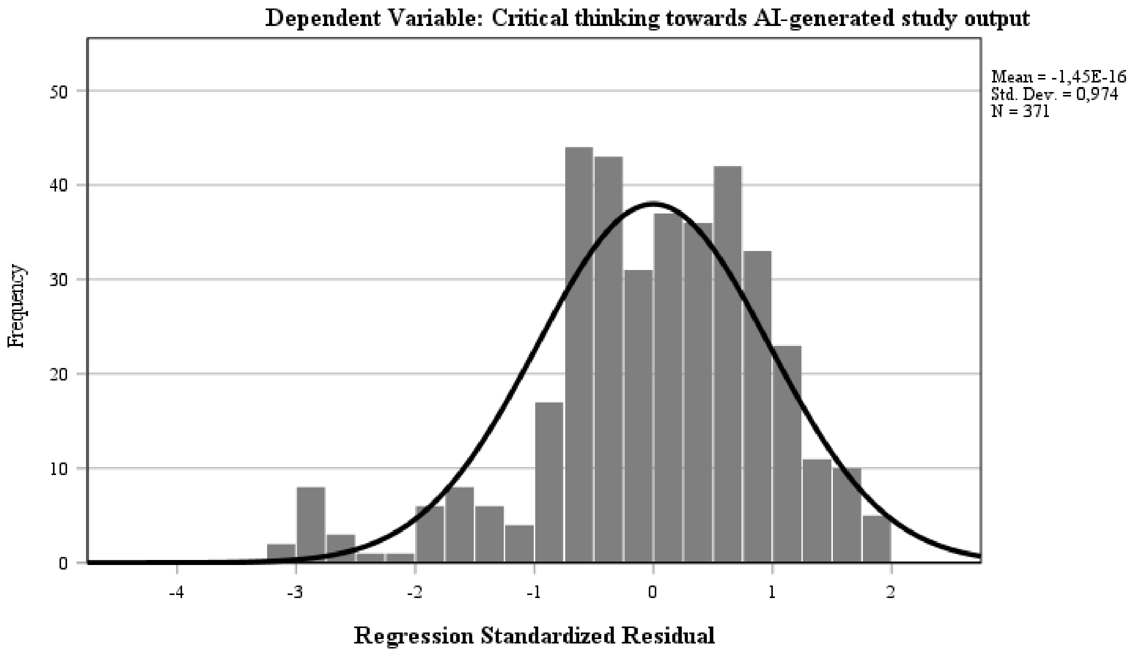
Reliability Statistics: Reverse-coded items excluded

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items	N of Cases
.85	.85	4	848

Appendix VI – Assumption Checks SPSS Output

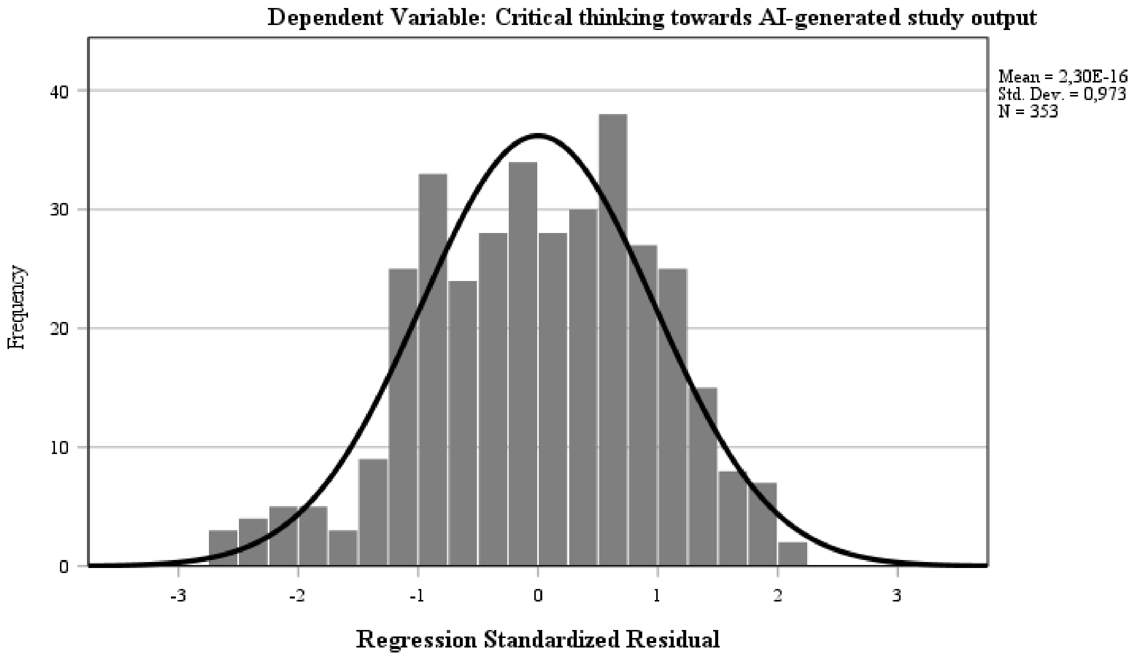
Graph 2

Histogram of Normal Distribution of Standardised Residuals RQ1 Before Outlier Removal



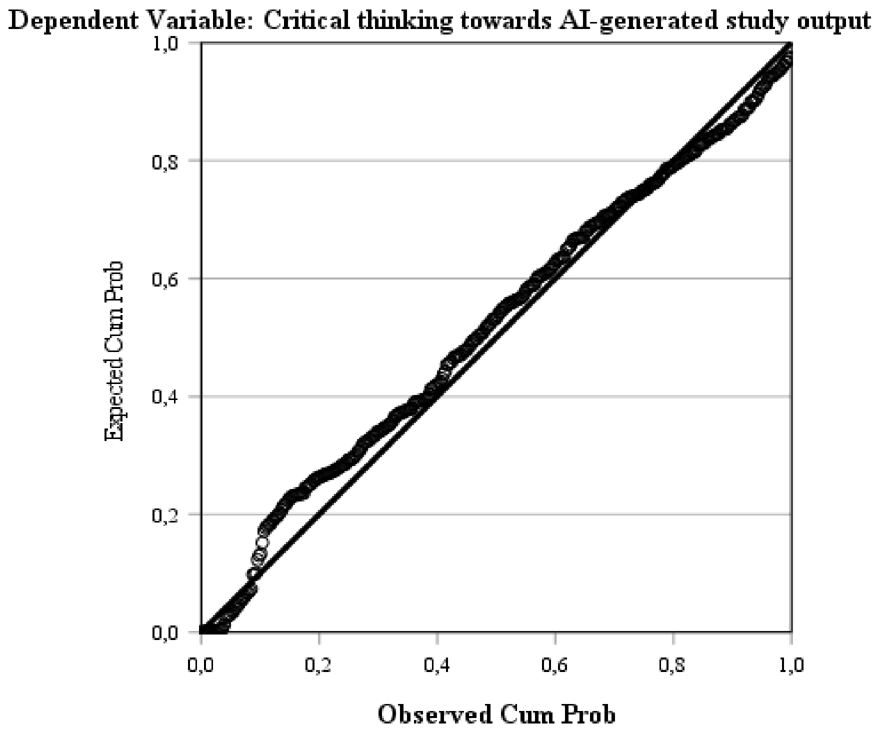
Graph 3

Histogram of Normal Distribution of Standardised Residuals RQ1 After Outlier Removal



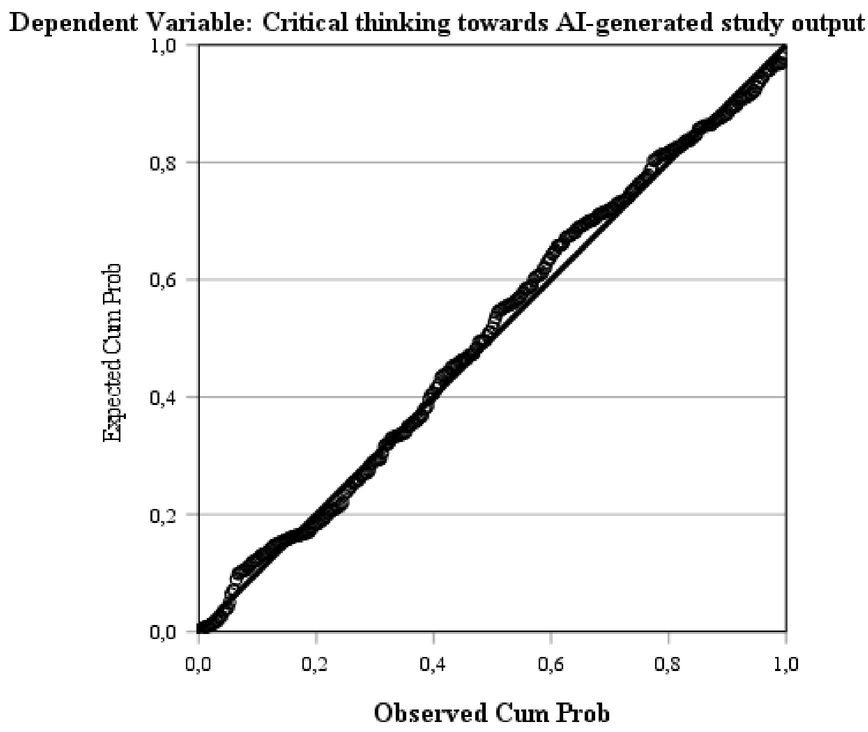
Graph 4

P-P Plot of Standardised Residuals RQ1 Before Outlier Removal



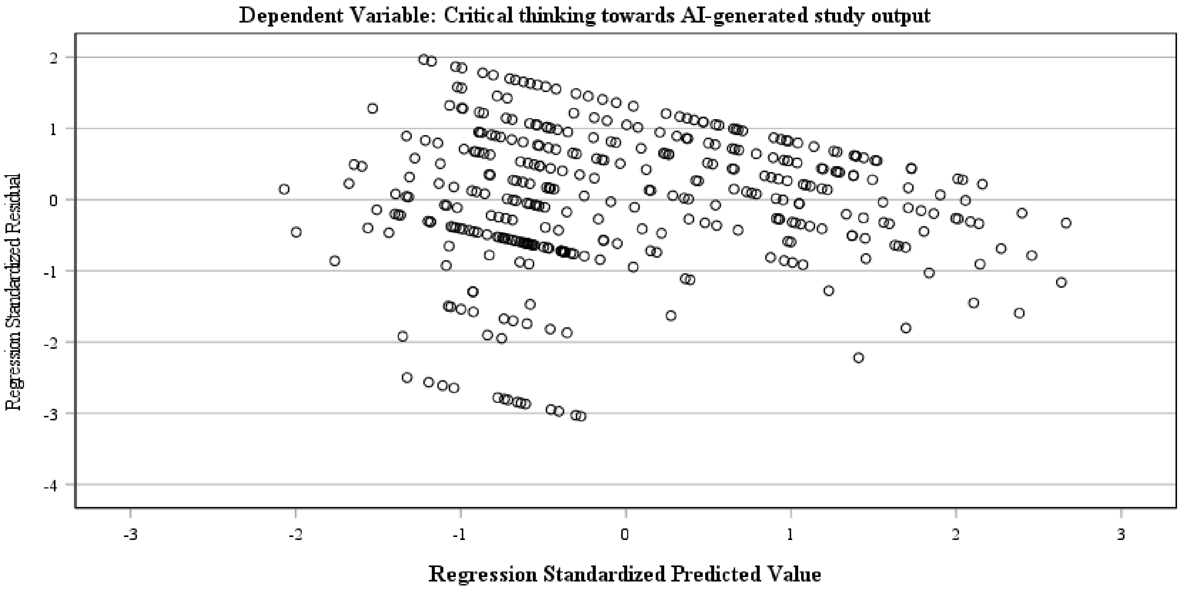
Graph 5

P-P Plot of Standardised Residuals RQ1 After Outlier Removal



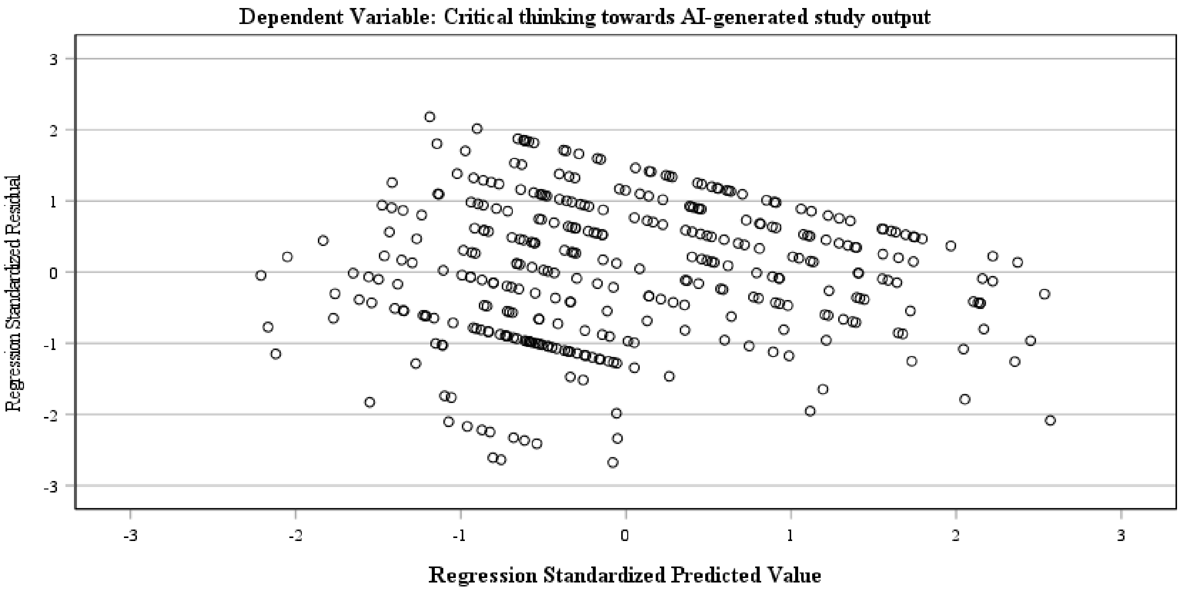
Graph 6

Scatterplot of Standardised Residuals and Predicted Values RQ1 Before Outlier Removal



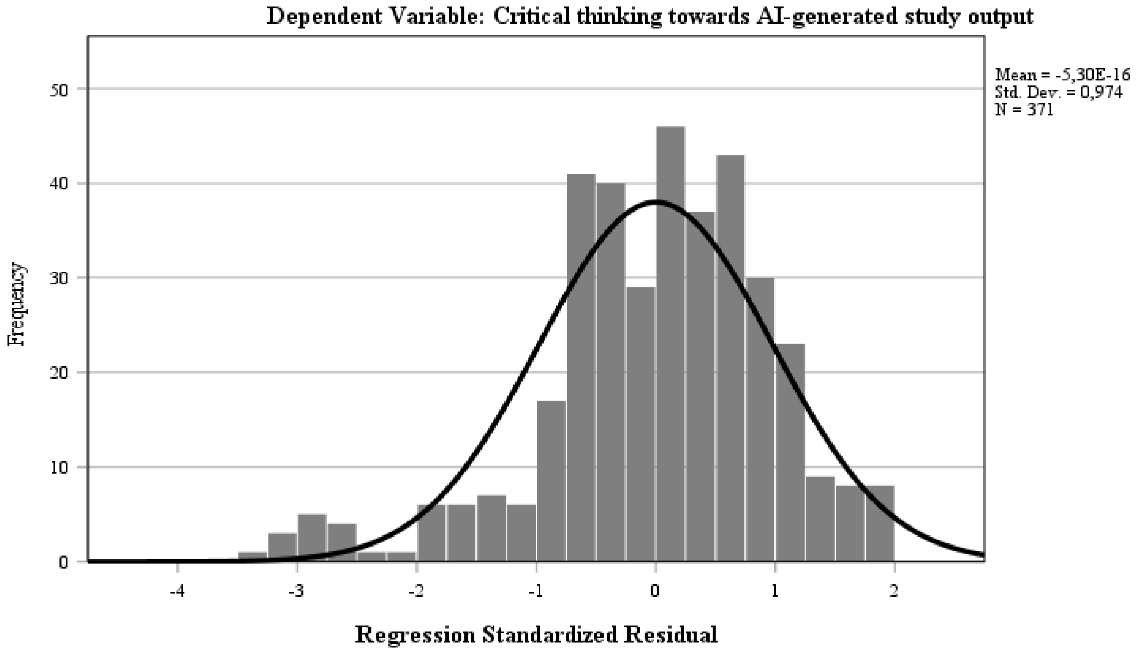
Graph 7

Scatterplot of Standardised Residuals and Predicted Values RQ1 After Outlier Removal



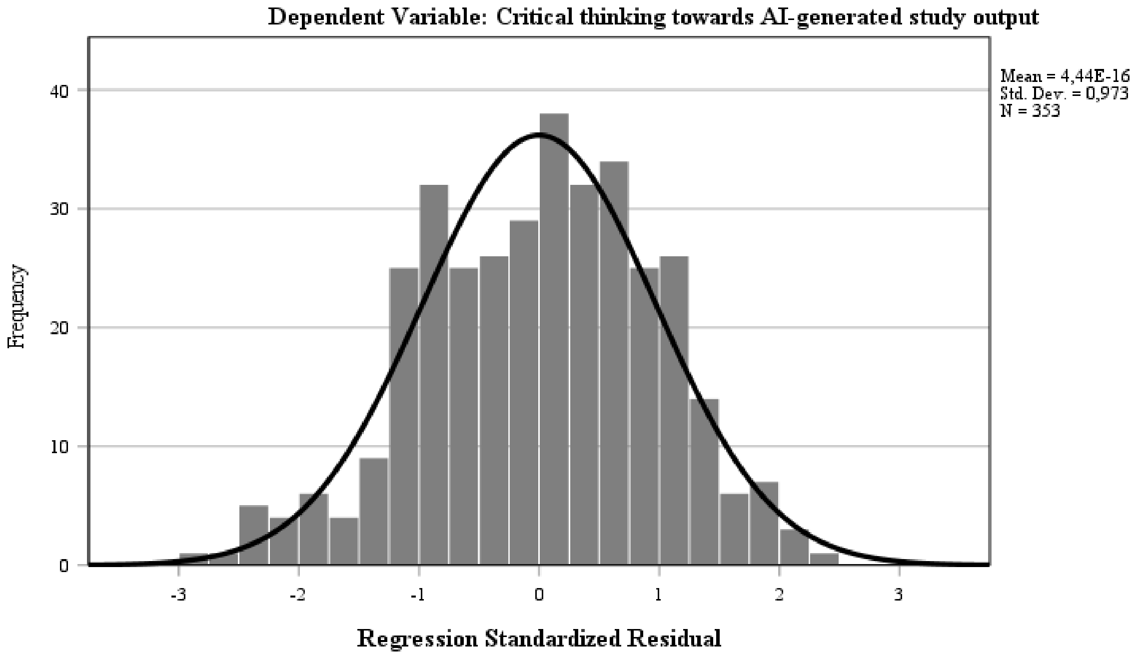
Graph 8

Histogram of Normal Distribution of Standardised Residuals RQ2 Before Outlier Removal



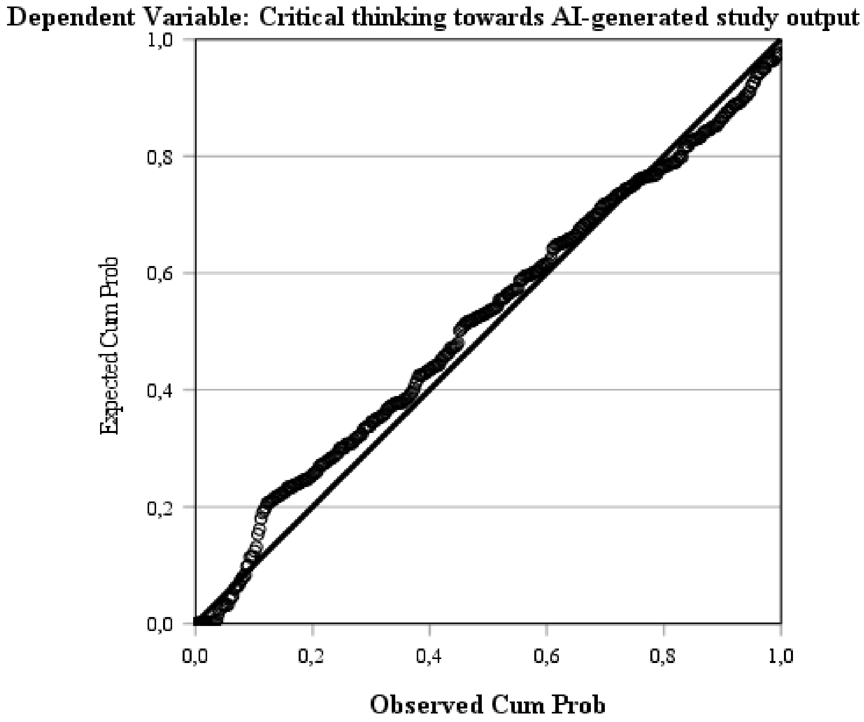
Graph 9

Histogram of Normal Distribution of Standardised Residuals RQ2 After Outlier Removal



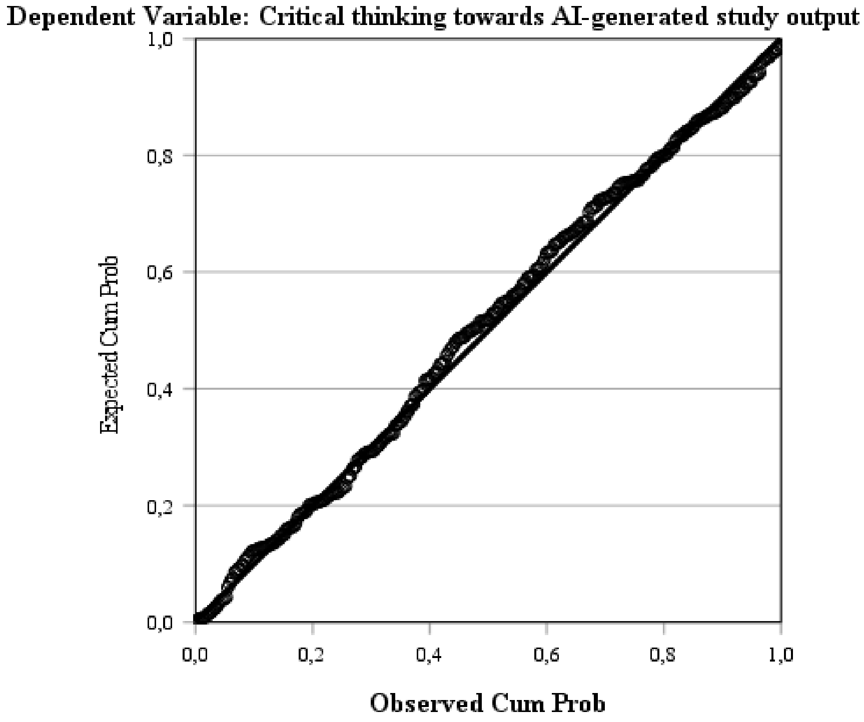
Graph 10

P-P Plot of Standardised Residuals RQ2 Before Outlier Removal



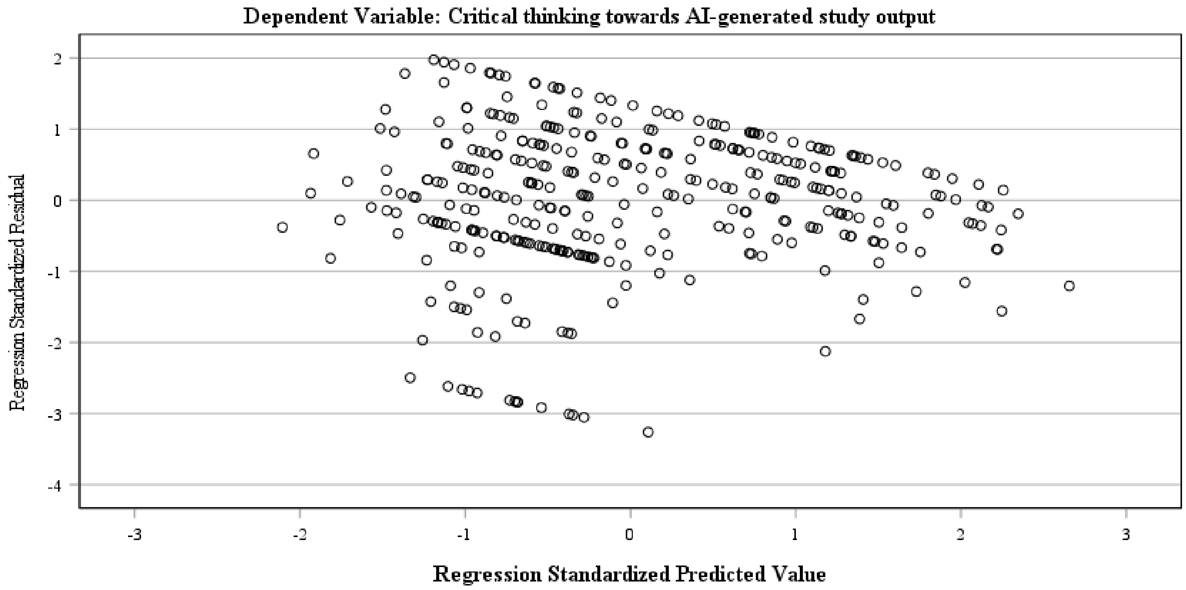
Graph 11

P-P Plot of Standardised Residuals RQ2 After Outlier Removal



Graph 12

Scatterplot of Standardised Residuals and Predicted Values RQ2 Before Outlier Removal



Graph 13

Scatterplot of Standardised Residuals and Predicted Values RQ2 After Outlier Removal

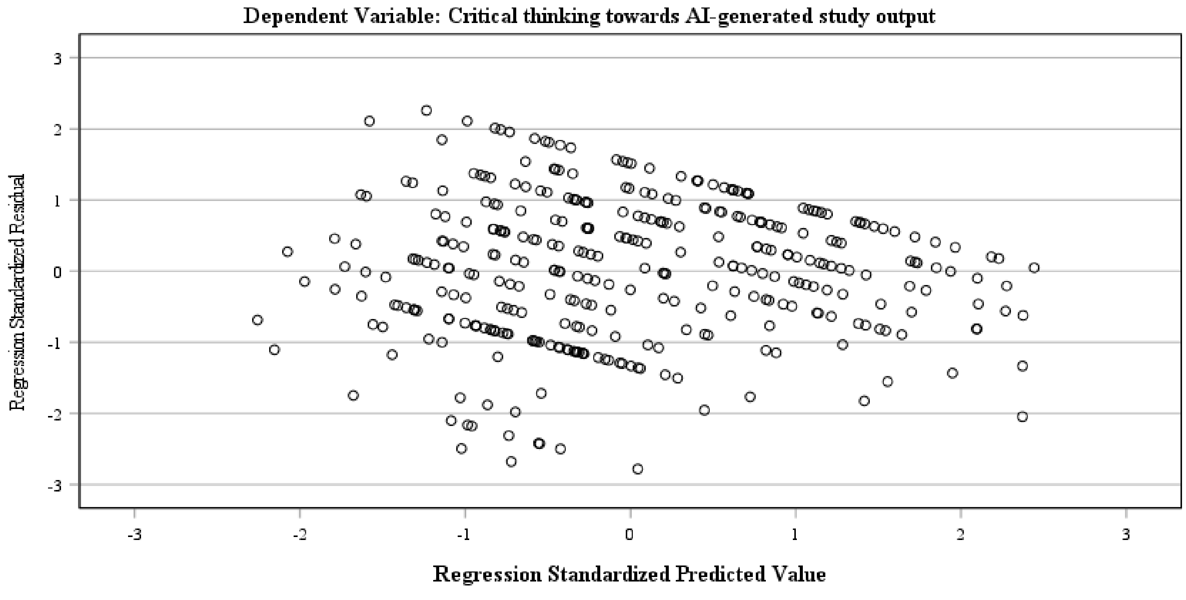


Table 17

Pearson Correlations Table of all Variables to Check for Absence of Multicollinearity

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1. Critical thinking towards AI-generated study output	--																				
2. Fluid intelligence (Raven Matrix)	.11*	--																			
3. Crystallised intelligence (Wechsler IV test)	.22**	.29**	--																		
4. Age	.02	-.04	-.11*	--																	
5. Academic accomplishment other degree	-.08	.004	-.10	-.03	--																
6. Academic accomplishment bachelor	.03	-.06	.04	.37**	-.18**	--															
7. Academic accomplishment master	-.03	.003	.01	.15**	-.05	-.19**	--														
8. I use GenAI LLMs for providing definitions and explanations of concepts	.34**	.002	.06	-.04	-.02	-.001	-.03	--													
9. I use GenAI LLMs for providing writing feedback on my context, language, or grammar	.31**	-.04	.01	-.03	.03	.07	-.02	.44**	--												
10. I use GenAI LLMs for generating ideas or solutions	.25**	.03	.04	-.04	.002	.05	-.10	.60**	.40**	--											
11. I use GenAI LLMs for evaluating my own ideas or solutions	.24**	.01	-.07	-.05	-.08	.002	-.09	.49**	.46**	.59**	--										
12. I use GenAI LLMs for answering practice questions or assignments	.13*	.03	<.001	-.13*	-.03	-.09	.03	.46**	.25**	.44**	.40**	--									
13. I use GenAI LLMs for generating practice questions or assignments	.10	.01	.03	-.04	-.04	.03	-.01	.30**	.35**	.37**	.40**	.48**	--								
14. I use GenAI LLMs for assistance with coding or programming tasks	.33**	.12*	.02	-.05	.01	.01	-.04	.48**	.22**	.44**	.33**	.42**	.29**	--							
15. I use GenAI LLMs for rewriting or paraphrasing my own texts	.31**	.08	.01	.05	-.01	-.003	.06	.50**	.51**	.49**	.45**	.39**	.27**	.38**	--						
16. I use GenAI LLMs for creating summaries of texts	.24**	.03	.04	-.04	.07	.01	-.07	.50**	.45**	.57**	.42**	.47**	.39**	.35**	.49**	--					
17. Openness to experience	.20**	.08	.26**	-.13*	-.05	-.02	.01	.03	-.002	-.07	.05	.04	.02	.08	.02	-.04	--				
18. Conscientiousness	-.03	.02	.01	-.03	.04	-.01	.11*	-.02	-.07	-.04	.07	.001	.09	-.08	-.05	-.12*	.22**	--			
19. Extroversion	.02	-.07	.11*	-.06	-.04	.08	-.03	.09	-.003	-.01	.01	.10	.11*	-.04	-.02	<.001	.44**	.14*	--		
20. Agreeableness	.12*	.03	.14**	.01	.04	.001	.13*	-.03	.03	-.09	.02	-.05	<.001	-.06	-.06	-.09	.37**	.23**	.30**	--	
21. Neuroticism	.02	-.03	.06	-.01	.09	.01	-.06	.04	.04	.04	.02	-.001	.03	-.02	.08	.01	.19**	.28**	.26**	.07	--

Note: $n = 353$. Each variable correlates with itself at a value of 1 ($r = 1$), portrayed by the dashes. A table of all the descriptives of all variables used in this study is included in Appendix VII.

* Correlation is significant at the 0.05 level (2-tailed, $p < .05$).

** Correlation is significant at the 0.01 level (2-tailed, $p < .01$).

Table 18

Collinearity Statistics for RQ1 to Check for Absence of Multicollinearity

	Collinearity Statistics	
Variables	Tolerance	VIF
Fluid intelligence (Raven Matrix)	.90	1.11
Age	.75	1.34
Academic accomplishment other degree	.82	1.22
Academic accomplishment bachelor	.71	1.40
Academic accomplishment master	.91	1.10
I use GenAI LLMs for providing definitions and explanations of concepts	.46	2.17
I use GenAI LLMs for providing writing feedback on my context, language, or grammar	.54	1.87
I use GenAI LLMs for generating ideas or solutions	.47	2.15
I use GenAI LLMs for evaluating my own ideas or solutions	.55	1.81
I use GenAI LLMs for answering practice questions or assignments	.60	1.66
I use GenAI LLMs for generating practice questions or assignments	.75	1.34

I use GenAI LLMs for assistance with coding or programming tasks	.63	1.60
I use GenAI LLMs for rewriting or paraphrasing my own texts	.53	1.89
I use GenAI LLMs for creating summaries of texts	.57	1.75
Openness to experience	.64	1.56
Conscientiousness	.77	1.30
Extroversion	.66	1.51
Agreeableness	.78	1.29
Neuroticism	.79	1.26

Note: n = 353.

Table 19

Collinearity Statistics for RQ2 to Check for Absence of Multicollinearity

	Collinearity Statistics	
Variables	Tolerance	VIF
Crystallised intelligence (Wechsler IV test)	.88	1.14
Age	.74	1.35
Academic accomplishment other degree	.82	1.23
Academic accomplishment bachelor	.71	1.40
Academic accomplishment master	.91	1.10
I use GenAI LLMs for providing definitions and explanations of concepts	.46	2.16

I use GenAI LLMs for providing writing feedback on my context, language, or grammar	.54	1.87
I use GenAI LLMs for generating ideas or solutions	.46	2.17
I use GenAI LLMs for evaluating my own ideas or solutions	.55	1.84
I use GenAI LLMs for answering practice questions or assignments	.60	1.66
I use GenAI LLMs for generating practice questions or assignments	.74	1.34
I use GenAI LLMs for assistance with coding or programming tasks	.64	1.57
I use GenAI LLMs for rewriting or paraphrasing my own texts	.53	1.87
I use GenAI LLMs for creating summaries of texts	.57	1.75
Openness to experience	.63	1.59
Conscientiousness	.77	1.30
Extroversion	.68	1.48
Agreeableness	.78	1.29
Neuroticism	.79	1.26

Note: n = 353.

Appendix VII – Sample and Variable Descriptives

Table 20

Socio-demographic Characteristics of Participants

Characteristic	Sample					
	N	*M*	%	*SD*	*Min.*	*Max.*
Gender, %					1	3
Male	193		54.7%			
Female	154		43.6%			
Other	6		1.7%			
Age (in years), *M* (*SD*)	353	25.01		3.59	18	34
Migration background, %					0	1
Childhood in Germany	198		56.1%			
Childhood not in Germany	27		7.6%			
Missing value	128		36.3%			

Note. $n = 353$. *M* = mean (for numeric variables) or % = percentage (for categorical variables); *SD* = Standard Deviation; *Min.* = Minimum; *Max.* = Maximum. 'Fachkraft 2030' shows no systematic deviations from German student population (Seegers et al., 2019).

Table 21

Descriptive Statistics of all Used Variables in the Current Study

Characteristic	Sample					
	N	*M*	%	*SD*	*Min.*	*Max.*
AI-CT, *M (SD)*	353	15.74		3.21	7	20
Fluid intelligence, *M (SD)*	353	3.04		1.34	0	6
Crystallised intelligence, *M (SD)*	353	3.92		2.41	0	12
Age (in years), *M* (*SD*)	353	25.01		3.59	18	34
Highest academic accomplishment, %					1	4
No degree	177		50.1%			
Other degree	16		4.5%			
Bachelor	142		40.2%			
Master	1		1.6%			

Intended study purpose of GenAI LLMs [a], %						
Providing definitions	231		65.4%		0	1
Providing feedback on writing	261		73.9%		0	1
Generating ideas or solutions	222		62.9%		0	1
Evaluating ideas or solutions	254		72.0%		0	1
Answering practice assignments	294		83.3%		0	1
Generating practice assignments	307		87.0%		0	1
Assisting programming tasks	271		76.8%		0	1
Rewriting or paraphrasing texts	238		67.8%		0	1
Generating text summaries	248		70.3%		0	1
Big Five personality traits, *M (SD)*						
Openness to experience	353	36.58		5.86	18	50
Conscientiousness	353	34.89		6.20	17	50
Extraversion	353	29.79		8.02	11	50
Agreeableness	353	38.83		6.07	18	50
Neuroticism	353	30.54		7.79	10	48

Note. $n = 353$. *M* = mean (for numeric variables) or % = percentage (for categorical variables); *SD* = Standard Deviation; *Min.* = Minimum; *Max.* = Maximum. 'Fachkraft 2030' shows no systematic deviations from German student population (Seegers et al., 2019).

[a] Reflects the number and percentage of participants answering applying this purpose, with value 0 representing 'no' and 1 representing 'yes'. Multiple purposes can be answered by each participant. This means that the % scores do not accumulate to 100%.

Appendix VIII – Plagiaatverklaring

Met ondertekening van deze overeenkomst verklaar ik, Luuk Oskam, student Pedagogische Wetenschappen van de Erasmus Universiteit Rotterdam, dat ik bij het schrijven van mijn scriptie over het onderwerp ‘kritisch denkvermogen in relatie tot AI-gegenereerde output voor studeren’ geen plagiaat heb gepleegd en dat de scriptie het resultaat is van mijn eigen werk en verwoord is in mijn eigen woorden, behoudens citaten. Daar waar mijn scriptie gebaseerd is op informatie dan wel ideeën van een ander zal ik die ander recht doen door naar diens geraadpleegde werk te verwijzen. Tevens verklaar ik dat ik te allen tijde verantwoordelijk blijf voor het bovenstaande.

Datum, plaats:

30 juni 2025, Burgemeester Oudlaan 50, 3062PA Rotterdam, Nederland

Handtekening: